CM2

Loss Reserving and Financial Engineering

Combined Materials Pack for exams in 2019

The Actuarial Education Company
on behalf of the Institute and Faculty of Actuaries

Subject CM2

2019 Study Guide

Introduction

This Study Guide has been created to help guide you through Subject CM2. It contains all the information that you will need before starting to study Subject CM2 for the 2019 exams and you may also find it useful to refer to throughout your Subject CM2 journey.

The guide is split into two parts:

- Part 1 contains general information about the Core Principles subjects
- Part 2 contains specific information about Subject CM2.

Please read this Study Guide carefully before reading the Course Notes, even if you have studied for some actuarial exams before.

Contents

1.1 Before you start

When studying for the UK actuarial exams, you will need:

- a copy of the **Formulae and Tables for Examinations of the Faculty of Actuaries and the Institute of Actuaries, 2nd Edition (2002)** – these are referred to as simply the *Tables.*
- a 'permitted' **scientific calculator** – you will find the list of permitted calculators on the profession's website. Please check the list carefully, since it is reviewed each year.

These are both available from the Institute and Faculty of Actuaries' eShop. Please visit **www.actuaries.org.uk**.

1.2 Core study material

This section explains the role of the Syllabus, Core Reading and supplementary ActEd text. It also gives guidance on how to use these materials most effectively in order to pass the exam.

Some of the information below is also contained in the introduction to the Core Reading produced by the Institute and Faculty of Actuaries.

Syllabus

The Syllabus for Subject CM2 has been produced by the Institute and Faculty of Actuaries. The relevant individual Syllabus Objectives are included at the start of each course chapter and a complete copy of the Syllabus is included in Section 2.2 of this Study Guide. We recommend that you use the Syllabus as an important part of your study.

Core Reading

The Core Reading has been produced by the Institute and Faculty of Actuaries. The purpose of the Core Reading is to assist in ensuring that tutors, students and examiners have clear shared appreciation of the requirements of the syllabus for the qualification examinations for Fellowship of the Institute and Faculty of Actuaries.

The Core Reading supports coverage of the syllabus in helping to ensure that both depth and breadth are re-enforced. It is therefore important that students have a good understanding of the concepts covered by the Core Reading.

The examinations require students to demonstrate their understanding of the concepts given in the syllabus and described in the Core Reading; this will be based on the legislation, professional guidance *etc* that are in force when the Core Reading is published, *ie* on 31 May in the year preceding the examinations.

Therefore the exams in April and September 2019 will be based on the Syllabus and Core Reading as at 31 May 2018. We recommend that you always use the up-to-date Core Reading to prepare for the exams.

Examiners will have this Core Reading when setting the papers. In preparing for examinations, students are advised to work through past examination questions and will find additional tuition helpful. The Core Reading will be updated each year to reflect changes in the syllabus, to reflect current practice, and in the interest of clarity.

Accreditation

The Institute and Faculty of Actuaries would like to thank the numerous people who have helped in the development of the material contained in this Core Reading.

ActEd text

Core Reading deals with each syllabus objective and covers what is needed to pass the exam. However, the tuition material that has been written by ActEd enhances it by giving examples and further explanation of key points. Here is an excerpt from some ActEd Course Notes to show you how to identify Core Reading and the ActEd material. **Core Reading is shown in this bold font.**

Note that in the example given above, the index *will* fall if the actual share price goes below the theoretical ex-rights share price. Again, this is consistent with what would happen to an underlying portfolio.

This is ActEd text

After allowing for chain-linking, **the formula for the investment index then becomes:**

$$I(t) = \frac{\sum_i N_{i,t} P_{i,t}}{B(t)}$$

This is Core Reading

where $N_{i,t}$ **is the number of shares issued for the** i**th constituent at time** t**;**

$B(t)$ **is the base value, or divisor, at time** t**.**

Copyright

1.3 ActEd study support

This section gives a description of the products offered by ActEd.

Successful students tend to undertake three main study activities:

1. *Learning* – initial study and understanding of subject material
2. *Revision* – learning subject material and preparing to tackle exam-style questions
3. *Rehearsal* – answering exam-style questions, culminating in answering questions at exam speed without notes.

Different approaches suit different people. For example, you may like to learn material gradually over the months running up to the exams or you may do your revision in a shorter period just before the exams. Also, these three activities will almost certainly overlap.

We offer a flexible range of products to suit you and let you control your own learning and exam preparation. The following table shows the products that we produce. Note that not all products are available for all subjects.

LEARNING	LEARNING & REVISION	REVISION	REVISION & REHEARSAL	REHEARSAL
Course Notes	Assignments Combined Materials Pack (CMP) Assignment Marking Tutorials Online Classroom	Flashcards	Revision Notes ASET	Mock Exam Mock Marking

The products and services are described in more detail below.

'Learning' products

Course Notes

The Course Notes will help you develop the basic knowledge and understanding of principles needed to pass the exam. They incorporate the complete Core Reading and include full explanation of the syllabus objectives, with worked examples and questions (including some past exam questions) to test your understanding.

Each chapter includes:

- the relevant syllabus objectives
- a chapter summary
- practice questions with full solutions.

Paper B Online Resources (PBOR)

The Paper B Online Resources (PBOR) will help you prepare for the computer-based paper. Delivered through a virtual learning environment (VLE), you will have access to worked examples and practice questions. PBOR will also include the Y Assignments, which are two exam-style assessments.

'Learning & revision' products

X Assignments

The Series X Assignments are written assessments that cover the material in each part of the course in turn. They can be used to both develop and test your understanding of the material.

Combined Materials Pack (CMP)

The Combined Materials Pack (CMP) comprises the Course Notes, PBOR and the Series X Assignments.

The CMP is available in **eBook** format for viewing on a range of electronic devices. eBooks can be ordered separately or as an addition to paper products. Visit **www.ActEd.co.uk** for full details about the eBooks that are available, compatibility with different devices, software requirements and printing restrictions.

X / Y Assignment Marking

We are happy to mark your attempts at the X and/or Y assignments. Marking is not included with the Assignments or the CMP and you need to order both Series X and Series Y Marking separately. You should submit your script as an attachment to an email, in the format detailed in your assignment instructions. You will be able to download your marker's feedback via a secure link on the internet.

Don't underestimate the benefits of doing and submitting assignments:

- Question practice during this phase of your study gives an early focus on the end goal of answering exam-style questions.
- You're incentivised to keep up with your study plan and get a regular, realistic assessment of your progress.
- Objective, personalised feedback from a high quality marker will highlight areas on which to work and help with exam technique.

In a recent study, we found that students who attempt more than half the assignments have significantly higher pass rates.

There are two different types of marking product: Series Marking and Marking Vouchers.

Series Marking

Series Marking applies to a specified subject, session and student. If you purchase Series Marking, you will **not** be able to defer the marking to a future exam sitting or transfer it to a different subject or student.

We typically provide full solutions with the Series Assignments. However, if you order Series Marking at the same time as you order the Series Assignments, you can choose whether or not to receive a copy of the solutions in advance. If you choose not to receive them with the study material, you will be able to download the solutions via a secure link on the internet when your marked script is returned (or following the final deadline date if you do not submit a script).

If you are having your attempts at the assignments marked by ActEd, you should submit your scripts regularly throughout the session, in accordance with the schedule of recommended dates set out in information provided with the assignments. This will help you to pace your study throughout the session and leave an adequate amount of time for revision and question practice.

The recommended submission dates are realistic targets for the majority of students. Your scripts will be returned more quickly if you submit them well before the final deadline dates.

Any script submitted *after* the relevant final deadline date will not be marked. It is your responsibility to ensure that we receive scripts in good time.

Marking Vouchers

Marking Vouchers give the holder the right to submit a script for marking at any time, irrespective of the individual assignment deadlines, study session, subject or person.

Marking Vouchers can be used for any assignment. They are valid for four years from the date of purchase and can be refunded at any time up to the expiry date.

Although you may submit your script with a Marking Voucher at any time, you will need to adhere to the explicit Marking Voucher deadline dates to ensure that your script is returned before the date of the exam. The deadline dates are provided with the assignments.

Tutorials

Our tutorials are specifically designed to develop the knowledge that you will acquire from the course material into the higher-level understanding that is needed to pass the exam.

We run a range of different tutorials including face-to-face tutorials at various locations, and Live Online tutorials. Full details are set out in our *Tuition Bulletin*, which is available on our website at **www.ActEd.co.uk**.

Regular and Block Tutorials

In preparation for these tutorials, we expect you to have read the relevant part(s) of the Course Notes before attending the tutorial so that the group can spend time on exam questions and discussion to develop understanding rather than basic bookwork.

You can choose ***one*** of the following types of tutorial:

- **Regular Tutorials** spread over the session.
- **A Block Tutorial** held two to eight weeks before the exam.

The tutorials outlined above will focus on and develop the skills required for the written Paper A examination. Students wishing for some additional tutor support working through exam-style questions for Paper B may wish to attend a Preparation Day. These will be available Live Online or face-to-face, where students will need to provide their own device capable of running Excel.

Online Classroom

The Online Classroom acts as either a valuable add-on or a great alternative to a face-to-face or Live Online tutorial, focussing on the written Paper A examination.

At the heart of the Online Classroom in each subject is a comprehensive, easily-searched collection of tutorial units. These are a mix of:

- teaching units, helping you to really get to grips with the course material, and
- guided questions, enabling you to learn the most efficient ways to answer questions and avoid common exam pitfalls.

The best way to discover the Online Classroom is to see it in action. You can watch a sample of the Online Classroom tutorial units on our website at **www.ActEd.co.uk**.

'Revision' products

Flashcards

For most subjects, there is ***a lot of material*** to revise. Finding a way to fit revision into your routine as painlessly as possible has got to be a good strategy. Flashcards are a relatively inexpensive option that can provide a massive boost. They can also provide a variation in activities during a study day, and so help you to maintain concentration and effectiveness.

Flashcards are a set of A6-sized cards that cover the key points of the subject that most students want to commit to memory. Each flashcard has questions on one side and the answers on the reverse. We recommend that you use the cards actively and test yourself as you go.

Flashcards are available in **eBook** format for viewing on a range of electronic devices. eBooks can be ordered separately or as an addition to paper products. Visit **www.ActEd.co.uk** for full details about the eBooks that are available, compatibility with different devices, software requirements and printing restrictions.

The following questions and comments might help you to decide if flashcards are suitable for you:

- Do you have a regular train or bus journey?

  Flashcards are ideal for regular bursts of revision on the move.

- Do you want to fit more study into your routine?

  Flashcards are a good option for 'dead time', eg using flashcards on your phone or sticking them on the wall in your study.

- Do you find yourself cramming for exams (even if that's not your original plan)?

  Flashcards are an extremely efficient way to do your pre-exam memorising.

If you are retaking a subject, then you might consider using flashcards if you didn't use them on a previous attempt.

'Revision & rehearsal' products

Revision Notes

Our Revision Notes have been designed with input from students to help you revise efficiently. They are suitable for first-time sitters who have worked through the ActEd Course Notes or for retakers (who should find them much more useful and challenging than simply reading through the course again).

The Revision Notes are a set of A5 booklets – perfect for revising on the train or tube to work. Each booklet covers one main theme or a set of related topics from the course and includes:

- Core Reading to develop your bookwork knowledge
- relevant past exam questions with concise solutions from the last ten years
- other useful revision aids.

ActEd Solutions with Exam Technique (ASET)

The ActEd Solutions with Exam Technique (ASET) contains our solutions to eight past exam papers, plus comment and explanation. In particular, it highlights how questions might have been analysed and interpreted so as to produce a good solution with a wide range of relevant points. This will be valuable in approaching questions in subsequent examinations.

'Rehearsal' products

Mock Exam

The Mock Exam consists of two papers. There is a 100-mark mock exam for the written Paper A examination and a separate mock exam for the computer-based Paper B exam. These provide a realistic test of your exam readiness.

Mock Marking

We are happy to mark your attempts at the mock exams. The same general principles apply as for the Assignment Marking. In particular:

- Mock Exam Marking applies to a specified subject, session and student. In this subject it covers the marking of both papers.
- Marking Vouchers can be used for each mock exam paper. Note that you will need two marking vouchers in order to have the two mock papers marked.

Recall that:

- marking is not included with the products themselves and you need to order it separately
- you should submit your script via email in the format detailed in the mock exam instructions
- you will be able to download the feedback on your marked script via a secure link on the internet.

1.4 Skills

Technical skills

The Core Reading and exam papers for these subjects tend to be very technical. The exams themselves have many calculation and manipulation questions. The emphasis in the exam will therefore be on *understanding* the mathematical techniques and applying them to various, frequently unfamiliar, situations. It is important to have a feel for what the numerical answer should be by having a deep understanding of the material and by doing reasonableness checks.

As a high level of pure mathematics and statistics is generally required for the Core Principles subjects, it is important that your mathematical skills are extremely good. If you are a little rusty you may wish to consider purchasing additional material to help you get up to speed. The course 'Pure Maths and Statistics for Actuarial Studies' is available from ActEd and it covers the mathematical techniques that are required for the Core Principles subjects, some of which are beyond A-Level (or Higher) standard. You do not need to work through the whole course in order – you can just refer to it when you need help on a particular topic. An initial assessment to test your mathematical skills and further details regarding the course can be found on our website at www.ActEd.co.uk.

Study skills

Overall study plan

We suggest that you develop a realistic study plan, building in time for relaxation and allowing some time for contingencies. Be aware of busy times at work, when you may not be able to take as much study leave as you would like. Once you have set your plan, be determined to stick to it. You don't have to be too prescriptive at this stage about what precisely you do on each study day. The main thing is to be clear that you will cover all the important activities in an appropriate manner and leave plenty of time for revision and question practice.

Aim to manage your study so as to allow plenty of time for the concepts you meet in these courses to 'bed down' in your mind. Most successful students will probably aim to complete the courses at least a month before the exam, thereby leaving a sufficient amount of time for revision. By finishing the courses as quickly as possible, you will have a much clearer view of the big picture. It will also allow you to structure your revision so that you can concentrate on the important and difficult areas.

You can also try looking at our discussion forum on the internet, which can be accessed at **www.ActEd.co.uk/forums** (or use the link from our home page at **www.ActEd.co.uk**). There are some good suggestions from students on how to study.

Study sessions

Only do activities that will increase your chance of passing. Try to avoid including activities for the sake of it and don't spend time reviewing material that you already understand. You will only improve your chances of passing the exam by getting on top of the material that you currently find difficult.

Ideally, each study session should have a specific purpose and be based on a specific task, *eg 'Finish reading Chapter 3 and attempt Practice Questions 1.4, 1.7 and 1.12'*, as opposed to a specific amount of time, *eg 'Three hours studying the material in Chapter 3'*.

Try to study somewhere quiet and free from distractions (*eg* a library or a desk at home dedicated to study). Find out when you operate at your peak, and endeavour to study at those times of the day. This might be between 8*am* and 10*am* or could be in the evening. Take short breaks during your study to remain focused – it's definitely time for a short break if you find that your brain is tired and that your concentration has started to drift from the information in front of you.

Order of study

We suggest that you work through each of the chapters in turn. To get the maximum benefit from each chapter you should proceed in the following order:

1. Read the Syllabus Objectives. These are set out in the box at the start of each chapter.
2. Read the Chapter Summary at the end of each chapter. This will give you a useful overview of the material that you are about to study and help you to appreciate the context of the ideas that you meet.
3. Study the Course Notes in detail, annotating them and possibly making your own notes. Try the self-assessment questions as you come to them. As you study, pay particular attention to the listing of the Syllabus Objectives and to the Core Reading.
4. Read the Chapter Summary again carefully. If there are any ideas that you can't remember covering in the Course Notes, read the relevant section of the notes again to refresh your memory.
5. Attempt (at least some of) the Practice Questions that appear at the end of the chapter.
6. Where relevant, work through the Paper B Online Resources for the chapter. You will need to have a good understanding of the relevant section of the paper-based course before you attempt the corresponding section of PBOR.

It's a fact that people are more likely to remember something if they review it several times. So, do look over the chapters you have studied so far from time to time. It is useful to re-read the Chapter Summaries or to try the Practice Questions again a few days after reading the chapter itself. It's a good idea to annotate the questions with details of when you attempted each one. This makes it easier to ensure that you try all of the questions as part of your revision without repeating any that you got right first time.

Once you've read the relevant part of the notes and tried a selection of questions from the Practice Questions (and attended a tutorial, if appropriate) you should attempt the corresponding assignment. If you submit your assignment for marking, spend some time looking through it carefully when it is returned. It can seem a bit depressing to analyse the errors you made, but you will increase your chances of passing the exam by learning from your mistakes. The markers will try their best to provide practical comments to help you to improve.

To be really prepared for the exam, you should not only know and understand the Core Reading but also be aware of what the examiners will expect. Your revision programme should include plenty of question practice so that you are aware of the typical style, content and marking structure of exam questions. You should attempt as many past exam questions as you can.

Active study

Here are some techniques that may help you to study actively.

1. Don't believe everything you read. Good students tend to question everything that they read. They will ask 'why, how, what for, when?' when confronted with a new concept, and they will apply their own judgement. This contrasts with those who unquestioningly believe what they are told, learn it thoroughly, and reproduce it (unquestioningly?) in response to exam questions.

2. Another useful technique as you read the Course Notes is to think of possible questions that the examiners could ask. This will help you to understand the examiners' point of view and should mean that there are fewer nasty surprises in the exam room. Use the Syllabus to help you make up questions.

3. Annotate your notes with your own ideas and questions. This will make you study more actively and will help when you come to review and revise the material. Do not simply copy out the notes without thinking about the issues.

4. Attempt the questions in the notes as you work through the course. Write down your answer before you refer to the solution.

5. Attempt other questions and assignments on a similar basis, *ie* write down your answer before looking at the solution provided. Attempting the assignments under exam conditions has some particular benefits:
   - It forces you to think and act in a way that is similar to how you will behave in the exam.
   - When you have your assignments marked it is *much* more useful if the marker's comments can show you how to improve your performance under exam conditions than your performance when you have access to the notes and are under no time pressure.
   - The knowledge that you are going to do an assignment under exam conditions and then submit it (however good or bad) for marking can act as a powerful incentive to make you study each part as well as possible.
   - It is also quicker than trying to write perfect answers.

6. Sit a mock exam four to six weeks before the real exam to identify your weaknesses and work to improve them. You could use a mock exam written by ActEd or a past exam paper.

You can find further information on how to study in the profession's Student Handbook, which you can download from their website at:

www.actuaries.org.uk/studying

Revision and exam skills

Revision skills

You will have sat many exams before and will have mastered the exam and revision techniques that suit you. However it is important to note that due to the high volume of work involved in the Core Principles subjects it is not possible to leave all your revision to the last minute. Students who prepare well in advance have a better chance of passing their exams on the first sitting.

Unprepared students find that they are under time pressure in the exam. Therefore it is important to find ways of maximising your score in the shortest possible time. Part of your preparation should be to practise a large number of exam-style questions under timed exam conditions as soon as possible. This will:

- help you to develop the necessary understanding of the techniques required
- highlight the key topics, which crop up regularly in many different contexts and questions
- help you to practise the specific skills that you will need to pass the exam.

There are many sources of exam-style questions. You can use past exam papers, the Practice Questions at the end of each chapter (which include many past exam questions), assignments, mock exams, the Revision Notes and ASET.

Exam question skill levels

Exam questions are not designed to be of similar difficulty. The Institute and Faculty of Actuaries specifies different skill levels that questions may be set with reference to.

Questions may be set at any skill level:

- Knowledge – demonstration of a detailed knowledge and understanding of the topic
- Application – demonstration of an ability to apply the principles underlying the topic within a given context
- Higher Order – demonstration of an ability to perform deeper analysis and assessment of situations, including forming judgements, taking into account different points of view, comparing and contrasting situations, suggesting possible solutions and actions, and making recommendations.

Command verbs

The Institute and Faculty of Actuaries use command verbs (such as ‘Define’, ‘Discuss’ and ‘Explain’) to help students to identify what the question requires. The profession has produced a document, ‘Command verbs used in the Associate and Fellowship written examinations’, to help students to understand what each command verb is asking them to do.

It also gives the following advice:

- The use of a specific command verb within a syllabus objective does not indicate that this is the only form of question which can be asked on the topic covered by that objective.
- The Examiners may ask a question on any syllabus topic using any of the agreed command verbs, as are defined in the document.

You can find the relevant document on the profession's website at:

https://www.actuaries.org.uk/studying/prepare-your-exams

1.5 The examination

What to take to the exam

IMPORTANT NOTE: The following information was correct at the time of printing, however it is important to keep up-to-date with any changes. See the profession's website for the latest guidance.

For the written exams the examination room will be equipped with:

- the question paper
- an answer booklet
- rough paper
- a copy of the *Tables*.

Remember to take with you:

- black pens
- a permitted scientific calculator – please refer to **www.actuaries.org.uk** for the latest advice.

Please also refer to the profession's website and your examination instructions for details about what you will need for the computer-based Paper B exam.

Past exam papers

You can download some past exam papers and Examiners' Reports from the profession's website at **www.actuaries.org.uk**. However, please be aware that these exam papers are for the pre-2019 syllabus and not all questions will be relevant.

1.6 Queries and feedback

Questions and queries

From time to time you may come across something in the study material that is unclear to you. The easiest way to solve such problems is often through discussion with friends, colleagues and peers – they will probably have had similar experiences whilst studying. If there's no-one at work to talk to then use our discussion forum at **www.ActEd.co.uk/forums** (or use the link from our home page at **www.ActEd.co.uk**).

Our online forum is dedicated to actuarial students so that you can get help from fellow students on any aspect of your studies from technical issues to study advice. You could also use it to get ideas for revision or for further reading around the subject that you are studying. ActEd tutors will visit the site from time to time to ensure that you are not being led astray and we also post other frequently asked questions from students on the forum as they arise.

If you are still stuck, then you can send queries by email to the relevant subject email address (see Section 2.6), but we recommend that you try the forum first. We will endeavour to contact you as soon as possible after receiving your query but you should be aware that it may take some time to reply to queries, particularly when tutors are away from the office running tutorials. At the busiest teaching times of year, it may take us more than a week to get back to you.

If you have many queries on the course material, you should raise them at a tutorial or book a personal tuition session with an ActEd tutor. Information about personal tuition is set out in our current brochure. Please email **ActEd@bpp.com** for more details.

Feedback

If you find an error in the course, please check the corrections page of our website (**www.ActEd.co.uk/paper_corrections.html**) to see if the correction has already been dealt with. Otherwise please send details via email to the relevant subject email address (see Section 2.6).

Each year our tutors work hard to improve the quality of the study material and to ensure that the courses are as clear as possible and free from errors. We are always happy to receive feedback from students, particularly details concerning any errors, contradictions or unclear statements in the courses. If you have any comments on this course please email them to the relevant subject email address (see Section 2.6).

Our tutors also work with the profession to suggest developments and improvements to the Syllabus and Core Reading. If you have any comments or concerns about the Syllabus or Core Reading, these can be passed on via ActEd. Alternatively, you can send them directly to the Institute and Faculty of Actuaries' Examination Team by email to **education.services@actuaries.org.uk**.

2.1 Subject CM2 – background

History

The Actuarial Mathematics subjects (Subjects CM1 and CM2) are new subjects in the Institute and Faculty of Actuaries 2019 Curriculum.

Subject CM2 is *Financial Engineering and Loss Reserving*.

Predecessors

The topics covered in the Actuarial Mathematics subjects (Subjects CM1 and CM2) cover content previously in Subjects CT1, CT5, CT8 and a small amount from Subjects CT4, CT6 and CT7:

- Subject CM1 contains material from Subjects CT1, CT4 and CT5.
- Subject CM2 contains material from Subjects CT8, CT6, CT1 and CT7.

Exemptions

You will need to have passed or been granted an exemption from Subject CT8 to be eligible for a pass in Subject CM2 during the transition process.

Links to other subjects

Concepts introduced in the following subjects are used in Subject CM2:

- CS1 – Actuarial Statistics 1
- CS2 – Risk Modelling and Survival Analysis
- CM1 – Actuarial Mathematics 1
- CB2 – Business Economics.

Topics in Subject CM2 are further built upon in the following subjects:

- Subject CP1 – Actuarial Practice
- CP2 – Modelling Practice
- SP5 – Investment and Finance Principles
- SP6 – Financial Derivatives Principles
- SP9 – Enterprise Risk Management Principles.

2.2 Subject CM2 – Syllabus and Core Reading

Syllabus

The Syllabus for Subject CM2 is given here. To the right of each objective are the chapter numbers in which the objective is covered in the ActEd course.

Aim

The aim of the Financial Engineering and Loss Reserving subject is to provide a grounding in the principles of modelling as applied to actuarial work – focusing particularly on stochastic asset liability models and the valuation of financial derivatives. These skills are also required to communicate with other financial professionals and to critically evaluate modern financial theories.

Competences

On successful completion of this subject, a student will be able to:

1. describe, interpret and discuss the theories on the behaviour of financial markets.
2. discuss the advantages and disadvantages of different measures of investment risk.
3. describe, construct, interpret and discuss the models underlying asset valuations.
4. describe, construct, interpret and discuss the models underlying liability valuations.
5. describe, construct, interpret and discuss the models underlying option pricing.

Syllabus topics

1.	Theories of financial market behaviour	(15%)
2.	Measures of investment risk	(15%)
3.	Stochastic investment return models	(10%)
4.	Asset valuations	(20%)
5.	Liability valuations	(20%)
6.	Option theory	(20%)

The weightings are indicative of the approximate balance of the assessment of this subject between the main syllabus topics, averaged over a number of examination sessions.

The weightings also have a correspondence with the amount of learning material underlying each syllabus topic. However, this will also reflect aspects such as:

- the relative complexity of each topic, and hence the amount of explanation and support required for it
- the need to provide thorough foundation understanding on which to build the other objectives
- the extent of prior knowledge which is expected
- the degree to which each topic area is more knowledge or application based.

Detailed syllabus objectives

1. Theories of financial market behaviour (15%)

1.1 Rational expectations theory (Chapter 1)

1.1.1 Discuss the three forms of the Efficient Markets Hypothesis and their consequences for investment management.

1.1.2 Describe briefly the evidence for or against each form of the Efficient Markets Hypothesis.

1.2 Rational choice theory (Chapters 2 and 3)

1.2.1 Explain the meaning of the term "utility function".

1.2.2 Explain the axioms underlying utility theory and the expected utility theorem.

1.2.3 Explain how the following economic characteristics of investors can be expressed mathematically in a utility function:

- non-satiation
- risk aversion, risk neutrality and risk seeking
- declining or increasing absolute and relative risk aversion

1.2.4 Discuss the economic properties of commonly used utility functions.

1.2.5 Discuss how a utility function may depend on current wealth and discuss state dependent utility functions.

1.2.6 Perform calculations using commonly used utility functions to compare investment opportunities.

1.2.7 State conditions for absolute dominance and for first and second-order dominance.

1.2.8 Analyse simple insurance problems in terms of utility theory

1.3 Behavioural economics (Chapter 3)

1.3.1 Describe the main features of Kahneman and Tversky's prospect theory critique of expected utility theory.

1.3.2 Explain what is meant by "framing", "heuristics" and "bias" in the context of financial markets and describe the following features of behaviour in such markets:

- the herd instinct
- anchoring and adjustment
- self-serving bias
- loss aversion
- confirmation bias
- availability bias
- familiarity bias

1.3.3 Describe the Bernartzi and Thaler solution to the equity premium puzzle.

2 Measures of investment risk (15%)

2.1 Properties of risk measures (Chapter 4)

2.1.1 Define the following measures of investment risk:

- variance of return
- downside semi-variance of return
- shortfall probabilities
- Value at Risk (VaR) / Tail VaR

2.1.2 Describe how the risk measures listed in 2.1.1 above are related to the form of an investor's utility function.

2.1.3 Perform calculations using the risk measures listed in.2.1.1 above to compare investment opportunities.

2.1.4 Explain how the distribution of returns and the thickness of tails will influence the assessment of risk.

2.2 Risk and insurance companies (Chapter 4)

2.2.1 Describe how insurance companies help to reduce or remove risk.

2.2.2 Explain what is meant by the terms "moral hazard" and "adverse selection".

3 Stochastic investment return models (10%)

3.1 Show an understanding of simple stochastic models for investment returns. (Chapter 5)

3.1.1 Describe the concept of a stochastic investment return model and the fundamental distinction between this and a deterministic model.

3.1.2 Derive algebraically, for the model in which the annual rates of return are independently and identically distributed and for other simple models, expressions for the mean value and the variance of the accumulated amount of a single premium.

3.1.3 Derive algebraically, for the model in which the annual rates of return are independently and identically distributed, recursive relationships which permit the evaluation of the mean value and the variance of the accumulated amount of an annual premium.

3.1.4 Derive analytically, for the model in which each year the random variable (1 + r) has an independent log-normal distribution, the distribution functions for the accumulated amount of a single premium and for the present value of a sum due at a given specified future time.

3.1.5 Apply the above results to the calculation of the probability that a simple sequence of payments will accumulate to a given amount at a specific future time.

4 Asset valuations (20%)

4.1 Mean-variance portfolio theory (Chapter 6)

4.1.1 Describe and discuss the assumptions of mean-variance portfolio theory.

4.1.2 Discuss the conditions under which application of mean-variance portfolio theory leads to the selection of an optimum portfolio.

4.1.3 Calculate the expected return and risk of a portfolio of many risky assets, given the expected return, variance and covariance of returns of the individual assets, using mean-variance portfolio theory.

4.1.4 Explain the benefits of diversification using mean-variance portfolio theory.

4.2 Asset pricing models (Chapter 8)

4.2.1 Describe the assumptions, principal results and uses of the Sharpe-Lintner-Mossin Capital Asset Pricing Model (CAPM).

4.2.2 Discuss the limitations of the basic CAPM and some of the attempts that have been made to develop the theory to overcome these limitations.

4.2.3 Perform calculations using the CAPM.

4.2.4 Discuss the main issues involved in estimating parameters for asset pricing models.

4.3 Single and multifactor models for investment returns (Chapter 7)

4.3.1 Describe the three types of multifactor models of asset returns:

- macroeconomic models
- fundamental factor models
- statistical factor models

4.3.2 Discuss the single index model of asset returns.

4.3.3 Discuss the concepts of diversifiable and non-diversifiable risk.

4.3.4 Discuss the construction of the different types of multifactor models.

4.3.5 Perform calculations using both single and multi-factor models.

4.4 Stochastic models for security prices (Chapters 9, 10 and 11)

4.4.1 Discuss the continuous time log-normal model of security prices and the empirical evidence for or against the model.

4.4.2 Explain the definition and basic properties of standard Brownian motion or Wiener process.

4.4.3 Demonstrate a basic understanding of stochastic differential equations, the Ito integral, diffusion and mean-reverting processes.

4.4.4 State Ito's Lemma and be able to apply it to simple problems.

4.4.5 Write down the stochastic differential equation for geometric Brownian motion and show how to find its solution.

4.4.6 Write down the stochastic differential equation for the Ornstein-Uhlenbeck process and show how to find its solution.

4.5 Models of the term structures of interest rates (Chapter 18)

4.5.1 Explain the principal concepts and terms underlying the theory of a term structure of interest rates.

4.5.2 Describe the desirable characteristics of models for the term-structure of interest rates.

4.5.3 Apply the term structure of interest rates to modelling various cash flows, including calculating the sensitivity of the value to changes in the term structure.

4.5.4 Describe, as a computational tool, the risk-neutral approach to the pricing of zero-coupon bonds and interest-rate derivatives for a general one-factor diffusion model for the risk-free rate of interest.

4.5.5 Describe, as a computational tool, the approach using state-price deflators to the pricing of zero-coupon bonds and interest-rate derivatives for a general one-factor diffusion model for the risk-free rate of interest.

4.5.6 Demonstrate an awareness of the Vasicek, Cox-Ingersoll-Ross and Hull-White models for the term-structure of interest rates.

4.5.7 Discuss the limitations of these one-factor models and show an awareness of how these issues can be addressed.

4.6 Simple models for credit risk (Chapter 19)

4.6.1 Define the terms credit event and recovery rate.

4.6.2 Describe the different approaches to modelling credit risk: structural models, reduced form models, intensity-based models.

4.6.3 Demonstrate a knowledge and understanding of the Merton model.

4.6.4 Demonstrate a knowledge and understanding of a two-state model for credit ratings with a constant transition intensity.

4.6.5 Describe how the two-state model can be generalised to the Jarrow-Lando-Turnbull model for credit ratings.

4.6.6 Describe how the two-state model can be generalised to incorporate a stochastic transition intensity.

5 Liability valuations (20%)

5.1 Ruin theory (Chapter 20)

5.1.1 Explain what is meant by the aggregate claim process and the cash-flow process for a risk.

5.1.2 Use the Poisson process and the distribution of inter-event times to calculate probabilities of the number of events in a given time interval and waiting times.

5.1.3 Define a compound Poisson process and calculate probabilities using simulation.

5.1.4 Define the probability of ruin in infinite/finite and continuous/discrete time and state and explain relationships between the different probabilities of ruin.

5.1.5 Describe the effect on the probability of ruin, in both finite and infinite time, of changing parameter values by reasoning or simulation.

5.1.6 Calculate probabilities of ruin by simulation.

5.2 Run-off triangles (Chapter 21)

5.2.1 Define a development factor and show how a set of assumed development factors can be used to project the future development of a delay triangle.

5.2.2 Describe and apply a basic chain ladder method for completing the delay triangle using development factors.

5.2.3 Show how the basic chain ladder method can be adjusted to make explicit allowance for inflation.

5.2.4 Describe and apply the average cost per claim method for estimating outstanding claim amounts.

5.2.5 Describe and apply the Bornhuetter-Ferguson method for estimating outstanding claim amounts.

5.2.6 Describe how a statistical model can be used to underpin a run off triangles approach.

5.2.7 Discuss the assumptions underlying the application of the methods in 5.2.1 to 5.2.6 above.

5.3 Value basic benefit guarantees using simulation techniques.

6 Option theory (20%)

6.1 Option pricing and valuations (Chapters 12 – 17)

6.1.1 State what is meant by arbitrage and a complete market.

6.1.2 Outline the factors that affect option prices.

6.1.3 Derive specific results for options which are not model dependent:

- show how to value a forward contract.
- develop upper and lower bounds for European and American call and put options.
- explain what is meant by put-call parity.

6.1.4 Show how to use binomial trees and lattices in valuing options and solve simple examples.

6.1.5 Derive the risk-neutral pricing measure for a binomial lattice and describe the risk-neutral pricing approach to the pricing of equity options.

6.1.6 Explain the difference between the real-world measure and the risk-neutral measure. Explain why the risk-neutral pricing approach is seen as a computational tool (rather than a realistic representation of price dynamics in the real world).

6.1.7 State the alternative names for the risk-neutral and state-price deflator approaches to pricing.

6.1.8 Demonstrate an understanding of the Black-Scholes derivative-pricing model:

- explain what is meant by a complete market.
- explain what is meant by risk-neutral pricing and the equivalent martingale measure.
- derive the Black-Scholes partial differential equation both in its basic and Garman-Kohlhagen forms.
- demonstrate how to price and hedge a simple derivative contract using the martingale approach.

6.1.9 Show how to use the Black-Scholes model in valuing options and solve simple examples.

6.1.10 Discuss the validity of the assumptions underlying the Black-Scholes model.

6.1.11 Describe and apply in simple models, including the binomial model and the Black-Scholes model, the approach to pricing using deflators and demonstrate its equivalence to the risk-neutral pricing approach.

6.1.12 Demonstrate an awareness of the commonly used terminology for the first, and where appropriate second, partial derivatives (the Greeks) of an option price.

Core Reading

The Subject CM2 Course Notes include the Core Reading in full, integrated throughout the course.

Further reading

The exam will be based on the relevant Syllabus and Core Reading and the ActEd course material will be the main source of tuition for students.

2.3 Subject CM2 – the course structure

There are four parts to the Subject CM2 course. The parts cover related topics and have broadly equal marks in the paper-based exam. The parts are broken down into chapters.

The following table shows how the parts, the chapters and the syllabus items relate to each other. The end columns show how the chapters relate to the days of the regular tutorials. We have also given you a broad indication of the length of each chapter. This table should help you plan your progress across the study session.

Part	***Chapter***	***Title***	***No of pages***	***Syllabus objectives***	***4 full days***
1	1	The Efficient Markets Hypothesis	23	1.1	1
	2	Utility theory	48	1.2	
	3	Stochastic dominance and behavioural finance	28	1.2, 1.3	
	4	Measures of investment risk	33	2.1, 2.2	
	5	Stochastic models of investment returns	33	3.1	
	6	Portfolio theory	40	4.1	
2	7	Models of asset returns	30	4.3	2
	8	Asset pricing models	30	4.2	
	9	Brownian motion and martingales	29	4.4	
	10	Stochastic calculus and Ito processes	55	4.4	
	11	Stochastic models of security prices	17	4.4	
	12	Characteristics of derivative securities	48	6.1	
3	13	The Greeks	18	6.1	3
	14	The binomial model	66	6.1	
	15	The Black-Scholes option pricing formula	42	6.1	
	16	The 5-step method in discrete time	37	6.1	
	17	The 5-step method in continuous time	51	6.1	
4	18	The term structure of interest rates	49	4.5	4
	19	Credit risk	44	4.6	
	20	Ruin theory	82	5.1	
	21	Run-off triangles	72	5.2	

2.4 Subject CM2 – summary of ActEd products

The following products are available for Subject CM2:

- Course Notes
- PBOR (including the Y Assignments)
- X Assignments – four assignments:
  - X1, X2: 80-mark tests (you are allowed 2¾ hours to complete these)
  - X3, X4: 100-mark tests (you are allowed 3¼ hours to complete these)
- Series X Marking
- Series Y Marking
- Online Classroom – over 150 tutorial units
- Flashcards
- Revision Notes
- ASET – four years' exam papers, *ie* eight papers, covering the period April 2014 to September 2017
- Mock Exam
- Mock Exam Marking
- Marking Vouchers.

We will endeavour to release as much material as possible but unfortunately some revision products may not be available until the September 2019 or even April 2020 exam sessions. Please check the ActEd website or email ActEd@bpp.com for more information.

The following tutorials are typically available for Subject CM2:

- Regular Tutorials (four days)
- Block Tutorials (four days)
- a Preparation Day for the computer-based exam.

Full details are set out in our *Tuition Bulletin*, which is available on our website at **www.ActEd.co.uk**.

2.5 Subject CM2 – skills and assessment

Technical skills

The *Actuarial Mathematics* subjects (Subjects CM1 and CM2) are very mathematical and have relatively few questions requiring wordy answers.

Exam skills

Exam question skill levels

In the CM subjects, the approximate split of assessment across the three skill types is:

- Knowledge – 20%
- Application – 65%
- Higher Order skills – 15%.

Assessment

Combination of a computer-based modelling assignment and a three-hour written examination.

2.6 Subject CM2 – frequently asked questions

Q: ***What knowledge of earlier subjects should I have?***

A: The Course Notes are written on the assumption that students have studied Subjects CM1, CS1 and CS2. Most students find CM2 quite a tough course and so a good grasp of the material in the earlier subjects is essential. Some of the material in CB2 is also relevant.

Q: ***What level of mathematics is required?***

A: Some of the maths required for this subject is quite advanced – up to degree standard. The techniques covered in Subjects CM1 and CS1 will be treated as assumed knowledge and the theory will build on these. You will find the course much easier if you feel comfortable with the mathematical techniques used in these earlier subjects and you can apply them confidently.

Q: ***What should I do if I discover an error in the course?***

A: If you find an error in the course, please check our website at:

www.ActEd.co.uk/paper_corrections.html

to see if the correction has already been dealt with. Otherwise please send details via email to **CM2@bpp.com**.

Q: ***Who should I send feedback to?***

A: We are always happy to receive feedback from students, particularly details concerning any errors, contradictions or unclear statements in the courses.

If you have any comments on this course in general, please email to **CM2@bpp.com**.

If you have any comments or concerns about the Syllabus or Core Reading, these can be passed on to the profession via ActEd. Alternatively, you can send them directly to the Institute and Faculty of Actuaries' Examination Team by email to **education.services@actuaries.org.uk**.

1

The Efficient Markets Hypothesis

Syllabus objectives

1.1 Rational expectations theory

1.1.1 Discuss the three forms of the Efficient Markets Hypothesis and their consequences for investment management.

1.1.2 Describe briefly the evidence for or against each form of the Efficient Markets Hypothesis.

0 Introduction

In simple terms, an *efficient* security market is one in which the price of every security fully reflects all available information and hence is equal to its 'true' investment value. According to the Efficient Markets Hypothesis security markets *are* efficient.

This basic idea has been extended to allow for different forms of market efficiency corresponding to different levels of information. These are:

- *weak form* – market prices reflect all of the information contained in historical price data
- *semi-strong form* – market prices reflect all publicly available information
- *strong form* – market prices reflect all information, whether or not it is publicly available.

The importance of market efficiency derives from the fact that if markets are *inefficient* then investors with better information may be able to obtain higher investment returns. If, however, markets are *efficient*, then it is not possible to identify under- or over-priced securities, which can then be traded to generate excess risk-adjusted returns. Hence, it is not worth trying to do so.

This chapter therefore:

- describes the three different definitions of market efficiency
- discusses the evidence for and against the different forms of the Efficient Markets Hypothesis – which turns out to be largely inconclusive.

1 Rational expectations theory

1.1 Background

From the 1930s until the early 1960s, there was a widespread folklore about how to make money on the stock market. The dominant theory, going back to Adam Smith in the 1700s, was that markets are essentially fickle, and that prices tend to oscillate around some true or fundamental value.

Starting with the seminal work by Benjamin Graham, traditional investment analysis involved detailed scrutiny of company accounts, to calculate fundamental values, and thus ascertain when a given investment is cheap or expensive. The objective would be to buy cheap stocks and sell expensive ones. Any excess performance thus obtained would be at the expense of irrational traders, who bought and sold on emotional grounds (a 'gut feeling', for example) and without the benefit of detailed analysis.

This detailed analysis is known as *fundamental analysis*.

By the 1960s, it had become clear that these supposedly foolproof methods of investment were not working. Strategies based on detailed analysis did not seem to perform any better than simple buy-and-hold strategies. Attempts to explain this phenomenon gave rise to the *Efficient Markets Hypothesis* (EMH), which claims that market prices already incorporate the relevant information. The market price mechanism is such that the active trading patterns of a small number of informed analysts can lead to accurate market prices. Uninformed (or 'cost-conscious', since actively trading incurs potentially unnecessary costs) investors can then take a free ride, in the knowledge that the research of others is keeping the market efficient.

This provides a strong argument in favour of the *passive* investment management style that we discuss below.

1.2 The three forms of the Efficient Markets Hypothesis

The academic literature has distinguished between different forms of the Efficient Markets Hypothesis, based on a finer dissection of exactly what constitutes *relevant information*. In particular, the following three forms of EMH are commonly distinguished:

Weak form EMH

The market price of an investment incorporates all information contained in the price history of that investment. Knowledge of a stock's price history cannot produce excess performance as this information is already incorporated in the market price. This form, if true, means that *technical analysis* (or chartism) techniques (*ie* analysing charts of prices and spotting patterns) will not produce excess performance.

Semi-strong form EMH

The market price of an investment incorporates all publicly available information. Knowledge of any public information cannot produce excess performance, as this information is already incorporated in the market price.

This form, if true, means that *fundamental analysis* techniques (*ie* analysing accounting statements and other pieces of financial information) will not produce excess performance.

Fundamental analysis uses information concerning the issuer of the security (*eg* turnover, profitability, liquidity, level of gearing) and general economic and investment conditions (*eg* real interest rates and inflation) in order to determine the 'true' or 'fundamental' value of a security and hence whether or not it is cheap or expensive.

Different stock exchanges have different levels of required disclosure of information. Hence it would be reasonable to expect different markets to have different levels of efficiency. For example, the New York Stock Exchange (NYSE), which requires a high level of disclosure, should be more efficient than a market with limited disclosure requirements.

There is also no commonly accepted definition of what constitutes publicly available information.

This can lead to problems when testing whether a particular market is actually efficient or not.

For example, unlike professional fund managers, private investors are unable to gain access to the senior management of companies.

Clearly, fund managers have an advantage in terms of being able to form an opinion on the competence of the management team and the strategy of the company. Fund managers are also increasingly utilising 'alternative data' (*eg* satellite images, web searches, social media *etc*) to generate excess performance.

Even if information is publicly available, there is a cost involved in obtaining the information quickly and accurately. Any advantage achieved by acting on price relevant information could well be eroded by the cost of obtaining and analysing that information.

In other words, the cost of obtaining additional information could outweigh the additional returns that it might generate. Note that a necessary requirement for efficiency as it has been defined above is that the costs of both acquiring the relevant information and trading on the basis of it should be equal to zero. This must be the case if investors are to trade until security prices do reflect *all* available information.

Note that just because information is publicly available, it does not mean that everybody has read and understood it, *eg* the contents of the Subject CM2 course. This could be because:

- for many people – who do not wish to be actuaries or investment specialists – the costs of buying a Subject CM2 course outweigh the benefits that it confers
- most of the population is unaware of either the existence or the benefits to be derived from studying Subject CM2, or both!

Strong form EMH

The market price of an investment incorporates all information, both publicly available and also that available only to insiders. Knowledge available only to insiders cannot produce excess performance as this information is already incorporated in market prices.

Stock markets around the world are subject to regulation. Often rules exist to prevent individuals with access to price sensitive information, which is not yet public, from using this information for personal gain. For example, senior management involved in merger and acquisition talks are often banned from trading in the stock of their company. Such rules would be unnecessary if strong form efficiency held.

Question

Explain why such rules would be unnecessary.

Solution

Such rules would be unnecessary because it would not be possible for senior management to use this information to obtain higher investment returns by trading in the stock of their own company. Thus, senior management would not be at an advantage compared to other investors, who would correspondingly not be disadvantaged by such trades.

However, one can argue that if senior management were allowed to trade their own company's stock, then strong form EMH would be possible. Hence the existence of these rules prevents strong EMH from occurring.

Trading on the basis of privileged information that is not publicly available is sometimes known as *insider trading* or *insider dealing*. If insider trading does not occur, then the strong form Efficient Markets Hypothesis cannot hold, as there is then no mechanism by which security prices can incorporate inside information.

Question

What is the relationship between the three forms of market efficiency?

Solution

Publicly available information is a subset of all information, whether publicly available or not. Consequently strong form efficiency implies semi-strong form efficiency, in the sense that if a market is strong form efficient, then it must also be semi-strong form efficient.

Similarly as historical price data is a subset of all publicly available information, so a market that is semi-strong form efficient must also be weak form efficient.

Active versus passive investment management

The question of whether or not markets are efficient has important implications for investment management. Active fund managers attempt to detect exploitable mispricings, since they believe that markets are not universally efficient. Passive fund managers simply aim to diversify across a whole market, perhaps because they do not believe they have the ability to spot mispricings.

According to the Efficient Markets Hypothesis, active investment management, with its active trading policy and consequent higher level of management fees, cannot be justified.

Question

Why can active management not be justified according to the EMH?

Solution

According to the Efficient Markets Hypothesis, active investment management cannot be justified because it is impossible to exploit the mispricing of securities in order to generate higher expected returns. Even if price anomalies exist, then the costs of identifying them and then trading will outweigh the benefits arising from the additional investment returns.

If active investment management cannot be justified, then a more appropriate investment strategy might be simply to match or 'track' the market by holding a portfolio whose performance will closely replicate that of the market as a whole. In this way the fund should yield approximately the same level of investment returns as the market, whilst also enjoying the benefits arising from both diversification and lower dealing costs. In practice this is often achieved by matching or tracking an index that is representative of the investment market in question. Such index tracking is a very important example of a *passive* investment management style.

If markets are inefficient, we would expect active managers with above average skill to perform better than passive managers. However, performance should be considered *net* of various fees and transaction costs (*eg* brokerage, market impact). To demonstrate an exploitable opportunity in the market, the opportunity should be sufficiently large to remain intact even after all these costs are taken into account.

An alternative definition of efficiency sometimes used is therefore that prices reflect all available information up to the point at which the marginal costs and benefits of that information are equal. If these marginal costs and benefits differ between investors, some investors may enjoy an advantage over their peers.

A further consequence is that those investment markets with the most freely available information and the lowest transaction costs are likely to be the most efficient. Thus, government bond markets tend to be more efficient than property markets.

The question of market efficiency therefore has a crucial bearing upon the choice of investment management style.

Question

Comment on the advantages that could be derived from 'insider trading' in a market that is strong form efficient.

Solution

If the market is strong form efficient, then there will be no advantage from insider trading because all the insider knowledge should be reflected in the current share price.

2 The evidence for or against each form of the Efficient Markets Hypothesis

2.1 Difficulties with testing the Efficient Markets Hypothesis

Tests of EMH are fraught with difficulty. There is a substantial body of literature proving the existence of mispricings, in contravention of EMH. Unfortunately, there is also a substantial body of literature providing evidence for various forms of EMH. Both schools of thought can cite a great deal of empirical evidence and an impressive wealth of statistical tests. It is reasonable to ask, from a philosophical point of view, how it could come about that we have categorical proof of mutually contradictory statements. One possible explanation is that many published tests make implicit and explicit, but possibly invalid, assumptions (for example, normality of returns, or stationarity of time series).

Consequently, a test that appears to disprove the Efficient Markets Hypothesis may actually be disproving something else.

We can note that whilst an apparent proof based upon historical data over one period of time might be valid for that particular period, it might not be valid for a subsequent time period, perhaps because the nature of the market or the available information has changed. We can also note that the parties involved in providing proof will have vested interests and may therefore be biased, publishing only those results that support their position.

Some of the differences are purely differences of terminology. For example, do we regard anomalies as disproving EMH, if transaction costs prevent their exploitation?

Thus, although it may in principle be possible to exploit temporary mispricings, it may not be possible in practice after appropriate allowance has been made for both transaction costs and the costs of obtaining information. Whether or not such a finding contradicts the EMH depends upon exactly how we define the EMH.

More subtle is the need to make an appropriate allowance for risk. The EMH is not contradicted by a strategy which produces higher profits than the market portfolio by taking higher risks. The market rewards investors for taking risks, so we expect, on average, a high-risk strategy to result in higher returns.

What would contradict the EMH is an investment strategy that provided returns over and above those necessary to compensate an investor for the risk they faced. Unfortunately, there is no universally agreed definition of risk, and no perfectly accurate way of measuring it.

We will consider a number of different measures of investment risk later in the course.

With these caveats in mind, we can now consider some empirical work.

Testing the weak form EMH

Using price history to try and forecast future prices, often using charts of historical data, is called *technical* or *chartist* analysis. Studies have failed to identify a difference between the returns on stocks using technical analysis and those from purely random stock selection after allowing for transaction costs. Credible challenges to Weak Form EMH took a long time to emerge, but recent econometric evidence suggests that stocks tend to exhibit short-run momentum (trending in the same direction) and medium-run mean-reversion (trending in opposite directions).

Question

Country X runs a national lottery in which the purchaser of a ticket selects six different numbers from 1 to 50 inclusive. If those same six numbers are then drawn randomly from a hat on live TV, the holder of the ticket wins a share of a large cash sum equal in value to the total ticket sales. Is the market for lottery tickets weak form efficient?

Solution

If the numbers drawn are truly random, then the market for lottery tickets is weak form efficient. This is because knowledge of the numbers that have been drawn in the past will not help you to predict the numbers that are likely to be drawn in the future, and thereby generate excess returns. It will also be semi-strong and strong form efficient, unless it is operated fraudulently.

Testing the semi-strong form EMH

The semi-strong form of the EMH has been where research has concentrated and where the debate is most fierce. We will consider tests of the EMH in two categories below: tests of informational efficiency and volatility tests.

Testing the strong form EMH

This is problematic as it requires the researcher to have access to information that is not in the public domain.

In order to decide if security prices do reflect all available information we ourselves need to have access to all available information – including information that is not publicly available.

However, studies of directors' share dealings suggest that, even with inside information, it is difficult to out-perform.

2.2 Informational efficiency

The EMH (in its various forms) states that asset prices reflect information. However, it does not explicitly tell us how new information affects prices.

For example, the speed and extent to which it does so.

It is also empirically difficult to establish precisely when information arrives. For example, many events are widely rumoured prior to official announcements.

An example here is a merger. Should tests of efficiency be based upon the official announcement date of the merger or the date at which rumours concerning the likelihood of the merger first started – or possibly some date in between?

Many studies show that the market over-reacts to certain events and under-reacts to other events. The over/under-reaction is corrected over a long time period. If this is true then traders could take advantage of the slow correction of the market, and efficiency would not hold.

Over-reaction to events

Some of the effects found by studies can be classified as over-reaction to events, for example:

1. **Past performance: past winners tend to be future losers and vice versa. The market appears to over-react to past performance.**

   Hence it might be possible to make excess profits by selling shares in firms that have performed well recently and buying those that have performed badly. This is sometimes referred to as a *contra-cyclical* investment policy.

2. **Certain accounting ratios appear to have predictive powers: *eg* companies with high earnings to price, cashflow to price and book value to market value (generally poor past performers) tend to have high future returns. Again, this is an example of the market apparently over-reacting to past growth.**

3. **Firms coming to the market: evidence from a number of major financial markets including the UK and the US appears to support the idea that stocks coming to the market by Initial Public Offerings and Seasoned Equity Offerings have poor subsequent long-term performance.**

Under-reaction to events

There are also well-documented examples of under-reaction to events:

1. **Stock prices continuing to respond to earnings announcements up to a year after their announcement. This is an example of under-reaction to information which is slowly corrected.**

2. **Abnormal excess returns for both the parent and subsidiary firms following a de-merger. This is another example of the market being slow to recognise the benefits of an event.**

3. **Abnormal negative returns following mergers (agreed takeovers leading to the poorest subsequent returns). The market appears to over-estimate the benefits from mergers and the stock price slowly reacts as the optimistic view is proved to be wrong.**

Anomalies

All these effects are often referred to as 'anomalies' in the EMH framework.

A fast-growing area of research in finance is Behavioural Finance, which investigates whether such anomalies arise due to behaviour of individual investors which departs from that predicted by models based on rational expectations. However, this approach is still controversial in some circles, with some academics unconvinced that 'irrational' behaviour is an important determinant of aggregate asset pricing.

Even if the market is efficient, pure chance is going to throw up some apparent examples of mispricings. We would expect to see as many examples of over-reaction as under-reaction. This is broadly consistent with the literature to date.

The 2007/09 financial crisis has led to many professors asking whether EMH (and other techniques discussed in CM2) should even form part of the syllabus. We should be critical of the theories and treat them as a structuring tool, a theoretical base rather than a dogma.

Even more important is the finding that the reported effects do not appear to persist over prolonged time periods and so may not represent exploitable opportunities to make excess profits. For example, the 'small companies effect' received attention in the early 1980s. This work showed the out-performance of small companies in the period 1960-80. However, if a strategy based on this evidence were to be implemented throughout the 1980s and early 1990s, the investor would have experienced abnormally low returns.

Other examples of anomalies, for example the ability of accounting ratios to indicate out-performance, are arguably proxies for risk (strategies exploiting these strategies are higher-risk than average). Once these risks have been taken into account, many studies, which claim to show evidence of inefficiency, turn out to be compatible with the EMH.

Question

Over the last five years, the shares of Company A have yielded an average investment return equal to twice that of Company B. Does this contradict the Efficient Markets Hypothesis?

Solution

Although Company A's shares have recently yielded more than Company B's shares, this does not contradict the Efficient Markets Hypothesis. This is because the EMH implies that it is not possible to identify shares that offer excess *risk-adjusted expected* returns. This is different from the situation described, which refers to actual *past* returns with no allowance being made for the relative riskiness of the two shares involved. Thus Company A may be inherently more risky than Company B.

2.3 Volatility tests

Several observers have commented that stock prices are 'excessively volatile'. By this they mean that the change in market value of stocks (observed volatility), could not be justified by the presence of news. This was claimed to be evidence of market over-reaction which was not compatible with efficiency.

Excessive volatility therefore arises when security prices are more volatile than the underlying fundamental variables that should be driving them.

The claim of 'excessive volatility' was first formulated into a testable proposition by Shiller in 1981. He considered a discounted cashflow model of equities going back to 1870. By using the actual dividends that were paid and some terminal value for the stock, he was able to calculate the *perfect foresight price*, the 'correct' equity price, if market participants had been able to predict future dividends correctly. The difference between the perfect foresight price and the actual price arises from the forecast errors of future dividends. If market participants are rational, we would expect no *systematic* forecast errors. Also if markets are efficient, broad movements in the perfect foresight price should be correlated with moves in the actual price as both react to the same news.

Shiller found strong evidence that the observed level of volatility in the S&P 500 stock index contradicted the EMH as such volatility was not in line with the subsequent fluctuations in the dividends.

In other words, Shiller found that actual security prices were more volatile than perfect foresight prices based upon the present value of future dividends.

However, subsequent studies, using different formulations of the problem, found that the violation of the EMH only had borderline statistical significance. Numerous criticisms were subsequently made of Shiller's methodology.

These criticisms covered:

- **the choice of terminal value for the stock price**
- **the use of a constant discount rate**
- **bias in estimates of the variances due to autocorrelation**
- **possible non-stationarity of the series, *ie* the series may have stochastic trends which invalidate the measurements obtained for the variance of the stock price.**

Although subsequent studies by many authors have attempted to overcome the shortcomings in Shiller's original work, there still remains the problem that a model for dividends and distributional assumptions are required. Some equilibrium models now exist which calibrate both to observed price volatility and observed dividend behaviour. However, the vast literature on volatility tests can at best be described as inconclusive.

2.4 Conclusion

The literature on testing the EMH is vast, and articles can be found to support whatever view you wish to take. It is possible to find research claiming incontrovertible evidence either for or against the EMH.

Question

An investment market is strong form efficient. Describe what would happen to the price of a company's shares if some positive information about the company becomes known. (Assume that nobody had known about this information in advance.)

Solution

1. The share price should go up.
2. This should happen immediately.
3. The share price should rise without bias, *ie* the market does not over-react or under-react.

Note that the answer to this question illustrates some of the ways that stock markets tend not to be fully efficient, *ie* information is not fed into share prices immediately and without bias.

Question

Is the following statement true or false?

'The semi-strong form of the Efficient Markets Hypothesis suggests that no investor will 'beat' the market in the long term.'

Solution

The laws of probability suggest that some investors will achieve returns in excess of the market even over the long term purely by chance. For example, they might happen to be holding a particular company's shares when some 'good' news is announced. However, the Efficient Markets Hypothesis suggests that no one will be able to do so systematically unless:

- they accept a higher level of risk than exhibited by the market as a whole, or
- they have inside information.

The chapter summary starts on the next page so that you can keep all the chapter summaries together for revision purposes.

Chapter 1 Summary

The Efficient Markets Hypothesis (EMH)

In an *efficient* security market the price of every security fully reflects all available and relevant information. The EMH states that security markets *are* efficient.

Three forms of the EMH are commonly distinguished:

1. *Weak form* – market prices incorporate all of the information contained in historical price data. If markets are weak form efficient, then technical analysis cannot be used to generate excess risk-adjusted returns.

2 *Semi-strong form* – market prices incorporate all publicly available information. If markets are semi-strong form efficient, then fundamental analysis cannot be used to generate excess risk-adjusted returns.

3. *Strong form* – market prices incorporate all information, whether or not it is publicly available. If markets are strong form efficient, then insider trading cannot be used to generate excess risk-adjusted returns.

In practice the level of efficiency depends on whether information is freely available, which in turn may depend on the level of disclosure required by regulation.

The importance of market efficiency derives from the fact that if markets are *inefficient* then investors with better information may be able to generate higher investment returns. If, however, they are efficient then active investment management is difficult to justify.

Tests of the EMH

Tests of the EMH are fraught with difficulty. Consequently, the empirical evidence is inconclusive concerning the extent to which security markets are in fact efficient in practice. However:

- studies of directors' share dealings suggest that, even with inside information, it is difficult to out-perform
- studies have failed to identify a difference between the returns on stocks selected using technical analysis and those from purely random stock selection
- research has concentrated on the semi-strong form of the EMH and in particular tests of informational efficiency and volatility tests.

Informational efficiency

Many studies show that the market over-reacts to certain events and under-reacts to other events. The over/under-reaction is corrected over a long time period. If this is true then traders could take advantage of the slow correction of the market, and efficiency would not hold.

Over-reaction to events

- Past winners tend to be future losers and the market appears to over-react to past performance.
- Certain accounting ratios appear to have predictive powers, an example of the market apparently over-reacting to past growth.
- Firms coming to the market have poor subsequent performance.

Under-reaction to events

- Stock prices continue to respond to earnings announcements up to a year after their announcement.
- Abnormal excess returns for both the parent and subsidiary firms following a de-merger.
- Abnormal negative returns following mergers.

Volatility tests

Shiller first formulated the claim of 'excessive volatility' into a testable proposition in 1981. He found strong evidence that the observed level of volatility contradicted the EMH. However, subsequent studies using different formulations of the problem found that the violation of the EMH only had borderline statistical significance. Numerous criticisms were subsequently made of Shiller's methodology. These criticisms covered:

- the choice of terminal value for the stock price
- the use of a constant discount rate
- bias in estimates of the variances due to autocorrelation
- possible non-stationarity of the series, *ie* the series may have stochastic trends that invalidate the measurements obtained for the variance of the stock price.

Chapter 1 Practice Questions

1.1 A researcher has analysed the annual returns of equity stocks in a particular country over a 10-year period and has made the following observations:

Exam style

(a) Annual market returns in consecutive years have a negative correlation of –0.25.

(b) The closing value of the index of the 100 stocks with the highest market capitalisation has been found to be 1% higher on average on Fridays than on Mondays.

(c) Announcements of changes in companies' dividend policies typically take three months to become fully reflected in the quoted share price.

(d) The prices of a particular subset of stocks have been consistently observed to fall immediately following a favourable announcement and to rise immediately following an unfavourable announcement.

Discussthese observations in the light of the EMH. [4]

1.2 Discuss the following statement:

Exam style

The existence of fund managers who sell their services based on their alleged ability to select over-performing sectors and stocks and so add value to portfolios demonstrates that capital markets are not efficient. [6]

1.3 (i) Describe what is meant by an 'efficient market'.

(ii) Describe the three different forms of the Efficient Markets Hypothesis.

(iii) Discuss the implications of the Efficient Markets Hypothesis.

1.4 At the quarterly meeting of the Auger Close Investment Club, four members are making proposals for new equity investment for the club.

Exam style

Anna wants to buy shares in Armadillo Adventures, claiming that they have performed poorly in recent weeks and are due an upturn.

Brian wants to invest in Biscuits-R-Us. They have recruited a new head of marketing, who has had success at other companies. Brian feels that this new appointment will have a positive effect on the firm.

Cathy selects shares at random. This quarter she is recommending the club buy into Cash 4 Kidneys PLC.

Dennis wants the club to buy shares in Diamond Dentists ('DD'). His brother works for a major health insurer and has insider information that DD's shares will rise sharply in the near future, when it is announced that his company has appointed DD as its 'dentist of choice'.

For each club member, describe how their share selection strategy would work in strongly efficient, semi-strongly efficient, weakly efficient and inefficient markets. [7]

1.5 (i) Explain what is meant by an 'excessively volatile' market. [2]

Exam style

(ii) Describe how you would test if a market is 'excessively volatile'. [7]

(iii) Explain the practical and conceptual difficulties in using a test of an excessively volatile market to establish whether or not a market is efficient. [4]

[Total 13]

1.6 (i) Explain the implications of the Efficient Markets Hypothesis for investment trading strategies.

(ii) Explain why investors will still wish to have as much information as possible concerning a company and its securities before investing in it even if the Efficient Markets Hypothesis applies.

Chapter 1 Solutions

1.1 (a) ***Annual market returns are negatively correlated***

This observation suggests that, over annual time periods, the market tends to systematically overreact to new information and hence that the market may not be semi-strong form efficient. [½]

In addition, trading rules could be developed based on this information that could generate excess, risk-adjusted returns, which suggests that this observation is inconsistent with the weak form of the EMH. [½]

(b) ***The index is higher on Fridays than on Mondays***

The observation suggests that there is a consistent tendency for prices on Fridays to be 'inflated', while prices on Mondays are 'depressed', *ie* there is a systematic bias present in the prices. [½]

Trading rules could be developed based on this information (*eg* buy on Monday, sell on Friday) that could generate excess, risk-adjusted returns, which suggests that this observation is inconsistent with the weak form of the EMH. [½]

(c) ***Announcements take three months to be reflected***

If the semi-strong form of the EMH holds, public dividend announcements should have an immediate effect on the share prices as the market should respond quickly and accurately to new information. [½]

This observation suggests that the market is not semi-strong form efficient. [½]

(d) ***Prices fall following a favourable announcement***

The prices are reacting when information is made public. This suggests that the prices have previously been distorted by insider information. [½]

Therefore, this observation contradicts the strong form of the EMH. [½]

[Total 4]

Note that, once a particular form of the EMH is contradicted, this also contradicts any of the stronger forms.

1.2 The Efficient Markets Hypothesis (EMH) suggests that it is not possible to achieve excess risk adjusted investment returns using investment strategies based only on certain subsets of information. The existence of fund managers who sell their services based on their alleged ability to select over-performing sectors and stocks does not demonstrate that capital markets are inefficient. [1]

In particular, the semi-strong form of the EMH suggests that excess risk-adjusted investment returns cannot be obtained using only publicly available information. [½]

In certain investment markets, it may therefore be possible (and legal) to achieve excess returns using privileged or inside information, which would not contradict the semi-strong form of the EMH. [½]

More generally, the EMH does not preclude managers achieving higher investment returns by adopting 'riskier' investment strategies and receiving due reward for the risks taken. [½]

It says precisely that it is not possible to develop investment strategies that yield excess risk-adjusted returns – though it is difficult to determine exactly how risk should be interpreted in this context. [½]

Some fund managers must necessarily achieve higher than average returns over a given short time period – *eg* several years. The point of the EMH is that managers cannot consistently achieve above excess returns. Moreover, they cannot guarantee to achieve excess returns over any particular time period. [1]

Finally, rather than reflecting any market inefficiency in contradiction of the EMH, the existence of such managers may instead reflect the following facts:

- Individual investors may be unaware of the EMH or choose not believe it and hence may be inclined to believe the claims of such managers and so place money with them. [½]
- Certain individual investors may choose to believe the claims of such managers, reflecting the fact that investment decisions are often made on the basis of subjective and emotional factors, in addition to, or instead of, on the basis of financial theory. [1]

For the above reasons, the existence of such fund managers does not therefore demonstrate that capital markets are inefficient. [½]

[Total 6]

1.3 (i) ***Definition of efficient market***

An efficient market is one in which every security's price equals its investment value at all times.

In an efficient market information is fully reflected in the price.

This means that share prices adjust instantaneously and without bias to new information.

(ii) ***Three forms of Efficient Markets Hypothesis***

The strong form requires that prices reflect all information that is currently known – whether or not it is publicly available.

The semi-strong form requires that prices reflect all information that is publicly available.

The weak form requires that prices fully reflect all information contained in the past history of prices.

(iii) ***Implications of the Efficient Markets Hypothesis***

The past history of prices is a subset of publicly available information, so a market must be weak form efficient if it is semi-strong form efficient. Similarly, if it is strong form efficient it must also be semi-strong and weak form efficient.

The Efficient Markets Hypothesis does not imply that beating the market is impossible, since investors could out-perform the market by chance, or by accepting above average levels of risk.

However, it does imply that it is not possible consistently to achieve superior risk-adjusted investment performance net of costs without access to superior information.

Weak form efficiency implies that it is impossible to achieve excess risk-adjusted investment returns purely by using trading rules based upon the past history of prices and trading volumes. It therefore suggests that technical analysis cannot be justified.

If only weak form efficiency applies, excess risk-adjusted returns are still possible by good fundamental analysis of public information.

The semi-strong form means that prices adjust instantaneously and without bias to newly published information. This implies that it is not possible to trade profitably on information gained from public sources. So neither fundamental analysis (without insider information) nor technical analysis will yield excess risk-adjusted returns.

Fundamental analysis may still, however, aid investors in selecting the investments that are most suitable for meeting their investment needs and objectives.

If the strong form is correct then the market reflects all known knowledge about the company and consequently excess risk-adjusted returns are possible only by chance. This implies that insiders cannot profit from dealing on inside information, *ie* insider trading is not profitable.

1.4 ***Anna***

Anna makes her recommendation based on the past price history of the investment. If weak form EMH holds, then the current share price already reflects the information contained in the past price history, so there would be no advantage in using this approach. [1]

Similarly, if the semi-strong or strong form of EMH holds, there is no advantage in using this approach. [½]

If the market was inefficient, Anna's strategy may be beneficial. [½]

Brian

Brian makes his recommendation based on company information that is in the public domain. If semi-strong form EMH holds, then the current share price already reflects relevant public information, so there would be no advantage in using this approach. [1]

Similarly, if the strong form of EMH holds, there is no advantage in using this approach. [½]

If the market is inefficient or only weak form efficient, Brian's strategy may be beneficial. [½]

Cathy

The approach of choosing stocks at random provides no advantage, whatever the level of market efficiency. [½]

If strong form EMH holds, this strategy is no worse than any other. [½]

Dennis

Dennis makes his recommendation based on insider information. If strong form EMH holds, then the current share price already reflects all relevant information, so there would be no advantage in using this approach. [1]

If the market is inefficient or weak or semi-strong form efficient, Dennis's strategy may be beneficial (though it could be questionable on ethical grounds). [1]

[Total 7]

1.5 (i) ***Excessively volatile markets***

An excessively volatile market is one in which the changes in the market values of stocks (the observed volatility) are greater than can be justified by the news arriving. This is claimed to be evidence of market over-reaction, which is not compatible with efficiency. [2]

(ii) ***Testing if a market is excessively volatile***

To test if a market is excessively volatile you need a long history of prices and cashflows for one of the securities in question – *eg* for the market in a particular equity, you would need many months or years of share prices and dividend payments. [1½]

A discounted cashflow model based on the actual dividends that were paid and some terminal value for the share could then be used to calculate a perfect foresight price for the equity. This would represent the 'correct' equity price if market participants had been able to predict future dividends correctly. [1½]

The difference between the perfect foresight price and the actual price arises from the forecast errors of future dividends. If market participants are rational, there should be no systematic forecast errors. [1½]

Also if markets are efficient, then broad movements in the perfect foresight price should be correlated with moves in the actual price as both are reacting to the same news and hence the same changes in the anticipated future cashflows. [1½]

If instead the actual price changes are greater, then this would suggest that the market in the particular equity is excessively volatile. [1]

[Total 7]

This was the approach adopted by Shiller.

(iii) ***Practical and conceptual difficulties***

These include:

- the difficulty of choosing an appropriate terminal value for the share price [1]
- the difficulty of choosing an appropriate discount rate at which to discount future cashflows – in particular, should it be constant? [1]
- possible biases in the estimates of the variances because of autocorrelation in the time series data used [1]
- possible non-stationarity of the time series data used, *ie* it may have stochastic trends which invalidate the measurements obtained for the variance of the stock price [1]
- the distributional assumptions underlying the statistical tests used might not be satisfied [½]
- the distributional characteristics of the share prices and dividends are unlikely to remain constant over a long period of time. [½]

[Maximum 4]

1.6 (i) ***Implications of the Efficient Markets Hypothesis***

The Efficient Markets Hypothesis implies that it is impossible, except by chance, to make abnormal profits using trading strategies that are based on only past share prices (weak form), publicly available information (semi-strong form) or any information (strong form).

In practice, however, the definition has sometimes been refined to preclude the possibility of systematically higher returns after allowing for transaction costs.

Market efficiency also implies that active investment management (which aims to enhance returns by identifying under- or over-priced securities) cannot be justified and consequently provides a rationale for passive investment management strategies, such as index tracking.

(ii) ***Information***

Even if markets are efficient, investors will still wish to have as much information as possible concerning a company and its securities in order to identify the characteristics of the shares, *eg* the volatility of returns, risk, income and capital growth *etc.* An appreciation of these will enable investors to make an informed decision whether or not to hold the security as part of a portfolio designed to meet their investment objectives.

2

Utility theory

Syllabus objectives

1.2 Rational choice theory

1. Explain the meaning of the term 'utility function'.
2. Explain the axioms underlying utility theory and the expected utility theorem.
3. Explain how the following economic characteristics of investors can be expressed mathematically in a utility function:
   - non-satiation
   - risk aversion, risk-neutrality and risk-seeking
   - declining or increasing absolute and relative risk aversion.
4. Discuss the economic properties of commonly used utility functions.
5. Discuss how a utility function may depend on current wealth and discuss state-dependent utility functions.
6. Perform calculations using commonly used utility functions to compare investment opportunities.
8. Analyse simple insurance problems in terms of utility theory.

0 Introduction

This chapter focuses on utility theory as applied to investment choices.

In economics, 'utility' is the satisfaction that an individual obtains from a particular course of action.

In Section 1 we introduce utility functions and the *expected utility theorem*. This provides a means by which to model the way individuals make investment choices.

Section 2 describes the properties that are normally considered desirable in utility functions to ensure that they reflect the actual behaviour of investors. Chief amongst these are:

- *non-satiation*, a preference for more over less, and
- *risk aversion*, a dislike of risk.

These ideas underlie the rest of the course.

Section 3 considers various methods of measuring risk aversion and the way in which risk aversion might vary with wealth. The concepts of *absolute risk aversion* and *relative risk aversion* are discussed.

Section 4 introduces some commonly used examples of utility functions, namely the *quadratic*, *log*, and *power* utility functions, and discusses the properties of each.

Section 5 describes how to deal with situations in which a single utility function is inappropriate. In such instances, it may be necessary to vary either the parameters or the form of the utility function according to the particular situation to be modelled. This leads to the idea of *state-dependent* utility functions.

In order to use the expected utility theory, we need an explicit utility function. In Section 6 we look at how we might go about constructing such a utility function.

Section 7 then uses these utility functions to solve problems involving insurance premiums. In particular, determining the maximum premium a policyholder is willing to pay and the minimum amount an insurer should charge.

Expected utility theory can be useful, but it is not without problems. In Section 8, we therefore consider the limitations of the expected utility theory for risk management purposes – in particular, the need to know the precise form and shape of the individual's utility function.

1 Utility theory

1.1 Introduction

In this section we use utility theory to consider situations that involve uncertainty, as will normally be the case where investment choices are concerned.

Uncertainty

In what follows, we assume any asset that yields uncertain outcomes or returns, *ie* any *risky* asset, can be characterised as a set of objectively known probabilities defined on a set of possible outcomes. For example, Equity A might offer a return to Investor X of either -4 % or +8% in the next time period, with respective probabilities of ¼ and ¾.

Question

Each year, Mr A is offered the opportunity to invest £1,000 in a risk fund. If successful, at the end of the year he will be given back £2,000. If unsuccessful, he will be given back only £500. There is a 50% chance of either outcome. Calculate the expected rate of return per annum on the investment.

Solution

We can calculate the expected rate of return as follows:

$$\frac{(0.5\times 2{,}000)+(0.5\times 500)}{1{,}000}-1=25\%$$

Given the uncertainty involved, the rational investor cannot maximise utility with complete certainty. We shall see that the rational investor will instead attempt to maximise *expected* utility by choosing between the available risky assets.

Utility functions

In the application of utility theory to finance and investment choice, it is assumed that a numerical value called the *utility* can be assigned to each possible value of the investor's wealth by what is known as a 'preference function' or 'utility function'.

Utility functions show the level of utility associated with different levels of wealth. For example, Investor X might have a utility function of the form:

$$U(w) = \log(w)$$

where w is the current or future wealth.

1.2 The expected utility theorem

Introduction

The theorem has two parts.

1. **The expected utility theorem states that a function, $U(w)$, can be constructed as representing an investor's utility of wealth, w, at some future date.**

2. **Decisions are made in a manner to maximise the expected value of utility given the investor's particular beliefs about the probability of different outcomes.**

In situations of uncertainty it is impossible to maximise utility with complete certainty. For example, suppose that Investor X invests a proportion a of his wealth in Equity A and places the rest in a non-interest-bearing bank account. Then his wealth in the next period cannot be predicted with complete certainty and hence neither can his utility.

It is possible, however, to say what his *expected* wealth equals. Likewise if the functional form of $U(w)$ is known, then it is possible to calculate his *expected* utility. The expected utility theorem says that when making a choice an individual should choose the course of action that yields the highest expected *utility* – and *not* the course of action that yields the highest expected wealth, which will usually be different.

Question

Derive an expression for the expectation of Investor X's next-period wealth if he invests a proportion a of his current wealth w in Equity A (which pays –4% or +8%, with respective probabilities ¼ and ¾) and the rest in a non-interest-bearing bank account.

Solution

$$
\begin{aligned}
E(w) &= (1-a)w + aw[0.25\times 0.96 + 0.75\times 1.08] \\
&= (1-a)w + 1.05aw \\
&= (1+0.05a)w
\end{aligned}
$$

The answer can also be arrived at directly by noting that the expected next-period wealth will be the initial wealth w, plus the expected return of 5% on the investment aw.

Calculating the expected utility

Suppose a risky asset has a set of N possible outcomes for wealth $(w_1,\ldots,w_N)$, each with associated probabilities of occurring of $(p_1,\ldots,p_N)$, then the *expected utility* yielded by investment in this risky asset is given by:

$$E(U) = \sum_{i=1}^{N} p_i\, U(w_i)$$

So the expected utility is a weighted average of the utilities associated with each possible individual outcome.

Question

State an expression for the expectation of the next-period utility of Investor X, again assuming that he invests a proportion a in Equity A and the rest in a non-interest-bearing bank account. He has the utility function $U(w) = \log(w)$.

Solution

$$E\left[U\left(w\right)\right]=0.25\left\{\log\left((1-0.04a)w\right)\right\}+0.75\left\{\log\left((1+0.08a)w\right)\right\}$$

Note that a *risk-free asset* is a special case of a risky asset that has a probability of one associated with the *certain* outcome, and zero probability associated with all other outcomes.

By combining an investor's beliefs about the set of available assets with a suitable utility function, we can determine the optimal investment portfolio for the investor, *ie* that which maximises expected utility in that period.

Question

Investor A has an initial wealth of $100, which is currently invested in a non-interest-bearing account, and a utility function of the form:

$$U(w)=\log(w)$$

where w is the investor's wealth at any time.

Investment Z offers a return of -18% or +20% with equal probability.

(i) What is Investor A's expected utility if nothing is invested in Investment Z?

(ii) What is Investor A's expected utility if they're entirely invested in Investment Z?

(iii) What proportion a of wealth should be invested in Investment Z to maximise expected utility? What is Investor A's expected utility if they invest this proportion in Investment Z?

Solution

(i) If nothing is invested in Investment Z, the expected utility is:

$$\log(100) = 4.605$$

(ii) If Investor A is entirely invested in Investment Z, the expected utility is:

$$0.5 \times \log(0.82 \times 100) + 0.5 \times \log(1.2 \times 100) = 4.597$$

(iii) If a proportion a of wealth is invested in Investment Z, the expected utility is given by:

$$E\left[U(w)\right]=0.5\left\{\log\left(100(1-0.18a)\right)\right\}+0.5\left\{\log\left(100(1+0.2a)\right)\right\}$$

$$=0.5\left\{\log\left(100-18a\right)\right\}+0.5\left\{\log\left(100+20a\right)\right\}$$

We differentiate with respect to a to find a maximum:

$$\frac{dE\left[U(w)\right]}{da}=0.5\times\frac{-18}{100-18a}+0.5\times\frac{20}{100+20a}$$

$$=\frac{-9}{100-18a}+\frac{10}{100+20a}$$

We then set equal to zero:

$$\frac{9}{100-18a}=\frac{10}{100+20a}$$

Solving, we find $a=0.2777$.

Checking to see if this gives a maximum:

$$\frac{d^2E\left[U(w)\right]}{da^2}=\frac{+9(-18)}{(100-18a)^2}+\frac{-10(20)}{(100+20a)^2}$$

This gives a negative value so it is a maximum.

Finding the expected utility from investing 27.77% in Investment Z:

$$E\left[U(w)\right]=0.5\left\{\log\left(100-18\times0.2777\right)\right\}+0.5\left\{\log\left(100+20\times0.2777\right)\right\}$$

$$=4.6066$$

1.3 Derivation of the expected utility theorem

The expected utility theorem can be derived formally from the following four axioms.

In other words, an investor whose behaviour is consistent with these axioms will always make decisions in accordance with the expected utility theorem.

1. Comparability

An investor can state a preference between all available certain outcomes.

In other words, for any two certain outcomes A and B, either:

A is preferred to B,

B is preferred to A,

or the investor is indifferent between A and B.

These preferences are sometimes denoted by:

$$U(\text{A}) > U(\text{B}),\ U(\text{B}) > U(\text{A}) \text{ and } U(\text{A}) = U(\text{B})$$

Note that A and B are examples of what we previously referred to as w_i.

2. Transitivity

If A is preferred to B and B is preferred to C, then A is preferred to C.

ie $\quad U(\text{A}) > U(\text{B}) \text{ and } U(\text{B}) > U(\text{C}) \quad \Rightarrow U(\text{A}) > U(\text{C})$

Also:

$$U(\text{A}) = U(\text{B}) \text{ and } U(\text{B}) = U(\text{C}) \quad \Rightarrow U(\text{A}) = U(\text{C})$$

This implies that investors are consistent in their rankings of outcomes.

3. Independence

If an investor is indifferent between two certain outcomes, A and B, then they are also indifferent between the following two gambles:

(i) A with probability *p* and C with probability (1 – *p*)

(ii) B with probability *p* and C with probability (1 – *p*).

Hence, if $U(\text{A}) = U(\text{B})$ (and of course $U(\text{C})$ is equal to itself), then:

$$p\,U(\text{A}) + (1-p)\,U(\text{C}) = p\,U(\text{B}) + (1-p)\,U(\text{C})$$

Thus, the choice between any two certain outcomes is independent of all other certain outcomes.

4. Certainty equivalence

Suppose that A is preferred to B and B is preferred to C. Then there is a unique probability, *p*, such that the investor is indifferent between B and a gamble giving A with probability *p* and C with probability (1 – *p*).

Thus if:

$$U(\text{A}) > U(\text{B}) > U(\text{C})$$

Then there exists a unique p ($0 < p < 1$) such that:

$$pU(A) + (1-p)U(C) = U(B)$$

B is known as the 'certainty equivalent' of the above gamble.

It represents the certain outcome or level of wealth that yields the same certain utility as the expected utility yielded by the gamble or lottery involving outcomes A and C.

The four axioms listed above are not the only possible set of axioms, but they are the most commonly used.

Question

Suppose that an investor is asked to choose between various pairs of strategies and responds as follows:

Choose between:	*Response*
B and D	B
A and D	D
C and D	indifferent
B and E	B
A and C	C
D and E	indifferent

Assuming that the investor's preferences satisfy the four axioms discussed above, how does the investor rank the five investments A to E?

Solution

From the responses we can note immediately that:

B > D, D > A, C = D, B > E, C > A, D = E

Hence, transitivity then implies that:

B > D > A
C = D = E

And so we have that: B > C = D = E > A

2 The expression of economic characteristics in terms of utility functions

In mainstream finance theory, investors' preferences are assumed to be influenced by their attitude to risk. We need to consider, therefore, how an investor's risk-return preference can be described by the form of their utility function.

The mathematical form of utility functions is normally assumed to satisfy desirable properties that accord with everyday observation about how individuals typically act in the face of uncertainty.

2.1 Non-satiation

As a basis to understanding risk attitudes, let us first assume that people prefer more wealth to less. This is known as the principle of non-satiation and can be expressed as:

$$U'(w) > 0$$

This is clearly analogous to individuals preferring more to less of a good or service in the standard choice between different bundles of goods in situations of certainty.

The derivative of utility with respect to wealth is often referred to as the *marginal utility of wealth*. Non-satiation is therefore equivalent to an assumption that the marginal utility of wealth is strictly positive.

2.2 Risk aversion

Attitudes to risk can now be expressed in terms of the properties of utility functions.

In particular, we can choose the form of the utility function that we use to model an individual's preferences according to whether or not the individual concerned likes, dislikes or is indifferent to risk.

Risk-averse investor

A *risk-averse* investor values an incremental increase in wealth less highly than an incremental decrease and will reject a fair gamble.

A *fair gamble* is one that leaves the expected wealth of the individual unchanged. Equivalently, it can be defined as a gamble that has an overall expected value of zero.

Question

Suppose that an unbiased coin is tossed once. Determine the fairness of a gamble in which you receive $1 if it lands heads up but lose $1 if it lands tails up.

Solution

The gamble is fair because your expected gain from accepting the gamble is zero and your expected wealth remains unchanged (though your actual wealth will of course change by $1).

Equivalently, the overall expected value of the gamble is equal to:

$$\tfrac{1}{2} \times (+\$1) + \tfrac{1}{2} \times (-\$1) = 0$$

Risk-averse investors derive less additional utility from the prospect of a possible gain than they lose from the prospect of an identical loss with the same probability of occurrence. Consequently they will not accept a fair gamble. They may, however, be willing to trade off lower expected wealth in return for a reduction in the variability of wealth. This is the basic principle underlying insurance.

It is normally assumed that investors are risk-averse and consequently that they will accept additional risk from an investment only if it is associated with a higher level of expected return. Hence, the importance of the trade-off between risk and return that is a feature of most investment decisions.

For a risk-averse investor, the utility function condition is:

$$\boldsymbol{U''(w) < 0}$$

In other words, for a risk-averse investor, utility is a (strictly) concave function of wealth, as shown in Figure 2.1.

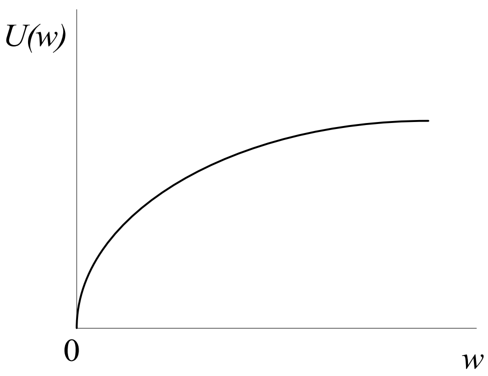


Figure 2.1 – A concave utility function for a risk-averse investor

This concavity condition means that the marginal utility of wealth (strictly) *decreases* with the level of wealth and consequently each additional dollar, say, adds less satisfaction to the investor than the previous one.

Question

Suppose that a risk-averse investor with wealth w is faced with the gamble described in the previous question. Show that this investor will derive less additional utility from the possible gain than that lost from the possible loss and hence that risk aversion is consistent with the condition $U''(w) < 0$.

Solution

The investor's certain utility if the gamble is rejected is $U(w)$. The investor's expected utility obtained by accepting the gamble is given by:

$$E(U) = ½U(w-1) + ½U(w+1)$$

The gamble is therefore rejected if:

$$½\,U(w-1) + ½\,U(w+1) \; < \; U(w)$$

$$\Leftrightarrow \qquad U(w-1) + U(w+1) \; < \; 2\,U(w)$$

$$\Leftrightarrow \qquad U(w+1) - U(w) \; < \; U(w) - U(w-1)$$

ie if the additional utility from winning the gamble is less than the loss of utility from losing the gamble. This will be the case if $U''(w) < 0$, which is by definition the case for a risk-averse investor.

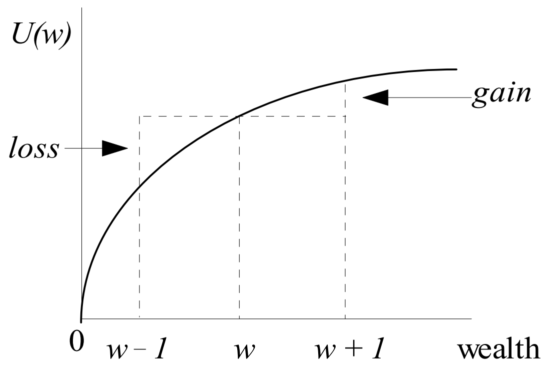


Figure 2.2 – Losses and gains in utility

Risk-seeking investor

A *risk-seeking* investor values an incremental increase in wealth more highly than an incremental decrease and will seek a fair gamble. The utility function condition is:

$$\boldsymbol{U''(w) > 0}$$

A risk-seeking or risk-loving investor will accept any fair gamble and may even accept some unfair gambles (that reduce expected wealth) because the potential increase in utility resulting from the possible gain exceeds the potential decrease in utility associated with the corresponding loss.

Question

What is the shape of the utility function of a risk-seeking investor?

Solution

A risk-seeking investor has a convex utility function, because $U''(w) > 0$. The utility function looks as follows:

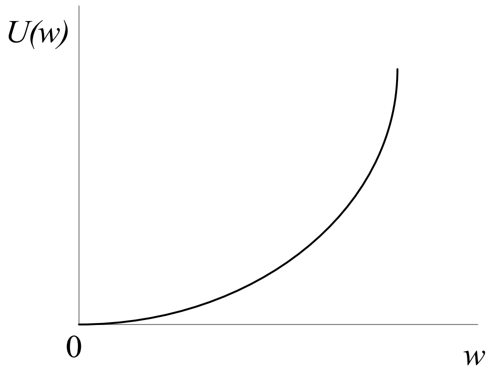


Figure 2.3 – A convex utility function

Risk-neutral investor

A *risk-neutral* investor is indifferent between a fair gamble and the *status quo*. In this case:

$$U''(w) = 0$$

Question

What can we say about the marginal utility of wealth of a risk-neutral investor?

Solution

For a risk-neutral investor, $U''(w) = 0$. Thus, $U'(w)$ is constant and so the marginal utility of wealth must itself be constant (and positive assuming non-satiation), so that each additional \$1 leads to the same change in utility, regardless of wealth.

The utility function of a risk-neutral investor is a *linear* function of wealth. Assuming non satiation, then $U'(w) > 0$ and so $U(w)$ increases with w for all w. Thus, the maximisation of expected utility is equivalent to the maximisation of expected *wealth*, in the sense that it will always lead to the same choices being made.

Question

(i) Ignoring any pleasure derived from gambling, a risk-averse person will:

A never gamble
B accept fair gambles
C accept fair gambles and some gambles with an expected loss
D none of the above

(ii) Ignoring any pleasure derived from gambling, a risk-neutral person will:

I always accept fair gambles
II always accept unfair gambles
III always accept better than fair gambles

A I and II are true
B II and III are true
C I only is true
D III only is true

(iii) A risk-loving person will:

I always accept a gamble
II always accept unfair gambles
III always accept fair gambles

A I and II are true
B II and III are true
C I only is true
D III only is true

Solution

(i) *A risk-averse person will not accept fair gambles*. However, they might accept a gamble where they expected, on average, to win. This would happen if the expected profit from gambling was sufficient to compensate them for taking on the risk. Therefore the correct answer is D.

(ii) *A risk-neutral person will be indifferent to accepting a fair gamble*, but will accept better than fair gambles. Therefore statement III is always true. Statement II is false. Statement I is false (we can't be certain that a person who is indifferent to the gamble will always accept it). The answer is thus D.

(iii) *A risk-loving person will be happy to accept fair gambles*. A risk-loving person will also accept *some* unfair gambles. However, if the odds are very unfair, even a risk-loving person will not accept a gamble. Thus the correct answer is D.

Note that this solution shows how risk aversion/neutrality/loving can be defined in terms of a person's attitude towards a fair gamble.

3 Measuring risk aversion

3.1 Introduction

In practice, we normally assume that an investor is risk-averse and by looking at the sign of $U''(w)$ we can deduce whether or not this is in fact the case.

The way risk aversion changes with wealth may also be of interest.

The degree of risk aversion is likely to vary with the investor's existing level of wealth. For example, we might imagine that wealthy investors are less concerned about risk.

3.2 Risk aversion and the certainty equivalent

Consider the certainty equivalent of a gamble. For a risk-averse individual this is higher than the actual likelihood of the outcome, *ie* the individual would need to receive odds higher than expected to accept this gamble.

Alternative definitions of the certainty equivalent

Note that we can distinguish between two different types of certainty equivalents depending upon the situation that we are considering, namely:

- The certainty equivalent of the portfolio consisting of the combination of the existing wealth *w and* the gamble *x*, which we can denote c_w.
- The certainty equivalent of the gamble *x* alone, c_x, which will also depend upon the existing level of wealth.

Thus, for a fair gamble and a risk-averse investor, it must be the case that $c_w < w$ and $c_x < 0$.

1. 'Additive' or 'absolute' gamble

Consider a gamble with outcomes represented by a random variable *x*, in which the sums won or lost are *fixed absolute amounts*. The actual sums won or lost are therefore independent of the value of initial wealth *w*. If the investor accepts the gamble, the resulting total wealth is $w + x$.

The certainty equivalent of the combined portfolio of initial wealth plus gamble, c_w, is then defined as the certain level of wealth that solves:

$$U(c_w) = E\left[U\left(w+x\right)\right]$$

The certainty equivalent of the gamble itself is equal to:

$$c_x = c_w - w$$

because we require that $U(w+c_x) = U(c_w)$ and U is a strictly increasing function.

c_x is negative for a fair gamble and its absolute value represents the maximum sum that the risk-averse investor would pay to avoid the risk.

Question

Suppose that an unbiased coin is tossed once, and a gamble exists in which an investor receives \$1 if it lands heads up but loses \$1 if it lands tails up. Further assume that:

- the investor has initial wealth of \$10 and
- a utility function of the form $U(w) = \sqrt{w}$.

Determine the investor's certainty equivalent for this gamble.

Solution

With an initial wealth of 10, the expected utility of total wealth is given by:

$$½ \times \left[\sqrt{11} + \sqrt{9}\right] = 3.1583$$

The certainty equivalent, c_w, of the initial wealth plus the gamble satisfies:

$$U(c_w) = 3.1583$$

$$\Rightarrow \qquad \sqrt{c_w} = 3.1583 \qquad \Rightarrow \qquad c_w = 3.1583^2 = 9.9749$$

Hence, the certainty equivalent of the gamble itself is:

$$c_x = c_w - w = 9.9749 - 10 = -0.0251$$

Note that c_x is negative. This means we would have to pay the investor to accept the gamble. Equivalently, the investor would be prepared to pay 0.0251 to avoid the gamble.

2. 'Multiplicative' or 'proportional' gamble

This is a gamble, with outcomes represented by a random variable *y*, in which the sums won or lost are all expressed as *proportions* of the initial wealth. If the investor accepts the gamble they therefore end up with a final wealth of $w \times y$. For example, in a fair gamble of this type, the investor might win 15% of their initial wealth (y = 1.15) with probability ¼ and lose 5% (y = 0.95) with a probability of ¾. Note that in this case, the actual sums won or lost therefore depend directly upon the value of initial wealth *w*, *ie* a larger *w* produces larger wins or losses.

In this instance, the certainty equivalent of total wealth including the proceeds of the gamble can be defined as the level of wealth that satisfies:

$$U(c_w) = E\left[U(w \times y)\right]$$

The certainty equivalent of the gamble alone is defined as before and is again negative for a fair gamble for a risk-averse investor.

The certainty equivalent and absolute risk aversion

If the absolute value of the certainty equivalent decreases with increasing wealth, the investor is said to exhibit declining *absolute* risk aversion. If the absolute value of the certainty equivalent increases, the investor exhibits increasing *absolute* risk aversion.

Here we have in mind:

- an *additive* gamble
- the certainty equivalent of the gamble alone, c_x .

If the investor's preferences exhibit decreasing (increasing) absolute risk aversion (ARA), then the absolute value of c_x decreases (increases) and the investor is prepared to pay a smaller (larger) absolute amount in order to avoid the risk associated with the gamble.

ie increasing / decreasing ARA $\Leftrightarrow$ increasing / decreasing $|c_x|$

The certainty equivalent and relative risk aversion

If the absolute value of the certainty equivalent decreases (increases) as a *proportion* of total wealth as wealth increases the investor is said to exhibit declining (increasing) *relative* risk aversion.

Here we are looking at:

- a multiplicative gamble
- the certainty equivalent of the gamble alone as a proportion of initial wealth, *ie* $\dfrac{c_x}{w}$.

If the investor's preferences exhibit decreasing (increasing) relative risk aversion (RRA), then the absolute value of c_x/w decreases (increases) with wealth *w.*

ie increasing / decreasing RRA $\Leftrightarrow$ increasing / decreasing $\left|\dfrac{c_x}{w}\right|$

Example

We shall see below that the log utility function $U(w) = \log(w)$ exhibits:

- decreasing absolute risk aversion
- constant relative risk aversion.

To see these results, consider an individual with an initial wealth of \$100, who faces a fair gamble that offers an equal chance of winning or losing \$20. In this case:

$$\begin{aligned} U(c_w) &= E\left[U(w+x)\right] \\ &= \frac{1}{2}\left[\log 120 + \log 80\right] \\ &= 4.5848 \end{aligned}$$

$\Rightarrow \quad c_w = e^{4.5848} = 97.980$ and $c_x = 97.980 - 100 = -2.020$

If instead the individual's initial wealth is \$200, then:

$$U(c_w) = \frac{1}{2}\left[\log 220 + \log 180\right] = 5.2933$$

$\Rightarrow \quad c_w = e^{5.2933} = 198.997$

and: $\quad c_x = 198.997 - 200 = -1.003$

The absolute value of c_x, the certainty equivalent of the (fair) gamble alone, has decreased with wealth (for a gamble with *fixed absolute proceeds*), as is the case with decreasing absolute risk aversion.

Let us now consider the case of a multiplicative gamble. Suppose the individual is offered an equal chance of winning or losing 20% of their initial wealth. If the initial wealth is \$100, then the investor could win or lose \$20. This is equivalent to our first example. We have seen that $c_w = 97.980$ and that $c_x = -2.020$. We can also find that:

$$\frac{c_x}{w} = \frac{-2.020}{100} = -0.0202$$

If the initial wealth is \$200, the investor could win or lose 20%, *ie* \$40. Then:

$$U(c_w) = \frac{1}{2}\left[\log 240 + \log 160\right] = 5.2779$$

$\Rightarrow \quad c_w = e^{5.2779} = 195.959$

And: $\quad c_x = 195.959 - 200 = -4.041$

Thus:

$$\frac{c_x}{w} = \frac{-4.041}{200} = -0.0202$$

ie the absolute value of c_x/w, the certainty equivalent of the (fair) gamble as a proportion of initial wealth, is invariant to wealth (for a gamble with *fixed percentage proceeds*), corresponding to constant relative risk aversion.

3.3 Risk aversion and the utility function

Absolute and relative risk aversion can be expressed in terms of the utility function as follows.

Absolute risk aversion is measured by the function

$$A(w) = \frac{-U''(w)}{U'(w)}$$

Relative risk aversion is measured by the function

$$R(w) = -w\frac{U''(w)}{U'(w)}$$

These are often referred to as the *Arrow-Pratt* measures of absolute risk aversion and relative risk aversion.

The above results concerning the relationship between the certainty equivalent and the measures of risk aversion arise because it can be shown that the:

- absolute value of the certainty equivalent of a fair gamble is proportional to $\dfrac{-U''(w)}{U'(w)}$
- absolute value of the certainty equivalent of a fair gamble expressed as a proportion of the investor's wealth is proportional to $-w\dfrac{U''(w)}{U'(w)}$.

The following table shows the relationships between the first derivatives of the above functions and declining or increasing absolute and relative risk aversion.

	Absolute risk aversion	***Relative risk aversion***
Increasing	$A'(w) > 0$	$R'(w) > 0$
Constant	$A'(w) = 0$	$R'(w) = 0$
Decreasing	$A'(w) < 0$	$R'(w) < 0$

3.4 Risk aversion and the investment choice

The way that risk aversion changes with wealth can be expressed in terms of the amount of wealth held as risky assets.

Investors who hold an increasing *absolute* amount of wealth in risky assets as they get wealthier exhibit declining *absolute* risk aversion. Investors who hold an increasing *proportion* of their wealth in risky assets as they get wealthier exhibit declining *relative* risk aversion.

In practice, it is often assumed that as wealth increases, so the absolute amount that a typical investor is willing to invest in risky assets will increase, *ie* that absolute risk aversion decreases with wealth.

It is not so clear cut as to whether we would expect the *proportion* of risky assets to increase or decrease. Consequently the assumption of constant relative risk aversion is sometimes made.

4 Some commonly used utility functions

4.1 The quadratic utility function

The general form of the quadratic utility function is

$$U(w) = a + bw + cw^2$$

Since adding a constant to a utility function or multiplying it by a constant will not affect the decision-making process, we can write the general form simply as:

$$U(w) = w + dw^2$$

Thus:

$$U'(w) = 1 + 2dw$$

and $$U''(w) = 2d$$

Therefore, if the quadratic utility function is to satisfy the condition of diminishing marginal utility of wealth (risk aversion), we must have $d < 0$.

The consequence of this is that the quadratic utility function can only satisfy the condition of non-satiation over a limited range of w:

$$-\infty < w < -\frac{1}{2d}$$

This constraint on the range of possible values for w is a significant limitation of using quadratic utility functions.

The absolute and relative risk aversion measures are given by:

$$A(w) = \frac{-U''(w)}{U'(w)} = \frac{-2d}{1+2dw}$$

$$A'(w) = \frac{4d^2}{(1+2dw)^2} > 0$$

and:

$$R(w) = w\frac{-U''(w)}{U'(w)} = \frac{-2dw}{1+2dw}$$

$$R'(w) = \frac{-2d}{1+2dw} + \frac{4d^2w}{(1+2dw)^2} = \frac{-2d}{(1+2dw)^2} > 0$$

Thus the quadratic utility function exhibits both increasing absolute and relative risk aversion.

Question

Draw the quadratic utility function over the range $0 \le w < -1/2d$ and show why it is valid only for $w < -1/2d$, for a non-satiated risk-averse investor.

Solution

For non-satiation we require:

$$U'(w) > 0, \ \ ie\ 1 + 2dw > 0$$

$$\Leftrightarrow \qquad 2dw > -1$$

$$\Leftrightarrow \qquad dw > -\tfrac{1}{2}$$

$$\Leftrightarrow \qquad w < -1/2d, \text{ as } d < 0 \text{ for a risk-averse investor.}$$

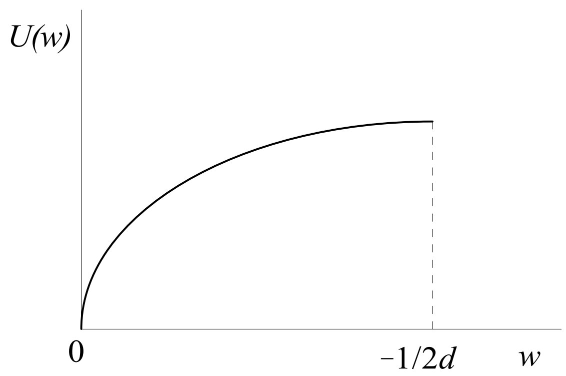


Figure 2.4 – The range of the quadratic utility function

4.2 The log utility function

The form of the log utility function is:

$$U(w) = \ln(w) \qquad\qquad (w > 0)$$

Thus:

$$U'(w) = \frac{1}{w}$$

and:

$$U''(w) = -\frac{1}{w^2}$$

Thus the log utility function satisfies the principle of non-satiation and diminishing marginal utility of wealth.

This is because we have assumed that the log utility function is defined only for positive values of w and $U'(w) = \frac{1}{w} > 0$, *ie* non-satiation and $U''(w) = \frac{-1}{w^2} < 0$, *ie* diminishing marginal utility of wealth, for $w > 0$.

The absolute and relative risk aversion measures are given by:

$$A(w) = \frac{-U''(w)}{U'(w)} = \frac{1}{w}$$

$$A'(w) = -\frac{1}{w^2} < 0$$

and:

$$R(w) = w\frac{-U''(w)}{U'(w)} = 1$$

$$R'(w) = 0$$

Thus the log utility function exhibits declining absolute risk aversion and constant relative risk aversion. This is consistent with an investor who keeps a constant proportion of wealth invested in risky assets as they get richer.

This investor will also invest an increasing *absolute* amount of wealth in risky assets.

Utility functions exhibiting constant relative risk aversion are said to be 'iso-elastic'.

Iso-elastic means that the elasticity of the marginal utility of wealth is constant with respect to wealth.

The use of iso-elastic utility functions simplifies the determination of an optimal strategy for a multi-period investment decision, because it allows for a series of so-called 'myopic' decisions. What this means is that the decision at the start of each period only considers the possible outcomes at the end of that period and ignores subsequent periods.

Thus, the individual's utility maximisation choice in each period is independent of all subsequent periods. The decision is said to be 'myopic' because it is short-sighted, *ie* it does not need to look to future periods.

4.3 The power utility function

The form of the power utility function is:

$$U(w) = \frac{w^{\gamma} - 1}{\gamma} \qquad (w > 0)$$

Thus:

$$U'(w) = w^{\gamma-1}$$

and:

$$U''(w) = (\gamma-1)w^{\gamma-2}$$

Thus for the power utility function to satisfy the principle of non-satiation and diminishing marginal utility of wealth we require $\gamma < 1$.

The absolute and relative risk aversion measures are given by:

$$A(w) = \frac{-U''(w)}{U'(w)} = -\frac{(\gamma-1)}{w}$$

$$A'(w) = \frac{(\gamma-1)}{w^2} < 0$$

and:

$$R(w) = w\frac{-U''(w)}{U'(w)} = -(\gamma-1)$$

$$R'(w) = 0$$

Thus, like the log utility function, the power utility function exhibits declining absolute risk aversion and constant relative risk aversion.

It is therefore also iso-elastic.

The power utility function, in the form given above, is one of a wider class of commonly used functions known as HARA (hyperbolic absolute risk aversion) functions. γ is the risk aversion coefficient.

This is because, for such functions, the absolute risk aversion is a hyperbolic function of wealth *w*. For example, in the case of the log utility function:

$$w \times A(w) = \text{constant}$$

Hence, a plot of *A(w)* against *w* describes a rectangular hyperbola.

Question

Suppose Investor A has a *power* utility function with $\gamma = 1$, whilst Investor B has a power utility function with $\gamma = 0.5$.

(i) Which investor is more risk-averse (assuming that $w > 0$)?

(ii) Suppose that Investor B has an initial wealth of 100 and is offered the opportunity to buy Investment X for 100, which offers an equal chance of a payout of 110 or 92. Will the Investor B choose to buy Investment X?

Solution

(i) ***Which investor is more risk-averse?***

Investor B is more risk-averse because they have a lower risk aversion coefficient γ. We can show this by deriving the absolute risk aversion and relative risk aversion measures for each investor.

For Investor A:

$$A(w) = R(w) = 0$$

ie Investor A is *risk-neutral*.

For Investor B:

$$A(w) = \frac{1}{2w} > 0, \quad R(w) = \frac{1}{2} > 0$$

Hence, Investor B is *strictly risk-averse* for all $w > 0$.

(ii) ***Will Investor B buy Investment X?***

If Investor B buys X, then they will enjoy an expected utility of:

$$0.5\left[2\left(\sqrt{110}-1\right)+2\left(\sqrt{92}-1\right)\right] = 18.08$$

If, however, they do not buy X, then their expected (and certain) utility is:

$$2\left(\sqrt{100}-1\right) = 18$$

Thus, as buying X yields a higher expected utility, the investor ought to buy it.

4.4 Other utility functions

As evidenced from the above, many different utility functions have appeared in literature whose role is to describe the manner in which an investor derives utility from given choices. None of the utility functions described above allows much freedom in calibrating the function used to reflect a particular investor's preferences.

Question

Consider the following utility function:

$$U(w) = -e^{-aw}, \quad a>0$$

Derive expressions for the absolute risk aversion and relative risk aversion measures. What does the latter indicate about the investor's desire to hold risky assets?

Solution

The utility function $U(w) = -e^{-aw}$ is such that:

$$U'(w) = ae^{-aw} \quad \text{and} \quad U''(w) = -a^2 e^{-aw}$$

Thus:

$$A(w) = \frac{-U''(w)}{U'(w)} = a > 0 \quad \text{and} \quad A'(w) = 0$$

and:

$$R(w) = \frac{-wU''(w)}{U'(w)} = aw > 0 \quad \text{and} \quad R'(w) = a > 0$$

Hence, as the absolute risk aversion is constant and independent of wealth the investor must hold the same *absolute* amount of wealth in risky assets as wealth increases. Both this, and the fact that the relative risk aversion increases with wealth, are consistent with a *decreasing proportion* of wealth being held in risky assets as wealth increases.

5 The variation of utility functions with wealth

5.1 Introduction

A further extension of the utility function is to consider wealth. It may not be possible to model an investor's behaviour over all possible levels of wealth with a single utility function. An obvious example is the quadratic utility function described above, which only satisfies the non-satiation condition over a limited wealth range. Sometimes this problem can be dealt with by using utility functions with the same functional form but different parameters over different ranges of wealth.

For example, the power utility function could be used to model preferences over all wealth levels, but with the value of the risk aversion coefficient γ changing with wealth. Sometimes, however, it may be necessary to go even further and use different functional forms over different ranges – by constructing *state-dependent utility functions*.

5.2 State-dependent utility functions

State-dependent utility functions can be used to model the situation where there is a discontinuous change in the state of the investor at a certain level of wealth.

They reflect the reality that the usefulness of a good or service to an individual, including wealth, may vary according to the circumstances of the individual. For example, the value of an umbrella depends upon whether or not we believe that it is going to rain over the next few hours or days. In a similar way, the utility that we derive from wealth may also reflect both our existing financial state and our more general circumstances in a way that cannot be captured by a simple functional form. We may therefore need to model preferences using a sophisticated utility function constructed by combining one or more of the standard functions discussed above – so that a different utility function effectively applies over different levels of wealth. A utility function of this kind may involve discontinuities and/or kinks.

Such a situation arises when we consider an insurance company that will become insolvent if the value of its assets falls below a certain level. At asset levels just above the insolvency position, the company will be highly risk-averse and this can be modelled by a utility function that has a discontinuity at the insolvency point.

However, the consequence of applying the same utility function when the company has just become insolvent would be that the company would be prepared to accept a high probability of losing its remaining assets for a chance of regaining solvency.

In other words, at the point of becoming (technically) insolvent the company is very risk-averse, being very keen to avoid this happening. Should it become (technically) insolvent, however, given that the damage (to its reputation or otherwise) has already been done, it may then be willing to take more risks in order to regain solvent status.

This is unlikely to reflect reality and so a different form of utility function would be required to model the company's behaviour in this state.

Question

Draw the utility function of the above company.

Solution

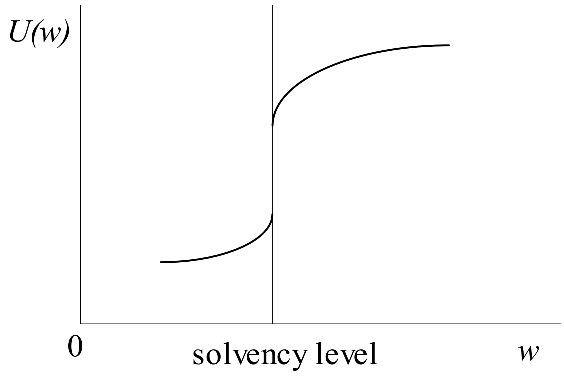


Figure 2.5 – A state-dependent utility function

ie the company is extremely risk-averse when just solvent, so the curve has a rapidly changing gradient. At the solvency level the curve is vertical as any slight increase in wealth leads to a large jump in utility.

Utility functions can also depend on states other than those, such as insolvency, which are determined by the level of wealth. Obvious examples for an individual include the differences between being healthy or sick, married or single. The state of an individual can also be affected by the anticipation of future events, *eg* if a legacy is expected.

Thus, under some circumstances an individual's utility might accurately be described by a function of the form $U(h,w)$ where h is an indicator of health.

6 Construction of utility functions

6.1 Introduction

In order to use a particular utility function, we need to calibrate the function so that it is appropriate to the particular individual to whom it applies. In other words, we need to find the values of the parameters, for example the value of γ in the power utility function, that apply to the individual.

One approach that has been proposed is to devise a series of questions that allow the shape of an individual's utility function to be roughly determined. A utility curve of a predetermined functional form can then be fitted by a least squares method to the points determined by the answers to the questions. The curve fitting is constrained by the requirement that the function has the desired economic properties (non-satiation, risk aversion and, perhaps, declining absolute risk aversion).

The student may be expected to show how a utility function can be constructed in general when there is a discrete set of outcomes and the axioms of this chapter apply.

6.2 Construction of utility functions by direct questioning

In theory, to determine an individual's utility function we could simply ask the individual what it is. However, in practice it is most unlikely that someone will be able to describe the mathematical form of their utility function.

6.3 Construction of utility functions by indirect questioning

An alternative procedure involves firstly, fixing two values of the utility function for the two extremes of wealth being considered. Secondly, the individual is asked to identify a certain level of wealth such that he or she would be indifferent between that certain level of wealth and a gamble that yields either of the two extremes with particular probabilities. The process is repeated for various scenarios until a sufficient number of plots is found.

Example

Suppose that we wish to determine the nature of an individual's utility function over the range of wealth $0 < w < 4$. One possible approach is to first fix $U(0) = 0$ and $U(4) = 1$. These are the first two points on the individual's utility function.

We could then ask the individual to identify the certain level of wealth, w' such that they would be indifferent between w' for certain and a gamble that yields each of 0 and 4 with equal probability, *ie* w' is the certainty equivalent of the gamble. The expected utility of the gamble is:

$$E[U]=\frac{1}{2}\left[U(0)+U(4)\right]=\frac{1}{2}[0+1]=\frac{1}{2}$$

If $w'=1.8$ say, then we know that $U(1.8)=0.5$ and thus have a third point on the utility function.

We could then repeat the exercise for a gamble involving equal probabilities of producing 1.8 and 4, which yields an expected utility of:

$$E[U]=\frac{1}{2}\left[U(1.8)+U(4)\right]=\frac{1}{2}\left[0.5+1\right]=0.75$$

If the certainty equivalent of this gamble is, say, 2.88, then we know that $U(2.88) = 0.75$, giving us a fourth point on the individual's utility function.

This process can be repeated until a sufficient number of points along the individual's utility curve have been identified and a plot of those points produced. Ordinary least squares regression or maximum likelihood methods can then be used to fit an appropriate functional form to the resulting set of values.

Another form of indirect questioning uses information on the premiums that a person is prepared to pay in order to gain an idea of the certainty equivalent of a particular risk.

Thus, we could ask a person what is the maximum that he would be prepared to pay for insurance with a given level of initial wealth and a given potential insurance situation. Points can then be derived on the utility function, which would give rise to the answers given. By repeating this questioning for different initial wealth levels, all the points on the person's utility function could be found.

Consider an example of a person with a house worth £100,000. Suppose that the owner is considering insurance against a variety of perils, each of which would destroy the house completely. These perils have different probabilities of occurring, and the owner has assessed the amount that they're prepared to pay to insure against each peril, as shown in the following table:

Peril	A	B	C	D
Loss (£K)	100	100	100	100
Probability	0.05	0.15	0.3	0.5
Premium owner is prepared to pay (£K)	20	40	60	80

We can use this table to find out some information about the owner's utility function:

- Fix two values of the utility function. For example, let us suppose that the owner derives utility of zero if they have no wealth, and utility of 1 if they suffer no loss at all, *ie* $U(100)=1,\ U(0)=0$, working in units of £1,000. This is legitimate, because by fixing two points we are just choosing a level and a scale for our measure of utility.

- Consider Peril A. The owner's utility of wealth with insurance will equal the expected utility of wealth without insurance, if they have paid the maximum premium they're prepared to pay. With insurance, the level of wealth is certain to be 80. Without it, it may be 100 with probability 0.95, or 0 with probability 0.05.

So:

$$U(80) = 0.05 \times U(0) + 0.95 \times U(100) = 0.95$$

and we have a point on the utility function.

Question

Show, using a similar argument, that $U(60) = 0.85$, and find two more points on the owner's utility function. Draw a rough sketch of the graph of the utility function.

Solution

Consider Peril B. With insurance against this peril, the owner's wealth will be 60. Without it, the level of wealth will either be 100 with probability 0.85 or 0 with probability 0.15. Equating the utilities of these two possibilities, we have

$$U(60) = 0.85 \times U(100) + 0.15 \times U(0) = 0.85 .$$

Similarly, considering Perils C and D, we obtain $U(40) = 0.7$ and $U(20) = 0.5$. So the utility function will look something like this:

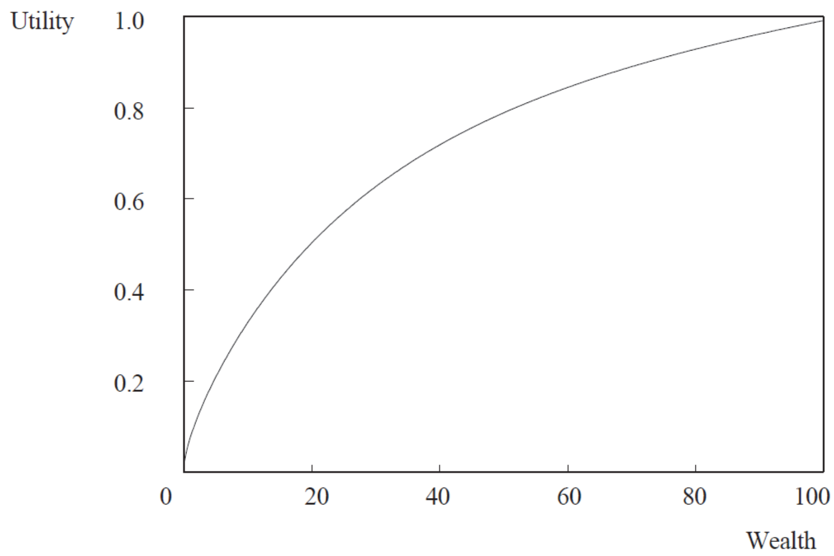


Figure 2.6 – A concave utility function

The complete utility function can be constructed by considering a large number of different scenarios, each of which contributes a point to the curve. Note that this particular function does appear to satisfy both the usual conditions $U'(w) > 0$ and $U''(w) < 0$.

7 Maximising utility through insurance

7.1 Introduction

Utility theory can be used to explain decisions such as purchasing insurance or buying a lottery ticket. Both of these activities are more likely to diminish the expected wealth of an individual. However, by purchasing insurance, one may be maximising expected utility.

A person who is risk averse will be prepared to pay more for insurance than the long-run average value of claims which will be made. Thus, insurance can be worthwhile for the risk averse policyholder even if the insurer has to charge a premium in excess of the expected value of claims in order to cover expenses and to provide a profit margin. An insurance contract is feasible if the minimum premium that the insurer is prepared to charge is less than the maximum amount that a potential policyholder is prepared to pay.

7.2 Finding the maximum premium

The maximum premium, *P*, which an individual will be prepared to pay in order to insure themselves against a random loss X is given by the solution of the equation:

$$E[U(a-X)] = U(a-P)$$

where a is the initial level of wealth.

Notice the similarity between this equation and the certainty equivalent relationship in Section 3.2. The individual is prepared to pay a certain amount P in order to avoid the uncertainty of the random loss X.

For example, consider an individual with a utility function of $U(x) = \sqrt{x}$ and current wealth of £15,000. Assume that this individual is at risk of suffering damages that are uniformly distributed up to 15,000. Then the individual's expected utility is:

$$E[U(a-X)] = \int_0^{15} \frac{1}{15}\sqrt{15-x}\,dx$$

$$= \left[\frac{-2}{3\times 15}(15-x)^{3/2}\right]_0^{15}$$

$$= 2.582$$

Then equating this to $U(a-P)$ gives:

$$U(a-P) = \sqrt{15-P} = 2.582$$

$$\Rightarrow \qquad P = 15 - 2.582^2 = 8.333$$

This individual would be willing to pay up to £8,833.33 for insurance that covers any loss.

This is well above the £7,500 expected loss.

7.3 Finding the minimum premium

The minimum insurance premium Q which an insurer should be prepared to charge for insurance against a risk with potential loss Y is given by the solution of the equation:

$$\boldsymbol{E[U(a+Q-Y)]=U(a)}$$

where a is the initial wealth of the insurer.

Question

An insurer with initial wealth of £2,000 and a utility of $U(x)=\log(x)$ is designing a policy to cover damages of £500 that occur with probability 0.5.

Calculate the minimum premium that the insurer can charge for the policy.

Solution

From the equation above we have:

$$E\left[U(2{,}000+Q-Y)\right]=U(2{,}000)$$

and with $U(x)=\log(x)$ the expectation becomes:

$$\begin{aligned}E\left[U(2{,}000+Q-Y)\right]&=0.5\log(2{,}000+Q-500)+0.5\log(2{,}000+Q)\\&=\log\left((1{,}500+Q)^{0.5}\right)+\log\left((2{,}000+Q)^{0.5}\right)\\&=\log\left((1{,}500+Q)^{0.5}\times(2{,}000+Q)^{0.5}\right)\\&=\log\left(\sqrt{(1{,}500+Q)\times(2{,}000+Q)}\right)\end{aligned}$$

Equating this with $U(2{,}000)=\log(2{,}000)$ yields:

$$\log\left(\sqrt{(1{,}500+Q)\times(2{,}000+Q)}\right)=\log\left(2{,}000\right)$$

$$\Rightarrow \quad (1{,}500+Q)\times(2{,}000+Q)=2{,}000^2$$

$$\Rightarrow \quad Q^2+3{,}500Q-1{,}000{,}000=0$$

Resulting in:

$$Q=\frac{-3{,}500\pm\sqrt{3{,}500^2+4\times1{,}000{,}000}}{2}\approx-1{,}750\pm2{,}015.56$$

Taking the positive root gives a minimum premium of £265.56.

8 Limitations of utility theory

The expected utility theorem is a very useful device for helping our thinking about risky decisions, because it focuses attention on the types of trade-offs that have to be made. However, the expected utility theorem has several limitations that reduce its relevance for risk management purposes:

1. **To calculate expected utility, we need to know the precise form and shape of the individual's utility function. Typically, we do not have such information.**

   Even using the questioning techniques described earlier in this chapter, it is still optimistic to assume that it will be possible to construct a utility function that accurately reflects an individual's preferences.

   Usually, the best we can hope for is to identify a general feature, such as risk aversion, and to use the rule to identify broad types of choices that might be appropriate.

2. **The theorem cannot be applied separately to each of several sets of risky choices facing an individual.**

3. **For corporate risk management, it may not be possible to consider a utility function for the firm as though the firm was an individual.**

   The firm is a coalition of interest groups, each having claims on the firm. The decision process must reflect the mechanisms with which these claims are resolved and how this resolution affects the value of the firm. Furthermore, the risk management costs facing a firm may be only one of a number of risky projects affecting the firm's owners (and other claimholders). The expected utility theorem is not an efficient mechanism for modelling the *interdependence* of these sources of risk.

Alternative decision rules that can be used for risky choices include those under mean-variance portfolio theory and stochastic dominance.

Both of these topics are covered in later chapters.

New theories of non-rational investment behaviour, known as behavioural finance, are also covered in this course.

The chapter summary starts on the next page so that you can keep all the chapter summaries together for revision purposes.

Chapter 2 Summary

The expected utility theorem

The *expected utility theorem* states that:

- a function, $U(w)$, can be constructed representing an investor's utility of wealth, w
- the investor faced with uncertainty makes decisions on the basis of maximising the *expected value* of utility.

Utility functions

The investor's risk-return preference is described by the form of their utility function. It is usually assumed that investors both:

- prefer more to less (*non-satiation*)
- are *risk-averse*.

Additionally, investors are sometimes assumed to exhibit decreasing *absolute risk aversion* – *ie* the *absolute* amount of wealth held in risky assets increases with wealth. In contrast, *relative risk aversion* indicates how the *proportion* of wealth held as risky assets varies with wealth.

Absolute and relative risk aversion are measured by the functions:

$$A(w) = \frac{-U''(w)}{U'(w)} \qquad\qquad R(w) = \frac{-wU''(w)}{U'(w)}$$

Amongst the utility functions commonly used to model investors' preferences are the:

- quadratic utility function
- log utility function
- power utility function.

State-dependent utility functions

Sometimes it may be inappropriate to model an investor's behaviour over all possible levels of wealth with a single utility function. This problem can be overcome either by using:

- utility functions with the same functional form but different parameters over different ranges of wealth, or
- *state-dependent utility functions*, which model the situation where there is a discontinuous change in the state of the investor at a certain level of wealth.

Construction of utility functions

One approach to constructing utility functions involves questioning individuals about their preferences.

The questioning may be direct or indirect. Indirect questioning may be framed in terms of how much an individual would be prepared to pay for insurance against various risks.

Maximum premium

With an initial wealth a, the maximum premium, P, that a policyholder would be willing to pay in order to avoid a potential loss, X, is given by:

$$E[U(a-X)] = U(a-P)$$

Minimum premium

With an initial wealth a, the minimum premium, Q, an insurer could charge to cover potential damages, Y, is given by:

$$E[U(a+Q-Y)] = U(a)$$

Limitations of utility theory

1. We need to know the precise form and shape of the individual's utility function.
2. The expected utility theorem cannot be applied separately to each of several sets of risky choices facing an individual.
3. For corporate risk management, it may not be possible to consider a utility function for the firm as though the firm was an individual.

Chapter 2 Practice Questions

2.1 An investor can invest in two assets, A and B:

Exam style

	A	B
expected return	6%	8%
variance	4%%	25%%

The correlation coefficient of the rate of return of the two assets is denoted by ρ and is assumed to take the value 0.5.

The investor is assumed to have an expected utility function of the form:

$$E_{\alpha}(U) = E(r_p) - \alpha\, Var(r_p)$$

where α is a positive constant and r_p is the rate of return on the assets held by the investor.

(i) Determine, as a function of α, the portfolio that maximises the investor's expected utility. [8]

(ii) Show that, as α increases, the investor selects an increasing proportion of Asset A. [1]
[Total 9]

2.2 Colin's preferences can be modelled by the utility function such that:

$$U'(w) = 3 - 2w, \ (w > 0).$$

(i) Determine the range of values over which this utility function can be satisfactorily applied.

(ii) Explain how Colin's holdings of risky assets will change as his wealth decreases.

(iii) Which of the following investments will he choose to maximise his expected utility?

Investment A		*Investment B*		*Investment C*	
outcome	*probability*	*outcome*	*probability*	*outcome*	*probability*
0.1	0.3	0	0.3	0.2	0.45
0.3	0.4	0.2	0.2	0.3	0.1
0.5	0.3	0.9	0.5	0.4	0.45

2.3 By considering the relationship $R(w) = w \times A(w)$, explain which of the following statements is true for a risk-averse individual.

1. If an investor's preferences display decreasing relative risk aversion then they must also display decreasing absolute risk aversion.

2. If an investor's preferences display decreasing absolute risk aversion then they must also display decreasing relative risk aversion.

2.4 Explain the four axioms that are required to derive the expected utility theorem.

2.5 Jenny has a quadratic utility function of the form $U(w) = w - 10^{-5}w^2$. She has been offered a job with Company X, in which her salary would depend upon the success or otherwise of the company. If it is successful, which will be the case with probability ¾, then her salary will be \$40,000, whereas if it is unsuccessful she will receive \$30,000.

Exam style

(i) Assuming that Jenny has no other wealth, state the salary range over which $U(w)$ is an appropriate representation of her individual preferences. [2]

(ii) Calculate the expected salary and the expected utility offered by the job. [2]

(iii) Suppose she was also to be offered a fixed salary by Company Z. Determine the minimum level of fixed salary that she would accept to work for Company Z in preference to Company X. [3]

(iv) Suppose that the owners of Company X are both risk-neutral and very keen that Jenny should join them and not Company Z. Determine whether the firm should agree to pay her a fixed wage, and, if so, how much. Comment briefly on your answer. [1]

[Total 8]

2.6 Suppose that Lance and Allan each have a log utility function and an initial wealth of 100 and 200 respectively. Both are offered a gamble such that they will receive a sum equal to 30% of their wealth should they win, whereas they will lose 10% of their wealth should they lose. The probability of winning is ¼.

(i) State whether or not the gamble is fair.

(ii) Calculate Lance's certainty equivalent for the gamble alone and comment briefly on your answer.

(iii) Repeat part (ii) in respect of Allan and compare your answer with that in part (ii).

(iv) Confirm that your comments in part (iii) apply irrespective of the individual's wealth.

2.7 Jayne's utility function can be described as $U(w) = \sqrt{w}$. She faces a potential loss of £100,000 in the event that her should house burn down, which has a probability of 0.01.

Exam style

(i) Calculate the maximum premium that Jayne would be prepared to pay to insure herself against the total loss of her house if her initial level of wealth was £140,000 and comment on your results. [3]

Suppose that UN Life plc has an initial wealth of £100 million and a utility function of the form $U(w) = w$.

(ii) Calculate the minimum premium UN Life plc would require in order to offer insurance to Jayne and comment on whether insurance is feasible in this instance. [3]

[Total 6]

2.8 An insurance company will be required to make a payout of £500 on a particular risk event, which is likely to occur with a probability of 0.4. The utility for any level of wealth, w, is given by:

Exam style

$$U(w) = 4{,}000 + 0.5w$$

The insurer's initial level of wealth is £6000. Calculate the minimum premium the insurer will require in order to take on the risk. [3]

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 2 Solutions

2.1 (i) ***Maximising the investor's expected utility***

Assuming that all of the investor's money is invested, and hence the portfolio weights sum to 1, the expected return and variance of a portfolio consisting of a proportion x_A of wealth held in Asset A, and a proportion $1-x_A$ of wealth held in Asset B are:

$$\begin{aligned} E_P &= x_A E_A + (1-x_A)E_B \\ &= 0.06x_A + 0.08(1-x_A) \\ &= 0.08 - 0.02x_A \end{aligned}$$ [1½]

and:

$$\begin{aligned} V_P &= x_A^2 V_A + x_B^2 V_B + 2x_A x_B \sigma_A \sigma_B \rho_{AB} \\ &= 0.0004x_A^2 + 0.0025(1-x_A)^2 + 0.0010x_A(1-x_A) \\ &= 0.0019x_A^2 - 0.0040x_A + 0.0025 \end{aligned}$$ [1½]

Therefore the investor's expected utility is:

$$\begin{aligned} E_\alpha(U) &= E(r_p) - \alpha Var(r_p) \\ &= 0.08 - 0.02x_A - \alpha\left(0.0019x_A^2 - 0.0040x_A + 0.0025\right) \end{aligned}$$ [2]

We can maximise this function of x_A by differentiating and setting to zero:

$$\frac{dE}{dx_A} = -0.02 - \alpha\left(0.0038x_A - 0.0040\right) = 0$$ [1]

$$\Leftrightarrow \qquad x_A = \frac{20\alpha - 100}{19\alpha}$$ [½]

or:

$$x_A = \frac{20}{19} - \frac{100}{19\alpha}$$ [½]

The second-order derivative is:

$$\frac{d^2E}{dx_A^2} = -0.0038\alpha < 0$$ [1]

which confirms that we have a maximum.

[Total 8]

(ii) ***Show that the investor selects an increasing proportion of Asset A***

Differentiating the formula for the optimal value of x_A in terms of α gives:

$$\frac{dx_A}{d\alpha} = \frac{100}{19\alpha^2} > 0$$ [1]

This confirms that as α increases, so x_A, the proportion of wealth held in Asset A, increases too.

2.2 (i) ***Range of wealth applicable***

Assuming *non-satiation*, which requires that $U'(w) > 0$, Colin's preferences can be modelled by this utility function provided that $0 < w < \frac{3}{2}$.

(ii) ***How Colin's holdings of risky assets vary with his wealth***

Differentiating the expression given in the question yields $U''(w) = -2$.

Thus, over the relevant range of *w*:

$$A(w) = \frac{2}{3-2w} > 0, \; A'(w) = \frac{4}{(3-2w)^2} > 0$$

and $$R(w) = \frac{2w}{3-2w} > 0, \; R'(w) = \frac{6}{(3-2w)^2} > 0$$

Hence, as Colin's wealth *decreases* the:

- *absolute amount* of his investment in risky assets will *increase* (as his absolute risk aversion decreases as his wealth decreases)

- *proportion* of his wealth that is invested in risky assets will *increase* (as his relative risk aversion decreases as his wealth decreases).

(iii) ***Colin's choice of investments***

Integrating the expression in the question gives Colin's utility function:

$$U(w) = a + 3w - w^2$$

As the properties of utility functions are invariant to linear transformations, we can set the arbitrary constant *a* equal to zero.

His expected utility from each of the investments is therefore as follows.

$$EU_A = 0.3\times U(0.1) + 0.4\times U(0.3) + 0.3\times U(0.5)$$

$$= 0.3\times 0.29 + 0.4\times 0.81 + 0.3\times 1.25$$

$$= 0.786$$

$$EU_B = 0.3\times U(0) + 0.2\times U(0.2) + 0.5\times U(0.9)$$

$$= 1.057$$

$$EU_C = 0.45\times U(0.2) + 0.1\times U(0.3) + 0.45\times U(0.4)$$

$$= 0.801$$

Thus, Colin will choose Investment B to maximise his expected utility.

2.3 The relationship between absolute risk aversion $A(w)$ and relative risk aversion $R(w)$ is such that:

$$R(w) = w\times A(w)$$

Differentiating with respect to wealth w gives:

$$\frac{\partial R}{\partial w} = A + w\frac{\partial A}{\partial w} \qquad (1)$$

Considering the first statement, Equation (1) tells us that if $\frac{\partial R}{\partial w} < 0$ and so relative risk aversion is decreasing, then it must also be the case that $\frac{\partial A}{\partial w} < 0$ (given that w and $A(w)$ are both positive for a risk-averse individual), *ie* $\frac{\partial R}{\partial w} < 0 \Rightarrow \frac{\partial A}{\partial w} < 0$.

An investor who displays decreasing relative risk aversion invests a larger proportion of wealth in risky assets as wealth increases. This also implies a larger monetary amount is invested in risky assets, ie decreasing absolute risk aversion.

Considering the second statement, then if $\frac{\partial A}{\partial w} < 0$, it does not follow that $\frac{\partial R}{\partial w}$ is necessarily negative. This will depend upon the relative magnitudes of $A(w)$, w and $\frac{\partial A}{\partial w}$. Thus, $\frac{\partial A}{\partial w} < 0$ does not imply that $\frac{\partial R}{\partial w} < 0$.

An investor who displays decreasing absolute risk aversion invests a larger monetary amount in risky assets as wealth increases. This does not necessarily equate to a larger percentage of wealth.

Hence, the first statement is true, whereas the second statement is false.

2.4 The expected utility theorem can be derived formally from the following four axioms:

1. Comparability

An investor can state a preference between all available certain outcomes.

2. Transitivity

If A is preferred to B and B is preferred to C, then A is preferred to C.

3. Independence

If an investor is indifferent between two certain outcomes, A and B, then he is also indifferent between the following two gambles:

(i) A with probability p and C with probability $(1-p)$; and

(ii) B with probability p and C with probability $(1-p)$.

4. Certainty equivalence

Suppose that A is preferred to B and B is preferred to C. Then there is a unique probability, p, such that the investor is indifferent between B and a gamble giving A with probability p and C with probability $(1-p)$.

B is known as the *certainty equivalent* of the above gamble.

2.5 (i) ***Salary range of utility function***

If: $U(w) = w - 10^{-5}w^2$

then: $U'(w) = 1 - 2w \times 10^{-5}$ [½]

and: $U''(w) = -2 \times 10^{-5}$ [½]

Now in order for Jenny to:

- prefer more to less, we require that $U'(w) > 0$, which in this case will be true for all $w < ½ \times 10^5$, *ie* $w < 50{,}000$
- be risk-averse, we require that $U''(w) < 0$, which in this case will be true for all $w > 0$.

Thus, the appropriate salary range is $w < \$50{,}000$. [1]

[Total 2]

(ii) ***Expected salary and expected utility***

Her expected salary is given by:

$$¾ \times 40{,}000 + ¼ \times 30{,}000 = \$37{,}500$$ [1]

Her expected utility is given by:

$$\tfrac{3}{4} \times [(40{,}000 - 10^{-5} \times (40{,}000)^2] + \tfrac{1}{4} \times [(30{,}000 - 10^{-5} \times (30{,}000)^2] = 23{,}250$$ [1]

[Total 2]

(iii) ***Minimum fixed salary***

The minimum level of salary, x say, is equal to the certainty equivalent of the job offer from Company X. [½]

This is given by:

$$U(x) = 23{,}250$$

$$x - 10^{-5}x^2 = 23{,}250$$

$$-x + 10^{-5}x^2 + 23{,}250 = 0$$ [½]

Using the formula for solving quadratic equations we find:

$$x = 63{,}229 \text{ or } 36{,}771$$ [1½]

As the first of these values is greater than the maximum salary available when Company X is successful it can be disregarded. Hence the minimum level of fixed salary that she would accept to work for Company Z is $36,771. [½]

[Total 3]

(iv) ***Should Company X offer a fixed salary?***

Yes – if they are risk-neutral, then they should offer Jenny a fixed salary in preference to a variable one. Jenny is risk-averse and therefore derives additional utility from the certainty offered by a fixed salary. [½]

Therefore, Company X will be able to entice Jenny to work for them in return for a salary of just (or strictly speaking slightly above) $36,771, instead of the expected salary of $37,500 in (i). [½]

[Total 1]

2.6 (i) ***Is the gamble fair?***

For any given initial level of wealth w, the expected value of the gamble is given by:

$$\tfrac{1}{4} \times 1.3w + \tfrac{3}{4} \times 0.9w - w = 0$$

Thus, the gamble is fair.

(ii) ***Lance's certainty equivalent of the gamble alone***

Lance's expected utility should he undertake the gamble is given by:

$$E[U] = \tfrac{1}{4}\log(130) + \tfrac{3}{4}\log(90) = 4.59174$$

Thus, his certainty equivalent for the initial wealth and the gamble is given by:

$$U(c_w) = \log(c_w) = 4.59174$$

$$\Rightarrow \quad c_w = e^{4.59174} = 98.666$$

and the certainty equivalent for the gamble alone is given by:

$$c_x = c_w - w = -1.334$$

This is negative because he is risk-averse.

The negative value of c_x means that Lance would have to be paid to accept the gamble.

(iii) ***Allan's certainty equivalent of the gamble alone***

Allan's expected utility should he undertake the gamble is given by:

$$E[U] = ¼ \log(260) + ¾ \log(180) = 5.28489$$

His certainty equivalent for the gamble alone is:

$$c_x = e^{5.28489} - 200 = -2.668$$

Comparing the two answers, we can see that the two certainty equivalents are equal to the same proportion of each individual's initial wealth. This is because the log utility function is consistent with preferences that exhibit constant *relative* risk aversion.

(iv) ***Relative risk aversion***

The constancy of relative risk aversion with a log utility function can be confirmed by differentiating it, *ie*:

If: $U(w) = \log(w)$

then: $U'(w) = \dfrac{1}{w}$ and $U''(w) = -\dfrac{1}{w^2}$

Thus: $R(w) = -w\dfrac{U''(w)}{U'(w)} = 1$ and $R'(w) = 0$

So, the log utility function exhibits constant relative risk aversion irrespective of w – though the log utility function is of course defined only for $w > 0$.

2.7 (i) ***Jayne's maximum premium***

Let P be the maximum insurance premium Jayne is prepared to pay and X be the loss she faces. Then Jayne's utility with insurance is:

$$U = \sqrt{140,000 - P} \qquad [½]$$

Whereas her expected utility without insurance is:

$$E[U(140{,}000-X)]=0.99\sqrt{140{,}000}+0.01\sqrt{40{,}000}=372.424$$ [½]

Equating these two expressions gives:

$$\sqrt{140{,}000-P}=372.424$$

ie $P=140{,}000-372.424^2=1{,}300$ [1]

The maximum premium of £1,300 exceeds the expected loss of £1,000. This is because Jayne is risk-averse. [1]

[Total 3]

(ii) ***UN Life's minimum premium***

Let Q be the minimum premium required by UN Life, then its utility without insurance is 100,000,000. [½]

Whereas its expected utility with insurance is:

$$E[U(100m+Q-X)]=0.99\times(100{,}000{,}000+Q)+0.01\times(99{,}900{,}000+Q)$$

$$=99{,}999{,}000+Q$$ [½]

Equating these two expressions gives:

$$100{,}000{,}000=99{,}999{,}000+Q$$

ie $Q=1{,}000$ [1]

So, the minimum premium required by UN Life is less than the maximum premium Jayne is prepared to pay, which means that the insurance contract is feasible. [1]

Notice that the minimum insurance premium that the insurance company will accept is equal to the expected value of the claim. This is because the insurance company is risk-neutral.

[Total 3]

2.8 The minimum premium Q is given by the equation:

$$E\left[U(a+Q-Y)\right]=U(a)$$ [1]

where a is the initial wealth and Y is the payout. In this case we have:

$$E\left[U(6{,}000+Q-Y)\right]=0.4U(6{,}000+Q-500)+0.6U(6{,}000+Q)$$

$$=0.4(4{,}000+0.5(5{,}500+Q))+0.6(4{,}000+0.5(6{,}000+Q))$$ [1]

$$=0.5Q+6{,}900$$

Equating this to $U(6,000) = 4,000 + 0.5 \times 6,000 = 7,000$ leads to:

$$0.5Q + 6,900 = 7,000$$

$\Rightarrow \quad Q = 200$ [1]

[Total 3]

3

Stochastic dominance and behavioural finance

Syllabus objectives

1.2 Rational choice theory

1.2.7 State conditions for absolute dominance and for first- and second-order dominance.

1.3 Behavioural economics

1.3.1 Describe the main features of Kahneman and Tversky's prospect theory critique of expected utility theory.

1.3.2 Explain what is meant by 'framing', 'heuristics' and 'bias' in the context of financial markets and describe the following features of behaviour in such markets:

- social influence and the herd instinct
- anchoring and adjustment
- self-serving bias
- loss aversion
- confirmation bias
- availability bias
- familiarity bias

1.3.3 Describe the Bernartzi and Thaler solution to the equity premium puzzle.

0 Introduction

This chapter focuses on the topics of *stochastic dominance* and *behavioural finance*.

Recall that the expected utility theorem suggests that a rational investor will aim to maximise their expected utility. In practice, however, it is not always possible to model choices with utility functions. Consequently, alternative approaches, such as *stochastic dominance*, may then be used to say something about an investor's choices without knowing the exact specification of the investor's utility function.

In addition, there is much experimental and empirical evidence to suggest that in practice investors are often *irrational*, in the sense that their actions are inconsistent with the predictions of utility theory. Instead, they appear to be subject to a number of biases and errors. Behavioural finance analyses these biases and their implications for financial decision making.

1 Stochastic dominance

1.1 Background

Absolute dominance **is said to exist when one investment portfolio provides a higher return than another in all possible circumstances. Clearly, this situation will rarely occur so we usually need to consider the *relative* likelihood of out-performance, *ie* stochastic dominance.**

We consider two investment portfolios, A and B, with cumulative probability distribution functions of returns F_A and F_B respectively.

So the probability that portfolio *i* yields an end-of-period wealth, *X*, less than or equal to *L* is given by:

$$P(X \leq L) = F_i(L) = \int_{-\infty}^{L} f_i(x)\,dx \quad \text{or} \quad P(X \leq L) = F_i(L) = \sum_{x=-\infty}^{L} p_i(x)$$

depending upon whether or not the distribution of possible outcomes is continuous or discrete. $f_i(x)$ or $p_i(x)$ is therefore the corresponding probability (density) function.

1.2 First-order stochastic dominance

The first-order stochastic dominance theorem states that, assuming an investor prefers more to less, A will dominate B (*ie* the investor will prefer portfolio A to portfolio B) if:

$F_A(x) \leq F_B(x)$, for all *x*, and

$F_A(x) < F_B(x)$ for some value of *x*.

This means that the probability of portfolio B producing a return below a certain value is never less than the probability of portfolio A producing a return below the same value, and exceeds it for at least some value of *x*.

For example, if two normal distributions have the same variance but different means, the one with the higher mean displays first-order stochastic dominance over the other.

Question

Assuming that portfolio A first-order stochastically dominates portfolio B, draw a diagram illustrating the relationship between $F_A(x)$ and $F_B(x)$, the respective cumulative probability distribution functions.

Solution

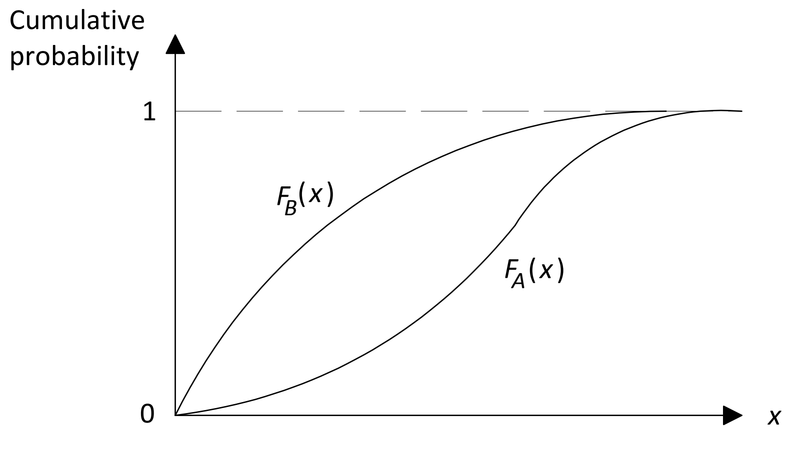


Figure 3.1: First-order stochastic dominance

The important points to note are that:

- $F_A(x)$ and $F_B(x)$ are both monotonically increasing functions of x
- $F_A(x)$ is never above (to the left of) $F_B(x)$.

Using first-order stochastic dominance to make investment decisions is similar to basing choices on the 'more to less' criterion discussed in the previous chapter. With this in mind, consider Asset X and Asset Z, which offer returns as follows according to whether or not there is a 'good' or 'poor' investment outcome.

	Asset X	*Asset Z*
Good outcome	6%	10%
Poor outcome	5%	8%

In this instance, it is clear that an investor who prefers more to less should choose Asset Z, which produces a higher return under both possible outcomes. Asset Z is said to *absolutely dominate* Asset X.

Suppose instead, however, that Assets X and Z offer the following possible outcomes with associated probabilities:

Asset X		*Asset Z*	
Return	*Probability*	*Return*	*Probability*
7%	½	8%	½
5%	½	6%	½

In this case the investor's choice is not quite as clear-cut because it is possible that Asset X *may* produce a higher actual return, although we suspect that the investor who prefers more to less should choose Asset Z, which offers a higher expected return. The first-order stochastic dominance theorem formalises the intuition behind the choice.

Consider the table below, which shows the cumulative probabilities of obtaining a return equal to or less than any particular value for each of these two assets.

	Cumulative probability	
Return	*Asset X*	*Asset Z*
5%	½	0
6%	½	½
7%	1	½
8%	1	1

Asset Z offers a (cumulative) probability of receiving any amount L or less that is never greater, and sometimes strictly less, than that offered by Asset X,

ie $P(\text{ Return} \le L \text{ for Asset Z}) \le P(\text{ Return} \le L \text{ for Asset X})$ for $L = 5, 6, 7, 8$

with the inequality being strict for $L = 5, 7$.

Equivalently:

$F_Z(x) \le F_X(x)$ for each of $x = 5, 6, 7, 8$, with

$F_Z(5) < F_X(5)$ and $F_Z(7) < F_X(7)$.

Hence, Asset Z (first-order stochastically) dominates Asset X. Consequently, an investor who prefers 'more to less' should choose Asset Z.

On the diagram below, the cumulative probability function for Asset Z, $F_Z(x)$, is never above that of Asset X, $F_X(x)$.

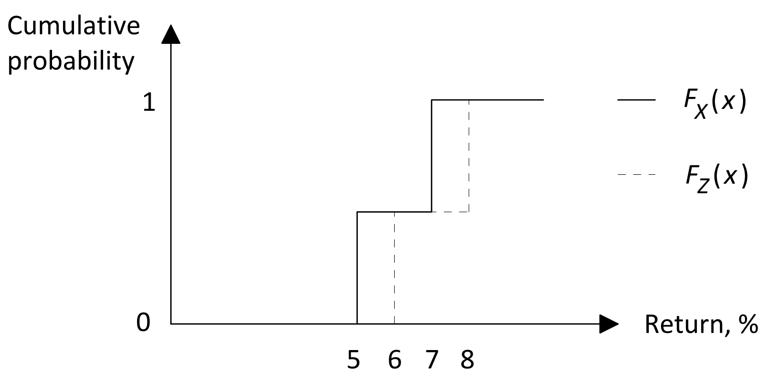


Figure 3.2: Assessing first-order stochastic dominance for Assets X and Z

A slightly different way of expressing the same idea is to say that Z (first-order stochastically) dominates X if Z can be obtained from X by shifting probability from lower to higher outcome levels.

Question

Consider two risky assets A and B. A yields $1 with probability ¼ and $2 with probability ¾. B yields $1 with probability ½ and $2 with probability ½.

(i) Explain why A first-order dominates B.

Now consider Asset C, which yields $1 with probability ¼ and $3 with probability ¾.

(ii) Determine whether C first-order dominates A or B.

Solution

(i) A can be obtained from B by shifting a probability of ¼ from the $1 outcome to the higher $2 outcome. Consequently A first-order dominates B.

(ii) C can be obtained from A by shifting ¾ from $2 to $3 and so C dominates A. C also dominates B because first-order stochastic dominance is *transitive*.

These results (C > A > B) are confirmed by probabilities shown in the table below.

Return	*Probability*			*Cumulative probability*		
	A	B	C	A	B	C
$1	¼	½	¼	¼	½	¼
$2	¾	½	0	1	1	¼
$3	0	0	¾	1	1	1

Often first-order stochastic dominance will not be a sufficiently strong criterion by which to choose between assets. On such occasions, we need to call upon the stronger criterion of second-order stochastic dominance.

1.3 Second-order stochastic dominance

The second-order stochastic dominance theorem applies when the investor is risk-averse, as well as preferring more to less.

In this case, the condition for A to dominate B is that

$$\int_a^x F_A(y)\,dy \le \int_a^x F_B(y)\,dy, \text{ for all } x,$$

with the strict inequality holding for some value of x, and where a is the lowest return that the portfolios can possibly provide.

The condition for second-order stochastic dominance is similar to that for first-order stochastic dominance, except that it is expressed in terms of the *sums* of the cumulative probabilities, rather than simply the cumulative probabilities themselves.

The interpretation of the inequality above is that a risk-averse investor will accept a lower probability of a given extra return at a low absolute level of return in preference to the same probability of extra return at a higher absolute level. In other words, a potential gain of a certain amount is not valued as highly as a loss of the same amount.

For example, if two normal distributions have the same mean but different variances, the one with the lower variance displays second-order stochastic dominance over the other.

Consider Assets U and V, which offer returns according to the table below:

	Probability	
Return	*Asset U*	*Asset V*
6%	¼	0
7%	¼	¾
8%	¼	0
9%	¼	¼

Question

Calculate $F_U(x)$ and $F_V(x)$ and explain why an investor cannot choose between Assets U and V on the basis of first-order stochastic dominance alone.

Solution

The cumulative probabilities are as follows.

Return	*Cumulative probability*	
	$F_U(x)$	$F_V(x)$
6%	¼	0
7%	½	¾
8%	¾	¾
9%	1	1

The choice between Assets U and V cannot be based upon first-order stochastic dominance alone, because neither asset dominates the other, *ie*:

$$F_U(6) > F_V(6), \text{ but } F_U(7) < F_V(7).$$

Alternatively, we could say that neither asset first-order stochastically dominates the other because we cannot obtain U from V by shifting probability from lower to higher outcome levels, nor can we obtain V from U by shifting probability from lower to higher outcome levels.

Although we cannot choose between U and V on the basis of first-order stochastic dominance, we can choose between them on the basis of second-order stochastic dominance.

Assets U and V both offer the same chance of a 9% return. In addition, U offers a wider spread of returns about 7% – *ie* a greater chance of an 8% return, but at the risk of a greater chance of obtaining only 6%. Assets U and V therefore both offer the same expected return of 7½%, but the variance of return is greater for Asset U. Thus, a risk-averse investor should choose Asset V.

An investor who bases their choices upon the second-order stochastic dominance theorem will make identical choices to those implied by non-satiation and risk aversion, assuming that they are able to make a choice.

We shall now show that Asset V (second-order stochastically) dominates Asset U.

Consider the following table that shows the *sums* of the cumulative probability functions, which is only a valid approach if the step sizes of the returns are all the same.

	Sum of cumulative probabilities	
Return	*U*	*V*
6%	¼	0
7%	¾	¾
8%	1½	1½
9%	2½	2½

In this case, V (second-order stochastically) dominates U because the sum of its cumulative probabilities is never greater than that of U and for one outcome is strictly less, *ie* :

$$\int_{a=6}^{x} F_V(y)\,dy \le \int_{a=6}^{x} F_U(y)\,dy, \text{ for } x = 6, 7, 8, 9$$

with the strict inequality holding for $6 < x < 7$.

According to the second-order stochastic dominance theorem, the investor should therefore always choose Asset V – which offers the same expected return as Asset U but with a lower variance.

Question

Draw a diagram showing $F_U(x)$ and $F_V(x)$ and use it to explain why Asset V (second-order stochastically) dominates Asset U.

Solution

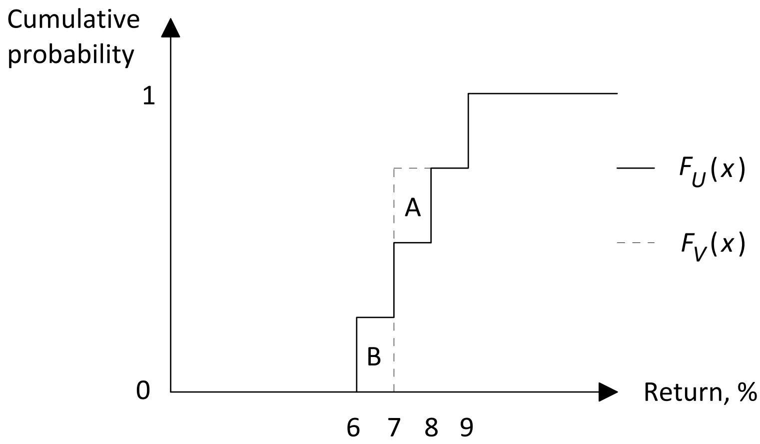


Figure 3.3: Assessing the stochastic dominance of Assets U and V

In this case, first-order stochastic dominance is insufficient to choose between the assets, because the cumulative probability graphs cross.

U is second-order stochastically dominated by V because the extra possibility of obtaining 8% (represented by the box marked A) is of less value to the investor than the possibility of avoiding 6% (represented by Box B).

The main advantage of using stochastic dominance is that it does not require explicit formulation of the investor's utility function, but can instead be used to make investment decisions for a wide range of utility functions.

The main disadvantages are that it:

- may be unable to choose between two investments, and
- generally involves pair-wise comparisons of alternative investments, which may be problematic if there is a large number of investments between which to choose.

2 Behavioural finance

2.1 Introduction

Although traditional economic theory assumes that investors always act rationally, *ie* with the aim of maximising expected utility, experimental and actual evidence suggests that this may not always be the case.

The field of behavioural finance looks at how a variety of mental biases and decision-making errors affect financial decisions. It relates to the psychology that underlies and drives financial decision-making behaviour.

2.2 Prospect Theory critique of expected utility theory

Since its inception, the expected utility theory (EUT) has drawn criticism from various quarters, primarily as a result of its axiomatic characterisation of preferences. These critics include Friedman and Savage (1948), who argued that an individual can have different utility functions (to be understood as attitudes to risk), depending on initial wealth. This observation was motivated by people's contrasting tendencies to buy insurance (risk-averse) and gamble (risk-seeking). In the Friedman and Savage approach, individuals are risk-seeking at low levels of wealth, and risk-averse at high levels of wealth.

Wealthy individuals may wish to preserve their current position, and thus be willing to pay for insurance in order to protect themselves against adverse events, even though they could withstand the financial consequences without insurance. This is risk-averse behaviour.

Less wealthy individuals, who consider that they have little to lose, may spend money on gambling in a bid to improve their financial position, rather than on insurance to protect the little they have. This is risk-seeking behaviour.

Similarly, Markowitz (1952) also criticised the general underpinnings of the EUT, arguing that utility should be measured relative to changes from a reference point rather than in absolute values of wealth.

The 'reference point' could be the current level of wealth.

Perhaps the most famous critique of the EUT emerged in the 1970s from a series of papers by two psychologists, namely Daniel Kahneman and Amos Tversky, culminating in their seminal 1979 paper on prospect theory.

Prospect theory was borne out of various laboratory experiments and sought to detail how human decision-making differs systematically from that predicted by EUT, and how human beings consistently violate the rationality axioms that form its basis. The model is descriptive: it tries to model real life choices, rather than optimal decisions.

There are two phases of decision-making described in prospect theory:

1. **Editing/framing phase – where outcomes of a decision are initially appraised and ordered.**
2. **Evaluation phase – choosing among the appraised options.**

Editing leads to a representation of the acts, outcomes and contingencies associated with a particular choice problem. It involves a number of basic operations that simplify and provide context for choice. The two basic operations involved in the editing process (others also exist) are:

- **Acceptance: people are unlikely to alter the formulation of choices presented.**
- **Segregation: people tend to focus on the most 'relevant' factors of a decision problem.**

Within this process, framing effects refer to the way in which a choice can be affected by the order or manner in which it is presented. Standard economic theory considers such transformations to be innocuous with no substantive impact on decisions.

Question

Identify which of the following two alternatives you would prefer:

1. \$30 for certain plus a 50% chance of losing \$20
2. \$10 for certain plus a 50% chance of winning \$20

Solution

The two alternatives are of course identical in that they both offer an equal chance of winning either \$30 or \$10. So, standard economic theory would tell us that people should be indifferent between the two alternatives.

However, experimental evidence based on similar choices suggests that people can and do view identical alternatives differently depending on how they are framed or worded. Thus, the proportion of people selecting each choice typically differs greatly from 50%.

As another example of the effect of framing, in the book:

- Plous S, (1993), *The psychology of judgement and decision making*, McGraw-Hill Inc

Plous describes an experiment in which people were asked the following questions about the length of a film (the same one) they had all recently watched:

- Question 1: How long was the movie?
- Question 2: How short was the movie?

The mean answer to the first question was 2 hours and 10 minutes, whereas that to the second question was 1 hour and 40 minutes!

Once prospects are edited, decision makers move on to the evaluation stage where they make their choice. Kahneman and Tversky observed a number of behavioural patterns in people when evaluating various alternatives. These include, most importantly:

- ***Reference dependence*: people derive utility from gains and losses measured relative to some reference point, rather than from absolute levels of wealth. This emerges from the idea that people are more attuned to changes in attributes rather than absolute magnitudes. This generates utility curves with a point of inflexion at the chosen reference point.**

- ***Loss aversion*: people are much more sensitive to losses (even small ones) than to gains of the same magnitude. In experiments, the pain from a loss is estimated to be twice as strong as the pleasure from an equivalent gain. Thus, utility curves are steeper in the domain of losses than they are in the domain of gains.**
- ***Endowment effects*: the endowment effect occurs when a person's preferences depend upon what they already possess. This implies that a person's preferences depend upon a certain reference point, perhaps determined by the person's possessions. Ownership itself creates satisfaction.**

In addition:

- ***Changing risk attitudes*: individuals tend to be risk averse in the domain of gains, and risk seeking when pondering losses.**
- ***Diminishing sensitivity*: as you gain (or lose) more, the marginal impact on utility of additional gains (losses) falls. Thus, a utility function is concave in the region of gains but convex in the region of losses.**
- ***Probability weighting*: people do not weight outcomes by their objective probabilities, but by transformed probabilities or decision weights. In general, weights are computed using a weighting function such that low probabilities are overweighted while high probabilities are underweighted.**
- ***Certainty effect*: when an outcome is certain and becomes less probable, the impact on your utility is greater than a similar reduction in probability for an outcome that was previously probable. For example, the impact on your utility of dropping the probability of earning $100 from 100% to 99% is more potent than dropping probability of earning $100 from 50% to 49%.**
- ***Isolation effect*: when choosing between alternatives, people often disregard components that the alternatives share and instead focus on what sets them apart, in order to simplify the decision. Since different choices can be decomposed in different ways, this invariably leads to inconsistent choices and preferences.**

To illustrate the point made above about 'changing risk attitudes', in their paper:

- Kahneman, D and Tversky, A (1979), Prospect theory: an analysis of decision under risk, *Econometrica 47*

Kahneman and Tversky describe an experiment in which they asked people to choose between two alternatives:

- Alternative 1: an 80% chance of winning $4,000 and a 20% chance of winning nothing
- Alternative 2: a 100% chance of winning $3,000.

Although the first alternative offers higher expected winnings ($3,200 *v* $3,000 for certain), 80% of people chose Alternative 2. This choice is consistent with the assumption of *risk aversion* that underpins expected utility theory. A risk-averse person may prefer a more certain outcome, even if the expected gains are lower (because the additional value derived from the extra certainty outweighs the additional value of the higher possible return).

The same people were then offered the following choice:

- Alternative 3: an 80% chance of losing $4,000 and a 20% chance of losing nothing
- Alternative 4: a 100% chance of losing $3,000.

Here 92% of people chose Alternative 3, even though the expected losses are greater (expected losses of $3,200 *v* a certain loss of $3,000). This evidence suggests that rather than being risk-averse, people may actually become *risk-seeking* when facing losses.

Question

Outline the key findings of prospect theory, and hence sketch a utility function of the form predicted by prospect theory.

Solution

Overall, prospect theory suggests that:

- Utility is based on gains and losses relative to some *reference point.*
- The reduction in utility from a loss is typically twice as much as the increase in utility from the same-sized monetary gain.
- People are typically *risk-averse* when considering gains relative to the reference point and *risk-seeking* when considering losses relative to the reference point.
- People experience diminishing sensitivity to gains and losses, so, for example, an initial gain has a greater impact on utility than a subsequent gain of the same size.

So, prospect theory would predict a utility function of the following form:

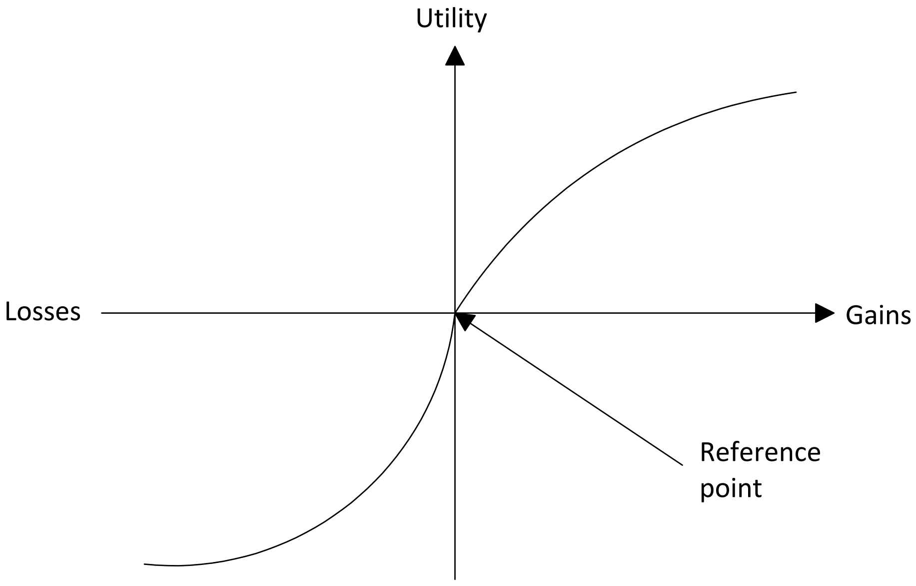


Figure 3.4: Prospect theory utility function

2.3 Heuristics and behavioural biases

Psychologists have long posited that the human brain consists of two separate systems of thinking. System 1 is the instinctive part of the brain that takes quick and effortless decisions, while System 2 is slow, deliberate and calculating, akin to the 'rational' thinker assumed by EUT. Which part of the brain is engaged depends largely on the decision context, including the environment and prevailing emotional state. Nonetheless, according to psychologists, System 1 is crucial to daily decisions. People essentially rely on System 1 unless System 2 regulates it.

Heuristics

System 1 involves dependence on shortcuts or 'heuristics' in order deliver quick decisions. These may, however, violate the axioms of rationality underlying EUT. Some of these heuristics include:

Anchoring and adjustment

'Anchoring and adjustment' is a term used to explain how people produce estimates. People start with an initial idea of the answer ('the anchor') and then adjust away from this initial anchor to arrive at their final judgement.

Thus, people may use experience or 'expert' opinion as the anchor, which they amend to allow for evident differences to the current conditions. The effects of anchoring are pervasive and robust and are extremely difficult to ignore, even when people are aware of the effect and aware that the anchor is ridiculous.

Even patently ridiculous anchor values have been shown to influence post-anchor estimates. For example, the result of the spin of a roulette wheel may be seen to influence people's estimates of the average daily temperature in London.

The anchor does not have to be related to the good. Nor does the anchor have to be consciously chosen by the consumer. If adjustments are insufficient, final judgments will reflect the (possibly arbitrary) anchors.

The following example is taken from the paper:

- Northcraft G B, and M A Neale , Experts, amateurs and real estate: an anchoring and adjustment perspective on property pricing decisions, *Organizational behaviour and human decision processes*, *39*.

An experiment was conducted in which a large number of estate agents were asked to value a property and come up with a recommended selling price. They were each provided with an information booklet containing a large volume of information concerning the property. The booklet was identical for all of the agents, except that four different versions were used, each with a different listed (*ie* suggested) price for the property.

It turned out that the average selling price recommended by the agents increased with the listed price as shown in the following table:

Listed price ($)	Average recommended selling price ($)
119,900 (Version 1)	117,745
129,900 (Version 2)	127,836
139,900 (Version 3)	128,530
149,900 (Version 4)	130,981

In the above example, the anchor value is the listed price. Thus, the agents' estimates were influenced by the 'anchor' or benchmark given to them in the form of the listed price. In addition, it turns out that the further the anchor value gets from the 'true' value, then the more it will pull people's estimates away from the true value.

Anchoring can have important implications for investment decisions, not least if individual investors rely on seemingly irrelevant yet salient data or statistics in order to guide their portfolio choices (*eg* expected returns from one-off investments in other unrelated industries, *etc*).

Representativeness

Decision-makers often use similarity as a proxy for probabilistic thinking.

This is the idea behind 'representativeness' – that a familiar scenario can be used as being representative of other similar cases. For example, an insurer may quote a home insurance premium with reference to another house on the same street.

Representativeness occurs because it is easier and quicker for our brain to compare a situation to a similar one (System 1) than assess it probabilistically on its own merits (System 2).

Representativeness is one of the most commonly-used heuristics and can, at times, work reasonably well. Nonetheless, similarity does not always adequately predict true probability, leading to irrational outcomes.

This is also related to the law of small numbers, where people assess the probability of something occurring based on its occurrence in a small, statistically-unrepresentative sample due to a desire to make sense of the uncertain situation (the name is an ironic play on the law of large numbers in statistics).

Representativeness can lead individuals to base their decision on whether to invest in a particular stock, or not, on the basis of its price over a few recent periods, rather than its long-term movement or the underlying fundamentals of the company.

Availability

This heuristic is characterised by assessing the probability of an event occurring by the ease with which instances of its occurrence can be brought to mind. Vivid outcomes are more easily recalled than other (perhaps more sensible) options that may require System 2 thinking.

This can lead to biased judgements when examples of one event are inherently more difficult to imagine than examples of another. For example, individuals living in areas that are prone to extreme weather events may only be compelled to take out home insurance after they have been directly affected by such events rather than beforehand, and may even cease their coverage after some time as the memory of the event subsides.

Another example here concerns the incidences of deaths due to car crashes and cancer. When asked to estimate the relative numbers of deaths due to each, people tend to overestimate the number of deaths due to car crashes perhaps because they receive more publicity and are easier to imagine.

Familiarity

This heuristic is closely-related to availability, and describes the process by which people favour situations or options that are familiar over others that are new. This may lead to an undiversified portfolio of investments if people simply put their money in industries or companies that they are familiar with rather than others in alternative markets or sectors.

Home-country bias refers to people's tendency to disproportionally invest in stocks from their home country, rather than forming an internationally-diversified portfolio.

Behavioural biases

The interaction between System 1 and System 2 can lead to several behavioural quirks or biases that have been documented by psychologists. These biases propose significant deviations from the rational outcomes proposed by standard economic thought, represented by EUT. Several have been proposed in the literature. Among these are:

Overconfidence

Overconfidence occurs when people systematically overestimate their own capabilities, judgement and abilities.

For example, if you ask 100 people if they are better than average drivers, then you might not be surprised if more than 50% of them reply 'yes'.

Moreover, studies show that the discrepancy between accuracy and overconfidence increases (in all but the simplest tasks) as the respondent becomes more knowledgeable! Accuracy increases to a modest degree but confidence increases to a much larger degree.

Overconfidence could therefore be a potentially serious problem in fields such as investment where most of the participants are likely to be highly knowledgeable. Moreover, the available evidence suggests that even when people are aware that they are overconfident they remain so.

This may, in turn, be a result of:

- ***Hindsight bias* – events that happen will be thought of as having been predictable prior to the event, events that do not happen will be thought of as having been unlikely prior to the event.**

  A possible example of the first type of event is the credit crunch of 2007/09.

  A possible example of the second type of event is when an underdog is heavily beaten in a sporting event. Although supporters may have had high hopes of an upset prior to the event, after the event a heavy defeat will always have seemed inevitable.

- ***Confirmation bias* – people will tend to look for evidence that confirms their point of view (and will tend to dismiss evidence that does not justify it). For example, Mauboussin (2009) shows how investors typically tend to overestimate their own investment results, by subconsciously focusing only on those assets in their portfolio whose results match their own optimistic self-perception.**

Self-serving bias

Closely related to overconfidence, self-serving bias occurs when people credit favourable or positive outcomes to their own capabilities or skills, while blaming external forces or others for any negative outcomes. This may be done in order to maintain a positive self-image and avoid what psychologists call 'cognitive-dissonance', which is the discomfort felt when there is a discrepancy between the perceived self and the actual self – as evidenced by outcomes.

So for example, having passed an exam, people may put this down to their own hard work and natural talent, rather than luck. However, when failing an exam, people may generate excuses (*eg* 'I wasn't feeling well', 'The room was too hot/cold/noisy…') to shield themselves from the truth of having not worked hard enough.

This type of behaviour is observed in investors when assessing their returns from investment. Furthermore, Doukas and Petmezas (2007) find that managers involved in acquisitions tend to credit themselves for any initial success of such deals, resulting in a larger number of deals which ultimately yield lower long-term returns for shareholders.

Status quo bias

Status quo bias is the inherent tendency of people to stick with their current situation, even in the presence of more favourable alternatives and even when no transaction costs are involved. (This was shown by Kahneman and Tversky in 1982.) A core reason why humans exhibit status quo bias is 'loss aversion' and 'endowment effects' as described above.

Herd behaviour

Herd behaviour describes the tendency of people to follow or mimic the actions and decisions taken by others, as a mechanism to deal with uncertain situations. The underlying rationale may be that others must know better (safety in numbers), learning or conformity preferences. However such decisions may lead to mass hysteria or delusion if initial actions are themselves biased.

Herd behaviour has been used to explain several issues observed in real-world financial markets, in particular stock market bubbles.

A 'bubble' occurs when the price of a particular good or share increases to an unsustainably high level. This can occur when well-publicised price rises lead to surges in demand, as investors 'follow the herd', hoping to benefit from further anticipated price rises. Those selling the good or share before the bubble bursts can make a lot of money, but those who fail to do so can incur heavy losses.

The first recorded bubble surrounded tulip bulbs in the Netherlands in the 1630s. A more recent example is the 'dotcom' bubble of the late 1990s.

2.4 A behavioural finance approach to the Equity Premium Puzzle

One of the most famous quandaries in finance is the so-called Equity Premium Puzzle, first identified by Mehta and Prescott in 1985. This puzzle is related to the gap in returns from risk-bearing stocks when compared to the returns earned from lower-risk government bonds and Treasury bills. Mehta (2008) estimates that over the last 116 years the average return on US equities has been approximately 7.67%, while the return on Treasury bills has been only 1.31% - an equity premium of 6.36%, with similar patterns observed in other countries around the world.

The standard explanation for this observed premium is that it reflects the difference in the level of risk associated with stocks relative to bonds and bills, with higher-risk securities requiring higher returns to attract investors. However, the equity premium levels observed in the data are far too large to be explained solely via risk aversion using the capital asset pricing models that are typically used in mainstream financial economics.

The Capital Asset Pricing Model is covered later in the course. It provides a link between the expected return from a security and its inherent level of risk.

The equity premium puzzle has led to a plethora of research papers attempting to find a plausible explanation for the gap in returns.

One of the most famous explanations for the puzzle was provided by Benartzi and Thaler in 1995, using the insights from behavioural finance. More specifically, the authors contend that *myopic loss aversion* can be used to explain the abnormally large discrepancy in returns across stocks and bonds.

Myopic loss aversion suggests that investors are less risk averse when faced with a multi-period series of 'gambles', and that the frequency of choice or length of reporting period will also be influential.

As its name suggests, myopic loss aversion relates to investors' aversion to short-term losses. The basic idea is that investors have been shown to be less 'risk-averse' when faced with a repeated series of 'gambles' than when faced with a single gamble.

This is because investors are generally extremely concerned by losses rather than equivalent gains, leading them to focus on very short-term returns and volatility rather than long-run earnings. Since stock prices are typically more volatile in the short run, this may dissuade myopic investors from buying stocks unless the returns premium on stocks is sufficiently high to compensate for this loss aversion. Thus, the substantial equity premia observed in the data across the world are so high since they take into account both risk aversion as well as this aversion to short-term losses.

So the excess return on stocks over bonds and bills comprises:

- extra return to compensate investors for the higher risk of equity investment (*ie* the greater uncertainty about the return that will ultimately be achieved), and
- additional extra return to overcome investors' loss aversion (*ie* the general unwillingness to expose themselves to potential losses of any kind).

The past equity premium is therefore consistent in this model with loss aversion, and an assumption that people evaluate their portfolios based on the last 12 months of returns (myopic, since investment decisions should rationally be made over longer timescales).

Many investment decisions relate to the longer term, *eg* an individual saving for retirement, or a life insurance company investing its reserves.

If an investor recognises that the investment strategy decision is in fact a series of repeated short-term gambles and consequently takes a long-term view when determining strategy, then they are likely to be less risk-averse than if they instead consider only the immediate short-term gamble and so take too short-term a view. In this latter case, they will tend to focus more on the short-term risk of loss than is necessarily in their best interests, the consequence being that their resulting portfolio ends up being overweight in less risky assets.

Question

Discuss the implications for a pension scheme's investment strategy of requiring it to report its financial position annually rather than triennially (assuming that assets are valued using market values).

Solution

If a pension fund has to report its financial position every year, it may be more averse to very short-term investment losses than if it had to report only every three years. The consequence of this might be to force the fund to invest in less volatile assets in order to reduce the risk of having to report a poor financial position. It might therefore result in the fund investing less heavily in equities, and possibly also less heavily in long-term bonds.

The chapter summary starts on the next page so that you can keep all the chapter summaries together for revision purposes.

Chapter 3 Summary

Stochastic dominance

Stochastic dominance offers an approach to modelling choices under uncertainty that does not require the use of explicit utility functions.

Given two investment portfolios, A and B, with cumulative probability distribution functions of returns F_A and F_B respectively:

The *first-order stochastic dominance theorem* states that A will be preferred to B if:

- the investor prefers more to less, $U'(x) > 0$ and
- $F_A(x) \leq F_B(x)$ for all x, with $F_A(x) < F_B(x)$ for at least one x.

The *second-order stochastic dominance theorem* states that A will be preferred to B if:

- the investor prefers more to less, $U'(x) > 0$,
- the investor is risk-averse, $U''(x) < 0$, and
- $\int_a^x F_A(y)\,dy \leq \int_a^x F_B(y)\,dy$, for all x, with the strict inequality holding for at least one x.

Behavioural finance

The field of behavioural finance relates to the psychology that underlies and drives financial decision-making behaviour.

Prospect theory details how human decision-making differs systematically from that predicted by expected utility theory, and how human beings consistently violate the rationality axioms that form the basis of the theory.

There are two phases of decision-making in prospect theory:

1. editing/framing – where outcomes are initially appraised and ordered
2. evaluation – where a choice is made from the appraised options.

The two basic operations involved in the editing process are:

1. acceptance – people are unlikely to alter the formulation of the choices presented
2. segregation – people tend to focus on the most 'relevant' factors of a problem.

In the evaluation phase of decision-making the following behavioural patterns are observed:

- reference dependence
- loss aversion
- endowment effects
- changing risk attitudes
- diminishing sensitivity
- probability weighting
- certainty effect
- isolation effect.

Prospect theory suggests that utility is based on gains and losses relative to some *reference point*, and people are typically *risk-averse* when considering gains relative to the reference point and *risk-seeking* when considering losses relative to the reference point.

A *heuristic* is a shortcut used by the brain to deliver a quick decision. Examples include:

- anchoring and adjustment
- representativeness
- availability
- familiarity.

The following *behavioural biases* have also been observed:

- overconfidence (perhaps as a result of hindsight bias or confirmation bias)
- self-attribution bias (or self-serving bias)
- status quo bias
- herd behaviour.

The *equity premium puzzle* (*ie* the fact that returns from equities exceed the returns from bonds and bills by more than is predicted by risk aversion alone) can be explained by myopic loss aversion.

Myopic loss aversion suggests that investors are much more concerned by losses than by equivalent gains, and so tend to focus on very short-term returns and volatility rather than long-run earnings. Investors therefore need to earn additional return on equities to overcome their aversion to the short-term losses typical of equity investment.

Chapter 3 Practice Questions

3.1 Define first-order and second-order stochastic dominance. Illustrate the definitions by sketching cumulative distribution functions of two random variables which represent the returns on two investments, one of which dominates the other.

3.2 Consider the two risky assets, A and B, with cumulative probability distribution functions:

Exam style

$$F_A(w) = w$$

$$F_B(w) = w^{1/2}$$

In both cases, $0 \le w \le 1$.

(i) Show that A is preferred to B on the basis of first-order stochastic dominance. [3]

(ii) Verify explicitly that A also dominates B on the basis of second-order stochastic dominance. [3]

[Total 6]

3.3 (i) Within the context of behavioural finance, explain fully what is meant by overconfidence. [4]

Exam style

The board of directors of an actively managed investment trust are concerned that the decisions of the trust's investment manager may be subject to overconfidence bias, which could adversely affect the performance of the trust.

(ii) Discuss possible actions that the board could take in order to try to limit the impact of the investment manager's overconfidence bias. [6]

[Total 10]

The solutions start on the next page so that you can separate the questions and solutions.

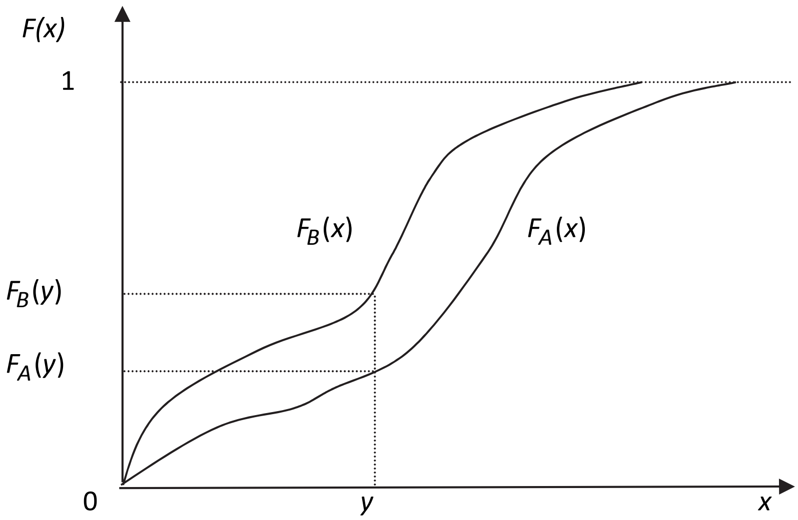
Chapter 3 Solutions

3.1 Assuming an investor prefers more to less, a distribution of investment returns A is said to exhibit *first-order stochastic dominance* over a distribution of investment returns B if:

$$F_A(x) \le F_B(x) \quad \text{for all } x\text{, and}$$

$$F_A(x) < F_B(x) \quad \text{for some value of } x.$$

In other words, the probability of B producing a return below a certain value is never less than the probability of A producing a return below the same value and exceeds it for at least some value of x.

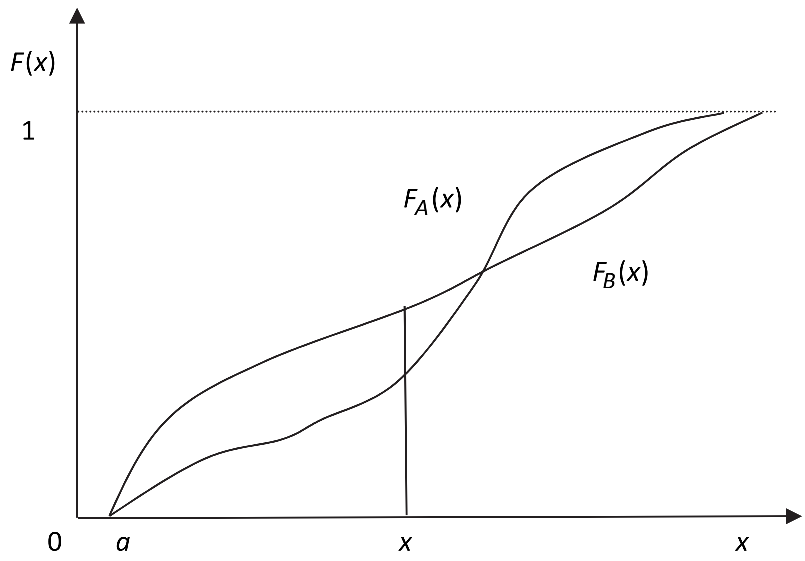


Here $F_A(x)$ *must never be above* $F_B(x)$. *Note, however, that the lowest possible value of x may be non-zero and even negative.*

Assuming an investor prefers more to less and is risk averse, a distribution of investment returns A is said to exhibit *second-order stochastic dominance* over a distribution of investment returns B if:

$$\int_a^x F_A(y)\,dy \le \int_a^x F_B(y)\,dy$$

for all x, with the strict inequality holding for some value of x, where a is the lowest return that the portfolios can possibly provide.



Here the area under $F_B(x)$ must never be less than the area under $F_A(x)$ for any value of x.

3.2 (i) ***First-order stochastic dominance***

A is preferred to B on the basis of first-order stochastic dominance if:

$F_A(w) \le F_B(w)$, for all $0 \le w \le 1$, and

$F_A(w) < F_B(w)$ for some value of w in this range. [1]

This is the case if:

$w \le w^{½}$ [½]

$\Leftrightarrow \quad w - w^{½} \le 0$

$\Leftrightarrow \quad w^{½}(w^{½} - 1) \le 0$

This clearly holds for all $0 \le w \le 1$, the equality being strict for $0 < w < 1$. Hence A first-order dominates B. [1½]

[Total 3]

Alternatively, we could draw the graphs of $F_A(w)$ and $F_B(w)$ over the range $0 \le w \le 1$ and note that the graph of $F_A(w)$ is below that of $F_B(w)$ for $0 < w < 1$ and equal for $w = 0,1$.

(ii) ***Second-order stochastic dominance***

A second-order dominates B if $G_A(w) \le G_B(w)$ for all $0 \le w \le 1$, with the strict inequality holding for some value of w, where $G(w) = \int_0^w F(y)dy$. [1]

This is the case if:

$$\int_0^w y\,dy \le \int_0^w y^{½}dy$$ [½]

$$\Leftrightarrow \quad \left[½y^2\right]_0^w \le \left[\tfrac{2}{3}y^{\frac{3}{2}}\right]_0^w$$

$$\Leftrightarrow \quad ½w^2 - \tfrac{2}{3}w^{\frac{3}{2}} \le 0$$

$$\Leftrightarrow \quad \tfrac{2}{3}w^{\frac{3}{2}}\,(\,¾w^{½} - 1) \le 0$$

This is true for all $0 \le w \le 1$ and strictly true for $0 < w \le 1$. Hence A does second-order dominate B. [1½]

[Total 3]

Alternatively, we note that $G(w) = \int_0^w F(y)dy$ *is the area under the graph of* $F(y)$. *We could plot graphs of* $G_A(w)$ *and* $G_B(w)$ *and note that the graph of* $G_A(w)$ *is below that of* $G_B(w)$ *for* $0 < w \le 1$ *and equal for* $w = 0$.

3.3 (i) ***Explain overconfidence***

Overconfidence occurs when people systematically overestimate their own capabilities, judgment and abilities. [1]

Moreover, studies show that the discrepancy between accuracy and overconfidence increases (in all but the simplest tasks) as the respondent is more knowledgeable. (Accuracy increases to a modest degree but confidence increases to a much larger degree.) [1]

This may be a result of:

- *Hindsight bias* – events that happen will be thought of as having been predictable prior to the event; events that did not happen will be thought of as having been unlikely ever to happen. [1]
- *Confirmation bias* – people will tend to look for evidence that confirms their point of view (and will tend to dismiss evidence that does not justify it). [1]

[Total 4]

(ii) ***Possible actions to limit the extent of overconfidence bias***

The board could require that all investment decisions made by the investment manager are reviewed by a second investment manager before being implemented. [1]

Alternatively, the management of the investment trust could be split equally between two investment managers. [1]

Either of these should reduce the impact of overconfidence bias, although the views of the second investment manager could be subject to similar overconfidence biases as those of the first manager. [1]

The investment manager could be sent on a training course about behavioural finance. This would make the manager aware of the possibility of overconfidence bias. [1]

However, providing training may have the opposite effect to that intended, making the investment manager feel more knowledgeable and aware of the issues and so even more confident. [1]

The board could place tighter constraints on the investment decisions taken by the investment manager, *eg* limits could be placed on the size of any transactions and/or on the size of holdings in individual companies or sectors. [1]

Limiting the manager's actions should limit the scope for biases in investment decisions but will also reduce the manager's scope for active investment management and possibly the returns achieved. [1]

[Maximum 6]

4

Measures of investment risk

Syllabus objectives

2.1 Properties of risk measures

2.1.1 Define the following measures of investment risk:

- variance of return
- downside semi-variance of return
- shortfall probabilities
- Value at Risk (VaR) / TailVaR.

2.1.2 Describe how the risk measures listed in 2.1.1 above are related to the form of an investor's utility function.

2.1.3 Perform calculations using the risk measures listed in 2.1.1 above to compare investment opportunities.

2.1.4 Explain how the distribution of returns and the thickness of tails will influence the assessment of risk.

2.2 Risk and insurance companies

2.2.1 Describe how insurance companies help to reduce or remove risk.

2.2.2 Explain what is meant by the terms 'moral hazard' and 'adverse selection'.

0 Introduction

In financial economics, it is often assumed that the key factors influencing investment decisions are 'risk' and 'return'. In practice, return is almost always interpreted as the *expected* investment return. However, there are many possible interpretations and different ways of measuring investment risk, of which the variance is just one, each of which corresponds to a different utility function.

This chapter outlines a small number of such measures, together with their relative merits, and then moves on to discuss how insurance can be used to reduce the impact of risk.

1 Measures of risk

1.1 Introduction

Most mathematical investment theories of investment risk use variance of return as the measure of risk.

Examples include (mean-variance) portfolio theory and the capital asset pricing model, both of which are discussed later in this course.

However, it is not obvious that variance necessarily corresponds to investors' perception of risk, and other measures have been proposed as being more appropriate.

Some investors might not be concerned with the mean and variance of returns, but simpler things such as the maximum possible loss. Alternatively, some investors might be concerned not only with the mean and variance of returns, but also more generally with other higher moments of returns, such as the *skewness* of returns. For example, although two risky assets might yield the same expectation and variance of future returns, if the returns on Asset A are positively skewed, whilst those on Asset B are symmetrical about the mean, then Asset A might be preferred to Asset B by some investors.

1.2 Variance of return

For a continuous distribution, **variance of return is defined as:**

$$\int_{-\infty}^{\infty} (\mu - x)^2 f(x)\, dx$$

where μ is the mean return at the end of the chosen period and $f(x)$ is the probability density function of the return.

'Return' here means the proportionate increase in the market value of the asset, *eg* $x = 0.05$ if the asset value has increased by 5% over the period.

The units of variance are '%%', which means 'per 100 per 100'.

eg $(4\%)^2 = 16\%\% = 0.16\% = 0.0016$

Question

Investment returns (% *pa*), X, on a particular asset are modelled using a probability distribution with density function:

$$f(x) = 0.00075\left(100 - (x-5)^2\right) \quad \text{where } -5 \le x \le 15$$

Calculate the mean return and the variance of return.

Solution

The density function is symmetrical about $x=5$. Hence the mean return is 5%. Alternatively, this could be found by integrating as follows:

$$E[X]=0.00075\int_{-5}^{15}x\left(100-(x-5)^2\right)dx$$

$$=0.00075\int_{-5}^{15}75x+10x^2-x^3dx$$

$$=0.00075\left[\frac{75}{2}x^2+\frac{10}{3}x^3-\frac{1}{4}x^4\right]_{-5}^{15}$$

$$=0.00075[7{,}031.25-364.5833]$$

$$=5$$

ie 5% *pa*.

The variance is given by:

$$Var(X)=0.00075\int_{-5}^{15}(5-x)^2\left(100-(x-5)^2\right)dx$$

$$=0.00075\int_{-5}^{15}100(x-5)^2-(x-5)^4dx$$

$$=0.00075\left[\frac{100}{3}(x-5)^3-\frac{1}{5}(x-5)^5\right]_{-5}^{15}$$

$$=0.00075[13{,}333.33-(-13{,}333.33)]$$

$$=20$$

ie 20%% *pa*.

Alternatively, we can calculate the variance using the formula: $Var(X)=E[X^2]-(E[X])^2$, where $E[X^2]$ can be found by integration to be 45%%.

For a discrete distribution, variance of return is defined as:

$$\sum_{x}(\mu-x)^2P(X=x)$$

where μ is the mean return at the end of the chosen period.

Question

Investment returns (% *pa*), X, on a particular asset are modelled using the probability distribution:

X	*Probability*
-7	0.04
5.5	0.96

Calculate the mean return and variance of return.

Solution

The mean return is given by:

$$E[X] = -7 \times 0.04 + 5.5 \times 0.96 = 5$$

ie 5% *pa*.

The variance of return is given by:

$$Var(X) = (5-(-7))^2 \times 0.04 + (5-5.5)^2 \times 0.96 = 6$$

ie 6%% *pa*.

Alternatively, we can calculate the variance using the formula: $Var(X) = E[X^2] - (E[X])^2$, where $E[X^2]$ is 31%%.

Variance has the advantage over most other measures in that it is mathematically tractable, and the mean-variance framework discussed in a later chapter leads to elegant solutions for optimal portfolios. Albeit easy to use, the mean-variance theory has been shown to give a good approximation to several other proposed methodologies.

Mean-variance portfolio theory can be shown to lead to optimum portfolios if investors can be assumed to have quadratic utility functions or if returns can be assumed to be normally distributed.

In an earlier chapter we discussed how the aim of investors is to maximise their expected utility. The mean-variance portfolio theory discussed in a later chapter assumes that investors base their investment decisions solely on the mean and variance of investment returns. This assumption is consistent with the maximisation of expected utility provided that the investor's expected utility depends only on the mean and variance of investment returns.

It can be shown that this is the case if:

- the investor has a quadratic utility function, and/or
- investment returns follow a distribution that is characterised fully by its first two moments, such as the normal distribution.

If, however, neither of these conditions holds, then we cannot assume that investors make choices solely on the basis of the mean and variance of return. For example, with more complex utility functions and non-normal return distributions investors may need to consider other features of the distribution of returns, such as skewness and kurtosis.

Question

Define both the skewness and the fourth central moment (called the *kurtosis*) of a continuous probability distribution.

Solution

The *skewness* of a continuous probability distribution is defined as the third central moment.

It is a measure of the extent to which a distribution is asymmetric about its mean. For example, the normal distribution is symmetric about its mean and therefore has zero skewness, whereas the lognormal distribution is positively skewed.

The *kurtosis* of a continuous probability distribution is defined as the fourth central moment.

It is a measure of how likely extreme values are to appear (*ie* those in the tails of the distribution).

1.3 Semi-variance of return

The main argument against the use of variance as a measure of risk is that most investors do not dislike uncertainty of returns as such; rather they dislike the possibility of low returns.

For example, all rational investors would choose a security that offered a chance of either a 10% or 12% return in preference to one that offered a certain 10%, despite the greater uncertainty associated with the former.

One measure that seeks to quantify this view is downside semi-variance (also referred to as simply semi-variance). For a continuous random variable, **this is defined as:**

$$\int_{-\infty}^{\mu} (\mu - x)^2 f(x)\, dx$$

Semi-variance is not easy to handle mathematically, and it takes no account of variability above the mean. Furthermore, if returns on assets are symmetrically distributed, semi-variance is proportional to variance.

Question

Investment returns (% *pa*), X, on a particular asset are modelled using a probability distribution with density function:

$$f(x) = 0.00075\left(100-(x-5)^2\right) \quad \text{where } -5 \le x \le 15$$

Calculate the downside semi-variance of return.

Solution

We saw in an earlier question that the variance of investment returns for this asset is 20%%. Since the continuous distribution $f(x)$ is symmetrical, the downside semi-variance is half the variance, *ie* 10%%.

For a discrete random variable, the downside semi-variance is defined as:

$$\sum_{x<\mu} (\mu - x)^2 P(X = x)$$

Question

Investment returns (% *pa*), X, on a particular asset are modelled using the probability distribution:

X	*probability*
-7	0.04
5.5	0.96

Calculate the downside semi-variance of return.

Solution

We saw in an earlier question that the mean investment return for this asset is 5%. So the downside semi-variance is given by:

$$\sum_{x<5} (5-x)^2 P(X=x) = (5-(-7))^2 \times 0.04 = 5.76$$

ie 5.76%% *pa*.

1.4 Shortfall probabilities

A shortfall probability measures the probability of returns falling below a certain level. For continuous variables, the risk measure is given by:

$$\textbf{Shortfall probability} = \int_{-\infty}^{L} f(x)\,dx$$

where L is a chosen benchmark level.

For discrete random variables, the risk measure is given by:

$$\text{Shortfall probability} = \sum_{x<L} P(X=x)$$

The benchmark level, *L*, can be expressed as the return on a benchmark fund if this is more appropriate than an absolute level. In fact, any of the risk measures discussed can be expressed as measures of the risk relative to a suitable benchmark which may be an index, a median fund or some level of inflation.

L could alternatively relate to some pre-specified level of surplus or fund solvency.

Question

Investment returns (% *pa*), X, on a particular asset are modelled using a probability distribution with density function:

$$f(x) = 0.00075\left(100-(x-5)^2\right) \quad \text{where } -5 \le x \le 15$$

Calculate the shortfall probability where the benchmark return is 0% *pa*.

Solution

The shortfall probability is given by:

$$\begin{aligned} P(X<0) &= 0.00075\int_{-5}^{0} 100-(x-5)^2\,dx \\ &= 0.00075\left[100x - \frac{1}{3}(x-5)^3\right]_{-5}^{0} \\ &= 0.00075\left[41.6667-\left(-166.6667\right)\right] \\ &= 0.15625 \end{aligned}$$

Question

Investment returns (% *pa*), X, on a particular asset are modelled using the probability distribution:

X	*probability*
-7	0.04
5.5	0.96

Calculate the shortfall probability where the benchmark return is 0% *pa*.

Solution

The shortfall probability is given by:

$$P(X<0)=0.04$$

The main advantages of the shortfall probability are that it is easy to understand and calculate.

The main drawback of the shortfall probability as a measure of investment risk is that it gives no indication of the magnitude of any shortfall (being independent of the extent of any shortfall).

For example, consider two securities that offer the following combinations of returns and associated probabilities:

Investment A: 10.1% with probability of 0.9 and 9.9% with probability of 0.1

Investment B: 10.1% with probability of 0.91 and 0% with probability of 0.09

An investor who chooses between them purely on the basis of the shortfall probability based upon a benchmark return of 10% would choose Investment B, despite the fact that it gives a much bigger shortfall than Investment A if a shortfall occurs.

1.5 Value at Risk

***Value at Risk (VaR)* generalises the likelihood of underperforming by providing a statistical measure of downside risk.**

For a continuous random variable, **Value at Risk can be determined as:**

$$\boldsymbol{VaR(X)=-t \quad \text{where} \quad P(X<t)=p}$$

VaR represents the maximum potential loss on a portfolio over a given future time period with a given degree of confidence, where the latter is normally expressed as $1-p$. So, for example, a 99% one-day VaR is the maximum loss on a portfolio over a one-day period with 99% confidence, *ie* there is a 1% probability of a greater loss.

Note that Value at Risk is a 'loss amount'. Therefore:

- a positive Value at Risk (a negative t) indicates a loss
- a negative Value at Risk (a positive t) indicates a profit
- Value at Risk should be expressed as a *monetary* amount and not as a percentage.

Question

Investment returns (% *pa*), X, on a particular asset are modelled using a probability distribution with density function:

$$f(x) = 0.00075\left(100-(x-5)^2\right) \quad \text{where } -5 \le x \le 15$$

Calculate the VaR over one year with a 95% confidence limit for a portfolio consisting of £100*m* invested in the asset.

Solution

We start by finding t, where $P(X<t)=0.05$:

$$\Rightarrow \qquad 0.00075\int_{-5}^{t} 100-(x-5)^2\,dx = 0.05$$

$$\Rightarrow \qquad 0.00075\left[100x-\frac{1}{3}(x-5)^3\right]_{-5}^{t} = 0.05$$

Since the equation in the brackets is a cubic in t, we are going to need to solve this equation numerically, by trial and error.

$$t=-3 \quad \Rightarrow \quad 0.00075\left[100x-\frac{1}{3}(x-5)^3\right]_{-5}^{-3} = 0.028$$

and $$t=-2 \quad \Rightarrow \quad 0.00075\left[100x-\frac{1}{3}(x-5)^3\right]_{-5}^{-2} = 0.06075$$

Interpolating between the two gives: $\quad t=-3+\dfrac{0.05-0.028}{0.06075-0.028}=-2.3$

In fact, the true value is $t=-2.293$. Since t is a percentage investment return per annum, the 95% Value at Risk over one year on a £100*m* portfolio is $£100m \times 2.293\% = £2.293m$. This means that we are 95% certain that we will not lose more than £2.293*m* over the next year.

For a discrete random variable, VaR is defined as:

$$VaR(X) = -t \quad \text{where } t=\max\left\{x : P(X<x) \le p\right\}$$

Question

Investment returns (% *pa*), X, on a particular asset are modelled using the probability distribution:

X	*Probability*
-7	0.04
5.5	0.96

Calculate the 95% VaR over one year with a 95% confidence limit for a portfolio consisting of £100*m* invested in the asset.

Solution

We start by finding t, where $t = \max\{x : P(X < x) \le 0.05\}$.

Now $P(X < -7) = 0$ and $P(X < 5.5) = 0.04$. Therefore $t = 5.5$.

Since t is a percentage investment return per annum, the 95% Value at Risk over one year on a £100*m* portfolio is $£100m \times -5.5\% = -£5.5m$. This means that we are 95% certain that we will not make profits of less than £5.5*m* over the next year.

VaR can be measured either in absolute terms or relative to a benchmark. Again, VaR is based on assumptions that may not be immediately apparent.

The problem is that in practice VaR is often calculated assuming that investment returns are normally distributed.

Question

Calculate the 97.5% VaR over one year for a portfolio consisting of £200*m* invested in shares. Assume that the return on the portfolio of shares is normally distributed with mean 8% *pa* and standard deviation 8% *pa*.

Solution

We start by finding t, where:

$$P(X < t) = 0.025, \text{ where } X \sim N(8, 8^2)$$

Standardising gives:

$$P\left(Z < \frac{t-8}{8}\right) = \Phi\left(\frac{t-8}{8}\right) = 0.025$$

Now $\Phi(-1.96) = 0.025$ from page 162 of the *Tables*, so: $\dfrac{t-8}{8} = -1.96 \;\Rightarrow\; t = -7.68$.

Since t is a percentage investment return per annum, the 97.5% Value at Risk over one year on a £200*m* portfolio is $£200m \times 7.68\% = £15.36m$. This means that we are 97.5% certain that we will not lose more than £15.36*m* over the next year.

Portfolios exposed to credit risk, systematic bias or derivatives may exhibit non-normal distributions. The usefulness of VaR in these situations depends on modelling skewed or fat-tailed distributions of returns, either in the form of statistical distributions (such as the Gumbel, Frechet or Weibull distributions) or via Monte Carlo simulations. However, the further one gets out into the 'tails' of the distributions, the more lacking the data and, hence, the more arbitrary the choice of the underlying probability becomes.

Hedge funds are a good example of portfolios exposed to credit risk, systematic bias and derivatives. These are private collective investment vehicles that often adopt complex and unusual investment positions in order to make high investment returns. For example, they will often short-sell securities and use derivatives.

If the portfolio in the previous question was a hedge fund then modelling the return using a normal distribution may no longer be appropriate. A different distribution could be used to assess the lower tail but choosing this distribution will depend on the data available for how hedge funds have performed in the past. This data may be lacking or include survivorship bias, *ie* hedge funds that do very badly may not be included.

The Gumbel, Frechet and Weibull distributions are three examples of extreme value distributions, which are used to model extreme events.

The main weakness of VaR is that it does not quantify the size of the 'tail'. Another useful measure of investment risk therefore is the *Tail Value at Risk*.

1.6 Tail Value at Risk (TailVaR) and expected shortfall

Closely related to both shortfall probabilities and VaR are the TailVaR (or TVaR) and expected shortfall measures of risk.

The risk measure can be expressed as the expected shortfall below a certain level.

For a continuous random variable, the expected shortfall is given by:

$$\textbf{Expected shortfall} = E\left[\max(L - X, 0)\right] = \int_{-\infty}^{L} (L - x) f(x)\, dx$$

where L is the chosen benchmark level.

If L is chosen to be a particular percentile point on the distribution, then the risk measure is known as the TailVaR.

The $(1-p)$ TailVaR is the expected shortfall in the p *th* lower tail. So, for the 99% confidence limit, it represents the expected loss *in excess of* the 1% lower tail value.

Question

Investment returns (% *pa*), X, on a particular asset are modelled using a probability distribution with density function:

$$f(x) = 0.00075\left(100-(x-5)^2\right) \quad \text{where } -5 \le x \le 15$$

Calculate the 95% TailVaR over one year for a portfolio consisting of £100*m* invested in the asset.

Solution

In a previous question, we calculated the 95% VaR for this portfolio to be £2.293m based on an investment return of –2.293%.

The expected shortfall in returns below –2.293% is given by:

$$\begin{aligned} E[\max(-2.293-X,0)] &= 0.00075\int_{-5}^{-2.293}(-2.293-x)\left(100-(x-5)^2\right)dx \\ &= 0.00075\int_{-5}^{-2.293}\left(-171.975-97.93x-7.707x^2+x^3\right)dx \\ &= 0.00075\left[-171.975x-48.965x^2-2.569x^3+0.25x^4\right]_{-5}^{-2.293} \\ &= 0.0462 \end{aligned}$$

On a portfolio of £100*m*, the 95% TailVaR is $£100m \times 0.000462 = £0.0462m$. This means that the expected loss *in excess of* £2.293*m* is £46,200.

For a discrete random variable, the expected shortfall is given by:

$$\text{Expected shortfall} = E\left[\max(L-X,0)\right] = \sum_{x<L}(L-x)P(X=x)$$

Question

Investment returns (% *pa*), X, on a particular asset are modelled using the probability distribution:

X	*Probability*
–7	0.04
5.5	0.96

Calculate the 95% TailVaR over one year for a portfolio consisting of £100*m* invested in the asset.

Solution

In a previous question, we calculated the 95% VaR for this portfolio to be –£5.5m based on an investment return of 5.5%.

The expected shortfall in returns below 5.5% is given by:

$$E[\max(5.5-X,0)] = \sum_{x<5.5}(5.5-x)P(X=x)$$

$$=(5.5-(-7))\times 0.04 = 0.5$$

On a portfolio of £100*m*, the 95% TailVaR is $£100m \times 0.005 = £0.5m$. This means that the expected reduction in profits *below* £5.5*m* is £0.5*m*.

However, TailVaR can also be expressed as the expected shortfall conditional on there being a shortfall.

To do this, we would need to take the expected shortfall formula and divide by the shortfall probability.

Other similar measures of risk have been called:

- **expected tail loss**
- **tail conditional expectation**
- **conditional VaR**
- **tail conditional VaR**
- **worst conditional expectation.**

They all measure the risk of underperformance against some set criteria. It should be noted that the characteristics of the risk measures may vary depending on whether the variable is discrete or continuous in nature.

Downside risk measures have also been proposed based on an increasing function of $(L-x)$, rather than $(L-x)$ itself in the integral above.

In other words, for continuous random variables, we could use a measure of the form:

$$\int_{-\infty}^{L} g(L-x)f(x)\,dx$$

Two particular cases of note are when:

1. $g(L-r)=(L-r)^2$ – this is the so-called *shortfall variance*
2. $g(L-r)=(L-r)$ – the *average or expected shortfall* measure defined above.

Note also that if $g(x)=x^2$ and $L = \mu$, then we have the semi-variance measure defined above.

Shortfall measures are useful for monitoring a fund's exposure to risk because the expected underperformance relative to a benchmark is a concept that is apparently easy to understand. As with semi-variance, however, no attention is paid to the distribution of outperformance of the benchmark, *ie* returns in excess of L are again completely ignored.

Question

Consider an investment whose returns follow a continuous uniform distribution over the range 0% to 10% *pa*.

(i) Write down the probability density function for the investment returns.

(ii) Calculate the mean investment return.

(iii) Calculate the variance and semi-variance measures of investment risk.

(iv) Calculate the shortfall probability and the expected shortfall based on a benchmark level of 3% *pa*.

Solution

Useful information about the continuous uniform distribution can be found on page 13 of the *Tables*, including the form of its probability density function, and formulae for its mean and variance.

(i) ***Probability density function***

Working in % units, the investment return follows a $U(0,10)$ distribution. So the probability density function is $f(x)=\dfrac{1}{10}$ for $0\le x\le 10$ and 0 otherwise.

If we work with the returns expressed in decimals instead, then $f(x)=10$ *for* $0\le x\le 0.10$ *and 0 otherwise.*

(ii) ***Mean***

The mean investment return is:

$$\frac{1}{2}(0+10)=5$$

ie 5% *pa*.

(iii) ***Variance and semi-variance***

The variance is given by:

$$\frac{1}{12}(10-0)^2=8.33$$

ie 8.33%% *pa*.

Alternatively, we could evaluate the variance using the integral:

$$\int_0^{10} \frac{(5-x)^2}{10}\,dx$$

Since the uniform distribution is symmetric, the semi-variance is equal to half the variance, *ie* 4.17%% *pa*.

Alternatively, we could evaluate the semi-variance using the integral:

$$\int_0^{5} \frac{(5-x)^2}{10}\,dx$$

(iv) ***Shortfall probability and expected shortfall***

The shortfall probability is given by:

$$SP = \int_0^3 \frac{1}{10}\,dx = \left[\frac{x}{10}\right]_0^3 = 0.3$$

The expected shortfall is given by:

$$ES = \int_0^3 \frac{(3-x)}{10}\,dx = \left[\frac{1}{10}\left(3x - 0.5x^2\right)\right]_0^3 = 0.45\,\%$$

2 Relationship between risk measures and utility functions

An investor using a particular risk measure will base their decisions on a consideration of the available combinations of risk and expected return. Given a knowledge of how this trade-off is made it is possible, in principle, to construct the investor's underlying utility function. Conversely, given a particular utility function, the appropriate risk measure can be determined.

For example, if an investor has a quadratic utility function, the function to be maximised in applying the expected utility theorem will involve a linear combination of the first two moments of the distribution of return.

In other words, if an investor has a quadratic utility function then their attitude towards risk and return can be expressed purely in terms of the mean and variance of investment opportunities.

Thus variance of return is an appropriate measure of risk in this case.

Question

(i) State the expected utility theorem.

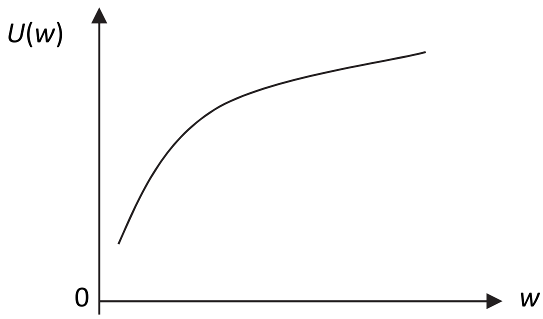
(ii) Draw a typical utility function for a non-satiated, risk-averse investor.

Solution

(i) The *expected utility theorem* states that:

- a function, $U(w)$, can be constructed representing an investor's utility of wealth, w
- the investor faced with uncertainty makes decisions on the basis of maximising the *expected value* of utility.

(ii)



Non-satiated investors prefer more wealth to less and so the graph slopes upwards, *ie* $U'(w) > 0$.

Risk-averse investors have diminishing marginal utility of wealth and so the slope of the graph decreases with w, *ie* $U''(w) < 0$.

If expected return and semi-variance below the expected return are used as the basis of investment decisions, it can be shown that this is equivalent to a utility function that is quadratic below the expected return and linear above.

Thus, this is equivalent to the investor being risk-averse below the expected return and risk-neutral for investment return levels above the expected return. Hence, no weighting is given to variability of investment returns above the expected return.

Use of a shortfall risk measure corresponds to a utility function that has a discontinuity at the minimum required return.

This therefore corresponds to the state-dependent utility functions discussed in a previous chapter.

Question

What is meant by a state-dependent utility function?

Solution

Sometimes it may be inappropriate to model an investor's behaviour over all possible levels of wealth with a single utility function. This problem can be overcome by using *state-dependent* utility functions, which model the situation where there is a discontinuous change in the state of the investor at a certain level of wealth.

3 Risk and insurance companies

3.1 Introduction

Individuals and corporations face risks resulting from unexpected events.

Risk-averse individuals can buy insurance to remove their exposure to risks. Being risk-averse, they will be willing to pay more for insurance than the expected cost of claims. The insurance company is willing to offer insurance primarily because it is able to spread its risks.

3.2 What to insure

Two considerations must be taken into account when assessing the effect of a risk: its likelihood and its severity. In a formal scenario, a risk matrix or graph is used as shown below.

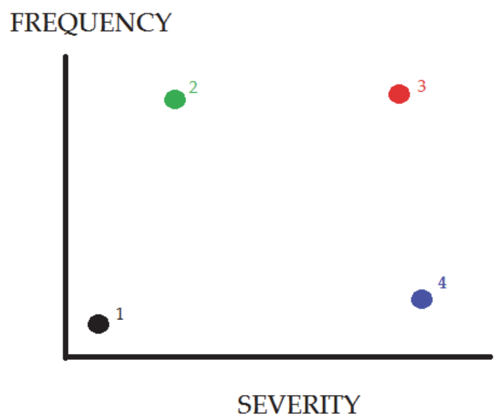


Figure 4.1

The figure above shows the frequency-severity dynamics of four possible events.

- **Event 1 is a low-frequency-low-severity event. Such an event does not warrant any worry to a corporation or individual.**

  For example, a solar eclipse occurs with low frequency, but is also accompanied by a low level of severity!

- **Event 2 is a high-frequency-low-severity event. Such an event would occur many times but at a low cost each time. The overall cost, due to high frequency, may be damaging. These types of events may need to be assessed on how they can be controlled.**

  For example, a smashed mobile phone screen is common but is typically limited in its extent.

- **Event 3 is a high-frequency-high-severity. Such events are to be avoided.**

  For example, car accidents involving fatalities happen relatively frequently and have a high severity.

- **Event 4 is a low-frequency-high-severity event. Such events tend to be insured.**

  For example, earthquake or hurricane damage.

In general, low severity events (such as events 1 and 2) are not generally insurable as the cost of management per claim is too expensive. However micro-insurance for poorer clients and new technologies that enable insuring small items over a short term (such as gadgets during a holiday) have been gaining ground.

3.3 Pooling resources

Insurance reduces the variability of losses due to adverse outcomes by pooling resources.

Consider a simple scenario of a property that has one hundredth chance of suffering £10,000 in damages (and 99% of no damages). The expected cost is 1% of £10,000 which is £100 while the VaR(99.5%) is £10,000.

Question

Explain why the 99.5% Value at Risk is £10,000.

Solution

Let X be the impact suffered by the property, so $P(X=0)=0.99$ and $P(X=-10{,}000)=0.01$. Then from the definition of Value at Risk for a discrete random variable we have:

$$VaR(X) = -t \text{ where } t = \max\{x : P(X<x) \le p\}$$

But $P(X<-10{,}000)=0$ and $P(X<0)=0.01$, therefore:

$$t = \max\{x : P(X<x) \le 0.005\} = -10{,}000$$

This means that the Value at Risk is £10,000.

If ten independent properties with similar characteristics are pooled together, the average cost is still £100 per property. At the extreme, the probability of all of them suffering the damage is 0.01^{10}. The VaR(99.5%) in this case is if one property suffers damage, that is £100 on average (using a Binomial distribution).

Question

Verify that the 99.5% Value at Risk in this case is £10,000.

Solution

Let X be the total impact suffered by the properties, so $X \sim Binomial(10, 0.01) \times (-10{,}000)$. Then from the definition of Value at Risk for a discrete random variable we have:

$$VaR(X) = -t \text{ where } t = \max\{x : P(X<x) \le p\}$$

If no properties suffer damage then $X=0$. The probability of this is:

$$P(X=0)=\binom{10}{0}\times 0.01^0(1-0.01)^{10}=0.99^{10}=0.9044$$

So: $P(X<0)=1-0.9044=0.0956$

If exactly one property suffers damage then $X=-10{,}000$. The probability of this is:

$$P(X=-10{,}000)=\binom{10}{1}\times 0.01^1(1-0.01)^9=\frac{10!}{9!1!}0.01^1\times 0.99^9=0.0914$$

So: $P(X<-10{,}000)=1-0.9044-0.0914=0.0042$

Therefore:

$$t=\max\{x:P(X<x)\le 0.005\}=-10{,}000$$

This means that the Value at Risk is £10,000, which equates to one property suffering damage.

In pooling resources, an insurer attempts to group insureds (being corporations or individuals) within homogeneous groups. In the case of an individual with the ability to influence into which group they fall, adverse selection can occur. If the insurer is also risk averse, then the insurance premium needs to include a margin to compensate the insurer for taking on the risk.

3.4 Policyholder behaviour

Adverse selection **describes the fact that people who know that they are particularly bad risks are more inclined to take out insurance than those who know that they are good risks.**

It arises because customers typically know more about themselves than the insurance company knows.

Adverse selection is sometimes called 'self-selection' or 'anti-selection'.

To try and reduce the problems of adverse selection, insurance companies try to find out lots of information about potential policyholders. Policyholders can then be put in small, reasonably homogeneous pools and charged appropriate premiums.

Moral hazard **describes the fact that a policyholder may, because they have insurance, act in a way which makes the insured event more likely.**

This is because having the insurance provides less incentive to guard against the insured event happening. For example, while driving to work one day you realise that you forgot to lock the front door of your house. If you didn't have any household contents insurance, you might decide to go back and lock it. If you had adequate insurance, you might decide to carry on to work. This difference in behaviour *caused by the fact that you are insured* is an example of 'moral hazard'.

Moral hazard makes insurance more expensive. It may even push the price of insurance above the maximum premium that a person is prepared to pay.

The chapter summary starts on the next page so that you can keep all the chapter summaries together for revision purposes.

Chapter 4 Summary

Measures of investment risk

Many investment models use *variance* of return as the measure of investment risk.

For a continuous random variable: $V = \int_{-\infty}^{\infty} (\mu - x)^2 f(x)\,dx$

For a discrete random variable: $V = \sum_{x} (\mu - x)^2 P(X = x)$

Variance has the advantage over most other measures that it:

- is mathematically tractable
- leads to elegant solutions for optimal portfolios, within the context of mean-variance portfolio theory.

The main argument against the use of variance as a measure of risk is that most investors do not dislike uncertainty of returns as such; rather they dislike the *downside risk* of low investment returns. Consequently, alternative measures of downside risk sometimes used include (in the continuous and then discrete cases):

- semi-variance of return: $\int_{-\infty}^{\mu} (\mu - x)^2 f(x)\,dx$ $\quad \sum_{x<\mu} (\mu - x)^2 P(X = x)$
- shortfall probability: $\int_{-\infty}^{L} f(x)\,dx$ $\quad \sum_{x<L} P(X = x)$

each of which ignores upside risk.

Value at Risk (VaR) represents the maximum potential loss on a portfolio over a given future time period with a given degree of confidence $(1-p)$. It is often calculated assuming that investment returns follow a normal distribution, which may not be an appropriate assumption.

For a continuous random variable, $VaR(X) = -t$, where $P(X < t) = p$.

For a discrete random variable, $VaR(X) = -t$, where $t = \max\{x : P(X < x) \le p\}$.

The expected shortfall, relative to a benchmark L is given by $E[\max(L - X, 0)]$.

For a continuous random variable, expected shortfall = $\int_{-\infty}^{L} (L - x) f(x)\,dx$.

For a discrete random variable, expected shortfall = $\sum_{x<L} (L - x) P(X = x)$.

When L is the VaR with a particular confidence level, the expected shortfall is known as *TailVaR*. TailVaR measures the expected loss *in excess* of the VaR.

It is also possible to calculate the expected shortfall and TailVaR *conditional* on a shortfall occurring by dividing through by the shortfall probability.

Relationship between risk measures and utility functions

If *expected return* and *variance* are used as the basis of investment decisions, it can be shown that this is equivalent to a quadratic utility function.

If *expected return* and *semi-variance* below the expected return are used as the basis of investment decisions, it can be shown that this is equivalent to a utility function that is quadratic below the expected return and linear above.

Use of a *shortfall risk measure* corresponds to a utility function that has a discontinuity at the minimum required return.

Using insurance to manage risk

Insurers decide which events to offer protection for based on the frequency and severity of the event.

The *pooling of resources* can be used to reduce an insurer's risk.

Adverse selection describes the fact that people who know that they are particularly bad risks are more inclined to take out insurance than those who know that they are good risks.

Moral hazard is the change in a policyholder's behaviour once insurance has been taken out, which makes the risk event more likely to occur.

Chapter 4 Practice Questions

4.1 Define the following measures of investment risk:

Exam style

(i) variance of return [1]

(ii) downside semi-variance of return [1]

(iii) shortfall probability [1]

(iv) Value at Risk. [1]

[Total 4]

4.2 Adam, Barbara and Charlie are all offered the choice of investing their entire portfolio in either a risk-free asset or a risky asset. The risk-free asset offers a return of 0% *pa*, whereas the returns on the risky asset are uniformly distributed over the range –5% to +10% *pa*. Assuming that each individual makes their investment choice in order to minimise their expected shortfall, and that they have benchmark returns of –2%, 0% and +2% *pa* respectively, who will choose which investment? Comment briefly on your answer.

4.3 (i) Define 'shortfall probability' for a continuous random variable.

(ii) An investor holds an asset that produces a random rate of return, R, over the course of a year. Calculate the shortfall probability using a benchmark rate of return of 1%, assuming:

(a) R follows a lognormal distribution with $\mu = 5\%$ and $\sigma^2 = (5\%)^2$

(b) R follows an exponential distribution with a mean return of 5%.

(iii) Explain with the aid of a simple numerical example the main limitation of the shortfall probability as a basis for making investment decisions.

4.4 Consider a zero-coupon corporate bond that promises to pay a return of 10% next period. Suppose that there is a 10% chance that the issuing company will default on the bond payment, in which case there is an equal chance of receiving a return of either 5% or 0%.

Exam style

(i) Calculate values for the following measures of investment risk:

(a) downside semi-variance

(b) shortfall probability based on the risk-free rate of return of 6%

(c) the expected shortfall below the risk-free return conditional on a shortfall occurring. [5]

(ii) Discuss the usefulness of downside semi-variance as a measure of investment risk for an investor. [3]

[Total 8]

4.5 An investor is contemplating an investment with a return of £ R , where:

Exam style

$$R = 250{,}000 - 100{,}000N$$

and N is a Normal $[1,1]$ random variable.

Calculate each of the following measures of risk:

(a) variance of return

(b) downside semi-variance of return

(c) shortfall probability, where the shortfall level is £50,000

(d) Value at Risk at the 95% confidence level

(e) Tail Value at Risk at the 95% confidence level, conditional on the VaR being exceeded.

[13]

Hint: For part (e), you may wish to use the formula for the truncated first moment of a normal distribution given on page 18 of the Tables.

4.6 (i) Explain the problem of adverse selection and how it might be dealt with by insurance companies. [2]

Exam style

(ii) Explain the problem of moral hazard and how it affects the price of insurance. [2]

[Total 4]

Chapter 4 Solutions

4.1 In the following we assume the investment return is given by a continuous random variable X with density function $f(x)$. This is the return over a chosen time period. Analogous formulae could be given for discrete or mixed cases.

(i) ***Variance***

$$\int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$

where $\mu = E[X]$ is the mean return for the chosen period. [1]

(ii) ***Downside semi-variance***

The downside semi-variance only takes into account returns below the mean return:

$$\int_{-\infty}^{\mu} (x-\mu)^2 f(x)dx$$ [1]

(iii) ***Shortfall probability***

A shortfall probability measures the probability of returns falling below a certain chosen benchmark level *L*:

$$P(X<L) = \int_{-\infty}^{L} f(x)dx$$ [1]

(iv) ***Value at Risk***

Value at Risk represents the maximum potential loss in value on a portfolio over a given future time period with a given degree of confidence. [1]

Alternatively, for a given confidence limit $(1-p)$:

$$VaR(X) = -t \quad \text{where} \quad P(X<t) = p$$

[Total 4]

4.2 The expected shortfall below a benchmark level *L* is defined as:

$$\int_{-\infty}^{L} (L-x)f(x)\,dx$$

For the risky asset, we have (working in percentage units):

$$f(x) = \begin{cases} \dfrac{1}{10-(-5)} = \dfrac{1}{15} & \text{if } -5<x<10 \\ 0 & \text{otherwise} \end{cases}$$

Adam

The expected shortfall of the *risky asset* is given by:

$$\int_{-5}^{-2} \frac{(-2-x)}{15}\,dx$$

$$= \frac{1}{15}\left[-2x - ½x^2\right]_{-5}^{-2}$$

$$= \frac{1}{15}\left[(4-2)-(10-12½)\right]$$

$$= 0.3\%$$

Alternatively, note that there is a chance of $1/5$ *that he will earn less than the benchmark, and in this case, the average shortfall will be 1½%. So the expected shortfall will be* $1/5 \times 1½\% = 0.3\%$.

The expected shortfall of the *risk-free* asset is 0%.

So Adam chooses the risk-free asset.

Barbara

The expected shortfall of the *risky asset* is given by:

$$\int_{-5}^{0} \frac{(-x)}{15}\,dx$$

$$= \frac{1}{15}\left[-½x^2\right]_{-5}^{-0}$$

$$= \frac{1}{15}\left[(0)-(-12½)\right]$$

$$= 0.833\%$$

The expected shortfall of the *risk-free* asset is again 0%.

So Barbara chooses the risk-free asset.

Charlie

The expected shortfall of the *risky asset* is given by:

$$\int_{-5}^{2} \frac{(2-x)}{15}\,dx$$

$$= \frac{1}{15}\left[2x - ½x^2\right]_{-5}^{2}$$

$$= \frac{1}{15}\left[(4-2)-(-10-12½)\right]$$

$$= 1.633\%$$

The expected shortfall of the *risk-free* asset is 2% × 1 = 2%.

So Charlie chooses the risky asset.

Thus, the expected shortfall increases with the benchmark return.

4.3 (i) ***Definition***

The shortfall probability for a continuous random variable, X, is:

$$P(X < L) = \int_{-\infty}^{L} f(x)\,dx$$

where L is the chosen benchmark level.

(ii)(a) ***Calculation based on lognormal distribution***

If $R \sim \log N(5,25)$, then the shortfall probability is:

$$P(R<1) = P\left(Z < \frac{\ln 1 - 5}{\sqrt{25}}\right) = P(Z<-1)$$

where $Z \sim N(0,1)$. Therefore:

$$P(R<1) = \Phi(-1) = 1-\Phi(1) = 1-0.84134 = 0.15866$$

(ii)(b) ***Calculation based on exponential distribution***

Since we know that $E(R) = 5$, this means that $R \sim Exp(0.2)$.

So the shortfall probability is:

$$P(R<1) = \int_0^1 0.2e^{-0.2x}\,dx = \left[-e^{-0.2x}\right]_0^1 = 1-e^{-0.2} = 0.18127$$

(iii) ***Main limitation***

The main limitation of the shortfall probability is that it ignores the *extent* of the shortfall below the benchmark L.

Thus, if L = 0, then the investor will prefer a gamble that offers +\$1 with probability 0.51 and –\$1,000,000 with probability 0.49 to one that that offers either \$1,000,000 or –\$2, each with probability of ½. This is somewhat unrealistic.

4.4 (i)(a) ***Downside semi-variance***

The expected return on the bond is given by:

$$0.90 \times 10\% + 0.05 \times 5\% + 0.05 \times 0\% = 9.25\%$$ [1]

So the downside semi-variance is equal to:

$$(9.25-5)^2 \times 0.05 \;+\; (9.25-0)^2 \times 0.05 \;=\; 5.18\%\% \qquad [1]$$

(i)(b) ***Shortfall probability***

The probability of receiving less than 6% is equal to the sum of the probabilities of receiving 5% and 0%, *ie* 0.10. [1]

(i)(c) ***Expected conditional shortfall***

The expected shortfall below the risk-free rate of 6% is given by:

$$(6-5) \times 0.05 \;+\; (6-0) \times 0.05 \;=\; 0.35\% \qquad [1]$$

The expected shortfall below the risk-free return *conditional on a shortfall occurring* is equal to:

$$\frac{\text{expected shortfall}}{\text{shortfall probability}} = \frac{0.35\%}{0.10} = 3.5\% \qquad [1]$$

[Total 5]

We can see this directly by noting that, given that there is a shortfall, it is equally likely to be 1% or 6%. So the expected conditional shortfall is 3½%.

(ii) ***Usefulness of downside semi-variance***

- It gives more weight to downside risk, *ie* variability of investment returns below the mean, which is likely to be of greater concern to an investor than upside risk. [½]
- In fact, it completely ignores risk above the mean. [½]
- This is consistent with the investor being risk-neutral above the mean, which is unlikely to be the case in practice. [½]
- The mean is an arbitrary benchmark, which might not be appropriate for the particular investor. [½]
- If investment returns are symmetrically distributed about the mean (as they would be, for example, with a normal distribution) then it will give equivalent results to the variance. [½]
- However, it is less mathematically tractable than the variance. [½]

[Total 3]

4.5 (a) ***Variance of return***

N has a Normal $[1,1]$ distribution, so R has a Normal distribution with mean $150{,}000$ and variance $100{,}000^2$, *ie* $R \sim N(150{,}000, 100{,}000^2)$.

So, the variance of return is $100{,}000^2 = 10^{10}$. [2]

(b) ***Downside semi-variance of return***

Any normal distribution is symmetrical about its mean, so that the downside semi-variance of return is equal to half of the variance, *ie* 5×10^9. [2]

(c) ***Shortfall probability, where the shortfall level is £50,000***

The shortfall probability below £50,000 is:

$$\begin{aligned} P(R < 50{,}000) &= P\left(\frac{R - 150{,}000}{100{,}000} < \frac{50{,}000 - 150{,}000}{100{,}000}\right) \\ &= \Phi\left(\frac{50{,}000 - 150{,}000}{100{,}000}\right) \\ &= \Phi(-1) \\ &= 1 - \Phi(1) = 1 - 0.84134 = 0.15866 \end{aligned}$$ [2]

(d) ***Value at Risk***

From the *Tables*:

$$\Phi(-1.6449) = 0.05$$

So, there is a 5% chance of the investment return R having a value less than:

$$\begin{aligned} R_{5\%} &= \mu_R - 1.6449\sigma_R \\ &= 150{,}000 - 1.6449 \times 100{,}000 \\ &= -14{,}490 \end{aligned}$$

So, the Value at Risk at the 95% confidence level is £14,490. [2]

(e) ***Tail Value at Risk***

The VaR is £14,490. So, the formula for the conditional TVaR is:

$$\frac{1}{0.05} \int_{-\infty}^{-14{,}490} (-14{,}490 - x) f(x)\,dx$$

where $f(x)$ is the *pdf* of a $N(150{,}000, 100{,}000^2)$ distribution. [1]

Splitting this into two integrals:

$$\frac{-14{,}490}{0.05} \int_{-\infty}^{-14{,}490} f(x)\,dx - \frac{1}{0.05} \int_{-\infty}^{-14{,}490} x\, f(x)\,dx$$ [½]

Evaluating the first of these integrals:

$$\int_{-\infty}^{-14,490} f(x)dx = P\left(N(150{,}000, 100{,}000^2) < -14{,}490\right)$$

$$= P\left(Z < \frac{-14{,}490 - 150{,}000}{100{,}000}\right)$$

$$= P(Z < -1.6449)$$

$$= 0.05$$ [½]

This is as expected since $-14{,}490$ *is the VaR at the 95% confidence level.*

Evaluating the second integral, using the formula for the truncated first moment of a normal random variable on page 18 of the *Tables*:

$$\int_{-\infty}^{-14,490} x\, f(x)dx = 150{,}000\left[\Phi(U') - \Phi(L')\right] - 100{,}000\left[\phi(U') - \phi(L')\right]$$ [½]

where:

$$U' = \frac{-14{,}490 - 150{,}000}{100{,}000} = -1.6449$$ [¼]

and:

$$L' = \frac{-\infty - 150{,}000}{100{,}000} = -\infty$$ [¼]

Now $\Phi(t)$ is the cumulative distribution function of the standard normal distribution, so:

$$\Phi(U') = \Phi(-1.6449) = 0.05$$ [¼]

and:

$$\Phi(L') = \Phi(-\infty) = 0$$ [¼]

Also, $\phi(t)$ is the probability density function of the standard normal distribution, which is stated on page 160 of the *Tables*. So:

$$\phi(U') = \phi(-1.6449) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-1.6449)^2} = 0.10313$$ [¼]

and:

$$\phi(L') = \phi(-\infty) = 0$$ [¼]

Putting these values into the formula:

$$\int_{-\infty}^{-14,490} x\,f(x)dx = 150,000\left[0.05-0\right] - 100,000\left[0.10313-0\right]$$

$$= -2,813 \qquad [½]$$

Putting all this together, the conditional TVaR at the 95% confidence level is:

$$\frac{-14,490}{0.05} \times 0.05 - \frac{1}{0.05} \times (-2,813) = £41,770 \qquad [½]$$

So, the overall expected loss, given that the VaR at the 95% confidence level is exceeded, is:

$$14,490 + 41,770 = £56,260$$

[Total 13]

4.6 (i) ***Adverse selection***

Adverse selection refers to the fact that people who know that they are particularly bad risks are more inclined to take out insurance than those who know that they are good risks. [1]

To try to reduce the problems of adverse selection, insurance companies try to find out information about potential policyholders. Policyholders can then be put into small, reasonably homogeneous groups and charged appropriate premiums. [1]

[Total 2]

(ii) ***Moral hazard***

Moral hazard describes the fact that a policyholder may, because they have insurance, act in a way which makes the insured event more likely to occur. [1]

Moral hazard makes insurance more expensive. It may even push the price of insurance above the maximum premium that a person is prepared to pay. [1]

[Total 2]

5

Stochastic models of investment returns

Syllabus objectives

3.1 Show an understanding of simple stochastic models for investment returns.

3.1.1 Describe the concept of a stochastic investment return model and the fundamental distinction between this and a deterministic model.

3.1.2 Derive algebraically, for the model in which the annual rates of return are independently and identically distributed and for other simple models, expressions for the mean value and the variance of the accumulated amount of a single premium.

3.1.3 Derive algebraically, for the model in which the annual rates of return are independently and identically distributed, recursive relationships which permit the evaluation of the mean value and the variance of the accumulated amount of an annual premium.

3.1.4 Derive analytically, for the model in which each year the random variable $(1+i)$ has an independent lognormal distribution, the distribution functions for the accumulated amount of a single premium and for the present value of a sum due at a given specified future time.

3.1.5 Apply the above results to the calculation of the probability that a simple sequence of payments will accumulate to a given amount at a specific future time.

0 Introduction

Financial contracts are often of a long-term nature. Accordingly, at the outset of many contracts there may be considerable uncertainty about the economic and investment conditions which will prevail over the duration of the contract.

Therefore, we might choose to develop models for future investment returns to help us study that inherent uncertainty.

A *deterministic* model is a model that provides an output based on one set of parameter and input variables. However, deciding which set of input variables to use may be a challenge. Thus, for example, if it is desired to determine premium rates on the basis of one fixed rate of return, it is nearly always necessary to adopt a conservative basis for the rate to be used in any calculations, subject to the premium being competitive.

The deterministic approach can provide only a single fixed answer to a problem. This answer will be correct only if the assumptions made about the future turn out to be correct.

When setting premium rates for insurance contracts, we need to make some assumption about future investment returns, as the premiums received will (at least, in part) be invested in order to meet the cost of future claims. Adopting a conservative basis for the rate of return means that the rate selected is lower than we actually expect to receive in the future. The difference between the assumed rate in the model and the expected rate provides a margin for uncertainty. However, if the assumed rate of return is too low, this will result in premiums that are too high to be competitive.

An alternative approach to recognising the uncertainty that in reality exists is provided by the use of *stochastic models*. In such models, no single rate is used and variations are allowed for by the application of probability theory.

Since we cannot specify in advance precisely what investment returns will be, in a stochastic model, we make an assumption about the statistical distribution of future investment returns. This enables us to consider the expected accumulated value of an investment at a future date and the variance of that accumulated value. As we have seen, considering the variance of the future value of a fund is one way of measuring the risk associated with our choice of investments.

The stochastic approach can give unreliable results if the statistical distribution used is not appropriate.

Possibly one of the simplest models is that in which each year the rate obtained is independent of the rates of return in all previous years and takes one of a finite set of values, each value having a constant probability of being the actual rate for the year.

For example, the effective annual rates of return that will apply during each of the next n years might be $i_1, i_2, \ldots, i_n$, where i_k, $k = 1, 2, \ldots, n$ are random variables with the following discrete distribution:

$$i_k = \begin{cases} 0.06 & \text{with probability } 0.2 \\ 0.08 & \text{with probability } 0.7 \\ 0.10 & \text{with probability } 0.1 \end{cases}$$

Question

Calculate the mean, j, and the standard deviation, s, of i_k.

Solution

The mean is:

$$j = E\left[i_k\right] = 0.2 \times 0.06 + 0.7 \times 0.08 + 0.1 \times 0.10 = 0.078$$

We can calculate the variance using the formula $Var(i_k) = E\left[i_k^2\right] - E^2\left[i_k\right]$:

$$E\left[i_k^2\right] = 0.2 \times 0.06^2 + 0.7 \times 0.08^2 + 0.1 \times 0.10^2 = 0.0062$$

$$\Rightarrow s^2 = Var(i_k) = E\left[i_k^2\right] - E^2\left[i_k\right] = 0.0062 - 0.078^2 = 0.000116 = 0.0108^2$$

So, the mean is 7.8% and the standard deviation is 1.08%.

Alternatively, the rate may take any value within a specified range, the actual value for the year being determined by some given probability density function.

For example we might assume that the annual rates of return are uniformly distributed between 5% and 10%, or that the annual growth factor $1+i$ follows a lognormal distribution with given parameter values, as we do in Section 2.

1 Simple models

1.1 Fixed rate model

At this stage we consider briefly an elementary example, which – although necessarily artificial – provides a simple introduction to the probabilistic ideas implicit in the use of stochastic rate models.

Suppose that an investor wishes to invest a lump sum of *P* into a fund with compound investment rate growth at a constant rate for *n* years. This constant investment return is not known *now*, but will be determined immediately after the investment has been made.

This model, where the effective annual rate of return is a single unknown rate i, which will apply throughout the next n years, is often known as the *fixed rate model*.

The accumulated value of the sum will, of course, be dependent on the investment rate. In assessing this value before the investment rate is known, it could be assumed that the mean rate will apply. However, the accumulated value using the mean rate will not equal the mean accumulated value. In algebraic terms:

$$P\left(1+\sum_{j=1}^{k}(i_j p_j)\right)^n \neq P\left(\sum_{j=1}^{k} p_j(1+i_j)^n\right)$$

where:

i_j **is the *j*th of *k* possible investment rates of return**

p_j **is the probability of the investment rate of return** i_j

The above result is easily demonstrated with the following simple numerical question.

Question

The returns from an investment are assumed to conform to the fixed rate model with the distribution of rates as specified below:

$$i_k = \begin{cases} 0.06 & \text{with probability } 0.2 \\ 0.08 & \text{with probability } 0.7 \\ 0.10 & \text{with probability } 0.1 \end{cases}$$

(i) Calculate the expected accumulated value at the end of 5 years of an initial investment of £5,000.

(ii) Calculate the accumulated value at the mean rate of return.

Solution

(i) The expected accumulated value is:

$$\begin{aligned}5{,}000E[(1+i_k)^5] &= 5{,}000(0.2\times 1.06^5 + 0.7\times 1.08^5 + 0.1\times 1.10^5)\\ &= 5{,}000\times 1.4572\\ &= £7{,}286\end{aligned}$$

(ii) The mean rate of return is:

$$E[i_k] = 0.2\times 0.06 + 0.7\times 0.08 + 0.1\times 0.10 = 0.078$$

Therefore the accumulated value at the mean rate of return is:

$$5{,}000\times 1.078^5 = £7{,}279$$

As expected, we see that the expected accumulated value is not equal to the accumulated value calculated at the expected rate of return.

Question

Calculate the variance of the accumulated value of the investment in the previous question.

Solution

The variance of the accumulated value, s^2, is:

$$\begin{aligned}s^2 &= 5{,}000^2(E[(1+i_k)^{10}] - (E[(1+i_k)^5])^2)\\ &= 5{,}000^2((0.2\times 1.06^{10} + 0.7\times 1.08^{10} + 0.1\times 1.10^{10})\\ &\qquad -(0.2\times 1.06^5 + 0.7\times 1.08^5 + 0.1\times 1.10^5)^2)\\ &= 5{,}000^2\times(2.128791 - 1.457226^2) = (£363)^2\end{aligned}$$

It's important to keep a few extra decimals in this last calculation to avoid losing accuracy, since the calculation involves subtracting two numbers of similar magnitude.

For the fixed rate model, the mean and variance of the accumulated value of an investment must be calculated from first principles.

1.2 Varying rate model

In our previous example the effective annual investment rate of return was fixed *throughout the duration* of the investment. A more flexible model is provided by assuming that over each *single year* the annual yield on invested funds will be one of a specified set of values or lie within some specified range of values, the yield in any particular year being independent of the yields in all previous years and being determined by a given probability distribution.

This model is often called the *varying rate model*. The main difference between this and the fixed rate model is that, in the varying rate model, the rates of return can be different in each future year, whereas, in the fixed rate model, the same (unknown) rate of return will apply in each future year.

Measure time in years. Consider the time interval $[0,n]$ subdivided into successive periods $[0,1], [1,2], \ldots, [n-1,n]$. For $t = 1,2,\ldots,n$ let i_t be the yield obtainable over the *t*th year, *ie* the period $[t-1,t]$. Assume that money is invested only at the beginning of each year. Let F_t denote the accumulated amount at time *t* of all money invested before time *t* and let P_t be the amount of money invested at time *t*. Then:

$$F_t = (1+i_t)(F_{t-1} + P_{t-1}) \text{ , for } t = 1,2,3,\ldots \qquad (1.1)$$

It follows from this equation that a single investment of 1 at time 0 will accumulate at time *n* to:

$$S_n = (1+i_1)(1+i_2)\ldots(1+i_n) \qquad (1.2)$$

Similarly, a series of annual investments, each of amount 1, at times $0,1,2,\ldots,n-1$ will accumulate at time *n* to:

$$\begin{aligned} A_n = {} & (1+i_1)(1+i_2)(1+i_3)\ldots(1+i_n) \\ & + (1+i_2)(1+i_3)\ldots(1+i_n) \\ & + \cdots \\ & + (1+i_{n-1})(1+i_n) \\ & + (1+i_n) \end{aligned} \qquad (1.3)$$

Note that A_n and S_n are random variables, each with its own probability distribution function.

For example, if the yield each year is 0.02, 0.04, or 0.06 and each value is equally likely, the value of S_n will be between 1.02^n and 1.06^n. Each of these extreme values will occur with probability $(1/3)^n$.

Question

Determine the probability that S_n will take the value $1.02 \times 1.04^{n-1}$.

Solution

This value of S_n will occur if the rate of return is 2% in one year and 4% in the remaining $n-1$ years. Since there are n different years in which the 2% could fall, the probability is:

$$n\times\frac{1}{3}\times\left(\frac{1}{3}\right)^{n-1}=\frac{n}{3^n}$$

In general, a theoretical analysis of the distribution functions for A_n and S_n is somewhat difficult. It is often more useful to use simulation techniques in the study of practical problems. However, it is perhaps worth noting that the moments of the random variables A_n and S_n can be found relatively simply in terms of the moments of the distribution for the yield each year. This may be seen as follows.

Moments of S_n

Let's consider the kth moment of S_n, where S_n denotes the accumulated value at time n of an initial investment of 1 made at time 0.

From Equation (1.2) we obtain:

$$(S_n)^k=\prod_{t=1}^{n}(1+i_t)^k$$

and hence:

$$E\left[S_n^k\right]=E\left[\prod_{t=1}^{n}(1+i_t)^k\right]$$

$$=\prod_{t=1}^{n}E\left[(1+i_t)^k\right] \qquad \textbf{(1.4)}$$

since (by hypothesis) $i_1,i_2,\dots,i_n$ are independent. Using this last expression and given the moments of the annual yield distribution, we may easily find the moments of S_n.

For example, suppose that the yield each year has mean j and variance s^2. Then, letting $k=1$ in Equation (1.4), we have:

$$E[S_n]=\prod_{t=1}^{n}E[(1+i_t)]$$

$$=\prod_{t=1}^{n}\left(1+E[i_t]\right)$$

$$=(1+j)^n \qquad \textbf{(1.5)}$$

since, for each value of t, $E[i_t]=j$.

With $k = 2$ in Equation (1.4) we obtain:

$$\begin{aligned} E[S_n^2] &= \prod_{t=1}^{n} E\left[(1+2i_t + i_t^2)\right] \\ &= \prod_{t=1}^{n} (1+2E[i_t]+E[i_t^2]) \\ &= (1+2j+j^2+s^2)^n \end{aligned} \tag{1.6}$$

since, for each value of t:

$$E[i_t^2] = (E[i_t])^2 + Var(i_t) = j^2 + s^2$$

The variance of S_n is:

$$\begin{aligned} Var(S_n) &= E[S_n^2] - (E[S_n])^2 \\ &= (1+2j+j^2+s^2)^n - (1+j)^{2n} \end{aligned} \tag{1.7}$$

from Equations (1.5) and (1.6).

Alternatively, we can write:

$$E\left[S_n^2\right] = \prod_{t=1}^{n} E[(1+i_t)^2] = \prod_{t=1}^{n} \left(\left(E\left[1+i_t\right]\right)^2 + Var(1+i_t)\right)$$

using a rearrangement of the variance formula: $Var(X) = E\left[X^2\right] - \left(E\left[X\right]\right)^2$.

Then, using the results that:

$$\left(E\left[1+i_t\right]\right)^2 = \left(1+E\left[i_t\right]\right)^2 = (1+j)^2 \quad \text{and} \quad Var(1+i_t) = Var(i_t) = s^2$$

for all values of t, we can write:

$$E\left[S_n^2\right] = \prod_{t=1}^{n} \left((1+j)^2 + s^2\right) = \left((1+j)^2 + s^2\right)^n$$

This means that:

$$Var(S_n) = \left((1+j)^2 + s^2\right)^n - (1+j)^{2n}$$

These arguments are readily extended to the derivation of the higher moments of S_n in terms of the higher moments of the distribution of the annual investment rates of return.

Question

Calculate the mean and variance of the accumulated value at the end of 25 years of an initial investment of £40,000, if the annual rate of return in year k is independent of that in any other year and $i_k \sim Gamma(16,200)$ for all k.

Solution

Since $i_k \sim Gamma(16,200)$, using the formulae for the mean and variance of the gamma distribution from page 12 of the *Tables*:

$$j = E[i_k] = \frac{\alpha}{\lambda} = \frac{16}{200} = 0.08$$

$$s^2 = Var(i_k) = \frac{\alpha}{\lambda^2} = \frac{16}{200^2} = (0.02)^2$$

So, the mean of the accumulated amount is:

$$40{,}000E[S_{25}] = 40{,}000(1+j)^{25} = 40{,}000 \times 1.08^{25} = £273{,}939$$

and the variance is:

$$\begin{aligned}
Var(40{,}000S_{25}) &= 40{,}000^2 Var(S_{25}) \\
&= 40{,}000^2\left([(1+j)^2 + s^2]^{25} - (1+j)^{50}\right) \\
&= 40{,}000^2[(1.08^2 + 0.02^2)^{25} - 1.08^{50}] \\
&= (£25{,}417)^2
\end{aligned}$$

Moments of A_n

Recall that A_n is a random variable that represents the accumulated value at time n of a series of annual investments, each of amount 1, at times $0,1,2,\dots,n-1$. $i_1, i_2, \dots, i_n$ are independent random variables, each with a mean j and a variance s^2.

From Equation (1.3):

$$\begin{aligned}
A_{n-1} = {} & (1+i_1)(1+i_2)\dots(1+i_{n-1}) \\
& + (1+i_2)\dots(1+i_{n-1}) \\
& \vdots \\
& + (1+i_{n-2})(1+i_{n-1}) \\
& + (1+i_{n-1})
\end{aligned}$$

Also:

$$\begin{aligned} A_n = (1+i_1)(1+i_2)\dots(1+i_{n-1})(1+i_n) \\ +(1+i_2)\dots(1+i_{n-1})(1+i_n) \\ \vdots \\ +(1+i_{n-1})(1+i_n) \\ +(1+i_n) \end{aligned}$$

It follows from Equation (1.3) (or from Equation (1.1)) that, for $n \geq 2$:

$$\boldsymbol{A_n = (1+i_n)(1+A_{n-1})} \qquad \textbf{(1.8)}$$

Equation (1.8) can also be deduced by general reasoning. A_{n-1} is the accumulated value at time $n-1$ of a series of annual payments, each of amount 1, at times $0, 1, 2, \dots, n-2$. The value, at time $n-1$, of the same series of payments together with an extra payment at time $n-1$ is $1+A_{n-1}$. Accumulating this value forward to time *n* gives $(1+i_n)(1+A_{n-1})$ and this is equivalent to A_n.

The usefulness of Equation (1.8) lies in the fact that, since A_{n-1} depends only on the values $i_1, i_2, \dots, i_{n-1}$, the random variables i_n and A_{n-1} are independent. (By assumption the yields each year are independent of one another.) Accordingly, Equation (1.8) permits the development of a recurrence relation from which may be found the moments of A_n. We illustrate this approach by obtaining the mean and variance of A_n.

Let:

$$\mu_n = E[A_n]$$

and let:

$$m_n = E[A_n^2]$$

Since:

$$A_1 = 1 + i_1$$

it follows that:

$$E[A_1] = E[1+i_1] = 1 + E[i_1] = 1 + j \quad \Rightarrow \quad \mu_1 = 1 + j$$

and:

$$m_1 = E[A_1^2] = E[(1+i_1)^2] = 1 + 2E[i_1] + E[i_1^2] \quad \Rightarrow \quad m_1 = 1 + 2j + j^2 + s^2$$

where, as before, *j* and s^2 are the mean and variance of the yield each year.

Taking expectations of Equation (1.8), we obtain (since i_n and A_{n-1} are independent):

$$\mu_n = (1+j)(1+\mu_{n-1}) \qquad n \geq 2$$

This equation, combined with initial value μ_1, implies that, for all values of n:

$$\mu_n = \ddot{s}_{\overline{n}|} \quad \text{at rate } j \tag{1.9}$$

We obtain this result by applying the recursive formula for μ_n repeatedly for each year:

$$\begin{aligned}
\mu_n &= (1+j)+(1+j)\mu_{n-1} \\
&= (1+j)+(1+j)^2[1+\mu_{n-2}] \\
&= (1+j)+(1+j)^2+(1+j)^2\mu_{n-2} \\
&= \dots \\
&= (1+j)+(1+j)^2+(1+j)^3+\cdots+(1+j)^n
\end{aligned}$$

Thus the expected value of A_n is simply $\ddot{s}_{\overline{n}|}$, calculated at the mean rate of return.

Recall that the symbol $\ddot{s}_{\overline{n}|}$ denotes the accumulated value at time n of payments of 1 at times 0, 1, 2,..., $n-1$. At the interest rate j, it is equal to:

$$\ddot{s}_{\overline{n}|} = (1+j)+(1+j)^2+(1+j)^3+\cdots+(1+j)^n = \frac{(1+j)^n - 1}{d}$$

where $d = \dfrac{j}{1+j}$.

Next, we consider the variance of A_n.

Since:

$$A_n^2 = (1+2i_n+i_n^2)(1+2A_{n-1}+A_{n-1}^2)$$

by taking expectations we obtain, for $n \geq 2$:

$$m_n = (1+2j+j^2+s^2)(1+2\mu_{n-1}+m_{n-1}) \tag{1.10}$$

As the value of μ_{n-1} is known (by Equation (1.9)), Equation (1.10) provides a recurrence relation for the calculation successively of $m_2, m_3, m_4, \dots$. The variance of A_n may be obtained as:

$$Var(A_n) = E[A_n^2] - (E[A_n])^2 = m_n - \mu_n^2 \tag{1.11}$$

In principle the above arguments are fairly readily extended to provide recurrence relations for the higher moments of A_n.

Question

A company considers that on average it will earn interest on its funds at the rate of 4% pa. However, the investment policy is such that in any one year the yield on the company's funds is equally likely to take any value between 2% and 6%.

For both single and annual premium accumulations with terms of 5, 10, 15, 20, and 25 years and single (or annual) investment of £1, find the mean accumulation and the standard deviation of the accumulation at the maturity date. (Ignore expenses.)

Solution

The annual rate of return is uniformly distributed on the interval $[0.02, 0.06]$. The corresponding probability density function is constant and equal to 25 (*ie* $1/(0.06-0.02)$). The mean annual rate of interest is clearly:

$$j = 0.04$$

and the variance of the annual rate of return is:

$$s^2 = \frac{1}{12}(0.06-0.02)^2 = \tfrac{4}{3}\times 10^{-4}$$

The formulae for the PDF, mean and variance of a uniform random variable are given on page 13 of the *Tables*.

We are required to find $E[A_n]$, $\left(Var(A_n)\right)^{\frac{1}{2}}$, $E[S_n]$, and $\left(Var(S_n)\right)^{\frac{1}{2}}$ for $n = 5$, 10, 15, 20 and 25.

Substituting the above values of j and s^2 in Equations (1.5) and (1.7), we immediately obtain the results for the single premiums.

For example:

$$E[S_5] = 1.04^5 = 1.21665$$

$$Var(S_5) = (1+0.08+0.04^2+\tfrac{4}{3}10^{-4})^5 - 1.04^{10} = 0.000913$$

$$\Rightarrow \quad \text{standard deviation } [S_5] = \sqrt{0.000913} = 0.03021$$

For the annual premiums we must use the recurrence relation (1.10) (with $\mu_{n-1} = \ddot{s}_{\overline{n-1|}}$ at 4%) together with Equation (1.11).

Equation (1.9) is used to calculate $E[A_n]$. For example:

$$E[A_5] = \ddot{s}_{\overline{5|}@4\%} = \frac{1.04^5 - 1}{0.04/1.04} = 5.63298$$

To calculate the standard deviation of A_5, we first need to calculate m_5 from the recursive formula:

$$m_n = (1+2j+j^2+s^2)(1+2\mu_{n-1}+m_{n-1})$$

starting with $\mu_1 = 1+j$ and $m_1 = 1+2j+j^2+s^2$.

The values required are tabulated below:

n	m_n	μ_n
1	1.08173	1.04
2	4.50189	2.1216
3	10.54158	3.24646
4	19.50853	4.41632
5	31.73933018	5.6329755

Therefore, using Equation (1.11):

$$Var(A_5) = m_5 - (\mu_5)^2 = 31.73933018 - 5.6329755^2 = 0.0089172$$

So the standard deviation is:

$$0.0089172^{½} = 0.09443$$

This answer is very sensitive to rounding.

The results are summarised in Table 1. It should be noted that, for both annual and single premiums, the standard deviation of the accumulation increases rapidly with the term.

Term (years)	Single investment £1		Annual investment £1	
	Mean accumulation (£)	Standard deviation (£)	Mean accumulation (£)	Standard deviation (£)
5	1.21665	0.03021	5.63298	0.09443
10	1.48024	0.05198	12.48635	0.28353
15	1.80094	0.07748	20.82453	0.57899
20	2.19112	0.10886	30.96920	1.00476
25	2.66584	0.14810	43.31174	1.59392

Table 1

This approach is also easily extended to provide recurrence relations for other series of investments. Let F_t again represent the accumulated amount at time *t* of all money invested *before* time *t* and let P_t be the amount of money invested at time *t*. We stated in Equation (1.1) that:

$$F_t = (1+i_t)(F_{t-1} + P_{t-1})$$

Therefore the mean of F_t can be found from the recursive relationship:

$$E[F_0] = 0$$

$$E[F_t] = (1+j)(E[F_{t-1}] + P_{t-1})$$

Question

An investor invests 1 unit at time $t=0$ and a further 2 units at time $t=2$. The expected rate of return in each year is 10%. Calculate the accumulated value of the fund at time $t=5$:

(i) using recursive formulae, and assuming the varying rate model applies

(ii) using the corresponding deterministic model.

Comment on your answers.

Solution

(i) Using a recursive approach, where F_t represents the accumulated amount at time *t* of all money invested *before* time *t*, we obtain:

$$E[F_0] = 0$$

$$E[F_1] = 1.1 \times [E[F_0] + P_0] = 1.1 \times (0+1) = 1.1$$

$$E[F_2] = 1.1 \times [E[F_1] + P_1] = 1.1 \times (1.1+0) = 1.21$$

$$E[F_3] = 1.1 \times [E[F_2] + P_2] = 1.1 \times (1.21+2) = 3.531$$

$$E[F_4] = 1.1 \times [E[F_3] + P_3] = 1.1 \times (3.531+0) = 3.8841$$

$$E[F_5] = 1.1 \times [E[F_4] + P_4] = 1.1 \times (3.8841+0) = 4.27251$$

(ii) Using the corresponding deterministic model, with an annual rate of return of 10%, gives:

$$1.1^5 + 2 \times 1.1^3 = 4.27251$$

So the corresponding deterministic model gives the same answer.

2 The lognormal distribution

In general a theoretical analysis of the distribution functions for A_n and S_n is somewhat difficult, even in the relatively simple situation when the yields each year are independent and identically distributed. There is, however, one special case for which an exact analysis of the distribution function for S_n is particularly simple.

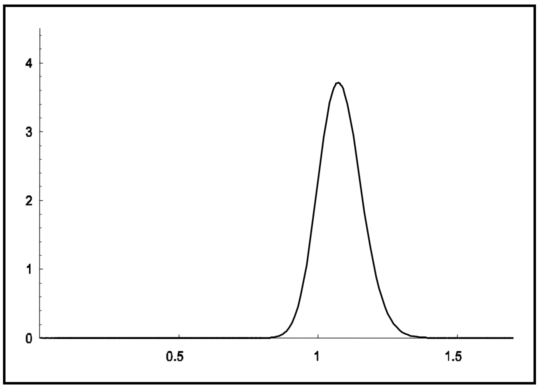
Due to the compounding effect of investment returns, the accumulated value of an investment bond grows multiplicatively. This makes the lognormal distribution a natural choice for modelling the annual growth factors $1+i$, since a lognormal random variable can take any positive value and has the following multiplicative property:

If $X_1 \sim \log N(\mu_1, \sigma_1^2)$ and $X_2 \sim \log N(\mu_2, \sigma_2^2)$ are independent random variables, then:

$$X_1 X_2 \sim \log N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

The graph below illustrates the shape of the PDF of a typical lognormal distribution used to model annual growth factors, $1+i$.

$\log N(0.075, 0.1^2)$



Question

Given that $1+i \sim \log N(0.075, 0.1^2)$, calculate the mean and standard deviation of the annual growth rate, i.

Solution

Using the formulae on page 14 of the *Tables*:

$$E[1+i] = e^{\mu + \frac{1}{2}\sigma^2} = e^{0.075 + \frac{1}{2}\times 0.1^2} = 1.0833$$

and: $$Var(1+i) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) = e^{2\times 0.075 + 0.1^2}(e^{0.1^2} - 1) = 0.1086^2$$

So, the annual growth rate will have a mean value of:

$$E[i]=8.33\%$$

and a standard deviation of:

$$sd(i)=sd(1+i)=10.86\%$$

Suppose that the random variable $\log(1+i_t)$ is normally distributed with mean μ and variance σ^2. In this case, the variable $(1+i_t)$ is said to have a lognormal distribution with parameters μ and σ^2.

So, if $\log(1+i_t)\sim N(\mu,\sigma^2)$, then $1+i_t \sim \log N(\mu,\sigma^2)$.

Equation (1.2) is equivalent to:

$$\log S_n = \sum_{t=1}^{n} \log(1+i_t)$$

The sum of a set of independent normal random variables is itself a normal random variable. Hence, when the random variables $(1+i_t)$ $(t \geq 1)$ are independent and each has a lognormal distribution with parameters and μ and σ^2, the random variable S_n has a lognormal distribution with parameters $n\mu$ and $n\sigma^2$.

Since the distribution function of a lognormal variable is readily written down in terms of its two parameters, in the particular case when the distribution function for the yield each year is lognormal we have a simple expression for the distribution function of S_n .

So $S_n \sim \log N(n\mu, n\sigma^2)$ or $\log S_n \sim N(n\mu, n\sigma^2)$, and the distribution function of S_n can therefore be written:

$$P(S_n \leq s)=\Phi\left(\frac{\log s - n\mu}{\sigma\sqrt{n}}\right)$$

Question

A man now aged exactly 50 has built up a savings fund of £400,000. In order to retire at age 60, he will require a fund of at least £600,000 at that time. The annual returns on the fund, i , are independent and identically distributed, with $1+i \sim \log N(0.075, 0.1^2)$.

Calculate the probability that, if the man makes no further contributions to the fund, he will be able to retire at age 60.

Solution

If the man makes no further contributions, his accumulated fund at age 60 will be $400,000S_{10}$.

So, the probability that the fund will be sufficient for him to retire is:

$$P(400,000S_{10} \geq 600,000) = 1 - P\left(S_{10} \leq \frac{600,000}{400,000}\right)$$

$$= 1 - \Phi\left(\frac{\log 1.5 - 10\mu}{\sigma\sqrt{10}}\right)$$

$$= 1 - \Phi(-1.0895) = \Phi(1.0895) = 0.862$$

Similarly for the present value of a sum of 1 due at the end of n years:

$$V_n = (1+i_1)^{-1} \dots (1+i_n)^{-1}$$

$$\Rightarrow \log V_n = -\log(1+i_1) - \cdots - \log(1+i_n)$$

Since, for each value of t, $\log(1+i_t)$ is normally distributed with mean μ and variance σ^2, each term on the right-hand side of the above equation is normally distributed with mean $-\mu$ and variance σ^2. Also the terms are independently distributed. So, $\log V_n$ is normally distributed with mean $-n\mu$ and variance $n\sigma^2$. That is, V_n has lognormal distribution with parameters $-n\mu$ and $n\sigma^2$.

So $V_n \sim \log N(-n\mu, n\sigma^2)$ or $\log V_n \sim N(-n\mu, n\sigma^2)$, and the distribution function of V_n can therefore be written:

$$P(V_n \leq s) = \Phi\left(\frac{\log s - (-n\mu)}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{\log s + n\mu}{\sigma\sqrt{n}}\right)$$

By statistically modelling S_n and V_n, it is possible to answer questions such as:

- **to a given point in time, for a specified confidence interval, what is the range of values for an accumulated investment**
- **what is the maximum loss which will be incurred with a given level of probability.**

It can also be noted that these techniques may be extended to calculate the risk metrics such as VaR, as introduced in a previous chapter, of a series of investments.

Question

The annual returns on a fund, i, are independent and identically distributed. Each year, the distribution of $1+i$ is lognormal with parameters $\mu = 0.075$ and $\sigma^2 = 0.025^2$.

Calculate the upper and lower quartiles for the accumulated value at the end of 5 years of an initial investment of £1,000.

Solution

By definition, the accumulated amount $1{,}000S_5$ will exceed the upper quartile u with probability 25%, *ie*:

$$0.75 = P(1{,}000S_5 \le u) = P(S_5 \le u/1{,}000)$$

So, using the formula for the distribution function of S_5:

$$0.75 = P(S_5 \le u/1{,}000) = \Phi\left(\frac{\log(u/1{,}000) - 5\mu}{\sigma\sqrt{5}}\right)$$

From page 162 of the *Tables*, we find that $\Phi(0.6745) = 0.75$. So, we must have:

$$\frac{\log(u/1{,}000) - 5\mu}{\sigma\sqrt{5}} = 0.6745 \quad ie \quad u = 1{,}000e^{5\mu + 0.6745\sigma\sqrt{5}} = £1{,}511$$

Similarly, the lower quartile is:

$$l = 1{,}000e^{5\mu - 0.6745\sigma\sqrt{5}} = £1{,}401$$

So far in this section, we have assumed that the annual growth factors in each year are independent, *ie* we have assumed that the varying rate model applies. We can also use a lognormal distribution for annual growth factors in conjunction with the fixed rate model.

Suppose that the rate of return on an investment, i, is currently unknown, but once determined, it will be the same in all future years. In this case:

$$S_n = (1+i)^n$$

If $1+i \sim \log N(\mu, \sigma^2)$, then:

$$\log(1+i) \sim N(\mu, \sigma^2)$$

and: $\log(1+i)^n = n\log(1+i) \sim N(n\mu, n^2\sigma^2)$

So: $S_n = (1+i)^n \sim \log N(n\mu, n^2\sigma^2)$

Note that the distribution of S_n obtained here is different to that derived for the varying rate model, where $S_n \sim \log N(n\mu, n\sigma^2)$, which applies in the case where the returns in each year are independent.

Question

A lump sum of \$14,000 will be invested at time 0 for 4 years at an annual rate of return, i. The rate of return, once determined, will be the same in each of the four years. $1+i$ has a lognormal distribution with mean 1.05 and variance 0.007.

Calculate the probability that the investment will accumulate to more than \$20,000 in 4 years' time.

Solution

We first need to find the values of the parameters for the lognormal distribution. Using the formulae for the mean and variance from the *Tables*:

$$e^{\mu+\frac{1}{2}\sigma^2} = 1.05 \qquad \text{(Equation 1)}$$

and: $$e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) = 0.007 \qquad \text{(Equation 2)}$$

Squaring Equation 1, we have:

$$\left(e^{\mu+\frac{1}{2}\sigma^2}\right)^2 = e^{2\mu+\sigma^2} = 1.05^2$$

Substituting this into Equation 2 gives:

$$1.05^2\left(e^{\sigma^2}-1\right) = 0.007 \quad \Rightarrow \quad \sigma^2 = \ln\left(\frac{0.007}{1.05^2}+1\right) = 0.006329$$

So, using Equation 1:

$$\mu = \ln(1.05) - \tfrac{1}{2}\sigma^2 = 0.04563$$

We know that $1+i \sim \log N(\mu,\sigma^2)$. Since we have a constant rate of return over the 4 years, the fixed rate model applies. Letting S_n denote the accumulated value at time n of an investment of 1 made at time 0, we have $S_n \sim \log N(n\mu, n^2\sigma^2)$.

We need the probability $P\left(14{,}000 S_4 > 20{,}000\right)$ where:

$$S_4 \sim \log N(4\mu, 16\sigma^2) = \log N(0.1825, 0.1013)$$

So:

$$P(14{,}000S_4 > 20{,}000) = P(S_4 > 1.429)$$

$$= P(\ln S_4 > 0.3567)$$

$$= P\left(Z > \frac{0.3567 - 0.1825}{\sqrt{0.1013}}\right)$$

$$= P(Z > 0.547)$$

$$= 1 - \Phi(0.547) = 0.2922$$

Chapter 5 Summary

A stochastic model of investment returns provides information about the distribution of financial outcomes. This distribution can be used to find best estimates and probabilities.

The varying rate model and the fixed rate model provide formulae for the mean and variance of the accumulated amount of a fund or the present value of a future payment.

Varying rate model (single premium)

$$E\left[S_n\right]=(1+j)^n \qquad Var(S_n)=[(1+j)^2+s^2]^n-(1+j)^{2n}$$

Fixed rate model (single premium)

$$E\left[S_n\right]=E\left[(1+i)^n\right] \qquad Var(S_n)=E\left[(1+i)^{2n}\right]-(E[(1+i)^n])^2$$

For the varying rate model, the variance and higher moments of the accumulated amount of a series of payments can be calculated using recursive formulae.

Varying rate model (annual premium)

$$E\left[A_n\right]=\ddot{s}_{\overline{n}|} \qquad \text{at rate } j$$

Recursive formulae for $E\left[A_n\right]$:

$$E\left[A_0\right]=0$$

and:

$$E\left[A_k\right]=(1+j)\left(1+E\left[A_{k-1}\right]\right) \qquad (k=1,2,\ldots,n)$$

Recursive formulae for $Var(A_n)$:

$$E\left[A_0^2\right]=0$$

and:

$$E\left[A_k^2\right]=\left((1+j)^2+s^2\right)\left(1+2E\left[A_{k-1}\right]+E\left[A_{k-1}^2\right]\right) \qquad (k=1,2,\ldots,n)$$

Then:

$$Var(A_n)=E\left[A_n^2\right]-\left(E\left[A_n\right]\right)^2$$

For the fixed rate model with annual premiums, calculate the mean and variance directly from the definitions.

The lognormal distribution can be used to model the annual growth factor, $1+i$. This allows probabilities to be determined in terms of the distribution function of the normal distribution.

The lognormal model formulae for the varying rate model are:

$$S_n \sim \log N(n\mu, n\sigma^2) \qquad P(S_n \le s) = \Phi\left(\frac{\log s - n\mu}{\sigma\sqrt{n}}\right)$$

$$V_n \sim \log N(-n\mu, n\sigma^2) \qquad P(V_n \le s) = \Phi\left(\frac{\log s - (-n\mu)}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{\log s + n\mu}{\sigma\sqrt{n}}\right)$$

Chapter 5 Practice Questions

5.1 Investment returns on a given fund are fixed over the next five years at either 4% *pa* with probability 0.2 or 5% *pa* with probability 0.8.

Calculate the standard deviation of the present value of a payment of £25,000 due to be made from the fund in 5 years' time.

5.2 A stochastic model of investment returns assumes that the annual rate of return during the next year will be 7.5% and that the rate of return in subsequent years will be at a fixed but unknown level with probabilities in accordance with the following probability distribution:

$$i = \begin{cases} 5.5\% & \text{with probability } 0.3 \\ 7.5\% & \text{with probability } 0.5 \\ 9.5\% & \text{with probability } 0.2 \end{cases}$$

Calculate the expected accumulated amount at the end of the fifth year of an initial investment of £20,000.

5.3 A stochastic model of investment returns assumes that the annual rates of return in different years are independent and identically distributed normal random variables with mean 8% and standard deviation 2%.

Calculate the mean and standard deviation of the accumulated value, at time 2, of an initial investment of £10,000.

5.4 An investment analyst wishes to model the annual rate of growth i of an investment fund using a probability distribution of the form:

$$i = \begin{cases} 5\% & \text{with probability } 3k \\ 7.5\% & \text{with probability } 1-4k \\ 12.5\% & \text{with probability } k \end{cases}$$

where k is a suitable constant.

Determine the maximum and minimum values that can be obtained for the mean and standard deviation of i using this family of distributions.

5.5 An investor intends to invest three lump sums in an investment account, one at the start of each of the next 3 years, and is interested in the amount to which the combined payments will have accumulated by the end of the third year. The amounts of the lump sums are shown in the table below.

Calculate the mean and variance of the accumulated amount assuming returns in each year are independent and the mean and standard deviation of the rate of return in each year are as shown in the table:

Year	Lump sum invested	Mean rate of return	SD of rate of return
1	£50,000	8%	2%
2	£30,000	7%	3%
3	£20,000	6%	4%

5.6 Exam style The annual returns, i, on a fund are independent and identically distributed, with a mean of 6% and a standard deviation of 3%. Each year, the distribution of $1+i$ is lognormal with parameters μ and σ^2.

(i) Calculate the values of μ and σ^2. [4]

(ii) Calculate the probability that the accumulation of a single investment of £1 will be greater than 110% of its expected value after 10 years. [4]

[Total 8]

5.7 Exam style £200 is invested for 12 years. In any year the yield on the investment will be 3% with probability 0.25, 5% with probability 0.6 and 6% with probability 0.15, and is independent of the yield in any other year.

(i) Calculate the mean accumulation at the end of 12 years. [2]

(ii) Calculate the standard deviation of the accumulation at the end of 12 years. [4]

[Total 6]

5.8 In any year, the rate of return on funds invested with a particular company has mean value i and standard deviation s, and is independent of the rates of return in all previous years.

Exam style

(i) Derive formulae for the mean and the variance of the accumulated value after n years of a single investment of 1 at time 0. [5]

(ii) Let i_t be the rate of return earned in the t th year. Each year the value of $(1+i_t)$ is lognormally distributed, with parameters $\mu = 0.04$ and $\sigma^2 = 0.09$.

(a) Show that n, the number of years that must elapse before the accumulation of a lump sum invested at time 0 has a 75% probability of at least doubling in size, satisfies:

$$0.04n - 0.2024\sqrt{n} - \ln 2 = 0$$

(b) Hence calculate the value of n. [7]

[Total 12]

5.9 The annual rates of return on an investment fund are assumed to be independent and identically distributed. Each year the distribution of $1+i$ is lognormal with parameters $\mu = 0.07$ and $\sigma^2 = 0.006$, where i is the annual yield on the fund.

Exam style

Calculate the amount that should be invested in the fund immediately to ensure an accumulated value of at least £500,000 in ten years' time with a probability of 0.99. [6]

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 5 Solutions

5.1 The present value is:

- $\dfrac{25{,}000}{1.04^5} = 20548.18$ with probability 0.2, or
- $\dfrac{25{,}000}{1.05^5} = 19588.15$, with probability 0.8.

The mean of these values is:

$$20{,}548.18 \times 0.2 + 19{,}588.15 \times 0.8 = 19{,}780.16 .$$

So, the standard deviation of the present value is:

$$\sqrt{20{,}548.18^2 \times 0.2 + 19{,}588.15^2 \times 0.8 - 19780.16^2} = £384$$

5.2 The probability distribution of the accumulated amount at the end of 5 years is:

$$S_5 = \begin{cases} 20{,}000 \times 1.075 \times 1.055^4 = £26{,}635 & \text{with probability } 0.3 \\ 20{,}000 \times 1.075 \times 1.075^4 = £28{,}713 & \text{with probability } 0.5 \\ 20{,}000 \times 1.075 \times 1.095^4 = £30{,}910 & \text{with probability } 0.2 \end{cases}$$

So the expected accumulated amount is:

$$E[S_5] = 0.3 \times 26{,}635 + 0.5 \times 28{,}713 + 0.2 \times 30{,}910 = £28{,}529$$

5.3 The mean accumulated amount at the end of 2 years is:

$$E[10{,}000 S_2] = 10{,}000 \times E[S_2] = 10{,}000(1+j)^2 = 10{,}000 \times 1.08^2 = £11{,}664$$

The variance of the accumulated amount at the end of 2 years is:

$$\begin{aligned} Var(10{,}000 S_2) &= 10{,}000^2 Var(S_2) \\ &= 10{,}000^2[((1+j)^2 + s^2)^2 - (1+j)^4] \\ &= 10{,}000^2[(1.08^2 + 0.02^2)^2 - 1.08^4] \\ &= 93{,}328 \end{aligned}$$

So the standard deviation is:

$$\sqrt{93{,}328} = £305.50$$

Note that we don't use the fact that the rates of return are normally distributed here. To apply the varying rate model, all we need to know is the mean and variance of the annual rate of return.

5.4 In order for this to be a sensible probability distribution, all the probabilities must lie between 0 and 1. So k must lie in the range $0 \le k \le 0.25$.

The mean growth rate is:

$$E[i] = 3k \times 0.05 + (1-4k) \times 0.075 + k \times 0.125 = 0.075 - 0.025k$$

So the minimum value (corresponding to $k = 0.25$) is 6.875% and the maximum value (corresponding to $k = 0$) is 7.5%.

The variance of the growth rate is:

$$Var(i) = E\left[i^2\right] - \left(E[i]\right)^2$$

$$= 3k \times 0.05^2 + (1-4k) \times 0.075^2 + k \times 0.125^2 - (0.075 - 0.025k)^2$$

$$= 0.004375k - 0.000625k^2$$

Differentiating with respect to k:

$$\frac{d}{dk} Var(i) = 0.004375 - 0.00125k$$

Setting this equal to 0, we see that the only turning point of the variance function is at:

$$k = 0.004375 / 0.00125 = 3.5$$

This is outside the permissible range of values of k. So the variance function must be monotonic over the range of interest to us. Therefore, the minimum standard deviation (corresponding to $k = 0$) is 0% and the maximum standard deviation (corresponding to $k = 0.25$) is 3.25%.

5.5 To find the mean of A_3, the accumulated amount at the end of the third year, we need to use the recursive relationships:

$$E[A_0] = 0$$

$$E[A_k] = (1 + j_k)\left(P_k + E[A_{k-1}]\right) \qquad (k = 1,2,3)$$

where P_k denotes the lump sum invested at the start of year k and j_k is the mean rate of return for year k.

This gives:

$$E[A_1] = 1.08(50{,}000 + 0) = 54{,}000$$

$$E[A_2] = 1.07(30{,}000 + 54{,}000) = 89{,}880$$

$$E[A_3] = 1.06(20{,}000 + 89{,}880) = 116{,}473$$

So the mean accumulated amount is £116,473.

To find the variance of A_3, we need to use the recursive relationships:

$$E[A_0^2]=0$$

$$E[A_k^2]=\left((1+j_k)^2+s_k^2\right)\left(P_k^2+2P_kE[A_{k-1}]+E[A_{k-1}^2]\right) \qquad (k=1,2,3)$$

where s_k is the standard deviation of the rate of return for year k.

This gives:

$$E[A_1^2]=(1.08^2+0.02^2)[50{,}000^2+2(50{,}000)(0)+0]=2{,}917{,}000{,}000$$

$$E[A_2^2]=(1.07^2+0.03^2)[30{,}000^2+2(30{,}000)(54{,}000)+2{,}917{,}000{,}000]$$
$$=8{,}085{,}910{,}600$$

$$E[A_3^2]=(1.06^2+0.04^2)[20{,}000^2+2(20{,}000)(89{,}880)+8{,}085{,}910{,}600]$$
$$=13{,}593{,}665{,}650$$

Then:

$$Var(A_3)=E[A_3^2]-(E[A_3])^2=13{,}593{,}665{,}650-116{,}473^2=(£5{,}264)^2$$

So the standard deviation of the accumulated amount is £5,264.

5.6 (i) ***Parameters of the lognormal distribution***

We know that $1+i\sim \log N(\mu,\sigma^2)$, and that $E[i]=0.06$ and $Var(i)=0.03^2$. Using the formulae for the mean and variance of the lognormal distribution from the *Tables*:

$$E[1+i]=1+E[i]=1.06=e^{\mu+\frac{1}{2}\sigma^2} \qquad \text{(Equation 1)} \qquad [1]$$

and: $$Var(1+i)=Var(i)=0.03^2=e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) \qquad \text{(Equation 2)} \qquad [1]$$

Squaring Equation 1, we have:

$$\left(e^{\mu+\frac{1}{2}\sigma^2}\right)^2=e^{2\mu+\sigma^2}=1.06^2$$

Substituting this into Equation 2 gives:

$$1.06^2\left(e^{\sigma^2}-1\right)=0.03^2 \;\Rightarrow\; \sigma^2=\ln\left(\frac{0.03^2}{1.06^2}+1\right)=0.00080068 \qquad [1]$$

So, using Equation 1:

$$\mu = \ln(1.06) - \tfrac{1}{2}\sigma^2 = 0.057869 \qquad [1]$$

[Total 4]

(ii) ***Probability***

Let S_{10} denote the accumulated value at time 10 of an investment of £1 at time 0. We require the probability $P(S_{10} > 1.1E[S_{10}])$.

As the rate of return in each year is independent of that in other years, using the varying rate model, we have:

$$E[S_{10}] = (1+j)^{10} = 1.06^{10} \quad \text{since} \quad j = E[i] = 0.06 \qquad [1]$$

The distribution of S_{10} under this model is:

$$S_{10} \sim \log N\left(10\mu, 10\sigma^2\right) = \log N(0.57869, 0.0080068) \qquad [1]$$

So:

$$\begin{aligned} P(S_{10} > 1.1E[S_{10}]) &= P\left(S_{10} > 1.1 \times 1.06^{10}\right) \\ &= P\left(\ln S_{10} > \ln\left(1.1 \times 1.06^{10}\right)\right) \\ &= P\left(Z > \frac{\ln\left(1.1 \times 1.06^{10}\right) - 0.57869}{\sqrt{0.0080068}}\right) \\ &= P(Z > 1.110) \\ &= 1 - \Phi(1.110) \\ &= 0.1335 \end{aligned} \qquad [2]$$

[Total 4]

5.7 (i) First of all, we need to find the mean, j, of the rate of return in each year:

$$j = (0.03 \times 0.25) + (0.05 \times 0.6) + (0.06 \times 0.15) = 0.0465 \qquad [1]$$

Let S_{12} represent the accumulated amount at time 12 of an investment of 1 at time 0. So:

$$E[200S_{12}] = 200(1+j)^{12} = 200(1.0465)^{12} = £345.06 \qquad [1]$$

[Total 2]

(ii) To calculate the variance of the accumulated amount, we first need to calculate the variance, s^2, of the rate of return in each year.

$$s^2 = (0.03^2 \times 0.25) + (0.05^2 \times 0.6) + (0.06^2 \times 0.15) - 0.0465^2 = 0.00010275 \qquad [1]$$

So:

$$Var\left(200S_{12}\right) = 200^2 Var(S_{12})$$

$$= 200^2 \left[\left((1+j)^2 + s^2 \right)^{12} - \left((1+j)^{12} \right)^2 \right]$$

$$= 200^2 \left[\left(1.0465^2 + 0.00010275 \right)^{12} - \left(1.0465 \right)^{24} \right]$$

$$= 134.13 \qquad [2]$$

So the standard deviation is £11.58. [1]

[Total 4]

5.8 (i) Let S_n be the accumulated value, after n years, of a single investment of 1 at time 0. Let i_t be the rate of return earned in the t th year. Then:

$$E\left[S_n\right] = E\left[(1+i_1)(1+i_2)\dots(1+i_n)\right] = E\left[1+i_1\right]E\left[1+i_2\right]\dots E\left[1+i_n\right] = (1+j)^n \qquad [1]$$

and:

$$E\left[S_n^2\right] = E\left[(1+i_1)^2(1+i_2)^2\dots(1+i_n)^2\right]$$

$$= E\left[(1+i_1)^2\right]E\left[(1+i_2)^2\right]\dots E\left[(1+i_n)^2\right] \qquad [1]$$

Now, since $E[X^2] = \left(E[X]\right)^2 + Var(X)$:

$$E\left[(1+i_k)^2\right] = \left(E[1+i_k]\right)^2 + Var(1+i_k) = \left(1+E[i_k]\right)^2 + Var(i_k) = (1+j)^2 + s^2 \qquad [1]$$

So:

$$E\left[S_n^2\right] = \left[(1+j)^2 + s^2\right]^n \qquad [1]$$

Hence:

$$Var(S_n) = \left[(1+j)^2 + s^2\right]^n - (1+j)^{2n} \qquad [1]$$

[Total 5]

(ii)(a) Now:

$$S_n = (1+i_1)(1+i_2)\ldots(1+i_n)$$

So:

$$\ln S_n = \ln\left[(1+i_1)(1+i_2)\ldots(1+i_n)\right]$$
$$= \ln(1+i_1) + \ln(1+i_2) + \cdots + \ln(1+i_n)$$ [1]

Since $\ln(1+i_t) \sim N(0.04, 0.09)$, using the additive property of independent normal distributions means that $\ln S_n \sim N(0.04n, 0.09n)$. [1]

The probability that the initial investment at least doubles is:

$$P(S_n \geq 2) = 0.75 \quad \Rightarrow \quad P(S_n < 2) = 0.25$$ [1]

Standardising gives:

$$P(S_n < 2) = P\left(Z < \frac{\ln 2 - 0.04n}{\sqrt{0.09n}}\right) = 0.25$$

where $Z \sim N(0,1)$, so using page 162 of the *Tables*, we have:

$$\frac{\ln 2 - 0.04n}{\sqrt{0.09n}} = -0.6745$$ [1]

Rearranging:

$$\ln 2 - 0.04n = -0.6745 \times 0.3\sqrt{n}$$

$$0.04n - 0.2024\sqrt{n} - \ln 2 = 0$$ [1]

(ii)(b) Letting $x = \sqrt{n}$, the equation becomes:

$$0.04x^2 - 0.2024x - \ln 2 = 0$$

Solving the quadratic equation:

$$x = \frac{0.2024 \pm \sqrt{0.2024^2 + 0.16\ln 2}}{0.08} = -2.3413, 7.4013$$ [1]

So:

$$n = 7.4013^2 = 54.8 \text{ years}$$ [1]

[Total 7]

5.9 Let X be the amount invested at time 0, and let i_k denote the yield obtained in year k, $k=1,2,...,10$. The accumulated value of the fund after 10 years is then:

$$X(1+i_1)(1+i_2)...(1+i_{10})=XS_{10}$$

This will be at least £500,000 if and only if:

$$S_{10} \geq \frac{500,000}{X} \qquad [1]$$

Since the yields are independent, it follows that:

$$S_{10} \sim \log N(10\times 0.07, 10\times 0.006)=\log N(0.7, 0.06) \qquad [1]$$

Hence:

$$P\left[S_{10} \geq \frac{500,000}{X}\right]=P\left[\ln S_{10} \geq \ln\left(\frac{500,000}{X}\right)\right]$$

$$=P\left[Z \geq \frac{\ln\left(\frac{500,000}{X}\right)-0.7}{\sqrt{0.06}}\right] \qquad [1]$$

Since this probability is 0.99, $\dfrac{\ln\left(\frac{500,000}{X}\right)-0.7}{\sqrt{0.06}}$ must be the lower 1% point of the standard normal distribution, *ie*:

$$\frac{\ln 500,000-\ln X-0.7}{\sqrt{0.06}}=-2.3263 \qquad [2]$$

Solving for X gives:

$$\ln X=\ln 500,000-0.7+2.3263\sqrt{0.06}=12.992188 \quad \Rightarrow \quad X=438,970.80$$

So the amount to be invested (rounded to the nearest £1) is £438,971. [1]

[Total 6]

6

Portfolio theory

Syllabus objectives

4.1 Mean-variance portfolio theory

4.1.1 Describe and discuss the assumptions of mean-variance portfolio theory.

4.1.2 Discuss the conditions under which application of mean-variance portfolio theory leads to the selection of an optimum portfolio.

4.1.3 Calculate the expected return and risk of a portfolio of many risky assets, given the expected return, variance and covariance of returns of the individual assets, using mean-variance portfolio theory.

4.1.4 Explain the benefits of diversification using mean-variance portfolio theory.

0 Introduction

This chapter introduces *mean-variance portfolio theory*, which is also called *modern portfolio theory (MPT)* or just *portfolio theory*. As well as being very important in its own right, MPT forms the basis of the capital asset pricing model discussed later in the course.

MPT assumes that investment decisions are based solely upon risk and return – more specifically the mean and variance of investment return – and that investors are willing to accept higher risk in exchange for higher expected return. This can be consistent with the maximisation of expected utility as outlined earlier in the course, if the investor is assumed to have a utility function that only uses the mean and variance of investment returns, such as the quadratic utility function. It can also be consistent if the distribution of investment returns is a function only of its mean and variance.

Based upon these and other assumptions, MPT specifies a method for an investor to construct a portfolio that gives the maximum return for a specified risk (variance), or the minimum risk for a specified return. Such portfolios are described as *efficient*.

Section 1 of the chapter focuses upon the means by which to determine:

- the *opportunity set* of available portfolios and the corresponding risk-return combinations between which the investor must choose and
- the set of efficient portfolios, the *efficient frontier*.

A rational investor who prefers more to less and is risk-averse will always choose an efficient portfolio.

Section 2 introduces the benefits of diversification and shows that if the returns from all the assets in a portfolio are statistically independent then the variance of the portfolio return tends towards zero when the number of assets in that portfolio is increased.

Note carefully that:

- Figures 6.1 to 6.3 included within this chapter are all part of the Core Reading for this topic.
- Within the context of mean-variance portfolio theory, risk is defined very specifically as the *variance* – or equivalently standard deviation – of investment returns. Elsewhere in this and other courses we discuss other possible measures or types of risk that might be relevant depending upon the exact context considered.

Some sections of this chapter have been adapted from lecture notes originally written by David Wilkie.

Question

List other possible types of risk that might be relevant in an investment context.

Solution

There are *many* other types of investment-related risk, which are discussed in detail in this and other subjects. Amongst the more important are:

- *default* or *credit risk* – the other party to an investment deal fails to fulfil their obligations
- *inflation risk* – inflation is higher than anticipated, so reducing real returns
- *exchange rate* or *currency risk* – exchange rates move in an unanticipated way
- *reinvestment risk* – stems from the uncertainty concerning the terms on which investment income can be reinvested
- *marketability risk* – the risk that you might be unable to realise the true value of an investment if it is difficult to find a buyer.

1 Portfolio theory

1.1 Introduction

Mean-variance portfolio theory, sometimes called modern portfolio theory (MPT), specifies a method for an investor to construct a portfolio that gives the maximum return for a specified risk, or the minimum risk for a specified return.

If the investor's utility function is known, then MPT allows the investor to choose the portfolio that has the optimal balance between return and risk, as measured by the variance of return, and consequently maximises the investor's expected utility.

However, the theory relies on some strong and limiting assumptions about the properties of portfolios that are important to investors. For example, it uses variance to measure risk and so penalises gains as well as losses. In the form described here the theory ignores the investor's liabilities, although it is possible to extend the analysis to include them.

By ignoring *actuarial risk* – the risk that the investor fails to meet his or her liabilities – the theory as presented here is severely limited. In its defence, MPT was the first real attempt to use statistical techniques to show the benefit of diversification for investors.

The application of the mean-variance framework to portfolio selection falls conceptually into two parts:

1. **First, the definition of the properties of the portfolios available to the investor – the opportunity set.**

   Here we are looking at the risk and return of the possible portfolios available to the investor.

2. **Second, the determination of how the investor chooses one out of all the feasible portfolios in the opportunity set.**

   That is, the determination of the investor's *optimal* portfolio from those available.

1.2 Assumptions of mean-variance portfolio theory

The application of mean-variance portfolio theory is based on some important assumptions:

- all expected returns, variances and covariances of pairs of assets are known
- investors make their decisions purely on the basis of expected return and variance
- investors are non-satiated
- investors are risk-averse
- there is a fixed single-step time period
- there are no taxes or transaction costs
- assets may be held in any amounts, *ie* short-selling is possible, we can have infinitely divisible holdings, and there are no maximum investment limits.

We will meet these assumptions again at various points throughout this chapter.

1.3 Specification of the opportunity set

In specifying the opportunity set it is necessary to make some assumptions about how investors make decisions. Then the properties of portfolios can be specified in terms of relevant characteristics. It is assumed that investors select their portfolios on the basis of:

- **the expected return and**
- **the variance of that return**

over a single time horizon. Thus all the relevant properties of a portfolio can be specified with just two numbers – the mean return and the variance of the return. The variance (or standard deviation) is known as the risk of the portfolio.

So, according to MPT, variance of return and expected return are all that matter – this is a key assumption. Other key factors that might influence the investment decision in practice are ignored. These include:

- the suitability of the asset(s) for an investor's liabilities
- the marketability of the asset(s)
- higher moments of the distribution of returns such as skewness and kurtosis
- taxes and investment expenses
- restrictions imposed by legislation
- restrictions imposed by the fund's trustees.

Finally, we should note that the length of the time horizon, which is likely to vary between investors, is not specified.

To calculate the mean and variance of return for a portfolio it is necessary to know the expected return on each individual security and also the variance/covariance matrix for the available universe of securities.

The variance/covariance matrix shows the covariance between each pair of the variables. So, if there are three variables, 1, 2 and 3 say, then the matrix has the form:

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

where c_{ij} is the covariance between variables *i* and *j* (for $i \neq j$), and c_{ii} is the variance of variable i. Since $c_{ij} = c_{ji}$, the matrix is symmetric about the leading diagonal.

Note that this means that with *N* different securities an investor must specify:

- *N* expected returns
- *N* variances of return
- $\dfrac{N(N-1)}{2}$ covariances.

Whilst estimates of the required parameters can be obtained using historical data, these are unlikely to prove reliable predictors of the future behaviour of investment returns and it may be necessary to adjust the historical estimates in the light of the other factors.

Question

Assuming that there are 350 shares in an equity index (as there are in the FTSE 350), calculate the number of items of data that need to be specified for an investor to apply MPT.

Solution

The required number of items of data is:

$$350+350+\frac{350\times 349}{2}=61{,}775$$

This requirement for an investor to make thousands of estimates of covariances is potentially a major limitation of mean-variance portfolio theory in its general form.

However, we will see in later chapters that:

- multifactor models and single-index models have been developed in an attempt to reduce the data requirements
- the capital asset pricing model avoids this problem.

Efficient portfolios

Two further assumptions about investor behaviour allow the definition of efficient portfolios.

We have already introduced these in the assumptions of mean-variance portfolio theory above.

The assumptions are:

- **Investors are never satiated. At a given level of risk, they will always prefer a portfolio with a higher return to one with a lower return.**
- **Investors dislike risk. For a given level of return they will always prefer a portfolio with lower variance to one with higher variance.**

This means that investors prefer 'more to less' and that they are risk averse.

A portfolio is *inefficient* if the investor can find another portfolio with the same (or higher) expected return and lower variance, or the same (or lower) variance and higher expected return.

A portfolio is *efficient* if the investor cannot find a better one in the sense that it has either a higher expected return and the same (or lower) variance or a lower variance and the same (or higher) expected return.

That is, an *efficient* portfolio is one that isn't *inefficient*. Thus, every portfolio – including those that consist of a single asset – is either efficient or inefficient.

Once the set of efficient portfolios has been identified all others can be ignored.

This is because an investor who is risk-averse and prefers more to less will never choose an inefficient portfolio. The set of efficient portfolios is known as the *efficient frontier*.

However, an investor may be able to rank efficient portfolios by using a utility function, as shown below.

This will determine the investor's optimal portfolio.

If we know an individual's utility function then we can describe their attitude towards risk and return. If the assumption that investors make their decisions purely on the basis of expected return and variance holds, then this attitude towards risk and return can equally be described by indifference curves. Indifference curves join points of equal expected utility in expected return-standard deviation space, *ie* portfolios that an individual is indifferent between.

Suppose an investor can invest in any of the N securities, $i = 1,\ldots N$. A proportion x_i is invested in security S_i.

Note that:

- x_i is a proportion of the total sum to be invested
- given infinite divisibility, x_i can take any real value, subject to the restriction that $\sum_{i=1}^{N} x_i = 1$
- we have not specified the nature of the N securities.

The return on the portfolio R_P is modelled as a random variable that is a linear combination of random variables, R_i, representing the return on security S_i, so that:

$$R_P = \sum_{i=1}^{N} x_i R_i$$

So the portfolio return is a weighted average of the individual security returns.

The expected return on the portfolio is:

$$E = E[R_P] = \sum_{i=1}^{N} x_i E_i$$

where E_i is the expected return on security S_i.

The variance is:

$$V = Var(R_P) = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i x_j C_{ij}$$

where C_{ij} is the covariance of the returns on securities S_i and S_j and we write $C_{ii} = V_i$.

So, the lower the covariance between security returns, the lower the overall variance of the portfolio. This means that the variance of a portfolio can be reduced by investing in securities whose returns are uncorrelated, *ie* by diversification.

The covariance C_{ij} is equal to $\rho_{ij}\sigma_i\sigma_j$, where:

σ_i = standard deviation of security S_i returns

σ_j = standard deviation of security S_j returns

ρ_{ij} = correlation coefficient between security S_i returns and security S_j returns

The case of two securities

We can derive expressions for the mean and the variance of portfolio returns in simple cases.

If there are just two securities, S_A and S_B, the above expressions reduce to:

$$E = x_A E_A + x_B E_B$$

and $$V = x_A^2 V_A + x_B^2 V_B + 2x_A x_B C_{AB}.$$

Question

Consider a portfolio consisting of equal holdings of two securities, S_x and S_y, where:

- the return on S_x is equally likely to be 5% or 10% *pa*
- the return on S_y is equally likely to be 10% or 20% *pa*.

(i) Calculate the means and variances of returns on each individual security.

(ii) Calculate the mean and variance of the return on the portfolio as a whole, given that the correlation coefficient of the two securities is:

(a) 1

(b) 0

(c) –1

(d) 0.7

(iii) Comment on your results.

Solution

(i) ***Means and variances for individual securities***

The mean returns on each security are:

$$E(S_x) = 0.5 \times 0.05 + 0.5 \times 0.10 = 0.075 \quad ie \ 7.5\%$$

$$E(S_y) = 0.5 \times 0.10 + 0.5 \times 0.20 = 0.15 \quad ie \ 15\%$$

The variances can be calculated using the formula:

$$V(S_i) = E(S_i^2) - [E(S_i)]^2 \quad \text{for} \quad i = x, y$$

where:

$$E(S_x^2) = 0.5 \times 0.05^2 + 0.5 \times 0.10^2 = 0.00625$$

$$E(S_y^2) = 0.5 \times 0.10^2 + 0.5 \times 0.20^2 = 0.025$$

So:

$$V(S_x) = 0.00625 - 0.075^2 = 0.000625 \quad ie \ (2.5\%)^2$$

$$V(S_y) = 0.025 - 0.15^2 = 0.0025 \quad ie \ (5\%)^2$$

(ii) ***Mean and variance of portfolio as a whole***

The mean return on the portfolio in each case is given by:

$$E(R_p) = 0.5 \times E(S_x) + 0.5 \times E(S_y)$$

ie $E(R_p) = 0.5 \times 0.075 + 0.5 \times 0.15 = 0.1125 \quad ie \ 11.25\%$

To find the variance of return on the portfolio we use the relationship:

$$V(R_p) = 0.5^2 V(S_x) + 0.5^2 V(S_y) + 2Cov(0.5S_x, 0.5S_y)$$

$$= 0.25V(S_x) + 0.25V(S_y) + 2(0.5)(0.5)Cov(S_x, S_y)$$

(a) $\boldsymbol{\rho_{xy} = 1}$

When the correlation coefficient equals 1, we have:

$$Cov(S_x, S_y) = 1 \times 0.025 \times 0.05 = 0.00125$$

So: $V(R_p) = 0.25V(S_x) + 0.25V(S_y) + 2(0.5)(0.5)Cov(S_x, S_y)$

$$= 0.25 \times 0.000625 + 0.25 \times 0.0025 + 0.5 \times 0.00125 = 0.00140625 \quad ie \ (3.75\%)^2$$

(b) $\boldsymbol{\rho_{xy} = 0}$

When the correlation coefficient equals 0, we have:

$$Cov(S_x, S_y) = 0 \times 0.025 \times 0.05 = 0$$

and: $V(R_p) = 0.25 \times 0.000625 + 0.25 \times 0.0025 = 0.00078125 \quad ie \ (2.795\%)^2$

(c) $\boldsymbol{\rho_{xy} = -1}$

When the correlation coefficient equals –1, we have:

$$Cov(S_x, S_y) = -1 \times 0.025 \times 0.05 = -0.00125$$

and: $V(R_p) = 0.25 \times 0.000625 + 0.25 \times 0.0025 - 0.5 \times 0.00125 = 0.00015625 \quad ie \ (1.25\%)^2$

(d) $\rho_{xy} = \mathbf{0.7}$

When the correlation coefficient = 0.7, we have:

$$Cov(S_x, S_y) = 0.7 \times 0.025 \times 0.05 = 0.000875$$

and: $V(R_p) = 0.25 \times 0.000625 + 0.25 \times 0.0025 + 0.5 \times 0.000875 = 0.00121875$ *ie* $(3.491\%)^2$

(iii) ***Comment***

The more closely correlated the investments in the portfolio, the larger the variance. The highest result was obtained when the securities were assumed to be perfectly correlated (*ie* the correlation coefficient was +1) and the lowest result when the securities were assumed to be perfectly negatively correlated (*ie* the correlation coefficient was –1).

This is to be expected. For example, with negative correlation, the potential deviations from the expected return of each security separately will tend to cancel each other out, giving a smaller overall portfolio deviation from the overall expected return of the portfolio.

We now return to the case where there are two securities, S_A and S_B.

As the proportion invested in S_A is varied, a curve is traced in $E-V$ space. The minimum variance can easily be shown to occur when:

$$\boldsymbol{x_A = \frac{V_B - C_{AB}}{V_A - 2C_{AB} + V_B}}$$

Question

Prove the above result.

Solution

The variance of the portfolio return is:

$$\begin{aligned} V_P &= x_A^2 V_A + x_B^2 V_B + 2x_A x_B C_{AB} \\ &= x_A^2 V_A + (1-x_A)^2 V_B + 2x_A(1-x_A)C_{AB} \end{aligned}$$

We want to choose the value for x_A that minimises the variance V_P. To do this, we differentiate and set to zero:

$$\frac{dV_P}{dx_A} = 2x_A V_A + 2(1-x_A)V_B(-1) + 2(1-x_A)C_{AB} - 2x_A C_{AB} = 0$$

$$\Leftrightarrow \quad 2x_A V_A - 2V_B + 2x_A V_B + 2C_{AB} - 4x_A C_{AB} = 0$$

$$\Leftrightarrow \quad x_A(V_A + V_B - 2C_{AB}) = V_B - C_{AB}$$

So, the minimum variance occurs when:

$$x_A = \frac{V_B - C_{AB}}{V_A - 2C_{AB} + V_B}$$

As an example, consider the case where:

$E_A = 4\%$ $\quad V_A = 4\%\%$ $\quad (\sigma_A = 2\%)$

$E_B = 8\%$ $\quad V_B = 36\%\%$ $\quad (\sigma_B = 6\%)$

We now let the correlation coefficient between the two securities vary by considering ρ_{AB} equal to –0.75, 0, and +0.75 in turn. The results are plotted in Figure 6.1, where the vertical axis represents expected values of return and the horizontal axis represents standard deviation of return. In this space $(E-\sigma)$ the curves representing possible portfolios of the two securities are hyperbolae. It is possible to plot the same results in $E-V$ space, where the lines would be parabolae.

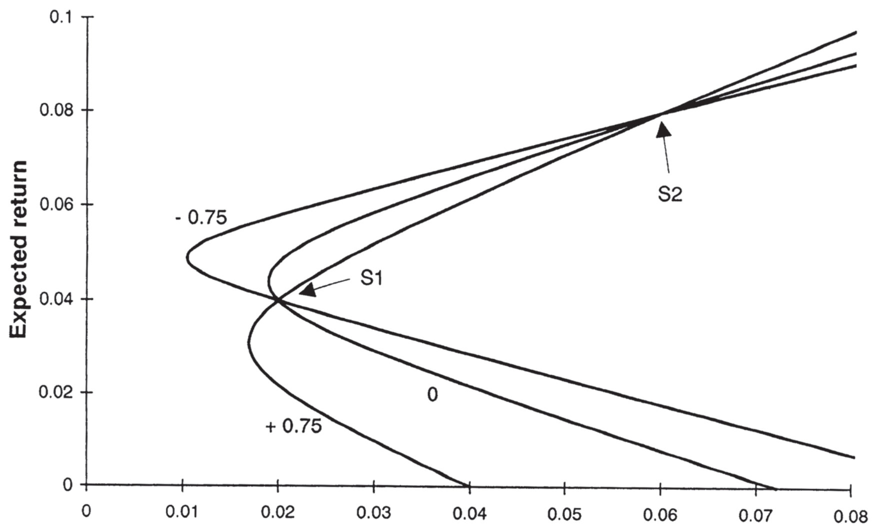


Figure 6.1

Note that S_A and S_B are denoted by S1 and S2 respectively in Figures 6.1 and 6.2, *ie* the points S1 and S2 represent the expected return and standard deviation for portfolios consisting entirely of S_A and S_B respectively.

Also note that in each case, the set of efficient portfolios consists only of those portfolios above the point of minimum variance.

Question

Calculate the co-ordinates of the point of minimum variance in Figure 6.1 in the case when $\rho_{AB} = 0$, and comment on this result.

Solution

If $\rho_{AB} = 0$, then the formula for the proportion of funds invested in asset A reduces to:

$$x_A = \frac{V_B}{V_A + V_B}$$

So in this case:

$$x_A = \frac{0.0036}{0.0004 + 0.0036} = 0.9 \quad \text{and} \quad x_B = 0.1$$

The resulting portfolio has expected return and variance given by:

$$E_P = x_A E_A + x_B E_B = 0.9 \times 0.04 + 0.1 \times 0.08 = 0.044 \quad \textit{ie} \ \ 4.4\%$$

and:

$$\begin{aligned}\sigma_P^2 &= x_A^2 V_A + x_B^2 V_B + 2 x_A x_B C_{AB} \\ &= 0.81 \times 0.0004 + 0.01 \times 0.0036 + 2 \times 0.9 \times 0.1 \times 0 \\ &= 0.01897^2\end{aligned}$$

ie $\sigma_P = 1.9\%$

This is less than both of the individual security standard deviations of 2% and 6%, illustrating the benefits of diversification.

Figure 6.2 below shows combinations of securities with correlation coefficients of +1, 0 and –1. For coefficients of +1 and -1, it is possible to obtain risk-free portfolios with zero standard deviation of return.

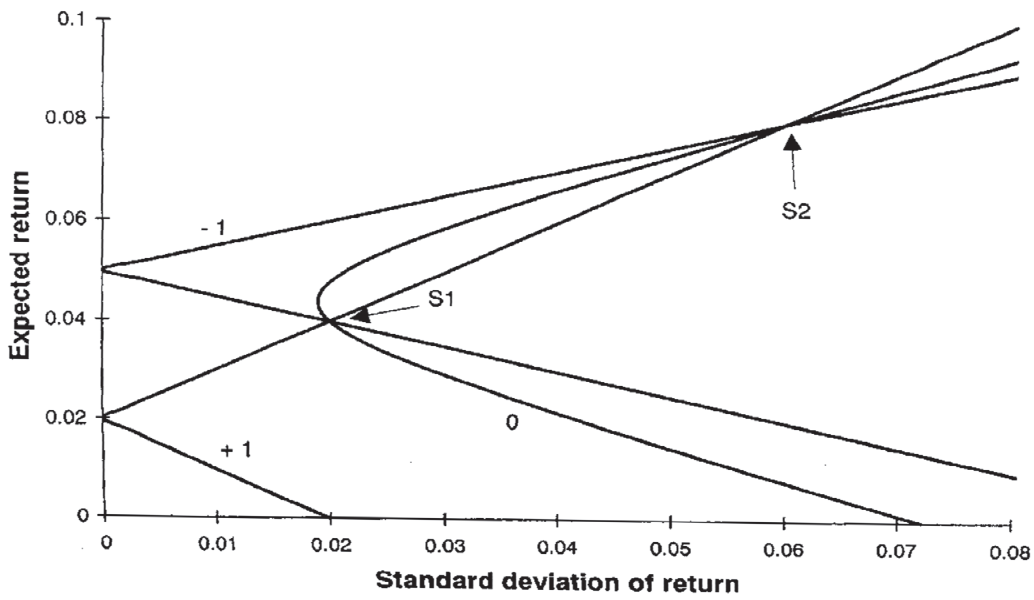


Figure 6.2

Note that if ρ is equal to +1 or –1, then there exists a *risk-free* portfolio with *V* = 0. If $\rho = -1$, it involves positive holdings of both securities, *ie* $x_1, x_2 > 0$. If $\rho = +1$, it involves a negative holding of S_B (S2) and a positive holding of S_A (S1).

Question

Show that with two assets the efficient frontier is a straight line in the case where $\rho = 1$, as shown in Figure 6.2.

Solution

We can use the formula for the expected return of the portfolio to express x_A in terms of E_P.

$$E_P = x_A E_A + x_B E_B = x_A E_A + (1 - x_A) E_B$$

$$\Leftrightarrow \qquad x_A = \frac{E_P - E_B}{E_A - E_B}$$

The variance of the portfolio return is:

$$V_P = x_A^2 V_A + x_B^2 V_B + 2x_A x_B C_{AB} = x_A^2 \sigma_A^2 + x_B^2 \sigma_B^2 + 2x_A x_B \sigma_A \sigma_B \rho$$

If $\rho = 1$ (*ie* there is perfect positive correlation), then this simplifies to:

$$V_P = x_A^2 \sigma_A^2 + x_B^2 \sigma_B^2 + 2x_A x_B \sigma_A \sigma_B = (x_A \sigma_A + x_B \sigma_B)^2$$

Taking square roots:

$$\begin{aligned}\sigma_P &= x_A\sigma_A + x_B\sigma_B \\ &= x_A\sigma_A + (1-x_A)\sigma_B \\ &= \frac{E_P - E_B}{E_A - E_B}\sigma_A + \frac{E_A - E_P}{E_A - E_B}\sigma_B \\ &= \frac{\sigma_A - \sigma_B}{E_A - E_B}E_P + \frac{\sigma_B E_A - \sigma_A E_B}{E_A - E_B}\end{aligned}$$

This is a straight line in (E_P, σ_P) space.

When there are *N* securities, the aim is to choose x_i to minimise *V* subject to the constraints:

$$\sum_{i=1}^{N} x_i = 1$$

and $E = E_P$, say, in order to plot the minimum-variance curve.

The aim is to choose the proportions to invest in each possible security in a way that minimises risk subject to the constraints that:

- all the investor's money is invested somewhere and
- the expected return on the portfolio is set equal to the desired level.

An alternative approach would be to *maximise E* subject to:

$$\sum_{i=1}^{N} x_i = 1 \quad \text{and} \quad V = V_P \text{ say}$$

However, the first approach is usually easier.

Note carefully that *E* and *V* without the subscripts are the *portfolio* expected return or variance, *ie* the quantities that we are optimising, and that E_P and V_P are the specified values used in the *constraints*.

Lagrangian multipliers

One way of solving such a minimisation problem is the method of Lagrangian multipliers.

The Lagrangian approach – in which we set up and optimise a suitable Lagrangian function – can be used to solve constrained optimisation problems such as we are considering here, provided that:

- they are static, *ie* involve only a single time period
- the constraints are all strict equalities.

The basic idea is that if we wish to:

- maximise or minimise some function $f(x_1,\ldots,x_N)$
- subject to a set of M constraints $g_j(x_1,\ldots,x_N)=c_j$, $j = 1, \ldots, M$
- by choice of N variables x_i , $i = 1, \ldots , N$

then we can set up a Lagrangian function of the form:

$$W = f(x_1,\ldots,x_N) - \sum_{j=1}^{M} \lambda_j \left(g_j(x_1,\ldots,x_N) - c_j \right)$$

It turns out that maximising/minimising $f(x_1,\ldots,x_N)$ subject to the constraints $g_j(x_1,\ldots,x_N)=c_j$, $j = 1,\ldots, M$ is the same as maximising/minimising W with respect to $x_1,\ldots,x_N$ and $\lambda_1,\ldots,\lambda_M$.

Thus we set the derivative of W with respect to each x_i and λ_j to zero. This gives us a set of $M+N$ equations in $M+N$ unknowns. Under suitable conditions these equations can be solved simultaneously to find the optimal values of the x_i , and also the λ_j . Note that the equations $\dfrac{\partial W}{\partial \lambda_j}=0$ are just the M constraint equations themselves.

In this instance we wish to:

- minimise the portfolio variance, V
- subject to the two constraints $\sum_{i=1}^{N} x_i = 1$ and $E = E_P$
- by choice of the securities x_i , $i = 1,2,\ldots,N$.

The Lagrangian function is:

$$\boldsymbol{W = V - \lambda(E - E_P) - \mu\left(\sum_{i=1}^{N} x_i - 1\right)}$$

or $$W = \sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}x_i x_j - \lambda\left(\sum_{i=1}^{N} E_i x_i - E_P\right) - \mu\left(\sum_{i=1}^{N} x_i - 1\right)$$

where:

- *V, E* and x_i are defined as above
- E_P and 1 are the constraining constants and
- λ and μ are known as the *Lagrangian multipliers*.

Question

Recall the example in the Core Reading involving securities S_A and S_B illustrated in Figure 6.1, in which:

$$E_A = 4\%, \quad V_A = 4\%\%$$

$$E_B = 8\%, \quad V_B = 36\%\%$$

Write down the Lagrangian function *W* in the case where the correlation coefficient is $\rho_{AB} = 0.75$.

Solution

Here the Lagrangian function is:

$$W = x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + 2x_A x_B \rho_{AB}\sigma_A\sigma_B - \lambda(E_A x_A + E_B x_B - E_P) - \mu(x_A + x_B - 1)$$

$$W = 4x_A^2 + 36x_B^2 + 2x_A x_B \times 0.75 \times 2 \times 6 - \lambda(4x_A + 8x_B - E_P) - \mu(x_A + x_B - 1)$$

ie $$W = 4x_A^2 + 36x_B^2 + 18x_A x_B - \lambda(4x_A + 8x_B - E_P) - \mu(x_A + x_B - 1)$$

To find the minimum, we set the partial derivatives of *W* with respect to all the x_i and λ and μ equal to zero. The result is a set of linear equations that can be solved.

The equations are linear because:

- the variance of portfolio returns is a quadratic function of the x_i and
- the constraint terms are linear in each of x_i, λ and μ.

Hence, the first-order derivatives of W contain powers of x_i no higher than one.

The partial derivative of *W* with respect to x_i is:

$$\frac{\partial W}{\partial x_i} = 2\sum_{j=1}^{N} x_j C_{ij} - \lambda E_i - \mu$$

To see where the $2\sum_{j=1}^{N} x_j C_{ij}$ term comes from in the above expression, we start from the fact that for an *N*-security portfolio, the variance is equal to:

$$V = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j C_{ij}$$

Writing out the covariance matrix in full gives:

$$\begin{bmatrix} x_1^2 C_{11} & x_1 x_2 C_{12} & \cdots & x_1 x_i C_{1i} & \cdots & x_1 x_N C_{1N} \\ x_2 x_1 C_{21} & x_2^2 C_{22} & \cdots & x_2 x_i C_{2i} & \cdots & x_2 x_N C_{2N} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_i x_1 C_{i1} & x_i x_2 C_{i2} & \cdots & x_i^2 C_{ii} & \cdots & x_i x_N C_{iN} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_N x_1 C_{N1} & x_N x_2 C_{N2} & \cdots & x_N x_i C_{Ni} & \cdots & x_N^2 C_{NN} \end{bmatrix}$$

If we differentiate with respect to x_i, the *i*th row of the matrix gives us the following terms:

(1) $x_1 C_{i1} + x_2 C_{i2} + \ldots + 2x_i C_{ii} + \ldots + x_N C_{iN}$

Also from the *i*th column, each of the other $j = 1,\ldots,i-1,i+1,\ldots,N$ rows, yields an $x_j C_{ji}$ term, *ie* the following terms:

(2) $x_1 C_{1i} + x_2 C_{2i} + \ldots + x_{i-1} C_{i-1,i} + x_{i+1} C_{i+1,i} + \ldots + x_N C_{Ni}$

Summing (1) and (2), and noting that $C_{ij} = C_{ji}$, we obtain:

$$2x_1 C_{i1} + 2x_2 C_{i2} + \ldots + 2x_i C_{ii} + \ldots + 2x_N C_{iN}$$

which we can write as:

$$2\sum_{j=1}^{N} x_j C_{ij}$$

The partial derivative of W with respect to λ is:

$$\frac{\partial \boldsymbol{W}}{\partial \lambda} = -\left(\sum_{i=1}^{N} \boldsymbol{E}_i \boldsymbol{x}_i - \boldsymbol{E}_P\right)$$

and with respect to μ is:

$$\frac{\partial \boldsymbol{W}}{\partial \mu} = -\left(\sum_{i=1}^{N} \boldsymbol{x}_i - 1\right)$$

Setting each of these to zero gives:

$$2\sum_{j=1}^{N} \boldsymbol{x}_j \boldsymbol{C}_{ij} - \lambda \boldsymbol{E}_i - \mu = 0 \quad \textbf{(one equation for each of } \boldsymbol{N} \textbf{ securities)}$$

$$\sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{E}_i = \boldsymbol{E}_P$$

$$\sum_{i=1}^{N} \boldsymbol{x}_i = 1$$

These $N+2$ equations in $N+2$ unknowns (first-order conditions) can be solved to find the optimal values of the security proportions, *ie* the x_i's, as functions of the portfolio expectation E_P. These functions can then be substituted into the expression for the portfolio variance, the resulting expression for the portfolio variance as a function of the portfolio expectation being the equation of the minimum variance curve. The top half of this curve, *ie* above the point of global minimum variance, is the efficient frontier.

Question

Write down the above conditions for the scenario in the previous question, where we had:

$$E_A = 4\%, \quad V_A = 4\%\%$$

$$E_B = 8\%, \quad V_B = 36\%\%$$

and $\rho_{AB} = 0.75$, giving a Lagrangian function of:

$$W = 4x_A^2 + 36x_B^2 + 18x_A x_B - \lambda(4x_A + 8x_B - E_P) - \mu(x_A + x_B - 1)$$

Solution

Differentiating the Lagrangian function gives the equations:

$$\frac{\partial W}{\partial x_A} = 8x_A + 18x_B - 4\lambda - \mu = 0$$

$$\frac{\partial W}{\partial x_B} = 18x_A + 72x_B - 8\lambda - \mu = 0$$

$$\frac{\partial W}{\partial \lambda} = -\left(4x_A + 8x_B - E_P\right) = 0$$

$$\frac{\partial W}{\partial \mu} = -\left(x_A + x_B - 1\right) = 0$$

We now generalise to any *E* and *V*.

This is instead of considering specific values of E_P and V_P. In other words, we now look at (E,V) as E is allowed to vary.

The solution to the problem shows that the minimum variance *V* is a quadratic in *E* and each x_i is linear in *E*.

The usual way of representing the results of the above calculations is by plotting the minimum standard deviation for each value of E_P as a curve in expected return-standard deviation $(E - \sigma)$ space.

Recall that this is what we did for the two-security case in Figures 6.1 and 6.2 earlier in the chapter. As V is quadratic in E, the resulting minimum-variance curve is a parabola in $E - V$ space, and the minimum-standard deviation curve is a hyperbola in $E - \sigma$. Hereafter we always consider $E - \sigma$ space unless stated otherwise.

In this space, with expected return on the vertical axis, the *efficient frontier* is the part of the curve lying above the point of the global minimum of standard deviation. All other possible portfolios are inefficient.

According to MPT, inefficient portfolios should never be chosen by a rational, non-satiated, risk-averse investor.

Question

Solve the set of equations obtained in the previous question to derive an expression for the efficient frontier when S_A and S_B are the only two securities available and they have a correlation coefficient equal to 0.75.

Solution

In order to minimise the portfolio variance V with respect to the portfolio expected return E we need to solve the following equations derived in the solution to the previous question:

$$\frac{\partial W}{\partial x_A} = 8x_A + 18x_B - 4\lambda - \mu = 0$$

$$\frac{\partial W}{\partial x_B} = 18x_A + 72x_B - 8\lambda - \mu = 0$$

$$\frac{\partial W}{\partial \lambda} = -\left(4x_A + 8x_B - E\right) = 0$$

$$\frac{\partial W}{\partial \mu} = -\left(x_A + x_B - 1\right) = 0$$

Here we need to find x_A and x_B as (linear) functions of E. In general terms, this requires us to solve the above 4 simultaneous equations in 4 unknowns x_A, x_B, λ and μ. In fact, in the two-security case we have here we can simply use the last two equations to obtain x_A and x_B (as we are not interested in solving for λ and μ).

These give:

$$x_A = \frac{8-E}{4}, \quad x_B = \frac{E-4}{4}$$

Substituting these back into the expression for the variance:

$$V = 4x_A^2 + 36x_B^2 + 18x_A x_B$$

and simplifying gives:

$$V = \frac{1}{8}\left(11E^2 - 68E + 128\right)$$

As expected, V is a quadratic function of E. In $E-\sigma$ space, the equation of the minimum-standard deviation curve is therefore:

$$\sigma = \sqrt{\frac{1}{8}\left(11E^2 - 68E + 128\right)}$$

This curve is shown in Figure 6.1.

The efficient frontier is that part of this curve above the point of global minimum standard deviation. As minimising the standard deviation is equivalent to minimising the variance, to find this point we can differentiate the expression for V with respect to E:

$$\frac{dV}{dE} = \frac{1}{8}\left(22E - 68\right) = 0 \quad \Rightarrow \quad E = \frac{68}{22} = 3.0909 \quad ie \;\; 3.0909\%$$

So the efficient frontier is:

$$\sigma = \sqrt{\frac{1}{8}\left(11E^2 - 68E + 128\right)} \quad \text{for} \quad E \geq 3.0909$$

Recall that in the MPT framework investors make their decisions purely on the basis of expected return and variance. If the returns on securities are normally distributed, these returns can be characterised by just the mean and the variance, so in this case MPT could be used to select optimal portfolios.

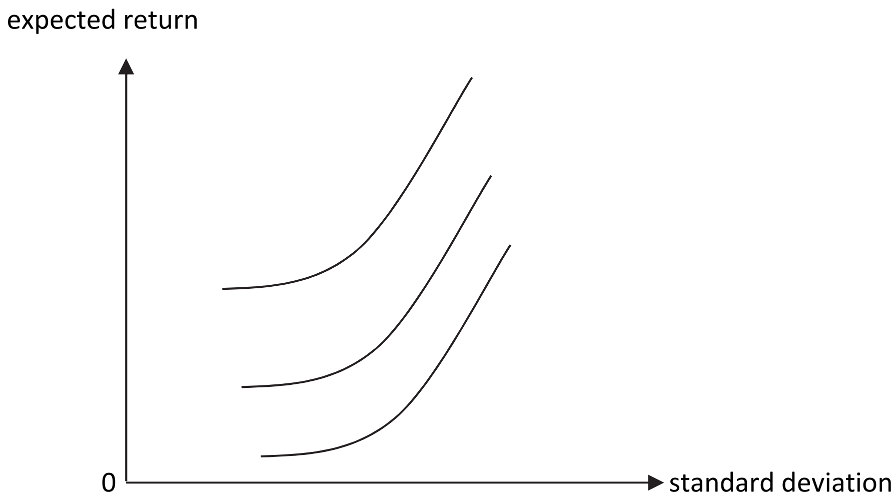
In fact, it can be shown that normality of returns is not a necessary condition for the selection of optimal portfolios. There is a more general class of distributions called the elliptically symmetrical family, which also result in optimality. All the distributions in this class have the property that the higher order moments can be expressed in terms of just their mean and variance.

1.4 Choosing an efficient portfolio

Recall from Section 1.3 that indifference curves join points of equal expected utility in expected return-standard deviation space, *ie* portfolios that an individual is indifferent between. Note that it is *expected utility* because we are considering situations involving uncertainty.

A series of indifference curves (curves which join all outcomes of equal expected utility) can be plotted in expected return-standard deviation space. Portfolios lying along a single curve all give the same value of expected utility and so the investor would be indifferent between them.

The following diagram shows typical indifference curves in $E - \sigma$ space.



Question

Explain why an investor's indifference curves slope upwards, and what determines their gradient.

Solution

An investor's indifference curves slope upwards because we are assuming that the investor is risk-averse and prefers more to less. Consequently, additional expected return yields extra utility, whereas additional risk reduces utility. Thus, any increase (decrease) in risk/standard deviation must be offset by an increase (decrease) in expected return in order to maintain a constant level of expected utility.

The gradient of the indifference curves is determined by the degree of the investor's risk aversion. The more risk-averse the investor, the steeper the indifference curves – as the investor will require a greater increase in expected return in order to offset any extra risk.

By combining the investor's indifference curves with the efficient frontier of portfolios, we can determine the investor's optimal portfolio, *ie* the portfolio that maximises the investor's expected utility.

Utility is maximised by choosing the portfolio on the efficient frontier at the point where the frontier is at a tangent to an indifference curve.

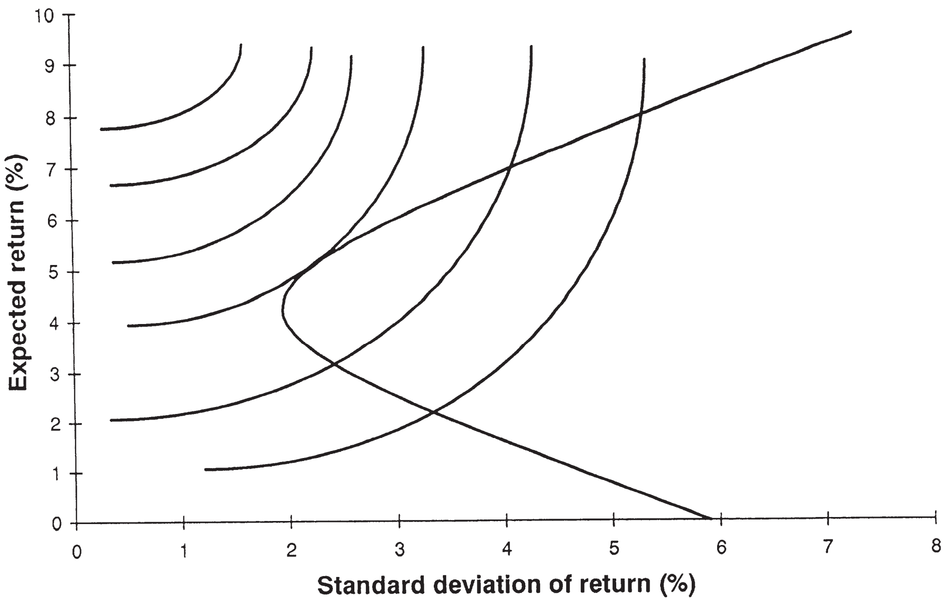


Figure 6.3

Question

Explain why the optimal portfolio on the efficient frontier is at the point where the frontier is at a tangent to an indifference curve.

Solution

The optimal portfolio occurs at the point where the indifference curve is tangential to the efficient frontier for the following two reasons:

1. The indifference curves that correspond to a higher level of expected utility are unattainable as they lie strictly above the efficient frontier.

2. Conversely, lower indifference curves that cut the efficient frontier are attainable, but correspond to a lower level of expected utility.

The highest attainable indifference curve, and corresponding highest level of expected utility, is therefore the one that is tangential to the efficient frontier. The optimal portfolio occurs at the tangency point, which is in fact the only attainable point on this indifference curve, which is why it is optimal.

For quadratic utility functions the process described above produces optimal portfolios whatever the distribution of returns, because expected utility is uniquely determined if we know the mean and variance of the distribution.

Recall from utility theory that, if the investor has a quadratic utility function, their attitude towards risk can be fully characterised by just the mean and the variance of return. Hence, when maximising expected utility by the choice of portfolio, the investor is concerned only with the first two moments of the investment returns yielded by the available portfolios and ignores all other factors.

Optimal portfolios are also produced for any utility function if investment returns are assumed to be normally distributed. This is important because investment returns are often modelled using a normal distribution.

If it is felt that the assumptions leading to a two-dimensional mean-variance type portfolio selection model are inappropriate, it is possible to construct models with higher dimensions. For example, skewness could be used in addition to expected return and a dispersion measure. It would then be necessary to consider an efficient *surface* in three dimensions rather than an efficient frontier in two. Clearly, the technique can be extended to more than three dimensions.

Although such models have been constructed, they do not appear to be widely used. The additional mathematical complexity, input data requirements and difficulty of interpretation building on the naïve assumptions are not compensated by real improvements in value added.

2 Benefits of diversification

Recall from Section 1.3 that the variance of the portfolio is:

$$V = Var(R_P) = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i x_j C_{ij}$$

This expression can be rewritten as:

$$\boldsymbol{V = \sum_{i=1}^{N} x_i^2 V_i + \sum_{j=1}^{N} \sum_{\substack{i=1 \\ i \neq j}}^{N} x_i x_j C_{ij}}$$

Recall also from Section 1.3 that, the lower (*ie* closer to zero) the covariance between security returns, the lower the overall variance of the portfolio. This means that the variance of a portfolio can be reduced by investing in securities whose returns are uncorrelated.

Where all assets are independent, the covariance between them is zero and the formula for variance becomes:

$$\boldsymbol{V = \sum_{i=1}^{N} x_i^2 V_i}$$

If we assume that equal amounts are invested in each asset, then with *N* assets the proportion invested in each is $1/N$. Thus:

$$\boldsymbol{V = \sum_{i=1}^{N} (1/N)^2 V_i = (1/N) \sum_{i=1}^{N} (1/N) V_i = \frac{\overline{V}}{N}}$$

where $\overline{V}$ represents the average variance of the stocks in the portfolio. As *N* gets larger and larger, the variance of the portfolio approaches zero. This is a general result – if we have enough *independent* assets, the variance of a portfolio of these assets approaches zero.

In other words, a lower variance, *ie* lower risk, can be achieved by diversification.

In general, the correlation coefficient and the covariance between assets is positive.

If you read the financial press or watch the business news, you'll usually find that the commentators think that the market as a whole is either doing well or badly. This suggests that investment returns tend to move together, *ie they are positively correlated*.

In these markets, the risk on the portfolio cannot be made to go to zero, but can be much less than the variance of an individual asset. With equal investment, the proportion invested in any one asset x_i is $1/N$ and the formula for the variance of the portfolio becomes:

$$\boldsymbol{V = \sum_{i=1}^{N} (1/N)^2 V_i + \sum_{j=1}^{N} \sum_{\substack{i=1 \\ i \neq j}}^{N} (1/N)(1/N) C_{ij}}$$

Factoring out $1/N$ from the first summation and $(N-1)/N$ from the second gives:

$$V = (1/N)\sum_{i=1}^{N}(1/N)\,V_i + \frac{(N-1)}{N}\sum_{j=1}^{N}\sum_{\substack{i=1 \\ i\neq j}}^{N}\frac{C_{ij}}{N(N-1)}$$

Replacing the variances and covariances in the summation by their averages $\overline{V}$ and $\overline{C}$, we have:

$$V = \frac{\overline{V}}{N} + \frac{N-1}{N}\overline{C}$$

Note that there are N variances and $N(N-1)$ covariances that make up the N^2 entries in the $N \times N$ covariance matrix.

The contribution to the portfolio variance of the variances of the individual securities goes to zero as *N* gets very large. However, the contribution of the covariance terms approaches the average covariance as *N* gets large.

So, as the number of assets in the portfolio is increased, the variance of the return on the portfolio gets closer to the average covariance of return between the pairs of assets in that portfolio.

The individual risk of securities can be diversified away, but the contribution to the total risk caused by the covariance terms cannot be diversified away.

Chapter 6 Summary

Assumptions underlying mean-variance portfolio theory

- All expected returns, variances and covariances of pairs of assets are known.
- Investors make their decisions purely on the basis of expected return and variance.
- Investors are non-satiated.
- Investors are risk-averse.
- There is a fixed single-step time period.
- There are no taxes or transaction costs.
- Assets may be held in any amounts, *ie* short-selling is possible, we can have infinitely divisible holdings, there are no maximum investment limits.

Definitions

- The *opportunity set* is the set of points in $E-\sigma$ space that are attainable by the investor based on the available combinations of securities.
- A portfolio is *efficient* if there is no other portfolio with either a higher mean and the same or lower variance, or a lower variance and the same or higher mean.
- The *efficient frontier* is the set of efficient portfolios in $E-\sigma$ space**.**
- *Indifference curves* join points of equal expected utility in $E-\sigma$ space.
- The *optimal portfolio* is the portfolio that maximises the investor's expected utility. It occurs where an indifference curve is tangential to the efficient frontier.

Derivation of efficient frontier – the case of two securities

The three equations used to derive the equation of the efficient frontier when only two securities are available are:

1. $x_A + x_B = 1$
2. $E_P = x_A E_A + x_B E_B$
3. $V_P = x_A^2 V_A + x_B^2 V_B + 2x_A x_B C_{AB}$

V_p is minimised when $x_A = \dfrac{V_B - C_{AB}}{V_A + V_B - 2C_{AB}}$.

Derivation of the efficient frontier – the case of *N* securities

Here we use the Lagrangian function:

$$W = \sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}x_i x_j - \lambda\left(\sum_{i=1}^{N} E_i x_i - E_P\right) - \mu\left(\sum_{i=1}^{N} x_i - 1\right)$$

This is differentiated with respect to each x_i, λ and μ to give:

$$\frac{\partial W}{\partial x_i} = 2\sum_{j=1}^{N} x_j C_{ij} - \lambda E_i - \mu$$

$$\frac{\partial W}{\partial \lambda} = E_P - \sum_{i=1}^{N} x_i E_i$$

$$\frac{\partial W}{\partial \mu} = 1 - \sum_{i=1}^{N} x_i$$

Setting these derivatives to zero and solving $N+2$ equations in $N+2$ unknowns then gives the equation of the efficient frontier in $E-\sigma$ space.

Benefits of diversification

Assuming an equal holding of each asset within a portfolio, the portfolio variance is given by:

$$V = \frac{\overline{V}}{N} + \left(\frac{N-1}{N}\right)\overline{C}$$

So, as the number of assets in the portfolio is increased, the variance of the return on the portfolio gets closer to the average covariance of return between the pairs of assets in that portfolio.

This means that if all the assets are independent then the variance of return on the portfolio tends towards zero, *ie* the risk can be diversified away completely.

$$V = \frac{\overline{V}}{N} \to 0$$

Chapter 6 Practice Questions

6.1 Explain what is meant by the following terms, in the context of mean-variance portfolio theory:

(i) efficient frontier

(ii) indifference curves

(iii) optimal portfolio.

6.2 An investor can invest in only two risky assets A and B. Asset A has an expected rate of return of 10% and a standard deviation of return of 20%. Asset B has an expected rate of return of 15% and a standard deviation of return of 30%. The correlation coefficient between the returns of Asset A and the returns of Asset B is 0.6.

(i) Calculate the expected rate of return if 20% of an investor's wealth is invested in Asset A and the remainder is invested in Asset B.

(ii) Calculate the standard deviation of return on the portfolio if 20% of an investor's wealth is invested in Asset A and the remainder is invested in Asset B.

(iii) Explain why an investor who invests 20% of his wealth in Asset A and the remainder in Asset B is risk-averse.

6.3 Using mean-variance portfolio theory, prove that the efficient frontier becomes a straight line in the presence of a risk-free asset.

6.4 Consider a portfolio, P, which consists of N assets held in equal proportions. Let R_P represent the return on the portfolio, and let R_i represent the return on asset i. The covariance of the return on asset i with that on asset j is C_{ij}.

(i) State the total number of data items needed to calculate $E\left[R_P\right]$ and $Var(R_P)$.

(ii) Write down an expression for $Var(R_P)$.

(iii) Using your expression from part (ii), show that the specific risk of the portfolio (*ie* the risk associated with the individual assets) tends to zero in a well-diversified portfolio.

6.5 Exam style

Consider two independent assets, Asset A and Asset B, with expected returns of 6% *pa* and 11% *pa* and standard deviations of returns of 5% *pa* and 10% *pa*, respectively. Let x_i denote the proportion of the portfolio invested in Asset i.

(i) If only Assets A and B are available, determine the equation of the efficient frontier in expected return-standard deviation space. [3]

A third Asset, Asset C, is risk-free and has an expected return of 4% *pa*. A Lagrangian function is to be used to calculate the equation of the new efficient frontier.

(ii) Write down, but do not solve, the five simultaneous equations that result from this procedure. [3]

(iii) Use your simultaneous equations to derive the relationship between x_A and x_B on the new efficient frontier. [2]

(iv) Hence derive the equation of the new efficient frontier in expected return-standard deviation space. [4]

[Total 12]

6.6 Exam style

Consider a world in which there are only 2 securities, 1 and 2, such that:

$$E_1 = 5\%, \quad V_1 = (10\%)^2$$

$$E_2 = 10\%, \quad V_2 = (20\%)^2$$

Let ρ denote the correlation coefficient between the returns yielded by the two securities.

(i) Derive the equation of the opportunity set in *E*–*V* space. [5]

(ii) Derive expressions for the portfolio expected return *E* and the portfolio proportion x_1 invested in Security 1 at the point of global minimum variance and hence comment briefly on how *E* and x_1 vary with ρ. [5]

[Total 10]

6.7 Exam style

(i) Describe in detail the assumptions underlying the use of mean-variance portfolio theory. [3]

Consider a two-security world in which the returns yielded by Assets 1 and 2 are perfectly positively correlated, though they have different expected returns.

(ii) Using the method of Lagrangian multipliers or otherwise, derive the equation of the efficient frontier in expected return-standard deviation space. [6]

(iii) Use your answer to part (ii) to:

(a) determine the gradient of the efficient frontier

(b) show that the efficient frontier is a straight line in expected return-standard deviation space that passes through the points representing Assets 1 and 2. [4]

[Total 13]

Chapter 6 Solutions

6.1 (i) ***Efficient frontier***

The efficient frontier is the line that joins the points in expected return-standard deviation space that represent efficient portfolios.

A portfolio is *efficient* if the investor cannot find a better one in the sense that it has either a higher expected return and the same (or lower) level of risk (measured in terms of standard deviation of returns) or a lower level of risk and the same (or higher) expected return.

(ii) ***Indifference curves***

Indifference curves in expected return-standard deviation space join together points representing all the portfolios that give the investor equal levels of expected utility, given the risk-return preferences of that particular investor. They slope upwards for a risk-averse investor.

(iii) ***Optimal portfolio***

The investor's optimal portfolio is the portfolio on the efficient frontier that gives the highest possible level of expected utility, given the investor's particular indifference curves.

It is represented by the point in expected return-standard deviation space where the efficient frontier is tangential to the highest attainable indifference curve.

6.2 (i) ***Expected return***

Let the expected return on the portfolio be denoted $E[R_P]$. Then:

$$E[R_P] = 0.2 \times 0.1 + 0.8 \times 0.15 = 0.14 \text{ or } 14\%$$

(ii) ***Standard deviation of return***

Let the standard deviation of return on the portfolio be denoted σ_P. Then:

$$\sigma_P^2 = x_A^2 \sigma_A^2 + x_B^2 \sigma_B^2 + 2 x_A x_B \rho \sigma_A \sigma_B$$

where the x's denote the proportion of the portfolio invested in each asset, and ρ is the correlation coefficient between the returns of Assets A and B.

Therefore:

$$\sigma_P^2 = 0.2^2 0.2^2 + 0.8^2 0.3^2 + 2 \times 0.2 \times 0.8 \times 0.6 \times 0.2 \times 0.3 = 0.07072$$

So: $\sigma_P = 0.26593 \text{ or } 26.6\%$

(iii) ***Explain why the investor is risk-averse***

In mean-variance portfolio theory, risk is measured by standard deviation of returns.

An investor who invests 20% of his wealth in Asset A rather than investing 100% in Asset B is demonstrating risk aversion, because they are choosing to reduce the level of risk at the cost of attaining a lower expected return.

6.3 The risk-free asset must sit on the efficient frontier because it provides the highest return for no risk, and it is not possible to have a risk (*ie* standard deviation) of less than zero.

Consider an efficient portfolio consisting of risky assets only with mean E_P and variance V_P, and a risk-free asset with return r and zero variance.

Let the proportion of an investor's wealth invested in the efficient portfolio of risky assets only be x_P, so that $1-x_P$ is the proportion invested in the risk-free asset, and let ρ be the correlation coefficient between the return on the risk-free asset and the return on the efficient portfolio of risky assets only.

The variance of a portfolio on the efficient frontier can then be expressed as:

$$V=(1-x_P)^2\times 0+x_P^2\sigma_P^2+2(1-x_P)x_P\times 0\times\sigma_P\rho=x_P^2\sigma_P^2$$

So: $\sigma=x_P\sigma_P$

Next, the formula for the expected return on the portfolio gives:

$$E=(1-x_P)r+x_PE_P \quad\Rightarrow\quad x_P=\frac{E-r}{E_P-r}$$

Substituting this into the above expression for the standard deviation we see that E is linear in σ since:

$$\sigma=\left(\frac{E-r}{E_P-r}\right)\sigma_P \quad\Rightarrow\quad E=r+\left(\frac{E_P-r}{\sigma_P}\right)\sigma$$

Hence the efficient frontier is a straight line in expected return-standard deviation space.

6.4 (i) ***Data items***

To work out $E\left[R_P\right]$, we would need to know the expected return of each of the N assets. So this requires N data items.

To work out $\text{var}(R_P)$, we would need to know:

- the variance of each of the N assets (N data items)
- the covariance of each different pair of assets (an additional $\dfrac{N(N-1)}{2}$ data items).

The number of covariances required is the number of different ways of choosing two assets from N.

So the total number of data items needed is:

$$N + N + \frac{N(N-1)}{2} = \frac{N(N+3)}{2}$$

(ii) ***Variance***

The proportion of the portfolio invested in each asset is $\frac{1}{N}$. So the variance of the portfolio is given by:

$$Var(R_P) = Var\left(\sum_{i=1}^{N}\left(\frac{1}{N}\right)R_i\right) = \sum_{i=1}^{N}\left(\frac{1}{N}\right)^2 V_i + \sum_{j=1}^{N}\sum_{\substack{i=1 \\ i\neq j}}^{N}\left(\frac{1}{N}\right)^2 C_{ij}$$

where V_i is the variance of asset i and C_{ij} is the covariance of the return on asset i with that on asset j.

This can alternatively be written as:

$$Var(R_P) = \sum_{i=1}^{N}\left(\frac{1}{N}\right)^2 V_i + 2\sum_{j=1}^{N}\sum_{\substack{i=1 \\ i> j}}^{N}\left(\frac{1}{N}\right)^2 C_{ij}$$

(iii) ***Effect of diversification on specific risk***

The expression for the variance can be re-written as:

$$\text{var}(R_P) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{V_i}{N}\right) + \frac{N-1}{N}\sum_{j=1}^{N}\sum_{\substack{i=1 \\ i\neq j}}^{N}\left(\frac{C_{ij}}{N(N-1)}\right)$$

Let $\overline{V}$ represent the average variance, and $\overline{C}$ represent the average covariance. Then:

$$\overline{V} = \sum_{i=1}^{N}\left(\frac{V_i}{N}\right) \quad \text{and} \quad \overline{C} = \sum_{j=1}^{N}\sum_{\substack{i=1 \\ i\neq j}}^{N}\left(\frac{C_{ij}}{N(N-1)}\right)$$

since there are N variances to average and $N(N-1)$ covariances in total.

Note, from (i), that there are only $\frac{N(N-1)}{2}$ *different covariances since* $C_{ij} = C_{ji}$.

So:

$$Var(R_P) = \frac{1}{N}\overline{V} + \frac{N-1}{N}\overline{C}$$

As $N \to \infty$, the contribution to the overall portfolio variance of the individual variances (through $\bar{V}$) tends to zero. So, the specific risk associated with the individual securities can be diversified away, but the contribution to the total risk (or variance) from the covariance terms cannot be diversified away.

6.5 (i) ***Efficient frontier of A and B***

We can express the portfolio proportions as functions of the portfolio expected return:

$$E = x_A E_A + x_B E_B = 6x_A + 11(1 - x_A)$$

$$\Leftrightarrow \qquad x_A = \frac{E-11}{6-11} = \frac{11-E}{5} \quad \text{and} \quad x_B = 1 - x_A = \frac{E-6}{5} \qquad [1]$$

The two assets are independent, so the portfolio variance is:

$$V = x_A^2 V_A + x_B^2 V_B$$

Substituting in the portfolio proportions we get:

$$V = 5^2\left(\frac{11-E}{5}\right)^2 + 10^2\left(\frac{E-6}{5}\right)^2$$

$$= 121 - 22E + E^2 + 4\left(E^2 - 12E + 36\right)$$

$$= 5E^2 - 70E + 265 \qquad [1]$$

Finally, we square root to get the equation for the opportunity set in expected return-standard deviation space:

$$\sigma = \sqrt{5E^2 - 70E + 265}$$

The efficient frontier is the part of this curve above the point at which the variance is minimised. To find this point we differentiate:

$$\frac{dV}{dE} = 10E - 70 = 0$$

So the efficient frontier is the part of the opportunity set where $E \geq \frac{70}{10} = 7\%$, *ie* the efficient frontier is $\sigma = \sqrt{5E^2 - 70E + 265}$, $E \geq 7\%$. [1]

[Total 3]

(ii) ***Simultaneous equations***

The Lagrangian function is given by:

$$V = 5^2 x_A^2 + 10^2 x_B^2 - \lambda\left(6x_A + 11x_B + 4x_C - E\right) - \mu\left(x_A + x_B + x_C - 1\right) \qquad [1]$$

We now differentiate the function with respect to its five parameters and set to zero:

$$\frac{\partial V}{\partial x_A} = 50x_A - 6\lambda - \mu = 0 \qquad (1)$$

$$\frac{\partial V}{\partial x_B} = 200x_B - 11\lambda - \mu = 0 \qquad (2)$$

$$\frac{\partial V}{\partial x_C} = -4\lambda - \mu = 0 \qquad (3)$$

$$\frac{\partial V}{\partial \lambda} = E - 6x_A - 11x_B - 4x_C = 0 \qquad (4)$$

$$\frac{\partial V}{\partial \mu} = 1 - x_A - x_B - x_C = 0 \qquad (5)$$

[2 in total for simultaneous equations]

[Total 3]

(iii) ***Relationship between holdings in A and B***

From Equation (3) in part (ii), we have:

$$\mu = -4\lambda$$

Substituting this into Equations (1) and (2) gives:

$$50x_A - 6\lambda + 4\lambda = 0 \;\Rightarrow\; x_A = \frac{2\lambda}{50} = \frac{8\lambda}{200}$$ [1]

and $$200x_B - 11\lambda + 4\lambda = 0 \;\Rightarrow\; x_B = \frac{7\lambda}{200} = \frac{7}{8}x_A$$ [1]

[Total 2]

(iv) ***New efficient frontier***

As a risk-free asset is available, the new efficient frontier is a straight line. This straight line gives the expected return and standard deviation of portfolios consisting of a combination of the risk-free Asset C and a risky portfolio consisting of Assets A and B. [1]

Part (iii) establishes that $x_B = \frac{7}{8}x_A$ along this line. The efficient portfolio consisting entirely of risky assets must have $x_C = 0$ and so at this point we have:

$$x_A = \frac{8}{15} \quad \text{and} \quad x_B = \frac{7}{15}$$

The expected return of this portfolio is:

$$E = x_A E_A + x_B E_B = \frac{8}{15} \times 6 + \frac{7}{15} \times 11 = 8.33\%$$ [1]

The standard deviation of this portfolio is:

$$\sigma = \sqrt{x_A^2 V_A + x_B^2 V_B} = \sqrt{\left(\frac{8}{15}\right)^2 5^2 + \left(\frac{7}{15}\right)^2 10^2} = 5.375\% \qquad [1]$$

Therefore, the efficient frontier is a straight line in expected return-standard deviation space joining the points $(\sigma = 0, E = 4)$ and $(\sigma = 5.375, E = 8.333)$:

$$E = 4 + \frac{8.333 - 4}{5.375 - 0}\sigma = 4 + 0.806\sigma \qquad [1]$$

[Total 4]

6.6 (i) ***Opportunity set***

The variance of a portfolio consisting of Securities 1 and 2 is given by:

$$V = 10^2 x_1^2 + 20^2 x_2^2 + 2(10)(20)\rho x_1 x_2 \qquad [1]$$

The portfolio expected return is given by:

$$E = 5x_1 + 10x_2 \qquad [1]$$

Since the portfolio is fully invested, we require that:

$$x_1 + x_2 = 1 \;\Rightarrow\; x_2 = 1 - x_1$$

Substituting the last equation into the previous one and rearranging gives:

$$x_1 = \frac{10 - E}{5} \;\Rightarrow\; x_2 = \frac{E - 5}{5} \qquad [1]$$

Substituting these back into the expression for the variance and simplifying gives the equation of the opportunity set as:

$$V = (20 - 16\rho)E^2 - 240(1-\rho)E + 800(1-\rho) \qquad [2]$$

[Total 5]

(ii) ***Expressions for E and x_1***

A first-order condition for the point of global minimum variance (as a function of ρ) can be found by differentiating the expression for the variance found in (i) and setting it equal to zero. Thus:

$$\frac{\partial V}{\partial E} = 2(20 - 16\rho)E - 240(1-\rho) = 0 \qquad [1]$$

Hence, we find that at this point, after simplifying:

$$E = \frac{30(1-\rho)}{5 - 4\rho} \qquad [1]$$

Substituting this into the expression for x_1 found in (i), and simplifying, then gives:

$$x_1 = \frac{4-2\rho}{5-4\rho} \qquad [1]$$

To see how E and x_1 vary with ρ, we can look at the derivatives, which can be found using the quotient rule. After simplifying, we get:

$$\frac{\partial E}{\partial \rho} = \frac{-30}{(5-4\rho)^2} \qquad [½]$$

$$\frac{\partial x_1}{\partial \rho} = \frac{6}{(5-4\rho)^2} \qquad [½]$$

Since $\frac{\partial E}{\partial \rho} < 0$, the portfolio expected return at the point of minimum global variance *decreases* as the correlation coefficient increases. In fact, E ranges from $6\frac{2}{3}\%$ when $\rho = -1$ through 6% when $\rho = 0$ to 0% when $\rho = 1$. [½]

Since $\frac{\partial x_1}{\partial \rho} > 0$, the portfolio proportion invested in Security 1 at the point of minimum global variance *increases* with the correlation coefficient. In fact, x_1 ranges from $66\frac{2}{3}\%$ when $\rho = -1$ through 80% when $\rho = 0$ to 200% when $\rho = 1$. [½]

[Total 5]

The portfolio proportion invested in Security 2 at the point of minimum global variance must correspondingly decrease with the correlation coefficient.

6.7 (i) ***Assumptions underlying mean-variance portfolio theory***

- All expected returns, variances and covariances of pairs of assets are known. [½]
- Investors make their decisions purely on the basis of expected return and variance. [½]
- Investors are non-satiated. [½]
- Investors are risk-averse. [½]
- There is a fixed single-step time period. [½]
- There are no taxes or transaction costs. [½]
- Assets may be held in any amounts, *ie* short-selling is possible, we can have infinitely divisible holdings, there are no maximum investment limits. [½]

[Maximum 3]

(ii) ***Equation of the efficient frontier***

In order to find the efficient frontier, we can set up the Lagrangian function. When the correlation coefficient between the two security returns is equal to one, this is given by (using the usual definitions):

$$W = x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_1\sigma_2 - \lambda(x_1E_1 + x_2E_2 - E_p) - \mu(x_1 + x_2 - 1) \qquad [1]$$

Therefore the first-order conditions are:

$$\frac{\partial W}{\partial x_1} = 2x_1\sigma_1^2 + 2x_2\sigma_1\sigma_2 - \lambda E_1 - \mu = 0$$

$$\frac{\partial W}{\partial x_2} = 2x_2\sigma_2^2 + 2x_1\sigma_1\sigma_2 - \lambda E_2 - \mu = 0$$

$$\frac{\partial W}{\partial \lambda} = x_1E_1 + x_2E_2 - E_p = 0$$

$$\frac{\partial W}{\partial \mu} = x_1 + x_2 - 1 = 0 \qquad [2]$$

Combining the last two equations gives the optimal proportions of the two assets as:

$$x_1 = \frac{E_p - E_2}{E_1 - E_2} \quad \text{and} \quad x_2 = 1 - x_1 = \frac{E_1 - E_p}{E_1 - E_2} \qquad [1]$$

Substituting these back into the expression for the variance of portfolio returns (*ie* the first three terms in the Lagrangian function) gives:

$$V_p = \left(\frac{1}{E_1 - E_2}\right)^2 \left[(E_p - E_2)^2\sigma_1^2 + (E_1 - E_p)^2\sigma_2^2 + 2(E_p - E_2)(E_1 - E_p)\sigma_1\sigma_2\right]$$

$$= \left(\frac{1}{E_1 - E_2}\right)^2 \left[\left\{(E_p - E_2)\sigma_1 + (E_1 - E_p)\sigma_2\right\}^2\right]$$

Hence, the standard deviation of portfolio returns equals:

$$\sigma_p = \left(\frac{1}{E_1 - E_2}\right)\left\{\left(E_p - E_2\right)\sigma_1 + (E_1 - E_p)\sigma_2\right\}$$

Thus:

$$\sigma_p = aE_p + b \quad \text{where} \quad a = \frac{\sigma_1 - \sigma_2}{E_1 - E_2} \quad \text{and} \quad b = \frac{\sigma_2E_1 - \sigma_1E_2}{E_1 - E_2}$$

This is therefore the equation of the efficient frontier in expected return-standard deviation space. [2]

[Total 6]

In fact, because we only have two risky assets, we can use the following quicker method, which avoids Lagrangians:

- *first write the equation* $x_1E_1 + x_2E_2 = E_P$ *in the form* $x_1E_1 + (1-x_1)E_2 = E_P$
- *solve for* x_1 *in terms of the* E*'s and then find* $x_2 = 1 - x_1$
- *write down the variance, which is:*

$$V_P = x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_1\sigma_2 = (x_1\sigma_1 + x_2\sigma_2)^2$$

- *square root to get:*

$$\sigma_P = x_1\sigma_1 + x_2\sigma_2$$

- *substitute the expressions for* x_1 *and* x_2.

(iii)(a) ***Gradient***

The efficient frontier is normally plotted with expected return on the vertical axis, so its gradient will be equal to $1/a$, *ie*:

gradient of the efficient frontier $= \dfrac{1}{a} = \dfrac{E_1 - E_2}{\sigma_1 - \sigma_2}$. [1]

(iii)(b) ***Points on efficient frontier***

When $E_P = E_1$, we have:

$$\begin{aligned}\sigma_p &= aE_1 + b \\ &= \left[\frac{\sigma_1 - \sigma_2}{E_1 - E_2}\right]E_1 + \left[\frac{\sigma_2E_1 - \sigma_1E_2}{E_1 - E_2}\right] \\ &= \sigma_1\frac{(E_1 - E_2)}{E_1 - E_2} + \sigma_2\frac{(E_1 - E_1)}{E_1 - E_2} \\ &= \sigma_1\end{aligned}$$

Hence (σ_1, E_1) lies on the efficient frontier, which is the point representing Asset 1. A similar argument shows that it also passes through the point representing Asset 2. [2]

We know that the efficient frontier is a straight line, since the slope is $1/a$, which is constant. [1]

[Total 4]

We have just shown that if two assets are perfectly positively correlated, then the efficient frontier is a straight line. This will also be the case if two assets are perfectly negatively correlated.

End of Part 1

What next?

1. Briefly **review** the key areas of Part 1 and/or re-read the **summaries** at the end of Chapters 1 to 6.
2. Ensure you have attempted some of the **Practice Questions** at the end of each chapter in Part 1. If you don't have time to do them all, you could save the remainder for use as part of your revision.
3. Attempt **Assignment X1**.

7

Models of asset returns

Syllabus objectives

4.3 Single and multifactor models for investment returns

4.3.1 Describe the three types of multifactor models of asset returns:

- macroeconomic models
- fundamental factor models
- statistical factor models.

4.3.2 Discuss the single-index model of asset returns.

4.3.3 Discuss the concepts of diversifiable and non-diversifiable risk.

4.3.4 Discuss the construction of the different types of multifactor models.

4.3.5 Perform calculations using both single and multifactor models.

0 Introduction

In the previous chapter, we looked at how mean-variance portfolio theory can be used to determine the efficient frontier, from which the investor can then determine the optimal portfolio. Unfortunately, this approach can be difficult to implement in practice due to computational difficulty and the amount and type of data required.

Much subsequent research has therefore been directed at developing means of simplifying the implementation process. This has led to the development of *multifactor* or *multi-index models* and the *single-index model* that we discuss in this chapter. These facilitate the determination of the efficient frontier with substantially less information than the standard mean-variance portfolio theory. In addition, they can be used to characterise the sensitivities of a security's returns to various factors. Consequently, they are very important tools for portfolio management, being used both to model and predict the future investment returns yielded by different assets.

This chapter also discusses the important ideas of *specific risk* – risk that can be diversified away – and *systematic risk* – which cannot. These ideas are discussed further in other subjects.

1 Multifactor models

1.1 Definition

A multifactor model of security returns attempts to explain the observed historical return by an equation of the form:

$$R_i = a_i + b_{i,1}I_1 + b_{i,2}I_2 + \ldots + b_{i,L}I_L + c_i$$

where:

- R_i is the return on security i
- a_i, c_i are the constant and random parts respectively of the component of return unique to security i
- $I_1, \ldots, I_L$ are the changes in a set of L factors which explain the variation of R_i about the expected return a_i
- $b_{i,k}$ is the sensitivity of security i to factor k.

Here the interpretation of the $b_{i,k}$'s is that if $b_{i,k}$ = 1.5 say, then an increase (decrease) of 1 in factor k is expected to produce an increase (decrease) of 1.5 in the return provided by security i. Conversely, a_i represents the expected value of that element of the investment return that is independent of the set of L factors and hence unique to security i.

The L factors are therefore the *systematic* factors that influence the returns on every security and the corresponding part of the total return which is equal to:

$$b_{i,1}I_1 + b_{i,2}I_2 + \ldots + b_{i,L}I_L$$

is referred to as the *systematic return* of security i. Conversely, the remaining element of the total return, which is specific to each individual security and independent of the returns on all other securities, namely:

$$a_i + c_i$$

is referred to as the *specific return* of security i.

When applying the multifactor model it is usual to assume that:

- $E[c_i] = 0$
- $Cov[c_i, c_j] = 0$ for all $i \neq j$.
- $Cov[c_i, I_k] = 0$ for all stocks and indices.

Note that:

- although the above equation for the model assumes that asset returns are linearly related to the factors or indices, this requirement is not as restrictive as it might first appear, as the factors themselves may be non-linear functions of the underlying variables, *eg* the log of inflation
- a_i and c_i are sometimes combined into a single parameter, with a non-zero expectation.

Question

Consider a two-factor model. If:

- the mean specific return is 1.0%
- the expected values of the two factors are 3.0% and 2.2% and
- the sensitivities of investment returns to each of the factors are 0.8 and –0.3 respectively,

what is the expected return predicted by the model?

Solution

In a two-factor model:

$$R_i = a_i + b_{i,1}I_1 + b_{i,2}I_2 + c_i$$

In this case:

$$a_i = 1.0 \quad b_{i,1} = 0.8 \quad b_{i,2} = -0.3$$

Thus:

$$\begin{aligned} E(R_i) &= E\left[a_i + b_{i,1}I_1 + b_{i,2}I_2 + c_i\right] \\ &= a_i + b_{i,1}E\left[I_1\right] + b_{i,2}E\left[I_2\right] + E\left[c_i\right] \\ &= 1.0 + 0.8 \times 3 - 0.3 \times 2.2 + 0 \\ &= 2.74\% \end{aligned}$$

The goal of the builders of such a model is to find a set of factors which explain as much as possible of the observed historical variation, without introducing too much 'noise' into predictions of future returns.

Multifactor models can be classified into three categories, depending on the type of factors used.

These are *macroeconomic*, *fundamental* and *statistical factor models*. We discuss each of these in the following sections.

1.2 Macroeconomic factor models

These use observable economic time series as the factors. They could include factors such as the annual rates of inflation and economic growth, short-term interest rates, the yields on long-term government bonds, and the yield margin on corporate bonds over government bonds.

The rationale here is that the price of, and hence the returns obtained from, a security should reflect the discounted present value to investors of the cashflows that it is expected to produce in the future. The macroeconomic variables mentioned above are the factors that we might typically expect to influence both the size of the future cashflows from security *i* and the discount rate used to value them. Note, however, that the more factors that are included, the more complex the model and so the harder it is to handle mathematically.

A related class of model uses a market index plus a set of industry indices as the factors.

Here, 'industry' refers to the company sectors, such as banking, energy, food, support services *etc*. So, security returns are assumed to reflect the influence of both market-wide and industry-specific effects.

Once the set of factors has been decided on, a time series regression is performed to determine the sensitivities for each security in the sample.

This kind of time series regression is a natural multivariable extension of the one-variable case considered in the single-index model discussed later in this chapter.

A common method used to determine how many factors to include is to start with relatively few, perform the regression and measure the residual (unexplained) variance. An extra factor is then added and the regression repeated. The whole process is repeated until the addition of an extra factor causes no significant reduction in the residual variance.

An alternative approach is to start with a more general model containing a large number of possible factors and then to remove those whose elimination does not significantly affect the explanatory power of the model – *ie* the size of the residual variance.

1.3 Fundamental factor models

Fundamental factor models are closely related to macroeconomic models, but instead of (or in addition to) macroeconomic variables the factors used are company-specific variables. These may include such fundamental factors as:

- **the level of gearing**
- **the price earnings ratio**
- **the level of research and development spending**
- **the industry group to which the company belongs.**

Again, the models are constructed using regression techniques.

Commercial fundamental factor models are available which use many tens of factors. They are used for risk control, by comparing the sensitivity of a portfolio to one of the factors with the sensitivity of a benchmark portfolio.

Suppose that you can find a portfolio that has similar sensitivities to similar factors as a benchmark portfolio. Then by holding that portfolio you should be able to closely replicate the performance of the benchmark. This technique could, for example, be used to construct a portfolio that follows or tracks the performance of an investment index without the investor needing to hold every individual constituent security of the index.

1.4 Statistical factor models

Statistical factor models do not rely on specifying the factors independently of the historical returns data. Instead a technique called principal components analysis can be used to determine a set of indices which explain as much as possible of the observed variance. However, these indices are unlikely to have any meaningful economic interpretation and may vary considerably between different data sets.

Question

Explain the distinction between systematic return and specific return.

Solution

The *systematic return* of security *i* is the element of the total return that arises due to the influences of the *L* factors that affect the returns on every security. It is equal to:

$$b_{i,1}I_1 + b_{i,2}I_2 + \ldots + b_{i,L}I_L$$

The *specific return* of security *i* is the element of the total return that is independent of the *L* factors and hence independent of the returns on all other securities. It is therefore specific or unique to security *i* and is equal to:

$$a_i + c_i$$

1.5 Construction of models

Principal components analysis is a technique used to investigate the relationship between a set of endogenous variables, such as the factors determining the investment return in a multifactor model. Within this context it can be used to:

- determine the relative significance of the various factors by analysing the variance-covariance matrix (between them) to determine which factors have the most influence upon the total variance of security returns

- combine groups of highly correlated factors into single factors or *principal components* that are much less highly correlated with each other – thereby reducing the number of factors in the model and improving the efficiency of the model.

For applications of multi-index models to portfolio selection problems, it is convenient if the factors used are uncorrelated (or orthogonal). Principal components analysis automatically produces a set of uncorrelated factors.

Question

Why is it convenient for the factors used to be uncorrelated?

Solution

Intuitively, the less correlated the factors are, the easier it is to disentangle the influences of each upon security returns. If they are highly correlated, then they act in unison and have insufficient independent variation to enable the model to isolate their separate influences.

Within a regression context, this problem is known as *multicollinearity* and it has the effect of making the coefficient estimates less efficient.

Where the factors are derived from a set of market indices or macroeconomic variables, it is possible to transform the original set into an orthogonal set which retains a meaningful economic interpretation.

Suppose, for example, we have two indices I_1 and I_2. I_1 could be a market index and I_2 an industry sector index. Two new, uncorrelated factors, I_1^* and I_2^*, can be constructed as follows:

First, let $I_1^* = I_1$.

This is done solely to keep our notation consistent in the final equation below.

We then carry out a linear regression analysis to determine the parameters γ_1 and γ_2 in the equation:

$$I_2 = \gamma_1 + \gamma_2 I_1^* + d_2$$

So γ_1 and γ_2 represent the intercept and the slope of the regression line and d_2 is the 'error' term, which by definition is uncorrelated with $I_1^* = I_1$.

We then set:

$$I_2^* = d_2 = I_2 - \left(\gamma_1 + \gamma_2 I_1^*\right).$$

By construction I_2^* is uncorrelated with I_1.

This is because $I_2^* = d_2$, which was the residual term in the previous equation. Mathematically:

$$\text{cov}(I_2^*, I_1) = \text{cov}\left(I_2 - (\gamma_1 + \gamma_2 I_1^*), I_1\right)$$

$$= \text{cov}(d_2, I_1) = \text{cov}(d_2, I_1^*) = 0$$

Changes in I_2^* can be interpreted as the change in the observed value of I_2 that cannot be explained by the observed change in I_1.

If there were a third index, a regression would be performed to determine the component of that index which could not be explained by the observed values of I_1 and I_2, and so on.

Question

A modeller has developed a two-factor model to explain the returns obtained from security *i*. It has the form:

$$R_i = 2 + 1.3I_1 + 0.8I_2 + c_i$$

However, the modeller is concerned that the two indices 1 and 2 may be correlated and so decides to re-express the model in terms of orthogonal factors. By regressing Index 1 on Index 2, the modeller obtains the following equation for the line of best fit:

$$I_1 = 0.8 + 0.3I_2$$

Use this information to re-express the two-factor model in terms of two orthogonal factors I_1^* and I_2^*.

Solution

In this case it is easier to carry out the process the other way round, because the modelled has regressed I_1 on I_2, not I_2 on I_1 as before.

So, first define two new variables:

(1) $I_2^* = I_2$

(2) $I_1^* = I_1 - 0.8 - 0.3I_2$

where I_1^* is equal to the residuals from the regression of I_1 on I_2, which by definition are uncorrelated.

We now need to re-express R_i in terms of the new variables I_1^* and I_2^*. It follows from (1) and (2) that:

$$I_1^* = I_1 - 0.8 - 0.3I_2^*$$

from which:

(3) $$I_1 = I_1^* + 0.8 + 0.3I_2^*$$

Using (1) and (3) to substitute for I_1 and I_2 in the original model then gives:

$$R_i = 2 + 1.3\left(I_1^* + 0.8 + 0.3I_2^*\right) + 0.8I_2^* + c_i$$

ie $$R_i = 3.04 + 1.3I_1^* + 1.19I_2^* + c_i$$

Note that in general the process can be carried out either way around, although a different, but equally acceptable, answer will be obtained in each case.

2 The single-index model

2.1 Definition

The single-index model as described below is a special case of the multifactor model that includes only one factor, normally the return on the investment market as a whole. It is based upon the fact that most security prices tend to move up or down with movements in the market as a whole. It therefore interprets such market movements as the major influence upon individual security price movements, which are consequently correlated only via their dependence upon the market.

The single-index model is sometimes also called the *market model*. Note that other single-index or one-factor models are possible, in which the single index is a variable other than the market.

The single-index model expresses the return on a security as:

$$R_i = \alpha_i + \beta_i R_M + \varepsilon_i$$

where:

- R_i **is the return on security** i
- α_i , β_i **are constants**
- R_M **is the return on the market**
- ε_i **is a random variable representing the component of** R_i **not related to the market.**

Question

How can we interpret α_i and β_i ?

Solution

α_i can be interpreted as the expected value of the component of security *i*'s return that is independent of the market's performance and specific to that particular security.

β_i quantifies the component of the security return that is directly related to movements in the market – so that if $\beta_i = x$, then security *i*'s return is expected to increase (decrease) by *x*% when the market return increases (decreases) by 1%.

Under the model, ε_i **is uncorrelated with** R_M **and** ε_i **is independent of** ε_j **for all** $i \neq j$**.**

It is also normal to set α_i such that $E(\varepsilon_i) = 0$ for $i = 1,\ldots, N$.

So:

- $E[\varepsilon_i]=0$
- $Cov[\varepsilon_i,\varepsilon_j]=0$ for all $i \neq j$
- $Cov[\varepsilon_i,R_M]=0$ for all i.

2.2 Results of the single-index model

For any particular security, α and β can be estimated by time series regression analysis.

In order to estimate α and β for security *i*, the historical returns produced over say t = 1, …, N monthly intervals for both security *i*, R_{it} , and the market, R_{Mt} say, are required. We can then use regression analysis based upon the equation:

$$R_i = \alpha_i + \beta_i R_M + \varepsilon_i$$

In each case, there is the usual problem that future values may differ from estimates of past values. An alternative approach is simply to use subjective estimates in the model – though even these are likely to be informed by estimates based on historical data.

The expected return and variance of return on security *i* and the covariance of the returns on securities *i* and *j* are given by:

$$\boldsymbol{E_i = \alpha_i + \beta_i E_M} \qquad \textbf{(7.1)}$$

$$\boldsymbol{V_i = \beta_i^2 V_M + V_{\varepsilon i}} \qquad \textbf{(7.2)}$$

and

$$\boldsymbol{C_{ij} = \beta_i \beta_j V_M} \qquad \textbf{(7.3)}$$

where $V_{\varepsilon i}$ is the variance of ε_i .

Question

Derive the first of the above results.

Solution

The expected return on security *i* is given by:

$$E(R_i) = E[\alpha_i + \beta_i R_M + \varepsilon_i]$$

By the linear additivity of expected values this can be written as:

$$E(R_i) = E(\alpha_i) + E(\beta_i R_M) + E(\varepsilon_i)$$

And since α_i and β_i are constants and α_i is chosen so that $E(\varepsilon_i) = 0$, we have:

$$E(R_i) = \alpha_i + \beta_i E_M$$

as required.

Let's now look at these equations in a little more detail.

Equation 7.1

The first of these three equations, (7.1), will be seen to be identical in form to the main result of the Capital Asset Pricing Model (CAPM) discussed in the next chapter. However, it should be emphasised that the single-index model is purely empirical and is not based on any theoretical relationships between β_i and the other variables.

This is in contrast to CAPM, which can be derived using economic theory.

Equation 7.2

The second equation, (7.2), models the variance of the return on security *i* as the sum of a term related to the variance of the return on the market and a term specific to security *i*. These two terms are usually called systematic and specific risk respectively.

***Systematic risk* can be regarded as relating to the market as a whole, while *specific risk* depends on factors peculiar to the individual security.**

So this result can be interpreted as: total risk equals systematic risk plus specific risk.

As shown in Chapter 6, in a diversified portfolio, consisting of a large number of securities, the contribution of the specific risk on each security, to the total risk of the portfolio, becomes very small. In this case, the contribution of each security to the portfolio's total risk is then only the systematic risk of that security.

So, diversification:

- can be used to reduce and ultimately eliminate specific risk
- leads to an averaging of systematic risk.

Thus it is only the systematic risk, measured by β_i of a security, that should be expected to be rewarded by increased return since this is non-diversifiable.

For this reason systematic risk is also sometimes referred to as non-diversifiable or market risk.

Investors can diversify away specific risk and do not therefore require compensation for accepting it. Specific risk is sometimes also referred to as *alpha*, *unsystematic*, *diversifiable* or *residual risk*.

Recall that α_i in the single-index model equation is the expected value of the component of security *i*'s return that is *independent* of the market's performance and *specific* to that particular security.

Question

Show that the variance of portfolio returns can be written as:

$$V_P = \sum_{i=1}^{N} x_i^2 V_{\varepsilon i} + \beta_p^2 V_M$$

and use this expression to show that:

- the contribution of the specific risk on each security to the total risk of the portfolio becomes very small as the number of securities increases and
- the contribution of each security to the portfolio's total risk is only the systematic risk of that security, *ie*:

$$\sigma_P \to \beta_p \sigma_M = \sigma_M \sum_{i=1}^{\infty} x_i \beta_i \text{ , as } N \to \infty$$

Solution

The variance of portfolio returns can be found as follows:

$$V_P = \sum_{i=1}^{N} x_i^2 V_i + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} x_i x_j C_{i,j}$$

where the x's are the portfolio weightings and the second summation is over $j \neq i$.

ie $$V_P = \sum_{i=1}^{N} x_i^2 (\beta_i^2 V_M + V_{\varepsilon i}) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} x_i x_j (\beta_i \beta_j V_M)$$

Combining all of the variance and covariance terms for the systematic risk then gives:

$$V_P = \sum_{i=1}^{N} x_i^2 V_{\varepsilon i} + \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \beta_i \beta_j V_M$$

Note that the second summation now includes the case $i = j$. Collecting together the i and j terms in this summation then gives:

$$V_P = \sum_{i=1}^{N} x_i^2 V_{\varepsilon i} + \left(\sum_{i=1}^{N} x_i \beta_i \right) \left(\sum_{j=1}^{N} x_j \beta_j \right) V_M$$

Therefore we have:

$$V_P = \sum_{i=1}^{N} x_i^2 V_{\varepsilon i} + \beta_P^2 V_M$$

where $\beta_P = \sum_{i=1}^{N} x_i \beta_i$ which is the required expression.

Now, suppose that equal amounts of money are invested in each of the *N* securities, so that $x_i = 1/N$ for all *i* = 1, …, *N*. Then the first term in the expression for V_P, which represents the contribution of the specific risk to the total portfolio variance, can be written as:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}V_{\varepsilon i}$$

ie $\frac{1}{N}$ times the average specific risk associated with each individual security $\overline{V}$. As *N* increases and the average specific risk remains unchanged, so this term will rapidly become smaller. Hence, as $N \to \infty$, the specific risk tends to zero and:

$$V_P \to \beta_P^2 V_M$$

ie $$\sigma_P \to \beta_P \sigma_M = \sigma_M \sum_{i=1}^{\infty} x_i \beta_i$$

In other words, the contribution of the specific risk on each security to the total risk of the portfolio becomes very small and the contribution of each security to the portfolio's total risk is therefore only the systematic risk of that security.

Equation 7.3

The third equation, (7.3), shows that, in this particular model, any correlation between the returns on two securities comes only from their joint correlation with the market as a whole. In other words, the only reason that securities move together is a common response to market movements; there are no other possible common factors.

In reality, however, we might expect shares within the same sector (*eg* banks) to tend to move together because of factors influencing that particular sector which might not affect other sectors (*eg* oil companies).

Question

If β_i doubles (with everything else remaining unchanged), then so does the expected return on security *i*. True or false?

Solution

Recall that the expected return on security *i* is given by:

$$E_i = \alpha_i + \beta_i E_M$$

So, provided α_i is non-zero, which is usually the case, if β_i doubles, the expected return will not also double.

ie $E_i' = \alpha_i + 2\beta_i E_M \neq 2(\alpha_i + \beta_i E_M) = 2E_i$

2.3 Data requirements

Although many studies have found that incorporating more factors into the model (for example industry indices) leads to a better explanation of the historical data, correlation with the market is the largest factor in explaining security price variation. Furthermore, there is little evidence that multifactor models are significantly better at forecasting the future correlation structure.

The use of the single-index model dramatically reduces the amount of data required as input to the portfolio selection process. For *N* securities, the number of data items needed has been reduced from $N(N+3)/2$ to $3N+2$.

Question

Derive these expressions for the number of data items.

Solution

In order to apply mean-variance portfolio theory with *N* securities, we need estimates of the following:

- *N* expected returns
- *N* variances
- ½ × *N* × (*N* – 1) correlation coefficients or covariances.

So, in total we need ½ × *N* × (*N* + 3) items of data. For example, if *N* = 200, then ½ × *N* × (*N* + 3) = 20,300.

Under the single-index model we need:

- *N* values for the α_i's
- *N* values for the β_i's
- *N* values for the $V_{\varepsilon i}$'s
- the expected return E_M and variance V_M for the market.

Thus, in total we need $3N+2$ items of data, which with 200 securities amounts to 602 items of data.

A primary use of both single-index models and multifactor models is to determine the expected return, variance and covariance of security returns, thereby enabling the investor to determine the efficient frontier. The single-index model in particular allows this to be done with a lot less information than is required with the standard mean-variance portfolio theory as described in the previous chapter.

Furthermore, the nature of the estimates required from security analysts conforms much more closely to the way in which they traditionally work.

Traditional investment analysis concentrates on estimating future performance, which in practice has normally meant the expected returns of securities, although increasing emphasis has also been placed on the risk or volatility of individual securities. Mean-variance portfolio theory, however, also requires estimates of each security's pairwise correlation with all other securities that may be included in the portfolio.

In addition, considerably simplified methods for calculating the efficient frontier have been developed under the single-index model although, with increasing computer power, this is of considerably less importance than it was when the model was first published.

Question

State the main uses of multifactor models and single-index models. What is their main limitation?

Solution

The main uses include:

- Determination of the investor's efficient frontier, as part of the derivation of the investor's optimal portfolio.
- Risk control – by enabling the investor to forecast the variability of portfolio returns both absolutely and relative to some benchmark. For example, by constructing a portfolio whose sensitivities to the relevant factors are the same as the benchmark, it is possible to reduce the risk of under- or over-performance compared to that benchmark.
- Performance analysis – by comparing the actual performance of the portfolio to that predicted by the model and based on the portfolio's actual exposure to the relevant factors over the period considered.
- Categorisation of investment styles – according to the extent of the exposure to particular factors.

The main limitation is that the construction of factor models is based on historical data that reflect conditions that may not be replicated in the future. Moreover, a model that does produce good predictions in one time period, may not produce good predictions in subsequent time periods.

Note that this particular limitation applies equally to many of the other investment models in common use, including the capital asset pricing model discussed in the next chapter, whose parameters are typically estimated using past data.

The chapter summary starts on the next page so that you can keep all the chapter summaries together for revision purposes.

Chapter 7 Summary

The multifactor model

A multifactor model of security returns attempts to explain the observed historical return by an equation of the form:

$$R_i = a_i + b_{i,1}I_1 + b_{i,2}I_2 + \dots + b_{i,L}I_L + c_i$$

where:

- R_i is the return on security *i*,
- a_i, c_i are the constant and random parts respectively of the component of return unique to security *i*
- $I_1, \dots, I_L$ are the changes in a set of *L* factors which explain the variation of R_i about the expected return a_i
- $b_{i,k}$ is the sensitivity of security *i* to factor *k*.

Types of multifactor model

Macroeconomic – the factors are the main macroeconomic variables such as interest rates, inflation, economic growth and exchange rates.

Fundamental – the factors will be company specifics such as P/E ratios, liquidity ratios and gearing measurements.

Statistical – the factors are not specific items initially. The method uses *principal components analysis* and historical returns on stocks to decide upon the factors.

Single-index model

This has just a single factor, which is usually the return on the market, R_M.

So: $R_i = \alpha_i + \beta_i R_M + \varepsilon_i$

It can then be shown that for any security *i*:

- $E_i = \alpha_i + \beta_i E_M$
- $V_i = \beta_i^2 V_M + V_{\varepsilon_i}$ *ie* total risk is the sum of *systematic risk* and *specific risk*
- $C_{ij} = \beta_i \beta_j V_M$ *ie* securities 'covary' only through their covariance with the market

Likewise, the variance of portfolio returns is equal to:

$$V_P = \sum_{i=1}^{N} x_i^2 V_{\varepsilon i} + \beta_p^2 V_M$$

where:

- x_i is the weight in security i
- $\beta_P = \sum_{i=1}^{N} x_i \beta_i$

Chapter 7 Practice Questions

7.1 A portfolio P consists of n assets, with a proportion x_i invested in asset i, $i=1,2,\ldots,n$ (so that $\sum_{i=1}^{n} x_i = 1$).

Exam style

(i) The annual returns R_P on this portfolio can be assumed to conform to the single-index model of asset returns. Write down an equation defining this model and show that:

$$\text{var}(R_P) = \beta_P^2 \,\text{var}(R_M) + \text{var}(\varepsilon_P)$$

where ε_P denotes the component of the portfolio return that is independent of movements in the market. [3]

(ii) Explain why the specific risk, $\text{var}(\varepsilon_P)$, is sometimes referred to as the 'diversifiable risk', giving an algebraic justification for your answer. [4]

(iii) Discuss the following statement:

'A portfolio with a beta of zero is equivalent to a risk-free asset.' [2]

[Total 9]

7.2 Show that in the single-index model of asset returns:

Exam style

$$E_i = \alpha_i + \beta_i E_M$$

$$V_i = \beta_i^2 V_M + V_{ei}$$

and $C_{ij} = \beta_i \beta_j V_M$

where V_{ei} is the variance of e_i. [8]

7.3 Consider the data in the table below, which relates to Securities 1, 2 and 3.

	Security		
	1	2	3
α_i	0.0	2.0	–2.2
β_i	1.1	0.6	2.0
$V_{\varepsilon i}$	2.2	1.3	1.2

You are given that:

- the expected return and standard deviation of the market return are 10 and 5 respectively
- the returns of each security can be modelled using an appropriate single-index model.

(i) Calculate:

(a) the expected return and standard deviation of return for each security

(b) the covariance of returns between each pair of securities.

(ii) Consider a portfolio which consists of Securities 1, 2 and 3 in equal proportions. Calculate:

(a) the variance of the portfolio

(b) the systematic risk of the portfolio

(c) the specific risk of the portfolio.

7.4 Distinguish between the three main classes of multifactor model.

7.5 Consider the single-index model of investment returns in which for any security i :

Exam style

$$R_i = \alpha_i + \beta_i R_m + \varepsilon_i$$

where $E(\varepsilon_i) = 0$, $E(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$, $E(R_m \varepsilon_i) = 0$ and R_m is the return on the market.

(i) Assuming that this model applies, derive expressions for the mean investment return on security i, and the mean investment return on a portfolio P, containing n securities, with a proportion x_i invested in security i. [3]

(ii) Show that $C_{iP} = \sum_{j=1}^{n} x_j C_{ij}$, where C_{iP} and C_{ij} are the covariance of investment returns between security i and portfolio P and securities i and j respectively. [2]

(iii) State a general expression for the variance σ_P^2 of portfolio P in terms of the covariances C_{ij}. [1]

(iv) Use your results from (ii) and (iii) to show that:

$$\beta_{iP} = \frac{\partial \sigma_P}{\partial x_i} \frac{1}{\sigma_P}$$

where $\beta_{iP} = \dfrac{C_{iP}}{\sigma_P^2}$ and comment briefly on this result. [7]

[Total 13]

Chapter 7 Solutions

7.1 (i) ***Equations for model and variance***

According to the single-index model:

$$R_P = \alpha_P + \beta_P R_M + \varepsilon_P \qquad [½]$$

Taking variances of both sides gives:

$$\begin{aligned}\text{var}(R_P) &= \text{var}(\alpha_P + \beta_P R_M + \varepsilon_P) \\ &= \beta_P^2 \text{var}(R_M) + \text{var}(\varepsilon_P) \qquad [1½]\end{aligned}$$

We have used the facts that:

- α_P is a constant [½]
- ε_P and R_M are uncorrelated. [½]

[Total 3]

(ii) ***Why specific risk is 'diversifiable'***

Modelling portfolio returns using the single-index model is usually based on the underlying assumption that each of the individual assets i also follows a single-index model of the form:

$$R_i = \alpha_i + \beta_i R_M + \varepsilon_i$$

where ε_i and ε_j are uncorrelated when $i \neq j$.

So: $$R_P = \sum_{i=1}^{n} x_i R_i = \sum_{i=1}^{n} x_i \alpha_i + \beta_P R_M + \sum_{i=1}^{n} x_i \varepsilon_i \qquad [1]$$

So ε_P corresponds to $\sum_{i=1}^{n} x_i \varepsilon_i$ and it follows (since ε_i and ε_j are uncorrelated) that:

$$\text{var}(\varepsilon_P) = \text{var}\left(\sum_{i=1}^{n} x_i \varepsilon_i\right) = \sum_{i=1}^{n} x_i^2 \text{var}(\varepsilon_i) \qquad [1]$$

If we select a portfolio with equal proportions of each asset, *ie* $x_i = \dfrac{1}{n}$, then:

$$\text{var}(\varepsilon_P) = \sum_{i=1}^{n} \frac{1}{n^2} \text{var}(\varepsilon_i) = \frac{1}{n} \times \left(\frac{1}{n} \sum_{i=1}^{n} \text{var}(\varepsilon_i)\right) = \frac{1}{n} \times \overline{V} \qquad [1]$$

where $\overline{V}$ denotes the average specific risk of the assets in the portfolio.

So, as $n \to \infty$, the variance $\text{var}(\varepsilon_P)$ will tend to zero. [½]

So this component of the variance of returns can be reduced to a very small level by selecting a sufficiently diversified portfolio. It is therefore often called the 'diversifiable risk'. [½]

[Total 4]

(iii) ***Zero beta versus risk-free***

The statement is not correct. [½]

A zero beta indicates no systematic risk. [½]

A portfolio will only be totally risk-free if it has zero variance, *ie* if $\text{var}(R_P)=0$. [½]

From the equation $\text{var}(R_P)=\beta_P^2\,\text{var}(R_M)+\text{var}(\varepsilon_P)$, we see that, even if $\beta_P=0$, the overall variance will only be zero if $\text{var}(\varepsilon_P)=0$, as well. [½]

But, any (non-trivial) portfolio of risky assets will have a non-zero specific risk, *ie* $\text{var}(\varepsilon_P)>0$. [½]

However, it is theoretically true that a well-diversified portfolio with a beta of zero will be approximately risk-free. [½]

[Maximum 2]

7.2 According to the single-index model, the return on Security *i* is given by:

$$R_i = \alpha_i + \beta_i R_M + e_i$$ [½]

where α_i and β_i are constants

R_M is the return on the market

e_i is a random variable representing the component of R_i not related to the market. [½]

By the linear additivity of expected values, we have:

$$E(R_i) = E(\alpha_i)+E(\beta_i R_M)+E(e_i)$$ [½]

Since α_i and β_i are constants and α_i is chosen so that $E(e_i)=0$, we have:

$$E(R_i) = \alpha_i + \beta_i E_M$$

as required. [½]

The variance of returns for Security i is:

$$V_i = \text{var}\left[\alpha_i + \beta_i R_M + e_i\right]$$ [½]

As α_i is constant, this is equal to:

$$V_i = \text{var}\left[\beta_i R_M + e_i\right]$$ [½]

Now, recall that the single-index model assumes that:

$$\text{cov}(e_i, R_M) = 0$$ [½]

Hence:

$$\Rightarrow \quad V_i = \text{var}\left[\beta_i R_M\right] + \text{var}\left[e_i\right]$$ [½]

ie $V_i = \beta_i^2 V_M + V_{ei}$

as required. [1]

The covariance between Securities i and j is given by:

$$C_{i,j} = \text{cov}\left[R_i, R_j\right]$$

$$= \text{cov}\left[\alpha_i + \beta_i R_M + e_i, \alpha_j + \beta_j R_M + e_j\right]$$ [½]

Again, since α_i is constant, this is equal to:

$$C_{i,j} = \text{cov}\left[\beta_i R_M + e_i, \beta_j R_M + e_j\right]$$ [½]

As before, recall that the single-index model assumes that:

$$\text{cov}(e_i, R_M) = 0$$ [½]

Hence:

$$C_{i,j} = \text{cov}\left[\beta_i R_M, \beta_j R_M\right] + \text{cov}\left[e_i, e_j\right]$$

$$= \beta_i \beta_j \text{cov}\left[R_M, R_M\right] + \text{cov}\left[e_i, e_j\right]$$ [½]

We also have the assumption that $\text{cov}(e_i, e_j) = 0$ when $i \neq j$.

So: $C_{i,j} = \beta_i \beta_j \text{cov}\left[R_M, R_M\right] = \beta_i \beta_j V_M$ [1]

[Total 8]

7.3 (i)(a) ***Calculate the expected return and standard deviation***

The expected return for each security is calculated using the equation:

$$E_i = \alpha_i + \beta_i E_M$$

Hence:

$$E_1 = 0 + 1.1 \times 10 = 11.0$$

$$E_2 = 2 + 0.6 \times 10 = 8.0$$

$$E_3 = -2.2 + 2.0 \times 10 = 17.8$$

The variance for each security is calculated using the equation:

$$V_i = \beta_i^2 V_M + V_{\varepsilon i}$$

Hence, the standard deviation for each security is given by:

$$\sigma_1 = \left(\beta_1^2 V_M + V_{\varepsilon 1}\right)^{½} = \left(1.1^2 \times 25 + 2.2\right)^{½} = (32.45)^{½} = 5.70$$

$$\sigma_2 = \left(\beta_2^2 V_M + V_{\varepsilon 2}\right)^{½} = \left(0.6^2 \times 25 + 1.3\right)^{½} = (10.3)^{½} = 3.21$$

$$\sigma_3 = \left(\beta_3^2 V_M + V_{\varepsilon 3}\right)^{½} = \left(2.0^2 \times 25 + 1.2\right)^{½} = (101.2)^{½} = 10.06$$

(i)(b) ***Calculate the covariances***

The covariance of returns between securities i and j is calculated using the equation:

$$C_{i,j} = \beta_i \beta_j V_M$$

Hence:

$$C_{1,2} = \beta_1 \beta_2 V_M = 1.1 \times 0.6 \times 25 = 16.5$$

$$C_{1,3} = \beta_1 \beta_3 V_M = 1.1 \times 2.0 \times 25 = 55.0$$

$$C_{2,3} = \beta_2 \beta_3 V_M = 0.6 \times 2.0 \times 25 = 30.0$$

(ii)(a) ***Variance of portfolio***

If R_P is the return on the portfolio and R_i is the return on security i, then:

$$R_P = \tfrac{1}{3}\left(R_1 + R_2 + R_3\right)$$

So, the variance is:

$$\begin{aligned}\text{var}(R_P) &= \tfrac{1}{9}\text{var}\left(R_1 + R_2 + R_3\right) \\ &= \tfrac{1}{9}\left\{\text{var}(R_1) + \text{var}(R_2) + \text{var}(R_3) + 2C_{1,2} + 2C_{1,3} + 2C_{2,3}\right\} \\ &= \tfrac{1}{9}\left\{32.45 + 10.3 + 101.2 + 2 \times 16.5 + 2 \times 55 + 2 \times 30\right\} \\ &= 38.55\end{aligned}$$

(ii)(b) ***Systematic risk***

The beta of the portfolio, β_P, is the weighted average of the betas of the individual securities:

$$\beta_P = \tfrac{1}{3}\left(\beta_1 + \beta_2 + \beta_3\right) = \tfrac{1}{3}\left(1.1 + 0.6 + 2\right) = \tfrac{3.7}{3}$$

The systematic risk is:

$$\beta_P^2 V_M = \left(\frac{3.7}{3}\right)^2 \times 25 = 38.028$$

(ii)(c) ***Specific risk***

The specific risk, $V_{\varepsilon P}$, can be calculated as the total portfolio variance (or risk) minus the systematic risk:

$$38.55 - 38.028 = 0.522$$

Alternatively, the specific risk of the portfolio can be calculated directly using the specific risks of the individual securities:

$$V_{\varepsilon P} = \text{var}\left(\tfrac{1}{3}\varepsilon_1 + \tfrac{1}{3}\varepsilon_2 + \tfrac{1}{3}\varepsilon_2\right) = \left(\tfrac{1}{3}\right)^2 V_{\varepsilon 1} + \left(\tfrac{1}{3}\right)^2 V_{\varepsilon 2} + \left(\tfrac{1}{3}\right)^2 V_{\varepsilon 3}$$

$$= \left(\tfrac{1}{3}\right)^2 \times 2.2 + \left(\tfrac{1}{3}\right)^2 \times 1.3 + \left(\tfrac{1}{3}\right)^2 \times 1.2 = 0.522$$

7.4 *Macroeconomic factor models* use observable economic time series as the factors.

They therefore include factors such as:

- annual rates of inflation
- economic growth
- short-term interest rates
- the yields on long-term government bonds
- the yield margin on corporate bonds over government bonds.

These are the macroeconomic variables that are assumed to influence security prices and returns in practice.

A related class of model uses a market index plus a set of industry indices as the factors.

Fundamental factor models are closely related to macroeconomic models but instead of (or in addition to) macroeconomic variables, they use company-specific variables.

These may include such fundamental factors as:

- the level of gearing
- the price earnings ratio
- the level of R&D spending
- the industry group to which the company belongs.

The commercially available fundamental factor models typically use many tens of factors.

Statistical factor models do not rely on specifying the factors independently of the historical returns data.

Instead a technique called principal components analysis can be used to determine a set of orthogonal indices that explain as much as possible of the observed variance.

However, the resulting indices are unlikely to have any meaningful economic interpretation and may vary considerably between different data sets.

7.5 (i) ***Mean investment returns***

The mean investment return on security *i*, E_i, is found from:

$$E_i = E[R_i] = E\left[\alpha_i + \beta_i R_m + \varepsilon_i\right]$$

Since α_i and β_i are both constants, $E\left[\alpha_i\right] = \alpha_i$ and $E\left[\beta_i\right] = \beta_i$. [½]

Also, by assumption, $E(\varepsilon_i) = 0$. [½]

So: $E_i = E\left[\alpha_i + \beta_i R_m + \varepsilon_i\right] = \alpha_i + \beta_i E_m$ [1]

where E_m is the mean return on the market.

For a portfolio *P* with portfolio weightings x_i , $i = 1,\ldots,n$, the mean investment return is given by:

$$E_P = E\left[\sum_{i=1}^{n} x_i R_i\right] = \sum_{i=1}^{n} x_i E_i = \sum_{i=1}^{n} x_i \left(\alpha_i + \beta_i E_m\right)$$ [½]

If we define:

$$\alpha_P = \sum_{i=1}^{n} x_i \alpha_i \quad \text{and} \quad \beta_P = \sum_{i=1}^{n} x_i \beta_i$$

then this can be written as:

$$E_P = \alpha_P + \beta_P E_m$$ [½]

[Total 3]

(ii) ***Covariance***

The covariance between security i and the portfolio is:

$$\begin{aligned} C_{iP} &= \text{cov}(R_i, R_P) \\ &= \text{cov}\left(R_i, \sum_{j=1}^{n} x_j R_j\right) \\ &= \sum_{j=1}^{n} x_j \,\text{cov}\left(R_i, R_j\right) \\ &= \sum_{j=1}^{n} x_j C_{ij} \end{aligned}$$ [2]

(iii) ***Variance of portfolio P***

We can write the portfolio variance in terms of the covariances, as follows:

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j C_{ij} \qquad [1]$$

(iv) ***Show the formula for β_{iP}***

Differentiating the left-hand side of the expression in (iii) with respect to x_i gives:

$$2\sigma_P \frac{\partial \sigma_P}{\partial x_i} \quad \ldots (1) \qquad [1]$$

The right-hand side of the expression in (iii) can be written as:

$$\sum_{i=1}^{n} x_i^2 C_{ii} + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} x_i x_j C_{ij} \qquad [1]$$

Differentiating this with respect to x_i gives:

$$2x_i C_{ii} + 2\sum_{\substack{j=1 \\ j\neq i}}^{n} x_j C_{ij} = 2\sum_{j=1}^{n} x_j C_{ij} \qquad [1½]$$

Note that here the second term gives a contribution from each summation.

Using the result from (ii) above, this can be written as:

$$2\,C_{iP} \quad \ldots (2) \qquad [½]$$

Hence, equating Equations (1) and (2) and cancelling the 2's gives:

$$\sigma_P \frac{\partial \sigma_P}{\partial x_i} = C_{iP} \qquad [1]$$

Dividing both sides of this equation by σ_P^2:

$$\frac{1}{\sigma_P}\frac{\partial \sigma_P}{\partial x_i} = \frac{C_{iP}}{\sigma_P^2} \qquad [½]$$

The left-hand side now matches the definition of β_{iP} given in the question. So we have:

$$\beta_{iP} = \frac{\partial \sigma_P}{\partial x_i}\frac{1}{\sigma_P} \qquad [½]$$

The definition of β_{iP} given here is:

$$\beta_{iP} = \frac{C_{iP}}{\sigma_P^2} = \frac{\text{cov}(R_i, R_P)}{V_P}$$

So β_{iP} represents the beta of security i relative to portfolio P. The equation we have derived shows us that it is equal to the proportionate change in the standard deviation of the portfolio returns when there is a small change in the portfolio weighting x_i. [1]

[Total 7]

8

Asset pricing models

Syllabus objectives

4.2 Asset pricing models

4.2.1 Describe the assumptions, principal results and uses of the Sharpe-Lintner-Mossin Capital Asset Pricing Model (CAPM).

4.2.2 Discuss the limitations of the basic CAPM and some of the attempts that have been made to develop the theory to overcome these limitations.

4.2.3 Perform calculations using the CAPM.

4.2.4 Discuss the main issues involved in estimating parameters for asset pricing models.

0 Introduction

Mean-variance portfolio theory, discussed previously, showed how an individual investor can characterise the relationship between risk and return for a particular security. The *capital asset pricing model* described in this chapter extends the ideas discussed previously, in an attempt to characterise the entire investment market on the assumption that investors behave exactly as predicted by those models.

The capital asset pricing model tells us about the relationship between risk and return in the security market as a whole, assuming that investors act in accordance with mean-variance portfolio theory and that the market is in equilibrium.

Note that asset pricing models are also sometimes referred to as *equilibrium models*, because they characterise the equilibrium outcome in investment markets.

1 The capital asset pricing model (CAPM)

1.1 Introduction

Portfolio theory can be applied by a single investor given their own estimates of security returns, variances and covariances. The capital asset pricing model developed by Sharpe, Lintner and Mossin introduces additional assumptions regarding the market and the behaviour of other investors to allow the construction of an equilibrium model of prices in the whole market.

Hence, the assumption here is that all investors select their investments by applying the ideas and assumptions underlying mean-variance portfolio theory.

Question

List the assumptions underlying mean-variance portfolio theory.

Solution

The assumptions of mean-variance portfolio theory are that:

- all expected returns, variances and covariances of pairs of assets are known
- investors make their decisions purely on the basis of expected return and variance
- investors are non-satiated
- investors are risk-averse
- there is a fixed single-step time period
- there are no taxes or transaction costs
- assets may be held in any amounts, (with short-selling, infinitely divisible holdings, no maximum investment limits).

If this is the case, then the introduction of some additional assumptions enables us to characterise how investors may act in aggregate and thereby construct an equilibrium model of security prices. The resulting capital asset pricing model (CAPM) tells us about the relationship between risk and return for security markets as a whole.

1.2 Assumptions

The extra assumptions of CAPM are:

- **All investors have the same one-period horizon.**
- **All investors can borrow or lend unlimited amounts at the same risk-free rate.**
- **The markets for risky assets are perfect. Information is freely and instantly available to all investors and no investor believes that they can affect the price of a security by their own actions.**
- **Investors have the same estimates of the expected returns, standard deviations and covariances of securities over the one-period horizon.**
- **All investors measure in the same 'currency' *eg* pounds or dollars or in 'real' or 'money' terms.**

A number of conditions need to be met for an investment market to be perfect – basically there must be no anomalies or distortions in the pricing of assets. The following are the basic requirements for a perfect market:

- There are many buyers and sellers, so that no one individual can influence the market price.
- All investors are perfectly informed.
- Investors all behave rationally.
- There is a large amount of each type of asset.
- Assets can be bought and sold in very small quantities, *ie* perfect divisibility.
- There are no taxes.
- There are no transaction costs.

Not all of the above assumptions are 100% realistic. However, the fact that the assumptions do not hold does not necessarily invalidate the CAPM, as it may nevertheless yield useful insight into the operation of security prices and returns. We must recognise, however, that it will be only an approximation.

1.3 Consequences of the extra assumptions

Given the extra assumptions above, we can build on the results of mean-variance portfolio theory to develop the standard form of the capital asset pricing model as follows.

1. **If investors have homogeneous expectations, then they are all faced with the same efficient frontier of risky securities.**

Question

Why?

Solution

If investors:

- have the same estimates of the expected returns, standard deviations and covariances of securities over the one-period horizon and
- are able to perform correctly all the requisite calculations

then they will all arrive at the same opportunity set and hence the same efficient frontier of risky securities.

2. If in addition they are all subject to the same risk-free rate of interest, the efficient frontier collapses to the straight line in $E-\sigma$ space which passes through the risk-free rate of return on the *E*-axis and is tangential to the efficient frontier for risky securities.

Question

Suppose there are only two portfolios A and B available to invest in. A is a portfolio of risky assets and B is a portfolio consisting of just one risk-free asset. Show that the efficient frontier must be a straight line.

Solution

Firstly, we can use the formula for the expected return of the portfolio to express x_A in terms of E_P:

$$E_P = x_A E_A + x_B E_B = x_A E_A + (1 - x_A) E_B$$

$$\Leftrightarrow \qquad x_A = \frac{E_P - E_B}{E_A - E_B}$$

Also, recall from the chapter on mean-variance portfolio theory that the variance of the portfolio return is:

$$V_P = x_A{}^2 V_A + x_B{}^2 V_B + 2 x_A x_B C_{AB}$$

Because Portfolio B is risk-free, $V_B = 0$ and $C_{AB} = 0$. So the above equation simplifies to:

$$V_P = x_A{}^2 \sigma_A^2 = \left(\frac{E_P - E_B}{E_A - E_B} \right)^2 \sigma_A^2$$

$$\Leftrightarrow \qquad \sigma_P = \left(\frac{E_P - E_B}{E_A - E_B}\right)\sigma_A = \frac{\sigma_A}{E_A - E_B}E_P - \frac{\sigma_A E_B}{E_A - E_B}$$

This is a straight line in (E_P, σ_P) space.

So, if there is a risk-free asset, the efficient frontier must be a straight line.

Question

Explain why the new efficient frontier must be at a tangent to the old efficient frontier of risky assets.

Solution

The last question shows that the new efficient frontier must be a straight line. We also know that it must intercept the *E*-axis at *r* because the risk-free asset is efficient. (You cannot find a portfolio with the same zero variance and a higher return.)

If this straight line isn't at a tangent then it either passes above the old efficient frontier or it passes below it. It cannot pass above since there is no portfolio that exists here. If it passes below then it is not an efficient frontier because you can find portfolios of risky assets that have a higher expected return for the same variance. Therefore the only possibility is that it must be at a tangent.

All investors face the *same* straight-line efficient frontier in expected return-standard deviation space.

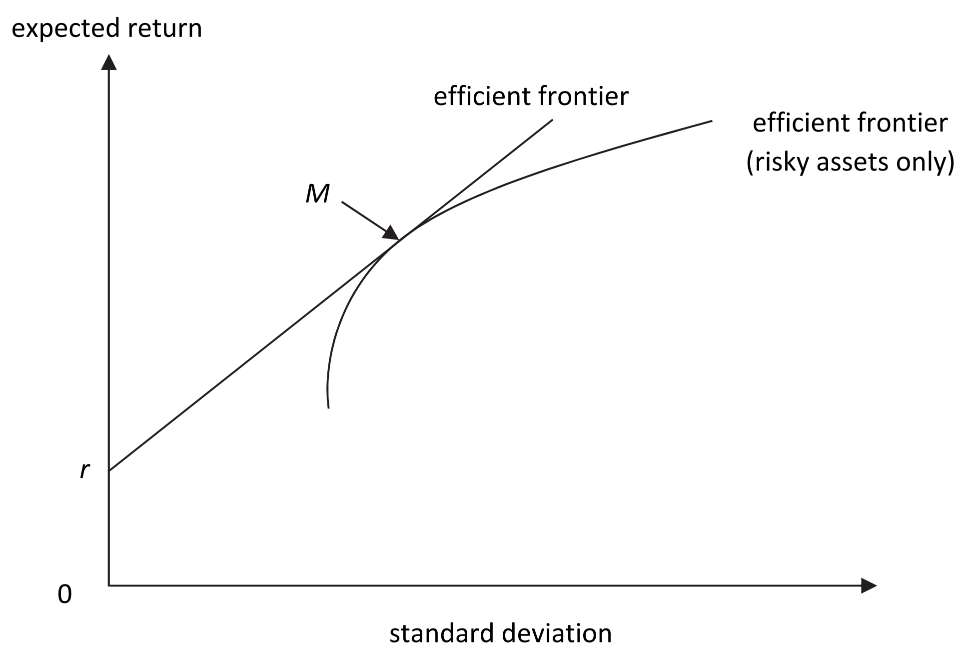


Figure 8.1: The efficient frontier with risk-free lending and borrowing

Note that:

- r represents the risk-free asset (which has zero standard deviation).
- The efficient frontier with a risk-free asset is *parabolic* in *expected return-variance* space. The straight line above in *expected return-standard deviation space* is a degenerate case of a hyperbola.
- The model does not require that investors all have the same attitude to risk, only that their views of the available securities are the same – and hence that the opportunity set is identical for all investors.

3. All rational investors will hold a combination of the risk-free asset and *M*, the portfolio of risky assets, at the point where the straight line through the risk-free return (on the *E*-axis) touches the original efficient frontier.

This is because if investors are rational they should only invest in efficient portfolios, which are located along the straight-line efficient frontier. Every investor should choose a portfolio of the form:

a% of the risk-free asset + $(100-a)$% of M

The portfolio of risky assets is shown as M on Figure 8.1. The choice of a depends on the investor's level of aversion to risk. Note that it could be negative for an investor who has low risk aversion, so that the investor's optimal portfolio lies on the efficient frontier to the right of M. This would mean that the investor is borrowing the risk-free asset and investing in the portfolio of risky assets M.

4. Because this is the portfolio held in different quantities by all investors, it must consist of all risky assets in proportion to their market capitalisation. It is commonly called the 'market portfolio'. The proportion of a particular investor's portfolio consisting of the market portfolio will be determined by their risk-return preference.

This is a key result of the CAPM. It is important to realise that *M is* the market portfolio, rather than the market portfolio just being a name we give to *M*.

Example

Suppose that the market of risky assets being considered is that consisting of FTSE 100 companies only and that there are 15 million investors. The market portfolio is then the FTSE 100 index itself. If Investor 1 has 5% of their portfolio of risky assets in Vodafone shares then Investor 2 must also have the same percentage and so on up to Investor 15,000,000. CAPM says that every single investor holds 5% of their portfolio of risky assets in Vodafone shares and so, because these investors cover the whole market, the market must hold 5% of its risky assets in Vodafone, *ie* Vodafone's market capitalisation would be 5% of all FTSE 100 shares.

If we apply this logic across all companies in the FTSE 100 then it becomes clear that every investor must hold shares in proportion to their market capitalisation. In this case every investor's portfolio of risky assets would be the FTSE 100 index, *ie* the market portfolio.

1.4 The separation theorem

The fact that the optimal combination of risky assets for an investor can be determined without any knowledge of their preferences towards risk and return (or their liabilities) is known as the *separation theorem*.

The separation theorem doesn't tell us whereabouts on the new efficient frontier an individual investor's portfolio will be. For this we need to know the investor's attitude towards risk and return, or equivalently the investor's utility function, in order to determine their preferred split between risky assets and the risk-free asset, *ie* the value of a.

However, we no longer have to make thousands of estimates of covariances in order to determine the portfolio of risky assets, because we know that it is always the market portfolio, M.

1.5 The capital market line

The straight line denoting the new efficient frontier is called the *capital market line.* Its equation is:

$$E_P - r = (E_M - r)\sigma_P / \sigma_M$$

where:

- E_P **is the expected return of any portfolio on the efficient frontier**
- σ_P **is the standard deviation of the return on portfolio P**
- E_M **is the expected return on the market portfolio**
- σ_M **is the standard deviation of the return on the market portfolio**
- r **is the risk-free rate of return.**

Thus, the expected return on any efficient portfolio is a linear function of its standard deviation. The factor $(E_M - r)/\sigma_M$ is often called the *market price of risk*.

Question

Derive the above equation of the capital market line.

Solution

We know that the points $(0,r)$ and (σ_M, E_M) are on the straight line and so its equation is:

$$\frac{E_P - r}{\sigma_P - 0} = \frac{E_M - r}{\sigma_M - 0}$$

We rearrange this to get the equation given for the capital market line, which can be found on page 43 of the *Tables*.

Note that:

- The expected return on any efficient portfolio *P* can be written as:

$$E_P = r + \left(\frac{E_M - r}{\sigma_M}\right)\sigma_P$$

ie expected return = risk-free rate + (market price of risk) × (amount of risk)

The '(price of risk) × (amount of risk)' term is sometimes referred to as the *risk premium*.

- The market price of risk is equal to the gradient of the capital market line in expected return-standard deviation space.

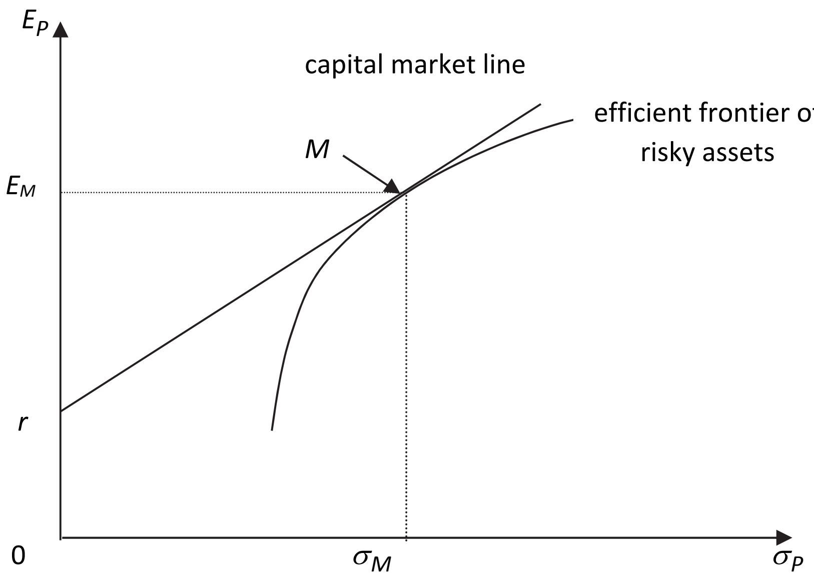


Figure 8.2: The capital market line

Hence, if the assumptions underlying the capital asset pricing model are true, then rational behaviour by investors should result in equilibrium security prices such that the expected return on any efficient portfolio is a *linear* function of its standard deviation. It is important to note that this result applies to efficient portfolios only and not to inefficient portfolios.

1.6 The security market line

It is also possible to develop an equation relating the expected return on *any* asset to the return on the market:

$$E_i - r = \beta_i (E_M - r)$$

where:

- E_i **is the expected return on security *i***
- r **is the return on the risk-free asset**
- E_M **is the expected return on the market portfolio**
- β_i **is the beta factor of security *i* defined as** $Cov[R_i, R_M]/V_M$ **.**

The above equation and definition of β_i can be found on page 43 of the *Tables*. The equation is often written as:

$$E_i = r + \beta_i (E_M - r)$$

This presentation emphasises that the expected return on any asset is again equal to the risk-free rate plus a risk premium, which here is derived from the beta of the security.

This is the equation of a straight line in $E - \beta$ space called the *security market line*. It shows that the expected return on any security can be expressed as a linear function of the security's covariance with the market as a whole. Since the beta of a portfolio is the weighted sum of the betas of its constituent securities, the security market line equation applies to portfolios as well as to individual securities.

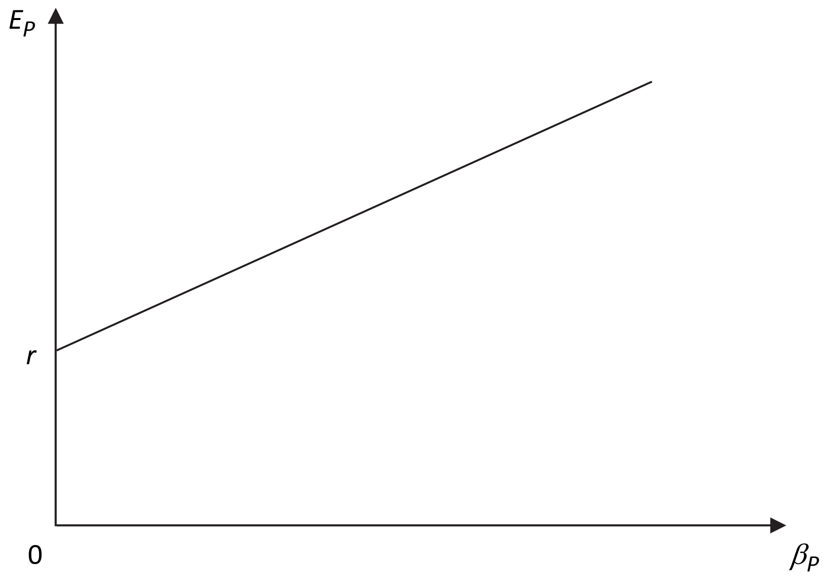


Figure 8.3: The security market line

Note that all portfolios, including those comprising a single security or asset, lie on the security market line whether or not they are efficient. Thus, according to the capital asset pricing model, the security market line relationship can be used to determine the expected return of *any* asset or portfolio from its beta. The expected return of a portfolio depends linearly upon its beta, which measures systematic risk and is independent of other non-systematic risk. Consequently, investors are rewarded only for systematic risk and not for non-systematic risk, precisely because they are able to diversify it away.

Question

Consider Security A, which has a standard deviation of investment returns of 4%. If:

- the standard deviation of the market return is 5%
- the correlation between A's return and that of the market is 0.75
- the risk-free rate is 5%
- and the expected return on the market is 10%

then calculate:

(i) the beta of Security A

(ii) Security A's expected return.

Solution

(i) ***Beta of Security A***

This is given by:

$$\beta_A = \frac{Cov(R_A, R_M)}{V_M} = \frac{\rho_{AM}\sigma_A\sigma_M}{\sigma_M^2}$$

$$= \frac{0.75 \times 0.04 \times 0.05}{0.05^2} = 0.6$$

(ii) ***Expected return of Security A***

This is given by:

$$\begin{aligned} E_A &= r + \beta_A(E_M - r) \\ &= 0.05 + 0.6(0.10 - 0.05) \\ &= 0.08 \end{aligned}$$

ie 8%.

Derivation of the security market line

One way to derive the security market line is to note that according to the capital asset pricing model, all investors will hold the market portfolio of risky assets as part of their overall investment portfolio. Consequently, each investor will hold a *very* well-diversified portfolio.

When discussing systematic and specific risk earlier in the course we saw that, as a portfolio becomes very well-diversified:

- the systematic risk of the portfolio tends towards a weighted average of the systematic risks of the constituent securities
- the non-systematic or specific risk tends to zero.

So, an investor who holds a well-diversified portfolio, should be concerned only with expected return and systematic risk. Hence portfolio decisions should be based only upon the expected return and the beta of the portfolio.

Consider two well-diversified portfolios, 1 and 2, with expected returns E_1 and E_2 and betas β_1 and β_2. If we construct a third portfolio consisting of equal holdings of Portfolios 1 and 2, then its expected return and beta will be equal to:

$$E_3 = ½\times(E_1 + E_2)$$

and $$\beta_3 = ½\times(\beta_1 + \beta_2)$$

Question

Why?

Solution

If this were not the case, then it would be possible to make an instantaneous, risk-free profit. For example, suppose that the beta relationship held, but that the expected return from Portfolio 3 was less than E_3. Then it would be possible to make risk-free profits by selling Portfolio 3 and using the proceeds to buy equal amounts of Portfolios 1 and 2. Starting with a zero initial sum, we could end up with a positive net expected return and hence a risk-free profit.

In practice, we would expect investors to notice the price anomaly and act exactly as above. Thus, the price of Portfolio 3 would be driven down (and hence its expected return up) and the prices of Portfolios 1 and 2 up (and hence their expected returns down) until the pricing anomaly was eliminated – with the security market line relationship again holding.

Recall that the capital asset pricing model is an *equilibrium* model. Therefore short-term deviations from the predicted expected returns may be possible when the market is out of equilibrium.

Hence, if we plot Portfolios 1, 2 and 3 in expected return-beta space then they must all lie on the same straight line. A similar argument applies to all portfolios and individual securities, the straight line in question being the security market line. The general form of the equation of a straight line in expected return-beta space is (for any security *i*):

$$E_i = a_0 + a_1\beta_i$$

Thus, to complete the derivation we need to determine the values of a_0 and a_1.

Question

What are the betas of the market portfolio and the risk-free asset? Use these beta values to complete the derivation of the security market line equation.

Solution

The beta factor of any portfolio *i* is defined as $Cov[R_i, R_M]/V_M$. Hence, for the market portfolio:

$$\beta_M = Cov[R_M, R_M]/V_M = V_M/V_M = 1$$

This must be the case since the return on the market is perfectly correlated with itself (*ie* the correlation coefficient equals one).

Conversely, the risk-free asset has, by definition, neither systematic nor specific risk and so its beta must be zero.

Now, the excess expected return on the market portfolio over and above the risk-free rate is $E_M - r$, whilst the excess systematic risk is $\beta_M = 1$. Hence, $E_M - r$ must also be the gradient of the security market line. As all portfolios lie on the security market line, for any portfolio with a beta β_P the excess expected return over and above the risk-free rate will equal $\beta_P \times (E_M - r)$.

Consequently, the *total* expected return on the same portfolio must be equal to:

$$E_P = \beta_P \times (E_M - r) + r$$

ie $\quad a_0 = r$ and $a_1 = E_M - r$

2 Limitations of CAPM

2.1 Limitations and empirical evidence

The limitations of the basic CAPM are well known and attempts have been made to overcome them since the model was first published. Most of the assumptions of the basic model can be attacked as unrealistic and, furthermore, empirical studies do not provide strong support for the model.

Models with less restrictive or unrealistic assumptions may therefore provide better predictions of the actual behaviour of security markets than the basic capital asset pricing model described above. A further subtle yet important issue is that if the model assumes that a particular factor such as tax does not exist, then it cannot tell us about the influence of taxation upon security markets. We might therefore wish to develop a model that explicitly introduces taxes.

There are basic problems in testing the model since, in theory, account has to be taken of the entire investment universe open to investors, not just capital markets.

An important asset of most investors, for example, is their human capital (*ie* the value of their future earnings).

In other words tests of the model should allow for every single asset that an investor could purchase and that yields an uncertain return to an investor, *eg* houses, works of art, training courses *etc*. It is therefore extremely difficult to test the validity of the capital asset pricing model if the returns on many assets cannot be observed. Moreover, even if actual returns are observable, the capital asset pricing model is expressed in terms of *expected* returns, which typically are not.

Nevertheless the studies that have been carried out do provide some evidence:

- in support of a linear relationship between return and systematic risk over long periods of time
- suggesting that return is *not* related to unsystematic or residual risk.

2.2 Extensions of the basic CAPM

Models have been developed which allow for decisions over multiple periods and for the optimisation of consumption over time to take account of this.

Other versions of the basic CAPM have been produced which allow for taxes and inflation, and also for a situation where there is no riskless asset.

Multi-period models

The basic capital asset pricing model assumes that investors are concerned only with a single-period time horizon. Various multi-period models have, however, been developed that consider how investors with preferences defined over a time horizon of more than one period make investment decisions over that time horizon. These models have shown that under certain assumptions, the investment decisions taken by investors with multi-period time horizons are consistent with those made by investors with just a single-period time horizon. Consequently, each single period can be considered in isolation and the results of the single period capital asset pricing model are equally valid (or not) in the multi-period context.

Amongst the multi-period models are:

- the *consumption capital asset pricing model* – which relates investment returns to the growth rate of per capita consumption
- one that allows for the uncertain *inflation* that will be present in a multi-period context – so that investors are concerned with *real* returns.

Model with taxes

The absence of taxes in the basic model means that investors are indifferent between income and capital gains. One extension of the basic capital asset pricing model therefore derives a somewhat more complicated equilibrium relationship allowing for differential taxation and based upon the means and variances of post-tax returns.

Zero-beta model

It can be shown that the absence of a risk-free asset does not alter the form of the security market line, the role of the risk-free asset simply being replaced by a 'zero-beta' portfolio, *ie* a portfolio of risky assets with a beta equal to zero. The equation of the security market line is then:

$$E_i = E_Z + (E_M - E_Z)\beta_i$$

where E_Z is the expected return of the zero-beta portfolio. This is sometimes referred to as the *zero-beta* version of the capital asset pricing model.

In the international situation there is no asset which is riskless for all investors (due to currency risks) so a model has been developed which allows for groups of investors in different countries, each of which considers their domestic currency to be risk-free.

3 Uses of CAPM

As its name suggests, the capital asset pricing model (CAPM) can be used to price assets, where these could be financial securities or other assets such as capital projects. If the beta of an asset can be estimated from past data, then the security market line can be used to estimate the prospective return that the asset should offer given its systematic risk, providing the economy is stable. This return can then be used to discount projected future cash flows and so price the asset. For a new asset with no history, the beta from a similar asset could be used.

CAPM can be used to estimate the expected return on a financial security given its exposure to the various risk factors modelled. This return can then be used to discount projected future cash flows and so price the security and determine if it appears to be under-valued or over-valued.

Question

How might you estimate the beta of a quoted share in practice?

Solution

To estimate the beta of the share you could obtain the returns on the share in question (R_i) over each of the last sixty months say, together with the monthly returns over the same period on an appropriate market-wide index (R_M). For example, in the UK you might use the FTSE All-Share Index.

The beta of the share could then be estimated as:

$$\beta_i = \frac{Cov(R_i, R_M)}{Var(R_M)}$$

This definition of β_i is given on page 43 of the *Tables*.

4 Estimating parameters for asset pricing models

The estimation of parameters is one of the most time-consuming aspects of stochastic asset modelling.

The simplest case is the purely statistical model, where parameters are calibrated entirely to past time series. Provided the data is available, and reasonably accurate, the calibration can be a straightforward and mechanical process.

Of course, there may not always be as much data as we would like, and the statistical error in estimating parameters may be substantial.

Consequently, the results obtained may lack credibility.

Furthermore, there is a difficulty in interpreting data which appears to invalidate the model being fitted. For example, what should be done when fitting a Gaussian model (discussed in Subjects CS1 and CS2) in the presence of large outliers in the data? Perhaps the obvious course of action is to reject the hypothesis of normality, and to continue building the model under some alternative hypothesis that accommodates extreme events.

The rejection of normality is a big step because non-normal models are generally much more difficult to apply and use than those based on the assumption of normality.

Question

Explain why the outliers might be important in an investment context.

Solution

The outliers are important because this is where the major financial risks lie. It is very low investment returns – *ie* returns in the lower tail of the distribution – that will lead to a fall in asset values and possible difficulties with regard to financial solvency.

In practice, a more common approach to outliers is to exclude them from the statistical analysis, and focus attention instead on the remaining residuals which appear more normal. The model standard deviation may be subjectively nudged upwards after the fitting process, in order to give some recognition to the outliers which have been excluded.

It has often been the practice in actuarial modelling to use the same data set to specify the model structure, to fit the parameters, and to validate the model choice. A large number of possible model structures are tested, and testing stops when a model which passes a suitable array of tests is found. Unfortunately, in this framework, we may not be justified in accepting a model simply because it passes the tests. Many of these tests (for example, tests of stationarity) can be weak and may not reject incorrect models.

Indeed, even if the 'true' model was not in the class of models being fitted, we would still end up with an apparently acceptable fit, because the rules say we keep generalising until we find one.

As we add more and more variables to a model, the model will necessarily fit the historical data more closely. Whether it is capable of a meaningful interpretation is another matter. Again, and as always, the modelling process relies heavily on the skill and judgement of the modeller.

This process of generalisation tends to lead to models which wrap themselves around the data, resulting in an understatement of future risk, and optimism regarding the accuracy of out-of-sample forecasts.

An alternative approach to this that is widely used by econometricians and which avoids this problem of 'data mining' is the *general-to-specific* approach. Here the modeller starts with a very general model, which includes all the variables (both current values and lagged values) that are thought likely to influence the variable being modelled. The model is then made more specific by eliminating variables (one at a time) that do not materially affect the significance of the fit to past data.

In the context of economic models, the calibration becomes more complex. The objective of such models is to simplify reality by imposing certain stylised facts about how markets would behave in an ideal world. This theory may impose constraints, for example, on the relative volatilities of bonds and currencies. Observed data may not fit these constraints perfectly. In these cases, it is important to prioritise the features of the economy that are most important to calibrate accurately for a particular application.

Thus, we need to decide which is the most important – fitting past data as accurately as possible or complying with economic theory. In practice, there will often be a trade-off between these two objectives.

Question

Why should the modeller not aim solely to fit the model as accurately as possible to past data?

Solution

The modeller should not aim solely to fit the model as accurately as possible to past data because the past might not be a very accurate indication of the future – *eg* if there are structural changes in investment and/or economic markets.

Chapter 8 Summary

Assumptions of the CAPM (including MPT assumptions)

- Investors make their decisions purely on the basis of expected return and variance. So all expected returns, variances and covariances of assets are known.
- Investors are non-satiated and risk-averse.
- There are no taxes or transaction costs.
- Assets may be held in any amounts.
- All investors have the *same* fixed one-step time horizon.
- All investors make the *same* assumptions about the expected returns, variances and covariances of assets.
- All investors measure returns consistently (*eg* in the same currency or in the same real/nominal terms).
- The market is perfect.
- All investors may lend or borrow any amounts of a risk-free asset at the same risk-free rate r.

The extra assumptions of CAPM from MPT move away from thinking about individual investors to assumptions about the entire economy. CAPM is an *equilibrium model*.

Results of the CAPM

- All investors have the same efficient frontier of risky assets.
- The efficient frontier collapses to a straight line in $E-\sigma$ space in the presence of the risk-free asset.
- All investors hold a combination of the risk-free asset and the same portfolio of risky assets *M.*
- *M* is the market portfolio – it consists of all assets held in proportion to their market capitalisation.

The *separation theorem* suggests that the investor's choice of portfolio of risky assets is independent of their utility function.

Capital market line

$$E_p = r + \frac{\sigma_P}{\sigma_M}(E_M - r)$$

Market price of risk

$$MPR = \frac{E_M - r}{\sigma_M}$$

Security market line

$$E_p = r + \beta_P(E_M - r)$$

where:

$$\beta_P = \frac{Cov\left(R_P, R_M\right)}{Var\left(R_M\right)}$$

The main limitations of the basic CAPM are that most of the assumptions are unrealistic and that empirical studies do not provide strong support for the model. However there is some evidence to suggest a linear relationship between expected return and systematic risk.

Estimating parameters for asset pricing models

Determining the parameter values to use in asset pricing models poses significant challenges. This is compounded by the fact that even the choice of model itself may be inappropriate.

Chapter 8 Practice Questions

8.1 Exam style

An investor has the choice of the following assets that earn rates of return as follows in each of the four possible states of the world:

State	Probability	Asset 1	Asset 2	Asset 3
1	0.2	5%	5%	6%
2	0.3	5%	12%	5%
3	0.1	5%	3%	4%
4	0.4	5%	1%	7%
Market capitalisation		10,000	17,546	82,454

Determine the market price of risk assuming CAPM holds.

Define all terms used. [6]

8.2 (i) Explain what is meant by specific risk and systematic risk in the CAPM.

(ii) Explain the meaning of the 'beta' of a share, and describe how you would calculate it for:

(a) a company

(b) a portfolio.

(iii) Explain how the beta for a portfolio can be used to determine the expected return for the portfolio.

(iv) Why might the beta calculated in (ii)(a) be inappropriate for practical use?

8.3 Exam style

(i) State the assumptions of the capital asset pricing model (CAPM). [5]

(ii) An investment market consisting of a risk-free asset and a very large number of stocks is such that, for modelling purposes, the market capitalisation of the k-th stock can be expressed as:

$$\frac{1}{2^k} \quad \text{where } k = 1,2,3,\ldots$$

The expected return on the k-th stock (expressed as a percentage) is:

$$25e^{-(k-1)} + 5\left(1 - e^{-(k-1)}\right)$$

Assuming that the CAPM assumptions hold, find the expected return on the portfolio of risky assets held by each investor. [5]

[Total 10]

8.4 Exam style (i) Within the context of the capital asset pricing model, explain what is meant by the 'market price of risk'. [3]

(ii) Show how the security market line relationship can be rearranged to give an expression for the expected return in terms of the market price of risk γ_M, and briefly interpret your answer. [3]

[Total 6]

8.5 Exam style (i) (a) State the equation of the security market line and, assuming that the market portfolio offers a return in excess of the risk-free rate, use it to derive the betas of the market portfolio and the risk-free asset.

(b) Draw a diagram of the security market line relationship.

(c) What does the security market line indicate about the relationship between risk and return? [9]

(ii) (a) By considering two points on the capital market line, determine its equation and comment briefly upon its applicability.

(b) Briefly interpret each of the terms in the relationship. [6]

[Total 15]

Chapter 8 Solutions

8.1 The market price of risk is:

$$\frac{E_M - r}{\sigma_M}$$

where:

- r is the risk-free interest rate
- E_M is the expected return on the market portfolio consisting of all risky assets
- σ_M is the standard deviation of the return on the market portfolio. [1]

Since Asset 1 always gives the same return of 5%, it is risk-free. So the risk-free interest rate is $r = 5\%$. [½]

Assets 2 and 3, with the capitalisations shown, constitute the market portfolio of risky assets. [½]

The total capitalisation of the market is $17{,}546 + 82{,}454 = 100{,}000$. The table below shows the possible returns on this market portfolio:

State	Probability	Return	Return (%)
1	0.2	$5\% \times 17{,}546 + 6\% \times 82{,}454 = 5{,}824.54$	5.82454%
2	0.3	$12\% \times 17{,}546 + 5\% \times 82{,}454 = 6{,}228.22$	6.22822%
3	0.1	$3\% \times 17{,}546 + 4\% \times 82{,}454 = 3{,}824.54$	3.82454%
4	0.4	$1\% \times 17{,}546 + 7\% \times 82{,}454 = 5{,}947.24$	5.94724%

[2]

So the expected return is:

$$E_M = 0.2 \times 5.82454\% + \cdots + 0.4 \times 5.94724\% = 5.794724\%$$ [½]

The variance of the returns is:

$$\sigma_M^2 = 0.2 \times (5.82454\%)^2 + \cdots + 0.4 \times (5.94724\%)^2 - (5.794724\%)^2$$

$$= 0.454020\%\% = (0.673810\%)^2$$ [1]

So the market price of risk is:

$$\frac{E_M - r}{\sigma_M} = \frac{5.794724\% - 5\%}{0.673810\%} = 1.179$$ [½]

This shows the extra expected return (over and above the risk-free rate) per unit of extra risk taken (as measured by the standard deviation) by investing in risky assets.

[Total 6]

8.2 (i) ***Specific and systematic risk***

The fluctuation (both up and down) of returns from a security can be broken into two components according to the extent to which:

- company/industry specific events cause the returns to vary independently of movements in the investment market as a whole (*ie specific risk*)
- the returns from the security move with the market as a whole (*ie systematic risk*).

Specific risk is the risk unique to a particular security that can be eliminated from a portfolio if the portfolio is suitably diversified. In terms of portfolio theory, it is the unrewarded risk.

Systematic risk cannot be diversified away.

(ii) ***Beta for a share***

A share's beta is a measure of its systematic risk. It is a coefficient that measures the extent to which the return from the security covaries with the return from the market of risky assets as a whole. It also indicates how the risk premium on the share compares to that for the market of risky assets as a whole.

If the returns on a particular share move more aggressively than the market, the share has a high beta (greater than 1). A defensive share that does not fluctuate as much as the market would have a coefficient below 1.

(a) ***To calculate it for a company***

Calculate the return on the share over a suitably large set of periods (say each month for 5 years) and also for the whole market.

Plot these values with market returns on the horizontal axis, and the corresponding share returns on the vertical axis, and find the gradient of the line of best fit (by least-squares regression). The beta is estimated as the gradient of the line.

We do, however, need to be wary of company- or industry-specific events that may have caused the historical beta of the company to change over the period of estimation. The estimate of the prospective beta may need to be adjusted accordingly.

(b) ***To calculate it for a portfolio***

The beta is the weighted average of the betas for the individual shares, weighted by the value of the holding for each of the shares. So to calculate it, we would calculate the weighted average using the betas as worked out above. Alternatively, repeat (a) using the returns on the portfolio.

(iii) ***Expected return for the portfolio***

Assuming the capital asset pricing model holds, the beta for the portfolio gives a guide as to how the portfolio's return is expected to differ from the market as a whole. The following data is needed:

- β_P the beta for the portfolio
- r the risk-free rate of return (for short-term periods, take this to be Treasury bill returns, for longer periods perhaps look at government bond yields)
- E_M the expected return on the market as a whole.

The expected return for the portfolio is given by the security market line equation as:

$$E_P = r + (E_M - r)\beta_P$$

Thus, for a portfolio with a beta of 1, the expected return on the portfolio is equal to the expected market return.

If the market is not efficient, the expected return may be higher or lower than this.

(iv) ***Why might the beta be inappropriate?***

1. An analysis over a limited time period may produce an estimate for the beta with some random bias. Empirical evidence suggests that betas for individual companies are not stable.

2. A company's beta may change over time as the company may have a shift in emphasis and management.

   - So the beta based on an historical analysis may not be appropriate for the company as it currently is.
   - Similarly, the current beta may not be appropriate for the future.

3. The assumptions underlying portfolio theory and CAPM may not hold exactly. So that even if we have an accurate estimate of beta we cannot use the equation in (iii) to estimate expected return.

8.3 (i) ***CAPM assumptions***

- Investors make their decisions purely on the basis of expected return and variance. So all expected returns, variances and covariances of assets must be known. [1]
- Investors are non-satiated. [½]
- Investors are risk-averse. [½]
- There are no taxes or transaction costs. [½]
- Assets may be held in any amounts. [½]
- All investors have the *same* fixed one-step time horizon. [½]

- All investors make the *same* assumptions about the expected returns, variances and covariances of assets. [½]
- All investors measure returns consistently (*eg* in the same currency or in the same real/nominal terms). [½]
- The market is perfect and in equilibrium. [½]
- All investors may lend or borrow any amounts of a risk-free asset at the same risk-free rate r. [½]

[Maximum 5]

(ii) ***Market portfolio***

If CAPM holds then the portfolio of risky assets held by each investor will be the market portfolio. [½]

With a very large number of stocks, the expected return on the market portfolio is:

$$E_M = \sum_{k=1}^{\infty} x_k E_k$$

where x_k is the holding in stock k and E_k is the expected return on that stock. [½]

In this case we have:

$$E_M = \sum_{k=1}^{\infty} x_k E_k = \sum_{k=1}^{\infty} \frac{1}{2^k}\left[25e^{-(k-1)} + 5\left(1-e^{-(k-1)}\right)\right] \qquad [1]$$

Simplifying, we get:

$$E_M = \sum_{k=1}^{\infty} \frac{1}{2^k}\left[5+20e^{-(k-1)}\right] = 5\sum_{k=1}^{\infty}\frac{1}{2^k} + 20\sum_{k=1}^{\infty}\frac{1}{2^k}e^{-(k-1)}$$

$$= \left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots\right)\times 5 + 20e\sum_{k=1}^{\infty}\left(\frac{1}{2e}\right)^k$$

$$= \left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots\right)\times 5 + 20e\left[\left(\frac{1}{2e}\right)+\left(\frac{1}{2e}\right)^2+\left(\frac{1}{2e}\right)^3+\ldots\right] \qquad [1½]$$

Using the formula for the infinite sum of a geometric progression, we get:

$$E_M = 5\times\frac{\frac{1}{2}}{1-\frac{1}{2}} + \frac{20e\left(\frac{1}{2e}\right)}{1-\left(\frac{1}{2e}\right)} = 5 + \frac{20e}{2e-1} = 17.25$$

So the expected return on the market is 17.25%. [1½]

[Total 5]

8.4 (i) ***Market price of risk***

Within the context of the capital asset pricing model, the *market price of risk* is defined as:

$$\frac{E_M - r}{\sigma_M}$$

where:

- E_M = the expected return on market portfolio
- r = the risk-free rate of return
- σ_M = the standard deviation of market portfolio returns. [1]

It is the additional expected return that the market requires in order to accept an additional unit of risk, as measured by the portfolio standard deviation of return. [1]

It is equal to the gradient of the capital market line in $E-\sigma$ space. [1]

[Total 3]

(ii) ***Security market line***

Recall the security market line relationship $E_i - r = \beta_i(E_M - r)$ and also that the beta of a security is defined as:

$$\beta_i = \frac{\text{cov}(R_i, R_M)}{\sigma_M^2}$$

If we define $\sigma_{iM} = cov(R_i, R_M)$, then the security market line can written as:

$$E_i = r + \frac{\sigma_{iM}}{\sigma_M^2}\left(E_M - r\right) = r + \frac{\sigma_{iM}}{\sigma_M}\left(\frac{E_M - r}{\sigma_M}\right) = r + \frac{\sigma_{iM}}{\sigma_M}\gamma_M \qquad [2]$$

As $\dfrac{\sigma_{iM}}{\sigma_M} = \dfrac{\rho_{iM}\sigma_i\sigma_M}{\sigma_M} = \rho_{iM}\sigma_i$ is a measure of the risk of portfolio i, the security market line states that the expected return on any portfolio can be expressed as the sum of the risk-free rate and the amount of risk multiplied by the market price of risk, *ie*:

expected return = risk-free rate + (market price of risk) $\times$ (amount of risk) [1]

[Total 3]

8.5 (i)(a) ***Equation of the security market line***

The security market line for any portfolio *P* is:

$$E_P = r + (E_M - r)\beta_P$$

where:

- E_P is the expected return on portfolio *P*
- r is the risk-free rate of return
- E_M is the expected return on the market portfolio
- β_P is the beta of the portfolio with respect to the market portfolio [1]

The security market line holds for all securities and portfolios. Thus, applying it to the market portfolio gives:

$$E_M = r + (E_M - r)\beta_M \quad \Rightarrow \quad E_M(1-\beta_M) = r(1-\beta_M)$$

Given that $E_M \neq r$, it must be the case that β_M = 1. [1]

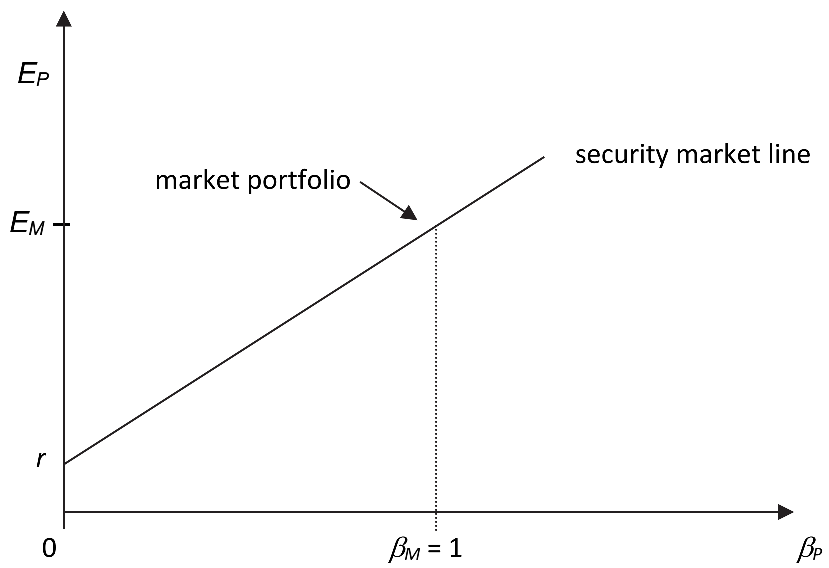
Similarly, applying the security market line relationship to the risk-free asset (with a beta of β_r) gives:

$$r = r + (E_M - r)\beta_r$$

ie $$0 = (E_M - r)\beta_r$$

Given that $E_M \neq r$, then it must be the case that β_r = 0, *ie* the risk-free asset has a beta of zero – which must be the case as it involves zero risk – systematic or otherwise. [1]

(i)(b) ***Diagram***



[2 for correct diagram]

(i)(c) ***What does the security market line indicate?***

The security market line relationship is of interest because:

- it enables us to determine the expected return on *any* asset or portfolio. This can be done if we can estimate the risk-free rate, the expected return on the market portfolio and the beta of the individual asset or portfolio. [1]

- it tells us that the expected return on any asset is equal to the risk-free rate plus a risk premium, which is a linear function of the systematic risk of the asset as measured by the beta factor. [1]

- it tells us that expected return does not depend on any other factors and in particular it is independent of the specific risk of an asset, which can be eliminated by diversification. [1]

The above results do of course depend upon the appropriateness or otherwise of the capital asset pricing model. [1]

[Total 9]

(ii)(a) ***Capital market line relationship***

The capital market line is the equation of the efficient frontier in (E,σ) space, which is a straight line. [½]

It passes through the risk-free asset with coordinates $(0,r)$ and the market portfolio, which has coordinates (σ_M,E_M). [½]

Thus, the gradient of the capital market line is equal to:

$$\frac{E_M-r}{\sigma_M-0}=\frac{E_M-r}{\sigma_M}$$

It has an intercept on the vertical axis at the risk-free rate r and consequently, for any efficient portfolio P, its equation must be:

$$E_P = r+\left(\frac{E_M-r}{\sigma_M}\right)\sigma_P \qquad [½]$$

The capital market line relationship only holds for efficient portfolios – those for which there is no other portfolio that offers either a higher expected return for a given risk or a lower risk for a given expected return – assuming that the capital asset pricing model itself applies. Efficient portfolios are always combinations of the risk-free asset and the market portfolio. [1½]

(ii)(b) ***Interpret the relationship***

r is the risk-free rate of return, *ie* the rate of return on a security that has a zero standard deviation of return. This is sometimes interpreted as the return on a Treasury bill. [1]

The quantity $\dfrac{E_M-r}{\sigma_M}$ is the *market price of risk*.

It can be interpreted as the extra expected return that can be gained by increasing the level of risk of an efficient portfolio by one unit. In this context, risk strictly means the standard deviation of investment returns. [1]

The second term in the relationship, $\left(\frac{E_M - r}{\sigma_M}\right)\sigma_P$, is known as the risk premium. It represents the additional return over and above the risk-free rate that can be obtained on a portfolio *P* by accepting risk, *ie* a non-zero portfolio standard deviation. [1]

[Total 6]

9

Brownian motion and martingales

Syllabus objectives

4.4 Stochastic models for security prices

4.4.2 Explain the definition and basic properties of standard Brownian motion (or Wiener process).

0 Introduction

Essentially, a stochastic process is a sequence of values of some quantity where the future values cannot be predicted with certainty. This and the following chapter are concerned with continuous-time stochastic processes that have applications in financial economics. These chapters are of a very mathematical nature and you may find some of it hard-going. It is more important that you gain a higher-level understanding than that you master the pure maths that underlies it. For example, you may find it useful to learn Ito's Lemma as a procedure rather than trying to understand the pure mathematical concepts.

The most important process studied here is the Wiener process, also known as Brownian motion, which is the subject of Section 1. These two terms will be used interchangeably. We define this as a process with continuous sample paths and independent and normally distributed increments. A Brownian motion is the continuous-time version of a random walk, as we will see. The graph in Section 1.2 shows a typical sample path.

If security prices can be modelled in some way in terms of Brownian motion, this will be useful for pricing certain types of options. This is discussed further in Parts 3 and 4 of the course.

Section 2 of this chapter introduces martingales. A martingale is a process whose current value is the best estimate of its future values. We will see later that martingale theory has important applications in relation to financial derivatives.

The notation used in financial economics generally is not standardised and similar notation can refer to different quantities: readers should check the definitions provided in each section. In particular, the value of a random stochastic process can be equivalently written as X_t or $X(t)$. Furthermore, standard Brownian motion can be denoted by B_t (as in the *Tables*), or W_t or Z_t as found throughout the Core Reading.

The Core Reading in this chapter is adapted from course notes written by Timothy Johnson.

1 Introduction to Brownian motion

1.1 Introduction

In 1895, Louis Bachelier embarked on a doctorate on the 'Theory of Speculation'. Bachelier's approach was fairly conventional at the time; he would model an asset price as a random walk. At the start of his thesis he argues that:

At a given instant the market believes neither in a rise nor in a fall of the true price.

Which means that:

The mathematical expectation of the speculator is zero.

His innovation was to consider the walk to be continuous, rather than a discrete-time random walk. This is analogous to moving from the binomial to the normal distribution.

Bachelier was unable to mathematically define the paths he discussed. A little later, in 1903–1904, Einstein used a similar model to represent the motion of atoms/molecules in a liquid. Einstein also failed to define the path he was working with. However, since his paper was used as evidence that atoms existed it became important in physics that the paths were rigorously defined. This was done by Norbert Wiener in 1921. Today physicists will refer to the paths used by Bachelier as 'Brownian motion', a physical process, while mathematicians refer to them as a 'Wiener process', which is a mathematical object.

The phenomenon of 'Brownian motion' is named after the nineteenth century botanist Robert Brown who observed the random movement of pollen particles in water. The path of a two-dimensional Brownian motion process bears a resemblance to the track of such pollen particles.

1.2 Definition of the Wiener process (standard Brownian motion)

A stochastic process W_t, $t \geq 0$ is a Wiener process if:

(i) $W_0 = 0$

(ii) W_t **has continuous sample paths.**

This means that the graph of W_t as a function of t doesn't have any breaks in it.

(iii) **For any $0 \leq s < t$ the increment $W_t - W_s$ is normally distributed, $W_t - W_s \sim N(0, t-s)$.**

This property shows that the increments are stationary in that their statistical properties rely on the size of the interval $t-s$. The concept of stationarity is discussed further in Subject CS2.

(iv) W_t **has independent increments, that is for any sequence of times $0 \leq t_1 < t_2 < \cdots < t_n$ we have that the increments $W_{t_n} - W_{t_{n-1}}, \ldots, W_{t_3} - W_{t_2}, W_{t_2} - W_{t_1}$ are independent random variables.**

Alternatively, $W_t - W_s$ is independent of $F_s^W = \sigma(W_{u \leq s})$, the natural filtration of W_s.

The natural filtration F_t^W represents the history of the process up to and including time t. This concept is covered in more detail in the martingale section of the chapter. F_t^W may be written as $\sigma(W_{u \le t})$ to denote that it is the filtration generated by the process W_t.

The fact that a Wiener Process has independent increments implies it is Markovian (in fact it is 'strong Markovian').

Intuitively, a Markov process is one where, if we know the latest value of the process, we have all the information required to determine the probabilities for the future values. Knowing the historical values of the process as well would not make any difference. Markov processes are also discussed in Subject CS2.

Since $W_t \sim N(0,t)$ it should be clear that $P(W_t < \infty,\ t < \infty) = 1$.

Property (iii) combined with property (i) gives us $W_t = W_t - W_0 \sim N(0,t)$, which results in $E[W_t] = 0$.

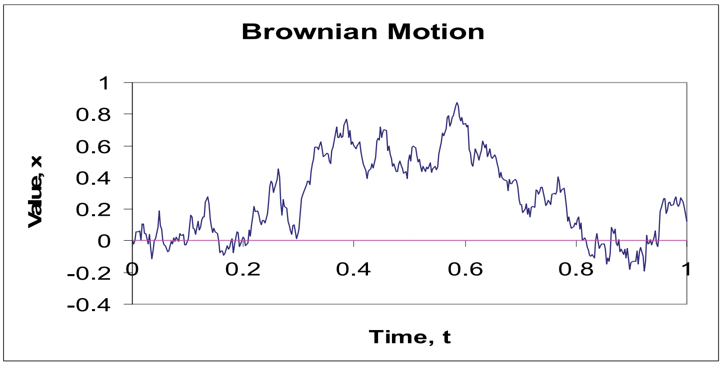


Figure 9.1: a typical sample path of Brownian motion

Brownian motion can be viewed as the continuous version of a simple symmetric random walk.

1.3 Brownian motion in general

Standard Brownian motion is a special case of the more general form of Brownian motion.

The term Brownian motion refers to a process $\{Z_t,\ t \ge 0\}$ that satisfies criteria (ii) and (iv) above, but with the distribution in criteria (iii) being replaced with $N\left(\mu \times (t-s), \sigma^2 \times (t-s)\right)$.

Here μ is the drift coefficient and σ is known as the diffusion coefficient (or volatility).

Standard Brownian motion is obtained when $\mu = 0$, $\sigma = 1$ and $Z_0 = 0$.

It turns out that Brownian motion is the *only process with stationary independent increments and continuous sample paths*. This is far from obvious and we won't prove it here.

The relationship between standard Brownian motion and Brownian motion is the same as the relationship between a standard normal distribution, $N(0,1)$, and a general $N(\mu,\sigma^2)$ distribution. A Brownian motion with given diffusion and drift coefficients can be constructed out of a standard Brownian motion $\{W_t, t \geq 0\}$ by setting:

$$Z_t = Z_0 + \sigma W_t + \mu t$$

Question

Let W_t be a standard Brownian motion. Prove that $Z_t = Z_0 + \sigma W_t + \mu t$ is a Brownian motion with diffusion coefficient σ and drift μ.

Solution

The second property of a Brownian motion – that it has continuous sample paths – is met because Z_t is driven by only time *t* and the continuous process W_t.

The increments of Z_t are independent of the past because:

$$Z_t - Z_s = \sigma(W_t - W_s) + \mu(t-s)$$

and we know that the increments $W_t - W_s$ have this property.

The third property we require is that:

$$Z_t - Z_s \sim N\left(\mu(t-s), \sigma^2(t-s)\right)$$

This follows because $W_t - W_s \sim N(0,(t-s))$ and therefore:

$$\begin{aligned} Z_t - Z_s &= \sigma(W_t - W_s) + \mu(t-s) \\ &\sim \sigma N(0, t-s) + \mu(t-s) \\ &\sim N\left(0, \sigma^2(t-s)\right) + \mu(t-s) \\ &\sim N\left(\mu(t-s), \sigma^2(t-s)\right) \end{aligned}$$

Question

How can a Brownian motion, Z_t, that has drift μ and diffusion parameter σ and a starting value of Z_0 be converted into a standard Brownian motion?

Solution

Just invert the relationship given:

$$W_t = \frac{Z_t - Z_0 - \mu t}{\sigma}$$

This is analogous to converting an observed value x from a general normal distribution, $N(\mu, \sigma^2)$, into a value from the standard normal distribution, $N(0,1)$, by calculating the standardised value:

$$z = \frac{x - \mu}{\sigma}$$

1.4 Properties of Brownian motion

Standard Brownian motion has a number of other properties inherited from the simple symmetric random walk. A simple symmetric random walk is a discrete-time stochastic process:

$$X_n = \sum_{i=1}^{n} Z_i \quad \text{where } Z_i = \begin{cases} +1 \text{ with probability } \frac{1}{2} \\ -1 \text{ with probability } \frac{1}{2} \end{cases}$$

The value of the process increases or decreases randomly by 1 unit (= 'simple') with equal probability (= 'symmetric').

If we reduce the step size progressively from 1 unit until it is infinitesimal (and rescale the X values accordingly), the simple symmetric random walk becomes standard Brownian motion. An important consequence of this is that a standard Brownian motion returns infinitely often to zero, or indeed any other level.

Many of the properties of standard Brownian motion can be demonstrated using the following decomposition. For $s < t$:

$$W_t = W_s + (W_t - W_s)$$

a decomposition in which the first term is known at time s and the second is independent of everything up to and including time s.

In calculations involving Brownian motion, we often need to split up W_t in this way, so that we can work with *independent* increments.

Covariance of a Wiener Process

An important characteristic of a process is the covariance between its value at $s \geq 0$ and $t > s$:

$$\begin{aligned} Cov(W_s, W_t) &= E[(W_s - E[W_s])(W_t - E[W_t])] \\ &= E[W_s(W_s + (W_t - W_s)] \end{aligned}$$

This follows from the fact that $E[W_t]=E[W_s]=0$, and then by applying the decomposition $W_t = W_s + (W_t - W_s)$.

By independence of increments:

$$\begin{aligned} Cov(W_s, W_t) &= E[W_s^2] + E[W_s]E[(W_t - W_s)] \\ &= Var(W_s) + 0 \\ &= s \end{aligned}$$

This follows from the fact that $E[W_s^2] = Var(W_s) + E^2[W_s] = s + 0$.

In general, $Cov(W_s, W_t) = \min\{s,t\}$.

The importance of this result is that, in fact, if a stochastic process has the property that:

$$Cov(X_s, X_t) = \min\{s,t\}$$

then the process X_t is a Wiener process (this is Lévy's Theorem).

Scaled Wiener process

Given a positive constant c and a Wiener process W_t define the stochastic process X_t by:

$$X_t = \sqrt{c}W_{t/c}$$

The 'clock' of the process X_t has been scaled by a factor c. For example, the process has been slowed down and magnified if $c > 1$ (and speeded up and shrunk if $c < 1$).

By applying Lévy's Theorem:

$$\begin{aligned} Cov(X_{t+u}, X_t) &= Cov\left(\sqrt{c}W_{\frac{t+u}{c}}, \sqrt{c}W_{\frac{t}{c}}\right) \\ &= c \times Cov\left(W_{\frac{t+u}{c}}, W_{\frac{t}{c}}\right) \\ &= c \times \min\left\{\frac{t+u}{c}, \frac{t}{c}\right\} \\ &= c\frac{t}{c} \\ &= t \end{aligned}$$

assuming $u > 0$.

Since $Cov(X_{t+u}, X_t) = \min\{t+u, t\}$, X_t is a Wiener process.

The scaled process could also be re-parameterised as:

$$X_t = \frac{1}{\sqrt{a}} W_{at}$$

with $a = \frac{1}{c}$.

Time-inverted Wiener process

Given a Wiener process W_t define the stochastic process X_t by:

$$X_t = tW_{1/t}$$

The time-inverted Wiener process is itself a Wiener process, as can be shown by Lévy's Theorem.

Let $u > 0$.

Then we have:

$$\begin{aligned} Cov(X_{t+u}, X_t) &= Cov\left((t+u)W_{\frac{1}{t+u}}, tW_{\frac{1}{t}}\right) \\ &= (t+u)t \times Cov\left(W_{\frac{1}{t+u}}, W_{\frac{1}{t}}\right) \end{aligned}$$

Since $1/(t+u) < 1/t$ and by the covariance of Wiener processes:

$$\begin{aligned} Cov(X_{t+u}, X_t) &= (t+u)t \times \min\left\{\frac{1}{t+u}, \frac{1}{t}\right\} \\ &= (t+u)t\frac{1}{t+u} \\ &= t \end{aligned}$$

Therefore X_t is also a Wiener process.

The time-inverted Wiener process is useful in proving limiting properties. For example, since $tW_{1/t}$ is a Wiener process, then:

$$\lim_{t\to\infty} \frac{W_t}{t} = \lim_{t\to\infty} W_{1/t} = W_0 = 0$$

Correlated Wiener processes

The process defined by:

$$Z_t = \rho W_t^a + \sqrt{1-\rho^2} W_t^b$$

where W_t^a and W_t^b are independent Wiener processes and $-1 \le \rho \le 1$ defines the correlation between Z_t and W_t^a.

It should be obvious that $E[Z_t]=0$.

Question

Show that $E[Z_t]=0$.

Solution

The process Z_t is only a weighted sum of two Wiener processes, both of which have zero expectation. Therefore we have:

$$E[Z_t]=E\left[\rho W_t^a+\sqrt{1-\rho^2}W_t^b\right]$$

$$=E\left[\rho W_t^a\right]+E\left[\sqrt{1-\rho^2}W_t^b\right]$$

$$=\rho E\left[W_t^a\right]+\sqrt{1-\rho^2}E\left[W_t^b\right]$$

$$=\rho\times 0+\sqrt{1-\rho^2}\times 0=0$$

The variance of the process is then given by:

$$Var(Z_t)=Var\left(\rho W_t^a+\sqrt{1-\rho^2}W_t^b\right)$$

Since ρW_t^a and $\sqrt{1-\rho^2}W_t^b$ are independent:

$$Var(Z_t)=\rho^2 Var\left(W_t^a\right)+\left(1-\rho^2\right)Var\left(W_t^b\right)$$

$$=\rho^2\times t+\left(1-\rho^2\right)\times t$$

$$=t$$

Similarly, the variance of the increment $Z_{t+u}-Z_t$ is u.

Question

Show that $Var(Z_{t+u}-Z_t)=u$ for $u>0$.

Solution

$$\begin{aligned}
Var(Z_{t+u}-Z_t) &= Var\left(\rho W^a_{t+u}+\sqrt{1-\rho^2}W^b_{t+u}-\rho W^a_t-\sqrt{1-\rho^2}W^b_t\right)\\
&= Var\left(\rho\left(W^a_{t+u}-W^a_t\right)+\sqrt{1-\rho^2}\left(W^b_{t+u}-W^b_t\right)\right)\\
&= Var\left(\rho\left(W^a_{t+u}-W^a_t\right)\right)+Var\left(\sqrt{1-\rho^2}\left(W^b_{t+u}-W^b_t\right)\right)\\
&= \rho^2\times(t+u-t)+(1-\rho^2)\times(t+u-t)\\
&= u
\end{aligned}$$

Furthermore, the covariance of Z_t and W^a_t is given by:

$$\boldsymbol{Cov(Z_t,W^a_t)=\rho t}$$

The correlation between Z_t and W^a_t is defined as:

$$\frac{Cov(Z_t,W^a_t)}{\sqrt{Var(Z_t)Var(W^a_t)}}=\rho$$

Non-differentiability of sample paths

There are various ways of proving the fact that the Wiener process is non-differentiable. Consider the following proof by contradiction. If the derivative $\frac{dW_t}{dt}$ existed, then we can say that:

$$\lim_{t-s\to 0}\left|\frac{W_t-W_s}{t-s}-\frac{dW_t}{dt}\right|<\varepsilon \quad\text{or}\quad \lim_{s-t\to 0}\left|\frac{W_s-W_t}{s-t}-\frac{dW_t}{dt}\right|<\varepsilon$$

The first inequality assumes that $t>s$, and the second covers the case when $s>t$. These statements come from the definition of a derivative as the convergence of a function's gradient:

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$$

If such a derivative existed, then we could find an arbitrarily small ε to measure the difference between the derivative and the gradient.

However, (in the case $t>s$):

$$\frac{W_t-W_s}{t-s}-\frac{dW_t}{dt}\sim N\left(-\frac{dW_t}{dt},\frac{1}{t-s}\right)$$

which will have a positive probability of being greater than ε, and so the statement is uncertain. In fact, since the variance increases as $t-s\to 0$, it never holds, almost surely.

That is to say, it's impossible to bound a normal distribution.

The fundamental theorem of calculus is that given a derivative, $f'(x)$, then the integral $f(x)$, is understood as:

$$f(b) = f(a) + \int_a^b f'(x)dx = f(a) + \int_a^b \frac{df(x)}{dx}dx = f(a) + \int_a^b df(x)$$

However, since the idea of $\frac{dW_t}{dt}$ is meaningless, the stochastic integral,

$$\int_a^b dW_t$$

cannot be handled using classical calculus and there appears to be no way of understanding how a Wiener process behaves between two times.

In other words, because a Wiener process isn't differentiable anywhere then it's not clear how integrals should be handled when the variable of integration is W_t. Stochastic integrals will be dealt with in the next chapter.

1.5 Geometric Brownian motion

As mentioned at the start of this chapter, Brownian motion was used by Bachelier to model the movements of the Paris stock exchange index.

However successful the Brownian motion model may be for describing the movement of market indices in the short run, it is useless in the long run, if only for the reason that a standard Brownian motion is certain to become negative eventually. It could also be pointed out that the Brownian motion model predicts that daily movements of size 100 or more would occur just as frequently when the process is at level 100 as when it is at level 10,000.

A more useful model is:

$$S_t = e^{Z_t}$$

where Z_t is the Brownian motion process $Z_t = Z_0 + \sigma W_t + \mu t$. Thus S_t, which is called geometric Brownian motion, is lognormally distributed with parameters $Z_0 + \mu t$ and $\sigma^2 t$. So the values of $\log S_t$ are normally distributed with mean $Z_0 + \mu t$ and variance $\sigma^2 t$.

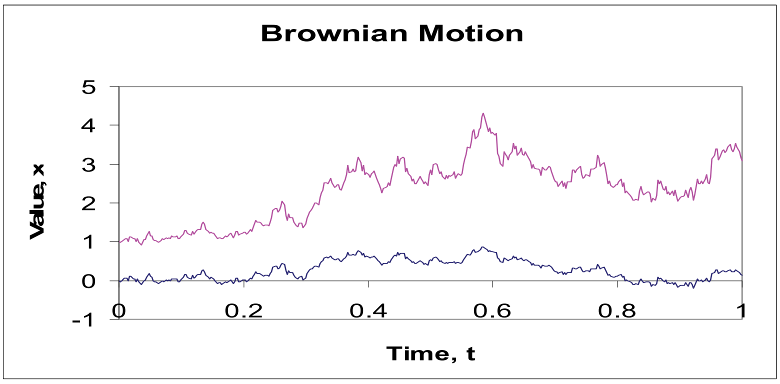


Figure 9.2: A Brownian motion, W_t, and a geometric Brownian motion, $S_t = e^{Z_t}$

The most important property of S_t is:

$$S_t \geq 0 \text{ for all } t$$

From the properties of the lognormal distribution we also have:

$$E[S_t] = \exp\left((Z_0 + \mu t) + ½\sigma^2 t\right) \text{ and } Var(S_t) = E^2[S_t]\left(\exp(\sigma^2 t) - 1\right)$$

Geometric Brownian motion features heavily in this course. For example, Black and Scholes' Nobel prize-winning formula for pricing European options assumes that the price of the underlying asset is a geometric Brownian motion.

The properties of S_t are less helpful than those of Brownian motion. For example, S_t has neither independent increments nor stationary increments.

The increments of S_t are of the form $S_t - S_s = e^{Z_t} - e^{Z_s}$.

But this is not so important because Z_t does possess these desirable properties. Analysis of path properties of S_t should involve first taking the logarithm of the observations, and then performing the analysis using techniques appropriate to Brownian motion.

The log-return $\log\frac{S_t}{S_s}$ from time s to time t is given by $\log\frac{S_t}{S_s} = \log\frac{e^{Z_t}}{e^{Z_s}} = Z_t - Z_s$.

It follows by the independent increments property of Brownian motion that the log-returns, and hence the returns themselves, are independent over disjoint time periods.

2 Martingales

2.1 Introduction

In simple terms, a martingale is a stochastic process for which its current value is the best estimate of its future value. So, the expected future value is the current value. Other ways of thinking of a martingale are that the expected change in the process is zero or that the process has 'no drift'.

Note

Throughout this course we will be using the word 'expected' in its statistical sense, rather than in the everyday sense.

Consider a person standing on a 'never-ending' ladder. Every minute they move up or down the ladder one step, depending on whether a tossed coin comes up heads or tails.

In the everyday sense of the word, after the next toss of the coin, we 'expect' them to move up or down (but we don't know which way). However, in the statistical sense, because $\frac{1}{2}\times(+1)+\frac{1}{2}\times(-1)=0$, we 'expect' them to stay exactly where they are, even though there's no way that that can happen!

The idea of martingales is consistent with the original equestrian term 'martingale', meaning a holster used to keep a horse 'pointing straight ahead'.

Their importance for modern financial theory cannot be overstated. In fact, the whole theory of pricing and hedging of financial derivatives is formulated in terms of martingales.

For this reason, it may be best to think of a martingale as being a random process that has 'no drift' because the idea of drift is more consistent with the way we think about real financial assets. We have already seen that it is possible to model the log of a share price, $\log S_t$, using Brownian motion with a drift μ. You can think of μ as being the rate of the long-term drift of the log of the share price. It is the underlying non-random trend. It should not come as a great surprise that when we remove this underlying non-random trend (or drift) and look at $\log S_t - \mu t$, we obtain a martingale.

Conditional expectation

The features of martingales rely on the application of conditional expectations.

The filtration F_t represents everything that can be known up to and including time *t*.

Some random variables will be known by time t. We say that X_t is F_t*-measurable* if the value of the process is known at time *t*, *ie* it belongs to F_t.

A stochastic process X_t, $t \geq 0$ is said to be *adapted* to the filtration F_t if X_t is F_t-measurable for all t.

If F_t is the filtration generated by X_t (as opposed to any other process), then it is known as the *natural filtration* of X_t and is denoted here by F_t^X. The following results will be used extensively.

(i) $E\left[E\left[X \mid F_t^X\right]\right] = E[X]$

(ii) if X is F_t-measurable, then $E\left[X \mid F_t\right] = X$

(iii) if X is independent of F_t, then $E[X \mid F_t] = E[X]$

(iv) any function of X_t is adapted to F_t^X.

2.2 Wiener processes are martingales

We have the following definitions of continuous-time martingales.

Given a filtered probability space (Ω, F, F_t, P), a stochastic process X_t is called a martingale with respect to the filtration, F_t, if:

- **X_t is adapted to F_t**
- **$E\left[|X_t|\right] < \infty$ for all t**
- **$E\left[X_t \mid F_s\right] = X_s$ for all $s \le t$**

The first condition is just a technicality to ensure that the process value can be known with certainty at time *t*, and the second is to guarantee that X_t is integrable. In most questions we are only concerned with the last condition and we'll assume the first two hold.

Question

In words, what does the $E[X_t \mid F_s]$ in the last condition mean?

Solution

$E[X_t \mid F_s]$ means the expected value of the process at time *t*, given that we are at time *s* and we know the history of the process up to and including time *s*.

Of all the properties of martingales, the most useful is also the simplest: a martingale has constant mean, *ie* $E[X_n] = E[X_0] = X_0$ for all n.

Given $s \le t$, a *supermartingale* is such that:

$$E\left[X_t \mid F_s\right] \le X_s$$

while a *submartingale* is such that:

$$E\left[X_t \mid F_s\right] \ge X_s$$

A supermartingale has either negative or zero drift, whereas a submartingale has either positive or zero drift. So a process which is both a supermartingale *and* a submartingale must therefore be a martingale.

Consider:

$$E\left[W_t \mid F_s^W\right] = E\left[W_s + (W_t - W_s) \mid F_s^W\right]$$
$$= E\left[W_s \mid F_s^W\right] + E\left[(W_t - W_s) \mid F_s^W\right]$$

Since increments are independent and $W_t - W_s \sim N(0, t-s)$:

$$E\left[W_t \mid F_s^W\right] = W_s$$

The Wiener process is a martingale with respect to its natural filtration, noting that $E\left[|W_t|\right] < \infty$ since $W_t < \infty$ almost surely.

Question

An asset's value at time t (in pence and measured in years) is denoted by A_t and fluctuates in value from day to day. Within these random fluctuations, there appears to be an underlying long-term trend, in that the asset's value is increasing by 2 pence on average each week.

(i) Assuming that there are exactly 52 weeks in a year, suggest a process based on A_t that you think might be a martingale.

(ii) Suppose that the price increments have a continuous uniform distribution such that $A_t - A_s \sim U[-52(t-s), 260(t-s)]$. Construct a martingale out of A_t.

Solution

(i) The value of the asset is increasing on average by 2 pence a week. Assuming that there are exactly 52 weeks in a year, this means the asset is 'drifting' by 104 pence a year. A martingale is a process without drift and so a good suggestion would be to remove this drift and consider the process:

$$A_t - 104t$$

(ii) Using the formula for the expected value of a uniform distribution from page 13 of the *Tables*, we have:

$$E\left[A_t \middle| F_s\right] = E\left[A_s + (A_t - A_s) \middle| F_s\right] = A_s + \frac{260(t-s) - 52(t-s)}{2} = A_s + 104(t-s)$$

So for every increase of $t-s$ in the time, the process is 'drifting' by $104(t-s)$.

One way to construct a martingale is to subtract the drift. Mathematically, we can subtract $104t$ from both sides of the last equation to get:

$$E\left[A_t - 104t \middle| F_s\right] = A_s - 104s$$

This now fits in with the definition given for a martingale in continuous time.

So the process $A_t - 104t$ is again a martingale.

2.3 Library of martingales

Now consider the stochastic process defined by $W_t^2 - t$.

$$E\left[W_t^2 - t \mid F_s^W\right] = E\left[(W_s + (W_t - W_s))^2 \mid F_s^W\right] - t$$

$$= E\left[W_s^2 \mid F_s^W\right] + 2E\left[W_s(W_t - W_s) \mid F_s^W\right] + E\left[(W_t - W_s)^2 \mid F_s^W\right] - t$$

$$= W_s^2 - s$$

so it is a martingale with respect to its natural filtration.

Note how the three expectations above were evaluated:

- $E\left[W_s^2 \mid F_s^W\right] = W_s^2$ since the value of W_s is known with certainty at time s
- $E\left[W_s(W_t - W_s) \mid F_s^W\right] = W_s E\left[(W_t - W_s) \mid F_s^W\right] = 0$ because the increments of a Wiener process have zero mean
- $E\left[(W_t - W_s)^2 \mid F_s^W\right] = Var(W_t - W_s) + E^2[W_t - W_s] = t - s + 0$ due to the statistical properties of Wiener process increments.

Question

Let W_t be a Wiener process. Determine whether the process W_t^2 is a supermartingale or a submartingale.

Solution

Given that $W_t^2 - t$ is a martingale with zero drift, then W_t^2 must have positive drift. This means that W_t^2 is a submartingale.

Finally, consider the stochastic process defined by $\exp\left(\lambda W_t - \frac{1}{2}\lambda^2 t\right)$.

The same approach is taken here as before; decompose the Wiener process, but also decompose time t into $s+(t-s)$.

$$E\left[\exp\left(\lambda W_t - \tfrac{1}{2}\lambda^2 t\right) \mid F_s^W\right] = E\left[\exp\left(\lambda(W_s + (W_t - W_s)) - \tfrac{1}{2}\lambda^2(s+(t-s)\right) \mid F_s^W\right]$$

$$= \exp\left(\lambda W_s - \tfrac{1}{2}\lambda^2 s\right) E\left[\exp\left(\lambda(W_t - W_s) - \tfrac{1}{2}\lambda^2(t-s)\right) \mid F_s^W\right]$$

To evaluate this expectation, consider a random variable Z where $Z \sim N\left(\mu, \sigma^2\right)$, then its moment generating function is defined as: $MGF_Z(t) = E[\exp(Z \times t)] = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Therefore when $t=1$ we have: $E[\exp(Z)] = e^{\mu + \frac{1}{2}\sigma^2}$.

From the MGF of a normal:

$$E[\exp(Z)] = \exp\left(\mu + \tfrac{1}{2}\sigma^2\right) = \exp\left(E[Z] + \tfrac{1}{2}Var(Z)\right)$$

If $\exp(Y_t)$, where $Y_t = \lambda(W_t - W_s) - \frac{1}{2}\lambda^2(t-s)$ is a normally distributed random variable, we have that:

$$E\left[\exp\left(\lambda(W_t - W_s) - \tfrac{1}{2}\lambda^2(t-s)\right) \mid F_s^W\right] = E[\exp(Y_t)]$$

$$= \exp\left(E[Y_t] + \tfrac{1}{2}Var(Y_t)\right)$$

$$= \exp\left(-\tfrac{1}{2}\lambda^2(t-s) + \tfrac{1}{2}\lambda^2(t-s)\right)$$

$$= 1$$

So:

$$E\left[\exp\left(\lambda W_t - \tfrac{1}{2}\lambda^2 t\right) \mid F_s^W\right] = \exp\left(\lambda W_s - \tfrac{1}{2}\lambda^2 s\right)$$

meaning that the process $\exp\left(\lambda W_t - \frac{1}{2}\lambda^2 t\right)$ is an F_t^W-martingale.

This result is a special case of the fact that:

$$\exp\left(\lambda\int_0^T f(t)dW_t - \tfrac{1}{2}\lambda^2\int_0^T (f(t))^2 dt\right)$$

is a martingale.

Chapter 9 Summary

Wiener process (standard Brownian motion)

A *Wiener process* is a stochastic process with the defining properties:

- $W_0 = 0$
- it has independent increments
- it has stationary increments
- it has normally distributed increments, *ie* $W_t - W_s \sim N(0, t-s)$
- it has continuous sample paths.

A Wiener process is the continuous-time analogue of a random walk.

Other properties of Wiener processes include:

- $\{W_t, t \geq 0\}$ is a *Markov process*
- $\{W_t, t \geq 0\}$ is a *martingale*, *ie* $E(W_t \mid F_s) = W_s$
- $\{W_t, t \geq 0\}$ returns infinitely often to 0, or indeed to any other level
- $Cov(W_s, W_t) = \min\{s, t\}$
- $\left\{X_t = \sqrt{c}W_{t/c}, t \geq 0\right\}$ is also a Wiener process (*scaling property*)
- $\left\{X_t = tW_{1/t}, t \geq 0\right\}$ is also a Wiener process (*time inversion property*)
- the sample path is not differentiable anywhere.

Brownian motion with drift

Brownian motion with drift is related to standard Brownian motion by the equation:

$$Z_t = Z_0 + \sigma W_t + \mu t$$

where σ is the *volatility* or *diffusion coefficient* and μ is the *drift*.

Geometric Brownian motion (lognormal model)

For modelling purposes a Brownian motion may have to be transformed, for example by taking logarithms. A useful model for security prices is *geometric Brownian motion*:

$$S_t = e^{Z_t}$$

where Z_t is the Brownian process $Z_t = Z_0 + \sigma W_t + \mu t$. Thus S_t is lognormally distributed with parameters $Z_0 + \mu t$ and $\sigma^2 t$.

Martingales

A *martingale* is a stochastic process such that:

- X_t is adapted to F_t
- $E\left[\left|X_t\right|\right] < \infty$ for all t
- $E[X_t \mid F_s] = X_s$ for all $s \le t$

Martingales are processes with no drift. In fact, it can be shown that a martingale has constant mean, *ie*:

$$E[X_n] = E[X_0] \text{ for all } n$$

Martingales constructed from Wiener processes

Various martingales can be constructed from Wiener processes, for example, W_t, $W_t^2 - t$ and $e^{\lambda W_t - \frac{1}{2}\lambda^2 t}$.

Chapter 9 Practice Questions

9.1 Assume that the spot rate of interest at time t, $S(t)$, can be modelled by $S(t)=e^{-2\mu W(t)}$, where $W(t)$ is a Brownian motion with drift coefficient μ and volatility coefficient 1 such that $W(0) = 0$.

Exam style

(i) Write down an expression for $W(t)$ in terms of a standard Brownian motion, $B(t)$. [1]

(ii) Show that $\{S(t): t > 0\}$ is a continuous-time martingale. [4]

[Total 5]

9.2 'Brownian motion is the only process with stationary independent increments and continuous sample paths.'

(i) Give mathematical definitions of each of the three underlined terms.

(ii) State the distribution of the increments for a standard Brownian motion.

9.3 (i) What is meant by saying that the process $\{Y_t\}$ is a martingale with respect to another process $\{X_t\}$?

Let B_t ($t \geq 0$) be a standard Brownian motion.

(ii) Show that B_t and $B_t^2 + kt$ are both martingales with respect to B_t, for a suitably chosen value of the constant k, which you should specify.

(iii) Show that there is a value of the constant c, which you should specify, such that $(a+bB_t)^2 + ct$ is a martingale with respect to B_t, where a and b are constants.

9.4 Let B_t ($t \geq 0$) be a standard Brownian motion process starting with $B_0 = 0$.

(i) What is the probability that B_2 takes a positive value?

(ii) What is the probability that B_2 takes a value in the interval $(-1,1)$?

(iii) Show that the probability that B_1 and B_2 both take positive values is $\frac{3}{8}$.

(iv) What is the probability that B_t takes a negative value at some time between $t = 0$ and $t = 2$?

9.5 Consider the statement: 'If you want to find the variance of $X = B(s) + B(t)$, where $s < t$, for a standard Brownian motion process, you can use the fact that $B(s)$ and $B(t)$ are independent to get $Var(X) = s + t$.'

(i) Explain why the statement is not correct, and find a correct expression for $Var(X)$.

(ii) Hence show that the general formula for $Var\big(B(t_1) + B(t_2)\big)$ when $t_1, t_2 > 0$ can be expressed as $t_1 + t_2 + 2\min(t_1, t_2)$.

9.6 Let B_t ($t \geq 0$) be a standard Brownian motion process starting with $B_0 = 0$.

Exam style

(i) Show that, when $s < t$, $E(B_s B_t) = s$. [3]

(ii) Hence find a general formula for the correlation coefficient $\rho(B_{t_1}, B_{t_2})$. [2]

[Total 5]

9.7 (i) Write down a formula for $E\left(e^{aX}\right)$ where $X \sim N\left(\mu, \sigma^2\right)$ and, by differentiating, or otherwise, derive an expression for $E\left(Xe^{aX}\right)$. [2]

Exam style

(ii) Show that:

$$X_t = \left(B_t - at\right)e^{aB_t - \frac{1}{2}a^2 t}$$

is a martingale, where B_t is a standard Brownian motion, and a is an arbitrary constant. You may assume that $E\left[\left|X_t\right|\right] < \infty$. [5]

[Total 7]

Chapter 9 Solutions

9.1 (i) ***General Brownian motion***

A general Brownian motion can be defined as:

$$W(t) = W(0) + \mu t + \sigma B(t)$$

where $B(t)$ is standard Brownian motion, μ is the drift, σ is the volatility coefficient and $W(0)$ is the value of general Brownian motion at time 0.

Since $\sigma = 1$ and $W(0) = 0$:

$$W(t) = \mu t + B(t) \qquad [1]$$

(ii) ***S(t) is a martingale***

We write $S(t)$ as S_t and $B(t)$ as B_t for neatness. Let $t > s$, then:

$$\begin{aligned}
E\left[S_t | F_s\right] &= E\left[e^{-2\mu(\mu t + B_t)} | F_s\right] \\
&= E\left[e^{-2\mu^2 t - 2\mu B_t} | F_s\right] \\
&= e^{-2\mu^2 t} E\left[e^{-2\mu B_t} | F_s\right] \\
&= e^{-2\mu^2 t} E\left[e^{-2\mu(B_t - B_s + B_s)} | F_s\right] \\
&= e^{-2\mu^2 t} e^{-2\mu B_s} E\left[e^{-2\mu(B_t - B_s)} | F_s\right] \\
&= e^{-2\mu^2 t} e^{-2\mu B_s} E\left[e^{-2\mu(B_t - B_s)}\right] \qquad [2]
\end{aligned}$$

The filtration F_s can be left out because of the independent increments property. Now $B_t - B_s \sim N(0, t-s)$ and so the expectation is of the form $E[e^{aX}]$ where $X \sim N(0, t-s)$ and $a = -2\mu$. Therefore, we can use the MGF of a normal distribution calculated at point -2μ to determine the expectation.

So, using the MGF formula, from page 11 of the *Tables*, we get:

$$\begin{aligned}
E\left[S_t | F_s\right] &= e^{-2\mu^2 t} e^{-2\mu B_s} e^{0(-2\mu) + \frac{1}{2}(t-s)(-2\mu)^2} \\
&= e^{-2\mu^2 s - 2\mu B_s} \qquad [1] \\
&= S_s
\end{aligned}$$

Finally we check that $E\left[\left|S_t\right|\right]<\infty$ for all values of t.

$$E[|S_t|]=E\left[e^{-2\mu^2t-2\mu B_t}\right]$$

$$=e^{-2\mu^2t}E\left[e^{-2\mu B_t}\right]$$

$$=e^{-2\mu^2t}M_{B_t}(-2\mu)$$

$$=e^{-2\mu^2t}e^{0(-2\mu)+\frac{1}{2}t(-2\mu)^2}=1<\infty$$

where: $M_{B_t}(-2\mu)$ is the MGF of B_t at -2μ. [1]

[Total 4]

9.2 (i) ***Mathematical definitions***

'Stationary increments' means that the distribution of B_t-B_s ($t>s$) depends only on $t-s$.

'Independent increments' means that B_t-B_s is independent of the filtration F_r whenever $r\leq s<t$.

'Continuous sample paths' means that the function $t\rightarrow B_t(\omega)$ for each particular realisation ω is a continuous function of t.

(ii) ***Distribution of the increments***

For a standard Brownian motion, B_t-B_s (with $t>s$) has a $N(0,t-s)$ distribution.

9.3 (i) ***What is a martingale?***

Strictly, we should say that the process $\{Y_t\}$ is a martingale with respect to the *filtration* $\{F_t\}$ of the process $\{X_t\}$, which means that:

$E[Y_t\,|\,F_s]=Y_s$ for all $s<t$

and $E[|Y_t|]<\infty$

and Y_t is adapted to F_t

(ii) ***Show that these processes are martingales***

If we use an s subscript to denote the expected value with respect to the filtration at time s, then we can write:

$$E_s[B_t]=E_s[(B_t-B_s)+B_s]=E_s[B_t-B_s]+E_s[B_s]$$

Since $B_t-B_s\sim N(0,t-s)$ and the value of B_s is known at time s, this gives:

$$E_s[B_t]=0+B_s=B_s$$

We have shown that the expected future value of B_t is equal to its current value (at time s). We also need to show that $E\left[\left|B_t\right|\right]<\infty$.

One way to do this is to note that $\left|x\right|<1+x^2$ for all values of x.

So: $$E\left[\left|B_t\right|\right]<E\left[1+B_t^2\right]=1+\text{var}(B_t)+\left[E(B_t)\right]^2=1+t+0^2<\infty$$

Therefore, B_t is a martingale (with respect to B_t).

Similarly:

$$\begin{aligned}E_s[B_t^2]&=E_s[\{(B_t-B_s)+B_s\}^2]\\&=E_s[(B_t-B_s)^2]+2E_s[(B_t-B_s)B_s]+E_s[B_s^2]\\&=\left\{\text{var}_s[B_t-B_s]+[E_s(B_t-B_s)]^2\right\}+2B_s\times E_s[B_t-B_s]+B_s^2\\&=(t-s)+0^2+0+B_s^2\\&=t-s+B_s^2\end{aligned}$$

So: $$E_s[B_t^2-t]=B_s^2-s$$

We can show that $E\left[\left|B_t^2-t\right|\right]<\infty$ for any value of t, by first noting that:

$$\left|x^2-k\right|<x^2+k \quad (k>0) \text{ for all values of } x.$$

So: $$E\left[\left|B_t^2-t\right|\right]<E\left[B_t^2+t\right]=\text{var}(B_t)+\left[E(B_t)\right]^2+t=t+0^2+t=2t<\infty$$

Since the expected future value of B_t^2-t is equal to its current value (at time s) and the expected value of its modulus is finite, B_t^2-t is a martingale with respect to B_t, and the required constant is $k=-1$.

(iii) ***Value of c to make the process a martingale***

We know that B_t and B_t^2-t are both martingales with respect to B_t.

So, if we use an s subscript to denote the expected value with respect to the filtration at time s, then:

$$E_s[B_t]=B_s$$

and $$E_s[B_t{}^2-t]=B_s{}^2-s \quad\Rightarrow\quad E_s[B_t{}^2]=B_s{}^2+t-s$$

Using these results, we then find that:

$$\begin{aligned}E_s[(a+bB_t)^2] &= E_s[a^2+2abB_t+b^2B_t^2] \\ &= a^2+2abE_s[B_t]+b^2E_s[B_t^2] \\ &= a^2+2abB_s+b^2(B_s^2+t-s) \\ &= (a+bB_s)^2+b^2(t-s)\end{aligned}$$

So: $E_s[(a+bB_t)^2-b^2t]=(a+bB_s)^2-b^2s$

ie $(a+bB_t)^2-b^2t$ is a martingale with respect to B_t. So the required value of the constant is $c=-b^2$.

The technical condition $E\left[\left|(a+bB_t)^2+ct\right|\right]<\infty$ will hold as in (ii).

9.4 (i) ***Probability that B_2 takes a positive value***

$B_2=B_2-B_0\sim N(0,2)$. Therefore, B_2 is equally likely to be positive or negative (and has zero probability of being exactly zero).

So: $P(B_2>0)=\frac{1}{2}$

(ii) ***Probability that B_2 takes a value in the interval $(-1,1)$***

$B_2=B_2-B_0\sim N(0,2)$.

Standardising:

$$P(-1<B_2<1)=\Phi\left(\frac{1}{\sqrt{2}}\right)-\Phi\left(-\frac{1}{\sqrt{2}}\right)=0.760-(1-0.760)=0.520$$

(iii) ***Probability that B_1 and B_2 both take positive values***

We can write the required probability as:

$$p=P(B_1>0,B_2>0)=P(B_1-B_0>0,B_2-B_1>-B_1)$$

If we now write $X=B_1-B_0$ and $Y=B_2-B_1$, then we know from the properties of Brownian motion that X and Y are independent, each with a $N(0,1)$ distribution.

So the required probability is:

$$p=P(X>0,Y>-X)$$

Since the range of values of Y depends on the value of X, we must use a double integral to evaluate this:

$$p = \int_{x=0}^{\infty} \int_{y=-x}^{\infty} \phi(x)\phi(y)\,dy\,dx$$

where the joint density function is expressed as the product of the individual density functions by independence.

So:

$$p = \int_{x=0}^{\infty} \phi(x)\left\{\int_{y=-x}^{\infty} \phi(y)\,dy\right\}dx$$

$$= \int_{x=0}^{\infty} \phi(x)\left\{\left[\Phi(y)\right]_{-x}^{\infty}\right\}dx$$

$$= \int_{x=0}^{\infty} \phi(x)\left\{1-\Phi(-x)\right\}dx$$

$$= \int_{x=0}^{\infty} \phi(x)\Phi(x)\,dx$$

Finally:

$$p = \int_0^{\infty} \phi(x)\Phi(x)\,dx = \left[\tfrac{1}{2}\{\Phi(x)\}^2\right]_0^{\infty} = \tfrac{1}{2}\left[1^2 - \left(\tfrac{1}{2}\right)^2\right] = \tfrac{3}{8}$$

(iv) ***Probability that B_t takes a negative value at some time between 0 and 2***

The probability is 1 because B_t will almost surely take a negative value at some point close to $t = 0$.

9.5 (i) ***Explain why the statement is not correct***

The statement is not correct because $B(s)$ and $B(t)$, which represent the value of the process at two different times, are not independent. In fact, they are positively correlated.

It is actually the increments $B(s)-B(0)$ and $B(t)-B(s)$ that are independent. The correct calculation can be done by expressing $X=B(s)+B(t)$ in terms of these increments:

$$Var(X)=Var\big(B(s)+B(t)\big)$$

$$=Var\big(2B(0)+2\{B(s)-B(0)\}+\{B(t)-B(s)\}\big)$$

$$=4Var\big(B(0)\big)+4Var\big(B(s)-B(0)\big)+Var\big(B(t)-B(s)\big)$$

$$=0+4s+(t-s)=3s+t$$

(ii) ***Show the general formula***

This $3s+t$ formula works if $s<t$. If $t<s$, we can swap the letters to get $3t+s$. If $s=t$, we have:

$$Var(X)=Var\big(B(s)+B(t)\big)=Var\big(2B(s)\big)=4Var\big(B(s)\big)=4s \text{ (or } 4t)$$

which agrees with either formula.

So a general formula would be $s+t+2\min(s,t)$, or if we're using t_1 and t_2 to denote the times, $t_1+t_2+2\min(t_1,t_2)$.

9.6 (i) ***Show that $E(B_sB_t)=s$***

If we express B_t in terms of the increment B_t-B_s, which is independent of the value of B_s, we get:

$$E[B_sB_t]=E[B_s\{(B_t-B_s)+B_s\}]$$

$$=E[B_s(B_t-B_s)]+E[B_s^2]$$

$$=E(B_s)E(B_t-B_s)+Var[B_s]+[E(B_s)]^2=0+s+0=s$$ [3]

(ii) ***Find a general formula for the correlation coefficient***

We can then calculate the covariance and correlation between these two values:

$$Cov(B_s,B_t)=E\big[B_sB_t\big]-E\big[B_s\big]E\big[B_t\big]=s$$

and $$\rho(B_s,B_t)=\frac{Cov(B_s,B_t)}{\sqrt{Var(B_s)Var(B_t)}}=\frac{s}{\sqrt{st}}=\sqrt{\frac{s}{t}}$$ [1]

This formula only applies when $s<t$. We can generalise this to cover any positive times t_1 and t_2, if we write it in the form:

$$\rho(B_{t_1},B_{t_2})=\sqrt{\frac{\min(t_1,t_2)}{\max(t_1,t_2)}}$$ [1]

[Total 2]

There are various alternative ways of writing this, eg $\rho(B_{t_1},B_{t_2})=\min\left(\sqrt{\frac{t_1}{t_2}},\sqrt{\frac{t_2}{t_1}}\right)$.

9.7 (i) ***Formulae for expectations***

$$E\left[e^{aX}\right]=e^{\mu a+\frac{1}{2}\sigma^2a^2}$$ [½]

This is the MGF of a normal random variable and can be found in the Tables.

We therefore have:

$$E\left[Xe^{aX}\right]=\frac{d}{da}E\left[e^{aX}\right]=\frac{d}{da}e^{\mu a+\frac{1}{2}\sigma^2a^2}=\left(\mu+\sigma^2a\right)e^{\mu a+\frac{1}{2}\sigma^2a^2}.$$ [1½]

[Total 2]

(ii) ***Show that*** X_t ***is a martingale***

$$E\left[X_t|F_s\right]=E\left[\left(B_t-at\right)e^{aB_t-0.5a^2t}|F_s\right]$$

$$=E\left[\left(B_s+B_t-B_s-at\right)e^{a\left(B_s+B_t-B_s\right)-0.5a^2t}|F_s\right]$$

$$=E\left[\left(B_s-at\right)e^{aB_s-0.5a^2t}e^{a\left(B_t-B_s\right)}|F_s\right]+E\left[\left(B_t-B_s\right)e^{aB_s-0.5a^2t}e^{a\left(B_t-B_s\right)}|F_s\right]$$ [2]

Now we can use the fact that we are conditioning on all the information known at time s. Any terms involving B_s can be taken outside the expectation, as can terms in s and t (which are fixed, not random points in time). This gives:

$$=\left(B_s-at\right)e^{aB_s-0.5a^2t}E\left[e^{a\left(B_t-B_s\right)}|F_s\right]+e^{aB_s-0.5a^2t}E\left[\left(B_t-B_s\right)e^{a\left(B_t-B_s\right)}|F_s\right]$$ [1]

We can then drop the conditions since the increments are independent of the past:

$$=\left(B_s-at\right)e^{aB_s-0.5a^2t}E\left[e^{a\left(B_t-B_s\right)}\right]+e^{aB_s-0.5a^2t}E\left[\left(B_t-B_s\right)e^{a\left(B_t-B_s\right)}\right]$$ [1]

Using part (i), and noting that $B_t-B_s\sim N(0,t-s)$ so that $\mu=0$ and $\sigma^2=t-s$, we therefore get:

$$=\left(B_s-at\right)e^{aB_s-0.5a^2t}e^{0.5a^2(t-s)}+e^{aB_s-0.5a^2t}a\left(t-s\right)e^{0.5a^2(t-s)}$$

$$=\left(B_s-as\right)e^{aB_s-0.5a^2s}$$

$$=X_s$$ [1]

as required. Note that the bounded condition is given in the question.

[Total 5]

10

Stochastic calculus and Ito processes

Syllabus objectives

4.4 Stochastic models for security prices.

4.4.3 Demonstrate a basic understanding of stochastic differential equations, the Ito integral, diffusion and mean-reverting processes.

4.4.4 State Ito's Lemma and be able to apply it to simple problems.

4.4.5 Write down the stochastic differential equation for geometric Brownian motion and show how to find its solution.

4.4.6 Write down the stochastic differential equation for the Ornstein-Uhlenbeck process and show how to find its solution.

0 Introduction

This chapter is concerned with stochastic calculus, in which continuous-time stochastic processes are described using stochastic differential equations. As is the case in a non-stochastic setting (*eg* in mechanics), these equations can sometimes be solved to give formulae for the functions involved. The solutions to these equations often involve *Ito integrals*, which we will look at in some detail. The other key result from stochastic calculus is Ito's Lemma. This is used to determine the stochastic differential equation for a stochastic process whose values are a function of another stochastic process.

Diffusions are a generalisation of Brownian motion in which the constraint that the increments are independent is dropped. However, a slightly weaker condition, known as the 'Markov property', is retained. Such processes can be thought of as Brownian motion where the drift and diffusion coefficients are variable.

The main example given is the *Ornstein-Uhlenbeck process*. It is mean-reverting, that is, when the process moves away from its long-run average value, there is a component that tends to pull it back towards the mean. For this reason the process can be used to model interest rates, which are usually considered to be mean-reverting.

As another example we will discuss *geometric Brownian motion*. This can be used to model share prices. Here the log of the share price is assumed to follow Brownian motion, so the model is sometimes known as the *lognormal model*.

The Core Reading in this chapter is adapted from course notes written by Timothy Johnson.

1 Stochastic calculus

1.1 Introduction

Newton originally developed calculus to provide the necessary mathematics to handle his laws describing the motion of bodies. His second law, for example, can be written as $F = ma$, where F is the force applied to a body, m its mass, and a is the resulting acceleration. The acceleration is the time derivative of velocity, which is in turn the time derivative of position. We therefore arrive at the differential equation:

$$F = m\ddot{x} = m\frac{d^2x}{dt^2}$$

This is the model of the motion of the body.

This is all very well so long as the path followed by a body or particle is sufficiently smooth to differentiate, as is generally the case in Newtonian mechanics. However, there are situations when the paths followed are not sufficiently smooth. For example, you may have studied impulses *eg* when two snooker balls collide. In these cases the velocity of the objects involved can change suddenly and $\frac{dx}{dt}$ is not a differentiable function.

As mentioned at the start of the previous chapter, the original Brownian motion referred to the movement of pollen grains suspended in a liquid. Each pollen grain is very light, and therefore jumps around as it is bombarded by the millions of molecules that make up the liquid, giving the appearance of a very random motion. A stochastic model of this behaviour is therefore appropriate. A deterministic model in terms of the underlying collisions wouldn't be very practical.

The sample paths of this motion are not sufficiently smooth however. As we have seen, they are differentiable nowhere. Therefore, a description of the motion as a differential equation in the usual sense is doomed to failure. In order to get around this, a new *stochastic calculus* has to be developed. This turns out to be possible and allows the formulation of stochastic differential equations (SDEs).

The stochastic differential equations that we deal with will be continuous-time versions of the equations used to define time series, *ie* stochastic processes operating in discrete time. For example, you may recall that a zero-mean random walk X_t can be defined by an equation of the form:

$$X_t = X_{t-1} + \sigma N_t \quad \text{or} \quad X_t - X_{t-1} = \sigma N_t$$

where N_t is a standard normal random variable. The N_t's in this equation are called *white noise*.

This is a stochastic difference equation: a 'difference' equation, since it involves the difference $X_t - X_{t-1}$, and 'stochastic' because the white noise terms are random. It can be 'solved' to give:

$$X_t = X_0 + \sigma \sum_{s=1}^{t} N_s$$

In continuous time, the analogue of a zero-mean random walk is a zero-mean Brownian motion, say Z_t. The change in this Brownian motion over a very short time period (in fact, an infinitesimal time period) will be denoted by $dZ_t = Z_{t+dt} - Z_t$. Since Brownian motion increments are independent, we can think of dZ_t as a continuous-time white noise. In fact, we have:

$$Cov(dZ_s, dZ_t) = \begin{cases} 0 & s \neq t \\ \sigma^2 dt & s = t \end{cases}$$

For a standard Brownian motion, W_t, this would be:

$$Cov(dW_s, dW_t) = \begin{cases} 0 & s \neq t \\ dt & s = t \end{cases}$$

We therefore have the stochastic differential equation:

$$dZ_s = \sigma dW_s$$

This can be solved by integrating both sides between 0 and *t* to give:

$$Z_t - Z_0 = \sigma \int_0^t dW_s$$

$$\Leftrightarrow \qquad Z_t = Z_0 + \sigma \int_0^t dW_s$$

Compare this to the discrete-time case $X_t = X_0 + \sigma \sum_{s=1}^{t} N_s$.

The analogy is that the dW_s process is considered as a continuous-time white noise process and, because we're working in continuous time, we need to integrate, rather than sum the terms. The existence, meaning and properties of such integrals are discussed in this section, together with some more interesting examples.

1.2 The Ito integral

When attempting to develop a calculus for Brownian motion and other diffusions, one has to face the fact that their sample paths are nowhere differentiable.

A direct approach to stochastic integrals like $\int_0^T Y_t dW_t$ is therefore doomed to failure.

An integral like this in which we are integrating with respect to Brownian motion is called an *Ito integral*. It is the fact that we are integrating with respect to the Brownian motion that is the problem. Integration of random variables with respect to a deterministic variable x can be dealt with in the standard way.

A quick review of some basic integration results and notation will be helpful.

An integral such as $\int_a^b dx = b - a$ should be very familiar. The integral sign can be interpreted simply as a summation. The summands are the dx expressions. These represent small changes in the value of x. Therefore, the integral just says that summing up all the small changes in x between a and b gives the total change $b-a$.

Similarly, the integral $\int_a^b df(x) = f(b) - f(a)$ just gives the total change in the function f as x varies between a and b. This notation may be less familiar, but this is what it means.

The integral $\int_a^b g(x) df(x)$ can be evaluated directly if $f(x)$ is a differentiable function since then we can use $\int_a^b g(x) df(x) = \int_a^b g(x) \frac{df}{dx} dx$. However, if we want to integrate $\int_0^T Y_t dW_t$ where Y_t is a (possibly random) function of t, and W_t is a standard Brownian motion, then we cannot apply the above method, as W_t is not differentiable.

However, such Ito integrals can be given a meaning for a suitable class of F_t-measurable random integrands Y_t. This involves a method of successive approximation by step functions.

We will illustrate this approach assuming that the integrand is not random, but just a deterministic function $f(t)$.

Ito integrals for deterministic functions

Firstly, when we integrate the constant function $f(t) \equiv 1$ we expect:

$$\int_0^T dW_t = \left[W_t\right]_0^T = W_T - W_0 = W_T$$

This is basically what integration means. We add up the (infinitesimal) increments dW_t to get the overall increment $W_T - W_0$. It is worth noting that the increments of Brownian motion are just normal random variables, and furthermore, the increments over disjoint time periods are independent. In 'adding' up the increments dW_t we are effectively summing independent normal random variables. Moreover, the increment dW_t should have a $N(0,dt)$ distribution. Again this is consistent with the value of the integral (W_t), which has a $N(0,t)$ distribution.

Also, any constant multiple of the integrand should just multiply the integral. For example:

$$\int_0^T 2dW_t = 2\left[W_t\right]_0^T = 2\left(W_T - W_0\right) = 2W_T$$

Finally, the integrals should add up in the usual way over disjoint time periods. For example, if we have the function:

$$f(t) = \begin{cases} 1 & 1 \le t < 2 \\ 2 & 2 \le t \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Then: $\displaystyle\int_1^3 f(t)dW_t = \int_1^2 1dW_t + \int_2^3 2dW_t = \left[W_t\right]_1^2 + \left[2W_t\right]_2^3 = 2W_3 - W_2 - W_1$

More generally:

$$\int_1^T f(t)dW_t = \begin{cases} \displaystyle\int_1^T 1dW_t = W_T - W_1 & 1 \le T < 2 \\ \displaystyle\int_1^2 1dW_t + \int_2^T 2dW_t = 2W_T - W_2 - W_1 & 2 \le T \le 3 \end{cases}$$

Question

What distribution does $\displaystyle\int_{1.5}^{2.5} f(t)dW_t$ have?

Solution

$$\int_{1.5}^{2.5} f(t)dW_t = \int_{1.5}^{2} 1dW_t + \int_{2}^{2.5} 2dW_t = \left[W_t\right]_{1.5}^{2} + \left[2W_t\right]_{2}^{2.5} = \left(W_2 - W_{1.5}\right) + 2\left(W_{2.5} - W_2\right)$$

Now we know that the two terms on the RHS are independent normal random variables with distributions $N(0,0.5)$ and $N(0,2)$ respectively. (Remember that constants square when taking the variance.) It follows that the original integral has distribution:

$$\int_{1.5}^{2.5} f(t)dW_t \sim N(0,2.5)$$

In summary, for these simple cases, we can think of the integration as summing independent normal variables.

It should be obvious from the above that we can integrate any function $f(t)$ that is piecewise constant by splitting the integral up into the constant pieces, and then adding up the answers (assuming this turns out to be finite).

Even without introducing random integrands, the problem of how to integrate more general functions, such as $\int_a^b f(t)dW_t$, remains.

This general integral can be thought of as the continuous-time limit of a summation. Consider discretising the interval $[a,b]$ using discrete times $\{t_0, t_1, \ldots, t_n\}$ where $t_0 = a$ and $t_n = b$. The infinitesimal increments can then be replaced by the finite increments $\Delta W_{t_i} = W_{t_i} - W_{t_{i-1}}$. Assuming n is large, $f(t)$ can be approximated by $f(t_{i-1})$ where $t_{i-1} < t \le t_i$. The integral above can then be defined as the limit of a summation:

$$\int_a^b f(t)dW_t = \lim_{n\to\infty} \sum_{i=1}^{n} f\left(t_{i-1}\right)\Delta W_{t_i}$$

What is the distribution of this integral? Again, the approximation helps. The distribution of each summand is known:

$$f\left(t_{i-1}\right)\Delta W_{t_i} \sim N\left(0, f^2\left(t_{i-1}\right)\left(t_i - t_{i-1}\right)\right)$$

And since the summands are independent we get:

$$\int_a^b f(t)dW_t \sim \lim_{n\to\infty} \sum_{i=1}^n N\left(0, f^2(t_{i-1})(t_i - t_{i-1})\right) \sim \lim_{n\to\infty} N\left(0, \sum_{i=1}^n f^2(t_{i-1})(t_i - t_{i-1})\right)$$

where we just assume this limit makes sense. In fact, as n gets large, the finite increment $t_i - t_{i-1}$ becomes the infinitesimal increment dt, and the summation becomes an integral.

Therefore:

$$\int_a^b f(t)dW_t \sim N\left(0, \int_a^b f^2(t)dt\right)$$

Once you get used to the notation, you needn't revert to the summation notation – just interpret the integral directly as a sum. For example, since $dW_t \sim N(0, dt)$, we must have $f(t)dW_t \sim N\left(0, f^2(t)dt\right)$ since we're just multiplying a normal random variable by a constant. (We are thinking of t as being fixed when we do this.) Furthermore, these random variables, for different values of t, are independent and, since independent normal random variables have an additive property, we arrive at:

$$\int_a^b f(t)dW_t \sim N\left(0, \int_a^b f^2(t)dt\right)$$

In the case where we have deterministic integrands, we have:

$$\boldsymbol{E\left[\int_a^b f(t)dW_t\right] = 0}$$

and, since:

$$\boldsymbol{E\left[f^2(t)\right] = f^2(t)}$$

we have that:

$$\boldsymbol{Var\left(\int_a^b f(t)dW_t\right) = \int_a^b f^2(t)dt}$$

One final property of the integral is that it is a martingale when considered as a process with respect to t, *ie* if we define the process $X_t = \int_a^t f(s)dW_s$. Intuitively, since the process has zero-mean increments it should continue 'straight ahead' on average.

Mathematically, if $u<t$:

$$E\left[X_t \mid F_u\right]=E\left[\int_a^u f(s)dW_s+\int_u^t f(s)dW_s \middle| F_u\right]=X_u+E\left[\int_u^t f(s)dW_s \middle| F_u\right]=X_u$$

The last equality follows because we know that $\int_u^t f(s)dW_s$ is normal with mean zero.

Ito integrals for stochastic functions

We now consider what happens when f is a function of both W_t and time t.

Kiyoshi Ito, working in isolation in Japan during the Second World War, realised that if a function of a Wiener process, $f(W_t,t)$ could be differentiated twice and was measurable with respect to the natural filtration of the Wiener process, and by using the fact that the process has independent increments, then the Ito integral:

$$\int_0^T f(W_t,t)dW_t = \lim_{n\to\infty}\sum_{i=1}^n f(W_{t_i},t_i)\left(W_{t_i}-W_{t_{i-1}}\right)$$

could be defined providing it was *square-integrable*:

$$E\left[\int_0^T f^2(W_t,t)dt\right]<\infty$$

where $t_0=0$ and $t_n=T$.

In particular:

$$E\left[\int_0^T f(W_t,t)dW_t\right]=0$$

and following from this:

$$Var\left(\int_0^T f(W_t,t)dW_t\right)=E\left[\left(\int_0^T f(W_t,t)dW_t\right)^2\right]=\int_0^T E\left[f^2(W_t,t)\right]dt$$

where the above result is called the *Ito isometry*, and:

$$E\left[\int_0^T f(W_t,t)dW_t \middle| F_S\right]=\int_0^S f(W_t,t)dW_t \quad \text{for } S<T.$$

The last statement tells us that Ito integrals are martingales.

Let g be a second function of W_t and time t.

Furthermore, the Ito integral is linear:

$$\int_0^T \left(f(W_t,t) + g(W_t,t) \right) dW_t = \int_0^T f(W_t,t)dW_t + \int_0^T g(W_t,t)dW_t$$

and the following product rule is satisfied:

$$E\left[\left(\int_0^T f(W_t,t)dW_t\right)\left(\int_0^T g(W_t,t)dW_t\right)\right] = \int_0^T E\left[f(W_t,t)g(W_t,t)\right]dt$$

2 Ito processes

An Ito process is a stochastic (random) process described in terms of a deterministic part and a random part as a *stochastic integral equation*:

$$X_T = X_0 + \int_0^T \mu(X_t,t)dt + \int_0^T \sigma(X_t,t)dW_t$$

To prevent the process 'exploding' (hitting $\pm\infty$ in finite time) some technical restrictions on the functions μ, which represents the drift, and σ, which represents the volatility, need to be imposed. A sufficient condition is:

$$|\mu(x)| + |\sigma(x)| \leq K(1+|x|)$$

for a constant *K*.

Ito processes are usually written in the shorthand:

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dW_t$$

and this form is often referred to as a *stochastic differential equation* (SDE), though Ito processes are a sub-class of stochastic processes (an SDE need not represent an Ito process).

Note that the SDE is just notational shorthand for the stochastic integral equation.

A number of SDEs are introduced here, which will then be solved later in the chapter.

Arithmetic Brownian motion

The process defined by:

$$dX_t = \mu dt + \sigma dW_t, \quad X_0 = x$$

for known constants μ and $\sigma > 0$ is an Ito process. It is sometimes known as a Brownian motion with drift or a Bachelier process.

This stochastic differential equation can be solved directly by integrating both sides between time 0 and *T*.

We have:

$$X_T = x + \int_0^T \mu dt + \int_0^T \sigma dW_t = x + \mu T + \sigma W_T$$

From this, we can deduce that:

- $E\left[X_T\right] = x + \mu T$
- $Var\left(X_T\right) = \sigma^2 Var\left(W_T\right) = \sigma^2 T$
- $X_T \sim N\left(x + \mu T, \sigma^2 T\right)$

Geometric Brownian motion (exponential process)

The SDE given by:

$$dX_t = \mu X_t dt + \sigma X_t dW_t\,,\ X_0 = x$$

for known constants μ and $\sigma > 0$, is known as geometric Brownian motion (GBM). The rate of change of X_t is proportional to X_t means that GBM never hits zero (or infinity by time inversion). It is not immediately apparent what the distributional properties of X_t will be, since the drift term, involving X_t is stochastic.

We will solve this SDE later in this chapter. Geometric Brownian motion is widely used to model asset prices.

The Ornstein-Uhlenbeck and the mean-reverting process

The SDE given by:

$$dX_t = -\gamma X_t dt + \sigma dW_t\,,\ X_0 = x$$

with $\sigma,\ \gamma > 0$ is known as the Ornstein-Uhlenbeck process.

Note that the drift term will pull this process to zero, while the volatility term is random. Adding the term $\gamma\mu$ to the Ornstein-Uhlenbeck SDE gives:

$$dX_t = \gamma(\mu - X_t)dt + \sigma dW_t$$

which is the mean-reverting process, which is drawn to the value of μ.

Mean-reverting process are ideal for modelling interest rates where one would expect the rate to fluctuate around a particular value rather than tending to grow indefinitely.

The Square root mean-reverting process

This process is also known as the Feller, or the Cox-Ingersoll-Ross (CIR) process.

The process defined by the SDE:

$$dX_t = \gamma(\mu - X_t)dt + \sigma\sqrt{X_t}\,dW_t$$

with $\sigma,\ \gamma,\ \mu > 0$ is, like the Ornstein-Uhlenbeck process, mean reverting. However, for the choice of parameters $\sigma^2 \le 2\gamma\mu$ the process is positive. If the process hits zero, its volatility disappears, and its drift is positive, the process deterministically moves away from zero and spends 'no time' at zero (*ie* the time spent at zero has measure zero). This is a very useful property in modelling asset prices.

This process is also very useful for modelling interest rates, which are generally required to remain positive. The CIR process will be dealt with later in the term structure of interest rates chapter.

2.1 Stochastic calculus

Taylor's theorem

We have shown that a stochastic calculus exists and satisfies certain properties. However, with standard calculus we have rules that allow us to integrate and differentiate, *eg* the product rule, the quotient rule, and the chain rule. The key result of stochastic calculus is Ito's Lemma.

This is the stochastic calculus version of the chain (function-of-a-function) rule and the only rule that we will need.

To be consistent with what is to follow, we will first derive the chain rule for standard calculus. Suppose we have a function-of-a-function $f(x_t)$ and we want to find $\frac{d}{dt}f(x_t)$. We first write down Taylor's theorem to second-order:

$$\delta f(x_t) = f'(x_t)\delta x_t + ½ f''(x_t)(\delta x_t)^2 + \cdots$$

You may find it helpful to refer to the formulae on page 3 of the *Tables* if you are unfamiliar with Taylor series.

Now dividing by δt and letting $\delta t \to 0$ gives:

$$\frac{df(x_t)}{dt} = f'(x_t)\frac{dx_t}{dt} + \lim_{\delta t \to 0} ½ f''(x_t)\frac{(\delta x_t)^2}{\delta t}$$

Since:

$$\lim_{\delta t \to 0}\frac{(\delta x_t)^2}{\delta t} = \lim_{\delta t \to 0}\frac{\delta x_t}{\delta t} \times \delta x_t = \frac{dx_t}{dt}\left(\lim_{\delta t \to 0}\delta x_t\right) = 0$$

the second term on the right-hand side must vanish, giving the chain rule:

$$\frac{df(x_t)}{dt} = f'(x_t)\frac{dx_t}{dt}$$

or in different notation:

$$df(x_t) = f'(x_t)dx_t$$

What does this become if we replace the function x_t by the non-differentiable Wiener process W_t? The analysis starts in much the same way. We can write Taylor's theorem to second-order as:

$$\delta f(W_t) = f'(W_t)\delta W_t + ½ f''(W_t)(\delta W_t)^2 + \cdots$$

Now, in the standard case, taking the limit $\delta t \to 0$ effectively involves replacing δ by d and ignoring second-order and higher-order terms. However, with Wiener processes, it turns out that the second-order term $(dW_t)^2$ cannot be ignored. In fact, it must be changed to dt, *ie* $(dW_t)^2 = dt$. This is not rigorous, but is a useful rule of thumb.

What we end up with is therefore:

$$df(W_t) = f'(W_t)dW_t + \tfrac{1}{2}f''(W_t)dt$$

This is Ito's Lemma for functions of Wiener processes, *ie* it tells us how to differentiate functions of standard Brownian motion. Note, however, that this statement must be interpreted in terms of integrals, since standard Brownian motion is not differentiable.

Question

Find the stochastic differential equation for W_t^2.

Solution

Applying the above formula we have:

$$d\left(W_t^2\right) = 2W_t dW_t + \tfrac{1}{2}2dt = 2W_t dW_t + dt$$

What does this actually mean? As we keep saying, this can only be interpreted sensibly in terms of integrals. If we integrate both sides of the SDE for W_t^2 in the previous question from 0 to s, say, we get:

$$\int_0^s d\left(W_t^2\right) = \int_0^s 2W_t dW_t + \int_0^s dt$$

The left-hand side and second term on the right-hand side can be evaluated:

$$\left[W_t^2\right]_0^s = W_s^2 = \int_0^s 2W_t dW_t + s$$

Finally, rearranging this equation tells us that:

$$\int_0^s W_t dW_t = \frac{1}{2}\left(W_s^2 - s\right)$$

This last example shows how Ito's Lemma can be used to evaluate Ito integrals.

The above version of Ito's Lemma only dealt with functions of Wiener processes. We will now generalise this to consider functions of X_t and t, where X_t is an Ito process for which $dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dW_t$, as defined earlier.

Ito's Lemma

At this point we can say little about the properties of Ito diffusions such as geometric Brownian motion, the Ornstein-Uhlenbeck process and the CIR process. We can begin to understand them if we can consider functions of the diffusions.

We call a function:

$$f(x,t) : \mathbb{R} \times [0,\infty) \to \mathbb{R}$$

an Ito function, $f(x,t) \in V$,

(V is the set of functions which have continuous derivatives in t, and continuous second derivatives in x)

if:

(i) $f(x,t) \in C^{2,1}$ **, *ie* it has a continuous second derivative with respect to x and is continuous in time**

(ii) $f(x,t)$ **is** $F^W \times B$ **-measurable**

Where B is a Borel measure. You may have met Borel measures if you have studied measure theory at university. However, we will not be using Borel measures further in this course.

(iii) $f(x,t)$ **is** F^W **-adapted**

Where W is a Wiener process.

(iv) It is square-integrable:

$$E\left[\int_S^T f^2(x,t)dt\right] < \infty$$

Then let the process X_t be defined by:

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dW_t$$

If the Ito process X_t was the input to an Ito function then:

$$f(X_T,T) = f(X_0,0) + \int_0^T \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)(X_t,t)dt + \int_0^T \left(\sigma\frac{\partial f}{\partial x}\right)(X_t,t)dW_t$$

This is Ito's Lemma (or Ito's formula).

The notation here means that the partial derivatives of the deterministic function f (and the functions μ and σ) are evaluated at the random point (X_t,t). On grounds of notational compactness it is also common to express Ito's Lemma in the following way:

$$f(X_T,T)=f(X_0,0)+\int_0^T\left(\frac{\partial f}{\partial t}+\mu(X_t,t)\frac{\partial f}{\partial X_t}+\frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f}{\partial X_t^2}\right)dt+\int_0^T\sigma(X_t,t)\frac{\partial f}{\partial X_t}dW_t$$

where we have used the notation $\frac{\partial f}{\partial t}$ to mean $\frac{\partial f}{\partial t}(X_t,t)$, and $\frac{\partial f}{\partial X_t}$ to mean $\frac{\partial f}{\partial x}(X_t,t)$ *etc.*

In shorthand this is sometimes written:

$$\boldsymbol{df(X_t,t)=\left(\frac{\partial f(x,t)}{\partial t}+\mu(X_t,t)\frac{\partial f(x,t)}{\partial x}+\frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f(x,t)}{\partial x^2}\right)dt+\sigma(X_t,t)\frac{\partial f(x,t)}{\partial x}dW_t}$$

This can be condensed to:

$$df(X_t,t)=\left(\frac{\partial f}{\partial t}+\mu(X_t,t)\frac{\partial f}{\partial X_t}+\frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f}{\partial X_t^2}\right)dt+\sigma(X_t,t)\frac{\partial f}{\partial X_t}dW_t$$

One interpretation of Ito's Lemma is that any (sufficiently well-behaved) function of an Ito process, is another Ito process, with drift and diffusion given by:

$$\frac{\partial f}{\partial t}+\frac{\partial f}{\partial X_t}\mu(X_t,t)+\frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f}{\partial X_t^2} \quad \text{and} \quad \frac{\partial f}{\partial X_t}\sigma(X_t,t)$$

Combining Ito's formula with the Taylor expansion of $f(x,t)$ we can deduce the following 'rules':

$$\boldsymbol{(dW_t)^2=dt\,,\quad dW_t dt=dtdW_t=(dt)^2=0}$$

These results can be summarised by the following multiplication table:

	dt	dW_t
dt	0	0
dW_t	0	dt

Using these rules, we can write Ito's Lemma in a third way:

$$\boldsymbol{df(X_t,t)=\frac{\partial f(x,t)}{\partial t}dt+\frac{\partial f(x,t)}{\partial x}dX_t+\frac{1}{2}\frac{\partial^2 f(x,t)}{\partial x^2}(dX_t)^2}$$

This is actually just an application of Taylor's theorem in two variables.

And we can derive the stochastic integration by parts formula. Given:

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dW_t \text{ and } dY_t = \eta(Y_t,t)dt + \nu(Y_t,t)dW_t$$

Notice that both processes are driven by the same Wiener process W_t.

Then:

$$d(X_tY_t) = X_tdY_t + Y_tdX_t + dX_tdY_t$$

This is another application of Taylor's theorem in two variables:

$$\begin{aligned}
d(X_tY_t) &= \frac{\partial(X_tY_t)}{\partial X_t}dX_t + \frac{1}{2}\frac{\partial^2(X_tY_t)}{\partial X_t^2}(dX_t)^2 + \dots \\
&\quad + \frac{\partial(X_tY_t)}{\partial Y_t}dY_t + \frac{1}{2}\frac{\partial^2(X_tY_t)}{\partial Y_t^2}(dY_t)^2 + \dots \\
&\quad + \frac{\partial(X_tY_t)}{\partial X_t\partial Y_t}dX_tdY_t + \dots \\
&= Y_tdX_t + X_tdY_t + dX_tdY_t
\end{aligned}$$

The dX_tdY_t term requires the use of the multiplication table to give:

$$\begin{aligned}
dX_tdY_t &= \big(\mu(X_t,t)dt + \sigma(X_t,t)dW_t\big)\big(\eta(Y_t,t)dt + \nu(Y_t,t)dW_t\big) \\
&= \mu(X_t,t)\eta(Y_t,t)\underbrace{(dt)^2}_{=0} + (\mu(X_t,t)\nu(Y_t,t) + \sigma(X_t,t)\eta(Y_t,t))\underbrace{dtdW_t}_{=0} \\
&\qquad + \sigma(X_t,t)\nu(Y_t,t)\underbrace{(dW_t)^2}_{=dt} \\
&= \sigma(X_t,t)\nu(Y_t,t)dt
\end{aligned}$$

Therefore we have:

$$\begin{aligned}
d(X_tY_t) &= X_t\big(\eta(Y_t,t)dt + \nu(Y_t,t)dW_t\big) + Y_t\big(\mu(X_t,t)dt + \sigma(X_t,t)dW_t\big) + dX_tdY_t \\
&= \big(X_t\eta(Y_t,t) + Y_t\mu(X_t,t) + \sigma(X_t,t)\nu(Y_t,t)\big)dt + \big(X_t\nu(Y_t,t) + Y_t\sigma(X_t,t)\big)dW_t
\end{aligned}$$

which involves an additional term to classical calculus $\big(d(fg) = fdg + gdf\big)$.

In classical calculus we have, using Taylor's theorem again:

$$d(fg) = \frac{\partial(fg)}{\partial f}df + \frac{1}{2}\frac{\partial^2(fg)}{\partial f^2}(df)^2 + \ldots$$

$$+\frac{\partial(fg)}{\partial g}dg + \frac{1}{2}\frac{\partial^2(fg)}{\partial g^2}(dg)^2 + \ldots$$

$$+\frac{\partial(fg)}{\partial f \partial g}dfdg + \ldots$$

$$= gdf + fdg + \underbrace{dfdg}_{=0}$$

Crucially here, the *dfdg* term is considered to be zero.

2.2 Applications of Ito's Lemma

Squared Brownian motion

Consider applying the function $f(x,t) = x^2 - t$ to the Ito diffusion defined by:

$$X_T = W_T = W_0 + \int_0^T dW_t \quad \Leftrightarrow \quad dX_t = dW_t$$

This example is very similar to one from earlier in the chapter when we found the stochastic differential of W_t^2, though the presentation is different.

In this case:

- $$\frac{\partial f(x,t)}{\partial t} = \frac{\partial}{\partial t}\left(x^2 - t\right) = -1$$
- $$\frac{\partial f(x,t)}{\partial x} = \frac{\partial}{\partial x}\left(x^2 - t\right) = 2x$$
- $$\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{\partial}{\partial x}(2x) = 2$$

so, using:

$$df(X_t,t) = \left(\frac{\partial f(x,t)}{\partial t} + \mu(X_t,t)\frac{\partial f(x,t)}{\partial x} + \frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f(x,t)}{\partial x^2}\right)dt + \sigma(X_t,t)\frac{\partial f(x,t)}{\partial x}dW_t$$

with $\mu(x,t) = 0$ and $\sigma(x,t) = 1$ we have:

$$df(X_t,t) = \left(-1 + 0 \times 2X_t + \frac{1}{2} \times 1 \times 2\right)dt + 1 \times 2X_t dW_t$$

$$= 2W_t dW_t$$

Or as a stochastic integral equation we have:

$$\boldsymbol{W_T^2 - T = 2\int_0^T W_t dW_t}$$

Which recovers our previous result of:

$$\int_0^T W_t dW_t = \frac{1}{2}\left(W_T^2 - T\right)$$

Question

A process X_t satisfies the stochastic differential equation:

$$dX_t = \sigma(X_t)dB_t + \mu(X_t)dt$$

where B_t is a standard Brownian motion.

Deduce the stochastic differential equation for the process X_t^3.

Solution

By Ito's Lemma:

$$d\left(X_t^3\right) = 3X_t^2\sigma(X_t)dB_t + \left(3X_t^2\mu(X_t) + 3X_t\sigma^2(X_t)\right)dt$$

Alternatively, Taylor's formula to second-order is given by:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \ldots$$

With the given notation, $df(x) = f(x+h) - f(x)$ and $h = dx$ this becomes:

$$df(x) = f'(x)dx + ½f''(x)(dx)^2 + \ldots$$

For the specific case of $f(x) = x^3$, and ignoring terms above order 2, gives:

$$dX_t^3 = 3X_t^2 dX_t + 3X_t(dX_t)^2$$

But $dX_t = \sigma(X_t)dB_t + \mu(X_t)dt$ so the multiplication table gives:

$$(dX_t)^2 = \sigma^2(X_t)dt$$

Finally:

$$dX_t^3 = 3X_t^2\left(\sigma\left(X_t\right)dB_t + \mu\left(X_t\right)dt\right) + 3X_t\sigma^2\left(X_t\right)dt$$

$$= 3X_t^2\sigma\left(X_t\right)dB_t + \left(3X_t\sigma^2\left(X_t\right) + 3X_t^2\mu\left(X_t\right)\right)dt$$

as Ito's Lemma gave us.

Geometric Brownian motion

Consider applying the function $f(x,t) = e^x$ to the Ito diffusion defined by:

$$X_T = X_0 + \int_0^T \left(\mu - \frac{1}{2}\sigma^2\right)dt + \int_0^T \sigma dW_t \quad \Leftrightarrow \quad dX_t = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t$$

It's important to note that the constants μ and σ here are not the same as the functions $\mu(X_t,t)$ and $\sigma(X_t,t)$ given in the definition of Ito's Lemma. Re-parameterising Ito's Lemma gives:

$$df(X_t,t) = \left(\frac{\partial f}{\partial t} + \eta(X_t,t)\frac{\partial f}{\partial X_t} + \frac{1}{2}\nu^2(X_t,t)\frac{\partial^2 f}{\partial X_t^2}\right)dt + \nu(X_t,t)\frac{\partial f}{\partial X_t}dW_t$$

where:

$$dX_t = \eta(X_t,t)dt + \nu(X_t,t)dW_t$$

So in this case if we let $\eta(x,t) = \left(\mu - \frac{1}{2}\sigma^2\right)$ and $\nu(x,t) = \sigma$ we get:

$$dX_t = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t$$

as required, and then Ito's Lemma gives:

$$df(X_t,t) = \left(\frac{\partial f}{\partial t} + \eta(X_t,t)\frac{\partial f}{\partial X_t} + \frac{1}{2}\nu^2(X_t,t)\frac{\partial^2 f}{\partial X_t^2}\right)dt + \nu(X_t,t)\frac{\partial f}{\partial X_t}dW_t$$

$$= \left(\frac{\partial f}{\partial t} + \left(\mu - \frac{1}{2}\sigma^2\right)\frac{\partial f}{\partial X_t} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial X_t^2}\right)dt + \sigma\frac{\partial f}{\partial X_t}dW_t$$

In this case:

- $$\frac{\partial f(x,t)}{\partial t} = \frac{\partial}{\partial t}\left(e^x\right) = 0$$
- $$\frac{\partial f(x,t)}{\partial x} = \frac{\partial}{\partial x}\left(e^x\right) = e^x = f(x,t)$$
- $$\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{\partial}{\partial x}\left(e^x\right) = e^x = f(x,t)$$

By using Ito's Lemma we have:

$$\begin{aligned} df(X_t,t) &= \left(\frac{\partial f}{\partial t} + \left(\mu - \frac{1}{2}\sigma^2\right)\frac{\partial f}{\partial X_t} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial X_t^2}\right)dt + \sigma \frac{\partial f}{\partial X_t} dW_t \\ &= \left(0 + \left(\mu - \frac{1}{2}\sigma^2\right) f(X_t,t) + \frac{1}{2}\sigma^2 f(X_t,t)\right)dt + \sigma f(X_t,t) dW_t \end{aligned}$$

So:

$$df(X_t,t) = f(X_t,t)\left(\mu dt + \sigma dW_t\right)$$

or, setting $S_t = f(X_t,t) = e^{X_t}$ **:**

$$dS_t = S_t\left(\mu dt + \sigma dW_t\right)$$

$$\Rightarrow \qquad \frac{dS_t}{S_t} = \mu dt + \sigma dW_t$$

This is geometric Brownian motion and implies that the SDE:

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

is equivalent to:

$$S_t = e^{X_t} \text{ with } dX_t = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t$$

or:

$$\begin{aligned} S_T &= S_0 \exp\left(\int_0^T \left(\mu - \tfrac{1}{2}\sigma^2\right)dt + \int_0^T \sigma dW_t\right) \\ &= S_0 \exp\left(\left(\mu - \tfrac{1}{2}\sigma^2\right)T + \sigma W_T\right) \end{aligned}$$

This is the solution to the geometric Brownian motion SDE $dS_t = S_t\left(\mu dt + \sigma dW_t\right)$, and is a standard share price model.

We can confirm this observation by applying the function $f(s,t) = \ln(s)$ to the Ito diffusion defined by:

$$S_T = S_0 + \int_0^T \mu S_t dt + \int_0^T \sigma S_t dW_t \quad \Leftrightarrow \quad dS_t = \mu S_t dt + \sigma S_t dW_t$$

This yields:

$$df(S_t, t) = \left(\mu - \tfrac{1}{2}\sigma^2\right)dt + \sigma dW_t = dX_t$$

showing the symmetry with the previous result.

Question

Let the process S_t evolve according to the SDE:

$$dS_t = S_t\left(\mu dt + \sigma dW_t\right)$$

By considering $\ln(S_t)$, show that:

$$S_T = S_0 \exp\left(\left(\mu - \tfrac{1}{2}\sigma^2\right)T + \sigma W_T\right)$$

Solution

Let $f(S_t) = \ln(S_t)$ and apply Taylor's theorem to give:

$$\begin{aligned}
df(S_t) &= \frac{\partial f}{\partial S_t} dS_t + \frac{1}{2}\frac{\partial^2 f}{\partial S_t^2}(dS_t)^2 + \dots \\
&= \frac{1}{S_t} dS_t - \frac{1}{2S_t^2}(dS_t)^2 + \dots \\
&= \frac{S_t}{S_t}\left(\mu dt + \sigma dW_t\right) - \frac{S_t^2 \sigma^2}{2S_t^2} dt \\
&= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t
\end{aligned}$$

Integrating both sides between time 0 and T gives:

$$f(S_T) - f(S_0) = \left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T$$

$$\Rightarrow \quad S_T = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T\right)$$

On this basis we can consider the distribution of $\ln(S_T)$ where S_t is a geometric Brownian motion:

$$E[\ln(S_T)] = \ln(S_0) + \left(\mu - \tfrac{1}{2}\sigma^2\right)T$$

and:

$$Var(\ln(S_T)) = Var\left(\int_0^T \sigma dW_t\right)$$
$$= \sigma^2 T$$

Question

Find the expected value and variance of $\ln S_T$.

Solution

Given that:

$$S_T = S_0 \exp\left(\left(\mu - \tfrac{1}{2}\sigma^2\right)T + \sigma W_T\right)$$

Then:

$$\ln S_T = \ln S_0 + \left(\mu - \tfrac{1}{2}\sigma^2\right)T + \sigma W_T$$

Since $E[W_T] = 0$, $Var(W_T) = T$ and all other terms are deterministic we have:

$$E[\ln S_T] = E\left[\ln S_0 + \left(\mu - \tfrac{1}{2}\sigma^2\right)T\right] + E[\sigma W_T]$$
$$= \ln S_0 + \left(\mu - \tfrac{1}{2}\sigma^2\right)T + 0$$

and:

$$Var(\ln S_T) = Var\left(\ln S_0 + \left(\mu - \tfrac{1}{2}\sigma^2\right)T\right) + Var(\sigma W_T)$$
$$= 0 + \sigma^2 T$$

as required.

In conclusion, the distribution of a geometric Brownian motion at a time T is lognormally distributed:

$$\ln(S_T) \sim N\left(\ln(S_0) + \left(\mu - \tfrac{1}{2}\sigma^2\right)T, \sigma^2 T\right)$$

The Ornstein-Uhlenbeck process

Consider applying the function $f(x,t) = xe^{\gamma t}$ to the Ito diffusion defined by:

$$X_T = X_0 - \int_0^T \gamma X_t dt + \int_0^T \sigma dW_t \quad \Leftrightarrow \quad dX_t = -\gamma X_t dt + \sigma dW_t$$

This is known as the Ornstein-Uhlenbeck process.

We need:

- $\dfrac{\partial f(x,t)}{\partial t} = \dfrac{\partial}{\partial t}\left(xe^{\gamma t}\right) = \gamma xe^{\gamma t} = \gamma f(x,t)$
- $\dfrac{\partial f(x,t)}{\partial x} = \dfrac{\partial}{\partial x}\left(xe^{\gamma t}\right) = e^{\gamma t}$
- $\dfrac{\partial^2 f(x,t)}{\partial x^2} = \dfrac{\partial}{\partial x}\left(e^{\gamma t}\right) = 0$

so:

$$\begin{aligned} df(X_t,t) &= \gamma X_t e^{\gamma t} dt + e^{\gamma t} dX_t \\ &= \sigma e^{\gamma t} dW_t \end{aligned}$$

When applying Ito's Lemma we use $\mu(X_t,t) = -\gamma X_t$ and $\sigma(X_t,t) = \sigma$, therefore we have:

$$\begin{aligned} df(X_t,t) &= \left(\frac{\partial f}{\partial t} + \mu(X_t,t)\frac{\partial f}{\partial X_t} + \frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f}{\partial X_t^2}\right)dt + \sigma(X_t,t)\frac{\partial f}{\partial X_t}dW_t \\ &= \left(\gamma X_t e^{\gamma t} - \gamma X_t e^{\gamma t} + \frac{1}{2}\sigma^2 \times 0\right)dt + \sigma e^{\gamma t} dW_t \\ &= \sigma e^{\gamma t} dW_t \end{aligned}$$

As a stochastic integral equation we have:

$$\int_0^T df(X_t,t) = \int_0^T \sigma e^{\gamma t} dW_t$$

$$\Rightarrow \quad f(X_T,T) - f(X_0,0) = \int_0^T \sigma e^{\gamma t} dW_t$$

$$\Rightarrow \quad X_T e^{\gamma T} = X_0 + \sigma \int_0^T e^{\gamma t} dW_t$$

$$\Rightarrow \quad X_T = X_0 e^{-\gamma T} + \sigma \int_0^T e^{-\gamma(T-t)} dW_t$$

Question

The stochastic differential equation $dX_t = -\gamma X_t dt + \sigma dW_t$ can also be solved using the integrating factor $e^{\gamma t}$.

Show that you get the same solution if you solve the SDE using an integrating factor.

Solution

We want to solve the equation:

$$dX_t = -\gamma X_t dt + \sigma dW_t$$

or $$dX_t + \gamma X_t dt = \sigma dW_t$$

Multiplying through by the integrating factor $e^{\gamma t}$ and then changing the dummy variable to *s* gives:

$$e^{\gamma s} dX_s + \gamma e^{\gamma s} X_s ds = \sigma e^{\gamma s} dW_s$$

The left-hand side is now the differential of a product. So we have:

$$d\left(e^{\gamma s} X_s\right) = \sigma e^{\gamma s} dW_s$$

Now we can integrate between 0 and *t* to get:

$$e^{\gamma t} X_t - e^{\gamma 0} X_0 = \sigma \int_0^t e^{\gamma s} dW_s$$

Finally, we can rearrange this to get the desired form:

$$X_t = X_0 e^{-\gamma t} + \sigma \int_0^t e^{-\gamma(t-s)} dW_s$$

More generally we have:

$$X_T = X_t e^{-\gamma(T-t)} + \sigma \int_t^T e^{-\gamma(T-u)} dW_u$$

On this basis we can examine the distributional properties of X_T. Note the Ito integral has a deterministic integrand and so:

$$\boldsymbol{E\left[X_T\right] = E\left[X_t e^{-\gamma(T-t)} + \sigma \int_t^T e^{-\gamma(T-u)} dW_u\right]}$$

$$\boldsymbol{= X_t e^{-\gamma(T-t)}}$$

This is because the expectation of an Ito integral is zero, *ie*:

$$E\left[\sigma\int_t^T e^{-\gamma(T-u)}dW_u\right]=0$$

And:

$$\boldsymbol{Var(X_T)=E\left[\left(X_T-E[X_T]\right)^2\right]}$$

$$=Var\left(X_te^{-\gamma(T-t)}+\sigma\int_t^T e^{-\gamma(T-u)}dW_u\right)$$

$$=\underbrace{Var\left(X_te^{-\gamma(T-t)}\right)}_{=0}+\sigma^2 Var\left(\int_t^T e^{-\gamma(T-u)}dW_u\right)+\underbrace{2Cov\left(X_te^{-\gamma(T-t)},\sigma\int_t^T e^{-\gamma(T-u)}dW_u\right)}_{=0}$$

By Ito isometry:

$$Var(X_T)=\sigma^2\int_t^T e^{-2\gamma(T-u)}du$$

$$=\frac{\sigma^2}{2\gamma}\left[e^{-2\gamma(T-u)}\right]_t^T$$

$$=\boldsymbol{\frac{\sigma^2}{2\gamma}\left(1-e^{-2\gamma(T-t)}\right)}$$

For large $(T-t)$ this is approximately $\boldsymbol{\frac{\sigma^2}{2\gamma}}$ while for small $(T-t)$ it is (unsurprisingly) close to zero.

Question

Why is this unsurprising?

Solution

$T-t$ is the length of time over which the process is being observed. When this quantity is small it means that there's little opportunity for the process value X_T to deviate very far from X_t. This behaviour is captured by having a variance close to zero.

The mean-reverting process

The mean-reverting process, defined by the SDE:

$$dY_t = \gamma(\mu - Y_t)dt + \sigma dW_t$$

is based on the Ornstein-Uhlenbeck process. In the mean-reverting process, the process is pulled back to some equilibrium level, μ at a rate determined by $\gamma > 0$. Note that this process can go negative.

We can investigate this by considering $e^{\gamma t}Y_t$ (noting that the equation $dx = \gamma x dt$ implies the solution $x = e^{\gamma t}$) then:

$$d\left(e^{\gamma t}Y_t\right) = \gamma e^{\gamma t}Y_t dt + e^{\gamma t}dY_t$$

$$= \gamma e^{\gamma t}Y_t dt + e^{\gamma t}\left(\gamma(\mu - Y_t)dt + \sigma dW_t\right)$$

so:

$$d\left(e^{\gamma t}Y_t\right) = \gamma\mu e^{\gamma t}dt + \sigma e^{\gamma t}dW_t$$

Changing the variable of integration and integrating both sides between times t and T gives:

$$\int_t^T d\left(e^{\gamma s}Y_s\right) = \int_t^T \gamma\mu e^{\gamma s}ds + \int_t^T \sigma e^{\gamma s}dW_s$$

which implies that:

$$e^{\gamma T}Y_T = e^{\gamma t}Y_t + \int_t^T \gamma\mu e^{\gamma s}ds + \int_t^T \sigma e^{\gamma s}dW_s$$

because there are no random variables in the integrands, these are straightforward, and:

$$Y_T = e^{-\gamma(T-t)}Y_t + \mu\left(1 - e^{-\gamma(T-t)}\right) + \sigma\int_t^T e^{-\gamma(T-s)}dW_s$$

This result implies that:

$$E\left[Y_T\right] = e^{-\gamma(T-t)}Y_t + \mu\left(1 - e^{-\gamma(T-t)}\right)$$

while

$$Var\left(Y_T\right) = E\left[\left(\sigma\int_t^T e^{-\gamma(T-s)}dW_s\right)^2\right]$$

By Ito isometry:

$$Var\left(Y_T\right)=\frac{\sigma^2}{2\gamma}\left(1-e^{-2\gamma(T-t)}\right)$$

Observe that this tells us that:

$$Y_T \sim N\left(e^{-\gamma(T-t)}Y_t+\mu\left(1-e^{-\gamma(T-t)}\right),\frac{\sigma^2}{2\gamma}\left(1-e^{-2\gamma(T-t)}\right)\right)$$

This process is used again in the chapter on the term structure of interest rates.

3 Appendix

The following text is Core Reading that gives additional background information. It is presented here, without explanation, for completeness.

3.1 Square root mean-reverting process

The process defined by the SDE:

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t$$

with $\sigma, \theta > 0$ is known as the CIR, Feller or 'square root mean-reverting' process. If parameters satisfy $\sigma^2 \le 2\kappa\theta$ the process is positive. If the process hits zero, its volatility disappears and its drift is positive, the process deterministically moves away from zero and spends 'no time' at zero (*ie* the time spent at zero has measure zero). This is a very useful property in modelling asset prices.

There is no closed form solution for X_t, unlike for the O-U process, however the moments of the process can be derived.

Consider the function $f(t,x) = e^{\kappa t}x$, and so:

$$d\left(e^{\kappa t}X_t\right) = \frac{\partial}{\partial t}\left(e^{\kappa t}X_t\right)dt + \frac{\partial}{\partial x}\left(e^{\kappa t}X_t\right)dX_t + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(e^{\kappa t}X_t\right)dX_t^2$$

$$= \kappa\theta e^{\kappa t}dt + e^{\kappa t}\sigma\sqrt{X_t}dW_t$$

$$\Rightarrow \qquad e^{\kappa T}X_T = X_0 + \int_0^T \kappa\theta e^{\kappa t}dt + \int_0^T e^{\kappa t}\sigma\sqrt{X_t}dW_t$$

since the expectation of an Ito integral is zero:

$$E\left[X_T\right] = e^{-\kappa T}X_0 + \theta(1 - e^{-\kappa T})$$

To calculate the variance, use the fact that $Var\left[X\right] = E\left[X^2\right] - E\left[X\right]^2$ and use $Y_t = e^{\kappa t}X_t$, from above we have:

$$dY_t = \kappa\theta e^{\kappa t}dt + e^{\kappa t}\sigma\sqrt{X_t}dW_t$$

Observe that: $\sqrt{y} = \sqrt{e^{\kappa t}x} = e^{\frac{1}{2}\kappa t}\sqrt{x}$

$$dY_t = \kappa\theta e^{\kappa t}dt + e^{\frac{1}{2}\kappa t}\sigma\sqrt{Y_t}dW_t$$

Since $y = xe^{\kappa t}$, it follows that $X_0 e^{\kappa 0} = Y_0$ and:

$$E\left[Y_t\right] = Y_0 + \theta(e^{\kappa t} - 1)$$

while:

$$d(Y_t^2) = 2Y_t dY_t + dY_t dY_t$$
$$= e^{\kappa t} Y_t (2\kappa\theta + \sigma^2) dt + 2e^{\frac{1}{2}\kappa t} \sigma Y_t^{\frac{3}{2}} dW_t$$

and, because the integral is finite, and we can put the expectation in the time integral and $X_0 = Y_0$:

$$E\left[Y_T^2\right] = X_0^2 + (2\kappa\theta + \sigma^2)\int_0^T e^{\kappa t} E\left[Y_t\right] dt$$
$$= X_0^2 + (2\kappa\theta + \sigma^2)\int_0^T e^{\kappa t}\left[X_0 + \theta(e^{\kappa t} - 1)\right] dt$$

so:

$$E\left[X_T^2\right] = e^{-2\kappa T} X_0^2 + \frac{2\kappa\theta + \sigma^2}{k}(X_0 - \theta)(e^{-\kappa T} - e^{-2\kappa T}) + \frac{2\kappa\theta + \sigma^2}{2k}\theta(1 - e^{-2\kappa T})$$

hence:

$$Var\left[X_T\right] = \left[e^{-2\kappa T} X_0^2 + \frac{2\kappa\theta + \sigma^2}{k}(X_0 - \theta)(e^{-\kappa T} - e^{-2\kappa T}) + \frac{2\kappa\theta + \sigma^2}{2k}\theta(1 - e^{-2\kappa T})\right]$$
$$- \left[e^{-\kappa T} X_0 + \theta(1 - e^{-\kappa T})\right]^2$$
$$= \frac{\sigma^2}{\kappa} X_0 (e^{-\kappa T} - e^{-2\kappa T}) + \frac{\theta\sigma^2}{2\kappa}(1 - 2e^{-\kappa T} + e^{-2\kappa T})$$

and:

$$\lim_{T \to \infty} Var\left[X_T\right] = \frac{\theta\sigma^2}{2\kappa}$$

3.2 Multi-dimensional Ito formula

An *m*-dimensional process is a stochastic process made up of *m* independent processes. On the basis of an *m*-dimensional process we can define an *n*-dimensional Ito process as:

$$dX_1 = \mu_1 dt + \sigma_{11} dW_1 + \ldots + \sigma_{1m} dW_m$$
$$\vdots \quad =$$
$$dX_n = \mu_n dt + \sigma_{n1} dW_1 + \ldots + \sigma_{nm} dW_m$$

Given a function $f(t,x) = f(t, x_1, x_2, \ldots, x_n)$ we have the multi-dimensional Ito formula:

$$df(t,x) = \frac{\partial}{\partial t} f(t,x) dt + \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(t,x) dX_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} f(t,x) dX_i dX_j$$

3.3 Changing the Measure and Girsanov's Theorem

The basic idea

A standard normal random variable, $X \sim N(0,1)$ has the familiar density:

$$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\}$$

and so:

$$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\} = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(x-\mu)^2\right\}\exp\left\{-\mu x+\frac{1}{2}\mu^2\right\}$$

This means that if $X \sim N(0,1)$ and we have a function $h(X)$ then:

$$\begin{aligned} E\left[h(X)\right] &= \int_{\mathbb{R}} h(x)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\}dx \\ &= \int_{\mathbb{R}} \left[h(x)\exp\left\{-\mu x+\frac{1}{2}\mu^2\right\}\right]\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(x-\mu)^2\right\}dx \end{aligned}$$

so defining:

$$g(y) = h(y)\exp\left\{-\mu y+\frac{1}{2}\mu^2\right\}$$

and given $Y \sim N(0,1)$:

$$E\left[h(X)\right] = \int_{\mathbb{R}} g(y)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(y-\mu)^2\right\}dy$$

We can call the original probability measure, the standard normal, Q and the new probability, $N(\mu,1)$ that has a non-zero mean, P so that:

$$\begin{aligned} E_Q\left[h(X)\right] &= \int_{\mathbb{R}} h(x)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\}dx \\ &= E_P\left[g(Y)\right] \\ &= E_P\left[h(Y)\exp\left\{-\mu Y+\frac{1}{2}\mu^2\right\}\right] \end{aligned}$$

Or there is a function, $Z(Y)$ such that:

$$E_Q\left[h(X)\right]=E_P\left[h(Y)Z(Y)\right]$$

or more specifically:

$$E_Q\left[h(X(\omega))\right]=E_P\left[h(Y(\omega))Z(\omega)\right]$$

and Z is a Radon-Nikodým Derivative.

When working in continuous time the state space is infinite, meaning that for any path, ω, $P(\omega)=Q(\omega)=0$ and so the definition of a Radon-Nikodým Derivative is:

$$Z(\omega)=\frac{dQ(\omega)}{dP(\omega)}$$

Bearing in mind that probability densities can evolve, we define:

$$Z_t=\frac{dQ_t}{dP_t}$$

but we retain the properties of the Radon-Nikodým Derivative.

- **$P(Z>0)=1$ (equivalently $P(Z<\infty)=1$),**
- **$E_P[Z]=1$,**
- **$E_Q[Y]=E_P[ZY]$,**
- **If Y is an F_t-measurable random variable on $0\le s\le t\le T<\infty$, then**

$$E_Q[Y\mid F_s]=\frac{1}{Z_s}E_P[YZ_t\mid F_s]$$

3.4 Change of a Wiener Process

A Wiener Process is $W_t \sim N(0,t)$ with the density:

$$\frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{1}{2}\frac{x^2}{t}\right\}$$

and so using the same arguments as above but with the substitutions $w\to\frac{x}{\sqrt{t}}$ and $\mu\to-\Lambda\sqrt{t}$, where Λ is any constant, we have that the Wiener Process has the density:

$$\frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{1}{2}\left(\frac{x}{\sqrt{t}}\right)^2\right\}=\frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{1}{2}\left(\frac{x+\Lambda t}{\sqrt{t}}\right)^2\right\}\exp\left\{\Lambda x+\frac{1}{2}\Lambda^2 t\right\}$$

This implies that:

$$\frac{\frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{1}{2}\left(\frac{x+\Lambda t}{\sqrt{t}}\right)^2\right\}}{\frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{1}{2}\left(\frac{x}{\sqrt{t}}\right)^2\right\}}=\frac{dQ}{dP}=\exp\left\{-\Lambda x-\frac{1}{2}\Lambda^2 t\right\}$$

Setting $X_t = W_t + \Lambda t$ we have that $X_t \sim N(\Lambda t, t)$, *ie* it is not a Wiener process. However, if we define the Radon-Nikodým Derivative:

$$Z_t = \frac{dQ}{dP} = \exp\left\{-\Lambda W_t - \frac{1}{2}\Lambda^2 t\right\}$$

we generate a new probability under which X_t is a Wiener process.

These ideas capture the Cameron-Martin-Girsanov (usually shortened to the 'Girsanov Theorem').

Given a Wiener Process W_t, $0 \le t \le T < \infty$, on (Ω, F, P) with F_t being the natural filtration of W_t. Let λ_t be an F_t adapted process. Define:

$$Z_t = \exp\left\{-\int_0^t \lambda_s dW_s - \frac{1}{2}\int_0^t \lambda_s^2 ds\right\}$$

and:

$$\hat{W}_t = W_t + \int_0^t \lambda_s ds$$

and assume that:

$$E_P\left[\int_0^T \lambda_s^2 Z_s^2 ds\right] < \infty$$

Then under the measure *Q* defined by:

$$Q(A) = \int_A Z(\omega) dP(\omega)$$

the process $\hat{W}_t$ is a Wiener Process and a *Q*-martingale.

3.5 The Martingale Representation Theorem

Consider a driftless Ito process:

$$dX_t = \phi(t, X_t)dW_t$$

then its integral equation, for $S \le T$ is:

$$X_T = X_S + \int_S^T \phi(t, X_t)dW_t$$

and so (informally):

$$E\left[X_T \mid F_S^W\right] = X_S + E\left[\int_S^T \phi(t, X_t)dW_t \mid F_S^W\right]$$

$$= X_S$$

and we have that the driftless Ito process is a martingale.

We have seen that the Ito processes $W_t^2 - t$, $\int W_t dt$ and $e^{-\phi W_t - \frac{1}{2}\phi^2 t}$ are F_t^W martingales and what is more, they can be described by driftless Ito processes:

$$W_T^2 - T = \int_0^T 2W_t dW_t$$

$$\int_0^T W_t dt = \int_0^T (T-t)dW_t$$

$$e^{W_T - \frac{1}{2}T} = 1 + \int_0^T e^{W_t - \frac{1}{2}t} dW_t$$

In fact, these are examples of a general theory, the Martingale Representation Theorem.

In the Martingale Representation Theorem, let W_t^n be an *n*-dimensional process. Suppose that M_t is a continuous $F_t^{W^n}$-martingale, then there exists a unique function, $\phi = \phi(t, W_t^n)$, that is $F_t^{W^n}$-adapted such that:

$$M_T = E\left[M_t\right] + \int_t^T \phi(u, W_u)dW_u^n$$

$$= M_0 + \int_0^T \phi(t, W_t)dW_t^n$$

The Martingale Representation Theorem can be understood by noting that, since M_t is a continuous $F_t^{W^n}$-martingale its movement depends completely on W_t^n, as such it is not unreasonable to believe that its motion is a function of the motion of W_t^n.

For example, consider an $F_t^{W^n}$-martingale $M_t = M(t, X_t) \in V$ with X_t being an Ito process defined by the SDE:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$$

In this case:

$$dM_t = \left[\frac{\partial M(t, X_t)}{\partial t} + \mu(t, X_t)\frac{\partial M(t, X_t)}{\partial X} + \frac{1}{2}\frac{\partial^2 M(t, X_t)}{\partial X^2}\sigma^2(t, X_t)\right]dt + \sigma(t, X_t)\frac{\partial M(t, X_t)}{\partial X}dW_t$$

Since M_t is a martingale, the drift term is zero:

$$= \sigma(t, X_t)\frac{\partial M(t, X_t)}{\partial X}dW_t$$

and so the $F_t^{W^n}$-adapted process, ϕ, associated with the Martingale Representation Theorem is given by:

$$\phi(t, W_t) = \sigma(t, X_t)\frac{\partial M(t, X_t)}{\partial X}$$

The chapter summary starts on the next page so that you can keep all the chapter summaries together for revision purposes.

Chapter 10 Summary

Ito integrals

Ito integrals of the form $\int_0^T f(W_t,t)\,dW_t$, where $f(W_t,t)$ is measurable with respect to the natural filtration of W_t:

- cannot be integrated directly
- can often be simplified using Ito's Lemma
- have a normal distribution, with mean zero and variance $\int_0^T E\left[f^2(W_t,t)\right]dt$.

Ito's Lemma

Ito's Lemma can be used to differentiate a function *f* of a stochastic process X_t.

If $dX_t = \mu dt + \sigma dW_t$, then:

- $$df(X_t) = \left(\mu\frac{\partial f}{\partial X_t} + \tfrac{1}{2}\sigma^2\frac{\partial^2 f}{\partial X_t^2}\right)dt + \sigma\frac{\partial f}{\partial X_t}dW_t$$

  This form is used when the new process depends only on the *values* of the original process.

- $$df(X_t,t) = \left(\mu\frac{\partial f}{\partial X_t} + \tfrac{1}{2}\sigma^2\frac{\partial^2 f}{\partial X_t^2} + \frac{\partial f}{\partial t}\right)dt + \sigma\frac{\partial f}{\partial X_t}dW_t$$

  This form is used when there is *explicit time-dependence*, *ie* the new process depends on the value of the original process *and* the time.

Alternatively, *Taylor's formula* to the second-order can be used to write:

$$df\left(X_t,t\right) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial X_t}dX_t + ½\frac{\partial^2 f}{\partial X_t^2}\left(dX_t\right)^2$$

into which $dX_t = \mu dt + \sigma dW_t$ can be substituted. The second-order terms can then be simplified using the multiplication table:

	dt	dW_t
dt	0	0
dW_t	0	dt

Ornstein-Uhlenbeck process

Mean-reverting financial quantities (such as interest rates) can be modelled using the *Ornstein-Uhlenbeck process*.

The process is defined by the SDE:

$$dX_t = -\gamma X_t dt + \sigma dW_t$$

where γ is a positive parameter.

It can be shown that the formula for the process itself is:

$$X_T = X_0 e^{-\gamma T} + \sigma \int_0^T e^{-\gamma(T-t)} dW_t$$

Also, the probability distribution of X_T is $N\left(X_0\, e^{-\gamma T}, \dfrac{\sigma^2}{2\gamma}(1-e^{-2\gamma T}) \right)$, and the long-term distribution is $N\left(0, \dfrac{\sigma^2}{2\gamma} \right)$.

The mean-reverting process

This process is a generalisation of the Ornstein-Uhlenbeck process, and is defined by the SDE:

$$dY_t = \gamma\left(\mu - Y_t \right) dt + \sigma dW_t$$

The solution is given by:

$$Y_T = e^{-\gamma(T-t)} Y_t + \mu\left(1 - e^{-\gamma(T-t)} \right) + \sigma \int_t^T e^{-\gamma(T-s)} dW_s$$

Geometric Brownian motion

Asset prices are often modelled using geometric Brownian motion.

The process is defined by the SDE:

$$dS_t = S_t\left(\mu dt + \sigma dW_t \right)$$

It can be shown that the formula for the process itself is:

$$S_T = S_0 \exp\left(\left(\mu - \tfrac{1}{2}\sigma^2 \right) T + \sigma W_T \right)$$

Chapter 10 Practice Questions

10.1 (i) Write down Ito's Lemma as it applies to a function $f(X_t)$ of a stochastic process X_t that satisfies the stochastic differential equation $dX_t = \sigma_t dB_t + \mu_t dt$, where B_t is a standard Brownian motion.

(ii) Hence find the stochastic differential equations for each of the following processes:

(a) $G_t = \exp(X_t)$

(b) $Q_t = X_t^2$

(c) $V_t = (1+X_t)^{-1}$

(d) $L_t = 100 + 10X_t$

(e) $J_t = \ln B_t$

(f) $K_t = 5B_t^3 + 2B_t$

10.2 Let B_t ($t \geq 0$) be a standard Brownian motion with $B_0 = 0$.

(i) By first writing down an expression for $d(B_s^2)$, show that

$$\int_0^t B_s dB_s = \tfrac{1}{2}(B_t^2 - t)$$

(ii) What is the expected value at time 0 of $\int_0^t B_s dB_s$?

(iii) What is the expected value at time u ($0 < u < t$) of $\int_0^t B_s dB_s$?

(iv) What can you say about the process $I_t = \int_0^t B_s dB_s$, based on your results from (i) and (iii)?

10.3 Find the mean and variance of the stochastic integral $I = \int_0^1 t\, dB_t$.

10.4 Let $\{X_t\}$ be a continuous-time stochastic process defined by the equation $X_t = \alpha W_t^2 + \beta$, where $\{W_t\}$ is a standard Brownian motion and α and β are constants.

By applying Ito's Lemma, or otherwise, write down the stochastic differential equation satisfied by X_t .

10.5 Let X_t be an Ito process that satisfies $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$ where B_t is a standard Brownian motion. Let $f(X_t, t)$ be a function of t and X_t.

(i) By considering Taylor's theorem, suggest a partial differential equation that must be satisfied by $f(X_t, t)$ in order that it is a martingale.

(ii) Verify that your equation holds when $f(X_t, t) = B_t^2 - t$.

(iii) Find $g(t)$ such that $B_t^3 + g(t)B_t$ is a martingale.

10.6 In the following, B_t denotes a standard Brownian motion.

Exam style

(i) Write down the general solution of the stochastic differential equation:

$$dX_t = -\gamma X_t dt + \sigma dB_t$$ [1]

(ii) Hence determine the solution of the stochastic differential equation:

$$dR_t = 0.8(4 - R_t)dt + dB_t$$

where $R_0 = 5$. [1]

(iii) Find the distribution of the process R_t at time t and in the long-term. [3]

[Total 5]

10.7 The Ito process, X_t, is defined by the stochastic differential equation:

Exam style

$$dX_t = 0.5X_t(1 - X_t)(1 - 2X_t)dt - X_t(1 - X_t)dB_t$$

where B_t is a standard Brownian motion, and $X_0 = 0.5$.

By considering the stochastic differential equation for the process $Y_t = \ln\left(\frac{1}{X_t} - 1\right)$, find X_t in terms of B_t. [7]

10.8 The market price of a certain share is being modelled as a geometric Brownian motion. The price S_t at time $t \geq 0$ satisfies the equation:

Exam style

$$\log_e \frac{S_t}{S_0} = \mu t + \sigma B_t$$

where $\{B_t, t \geq 0\}$ is a standard Brownian motion and μ and σ are constants.

(i) Show that the stochastic differential dS_t can be written in the form:

$$\frac{dS_t}{S_t} = c_1 dB_t + c_2 dt,$$

where c_1 and c_2 are constants you should specify. [4]

(ii) Derive expressions for $E[S_t]$ and $Var\left(S_t\right)$. [4]

(iii) Derive expressions for $Cov\left(S_{t_1}, S_{t_2}\right)$ and $E[S_{t_2} \mid S_{t_1}]$ where $0 < t_1 < t_2$. [6]

(iv) By using your expression for $E[S_{t_2} \mid S_{t_1}]$, write down a function of S_t that is a martingale. [1]

[Total 15]

10.9 Let S_t be a geometric Brownian motion process defined by the equation $S_t = \exp(\mu t + \sigma W_t)$, where W_t is a standard Brownian motion and μ and σ are constants.

Exam style

(i) Write down the stochastic differential equation satisfied by $X_t = \log_e S_t$. [1]

(ii) By applying Ito's Lemma, or otherwise, derive the stochastic differential equation satisfied by S_t. [3]

(iii) The price of a share follows a geometric Brownian motion with $\mu = 0.06$ and $\sigma = 0.25$ (both expressed in annual units). Find the probability that, over a given one-year period, the share price will fall. [3]

[Total 7]

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 10 Solutions

10.1 (i) ***Ito's Lemma***

Ito's Lemma states that $f(X_t)$ satisfies the SDE:

$$df(X_t) = \sigma_t f'(X_t) dB_t + [\mu_t f'(X_t) + \tfrac{1}{2}\sigma_t^2 f''(X_t)]dt$$

As a 'technicality', the function f has to be twice differentiable, in order for the RHS to make sense.

(ii)(a) ***Stochastic differential equation for G_t***

Here the function we are applying Ito's Lemma to is $f(x) = e^x$, with $f'(x) = e^x$ and $f''(x) = e^x$.

So we get:

$$\begin{aligned} dG_t &= \sigma_t e^{X_t} dB_t + [\mu_t e^{X_t} + \tfrac{1}{2}\sigma_t^2 e^{X_t}]dt \\ &= \sigma_t G_t dB_t + [\mu_t + \tfrac{1}{2}\sigma_t^2]G_t dt \end{aligned}$$

(ii)(b) ***Stochastic differential equation for Q_t***

Here the function we are applying Ito's Lemma to is $f(x) = x^2$, with $f'(x) = 2x$ and $f''(x) = 2$.

So we get:

$$dQ_t = 2\sigma_t X_t dB_t + [2\mu_t X_t + \sigma_t^2]dt$$

We can write this entirely in terms of the new process as:

$$dQ_t = 2\sigma_t Q_t^{½} dB_t + [2\mu_t Q_t^{½} + \sigma_t^2]dt$$

(ii)(c) ***Stochastic differential equation for V_t***

Here the function we are applying Ito's Lemma to is $f(x) = (1+x)^{-1}$, with $f'(x) = -(1+x)^{-2}$ and $f''(x) = 2(1+x)^{-3}$. So we get:

$$\begin{aligned} dV_t &= -\sigma_t (1+X_t)^{-2} dB_t + [-\mu_t (1+X_t)^{-2} + \sigma_t^2 (1+X_t)^{-3}]dt \\ &= -\sigma_t V_t^2 dB_t + [-\mu_t V_t^2 + \sigma_t^2 V_t^3]dt \end{aligned}$$

(ii)(d) ***Stochastic differential equation for L_t***

Here the function we are applying Ito's Lemma to is $f(x) = 100 + 10x$, with $f'(x) = 10$ and $f''(x) = 0$. So we get:

$$dL_t = 10\sigma_t dB_t + (10\mu_t + 0)dt = 10(\sigma_t dB_t + \mu_t dt)$$

(ii)(e) ***Stochastic differential equation for J_t***

Since J_t is a function of standard Brownian motion, B_t, rather than X_t, when applying Ito's Lemma, we note:

$$dB_t = 1 \times dB_t + 0 \times dt$$

Equivalently, $\mu_t = 0$ and $\sigma_t = 1$.

Here the function we are applying Ito's Lemma to is $f(x) = \ln(x)$, with $f'(x) = x^{-1}$ and $f''(x) = -x^{-2}$. So we get:

$$\begin{aligned} dJ_t &= 1 \times B_t^{-1} dB_t - \tfrac{1}{2} \times 1 \times B_t^{-2} dt \\ &= B_t^{-1} dB_t - \tfrac{1}{2} B_t^{-2} dt \end{aligned}$$

(ii)(f) ***Stochastic differential equation for K_t***

K_t is a function of standard Brownian motion, B_t, so $\mu_t = 0$ and $\sigma_t = 1$.

Here the function we are applying Ito's Lemma to is $f(x) = 5x^3 + 2x$, with $f'(x) = 15x^2 + 2$ and $f''(x) = 30x$. So we get:

$$\begin{aligned} dJ_t &= 1 \times (15B_t^2 + 2) dB_t + \tfrac{1}{2} \times 1 \times (30B_t) dt \\ &= (15B_t^2 + 2) dB_t + 15B_t dt \end{aligned}$$

10.2 (i) ***Expression for $d(B_s^2)$***

Either by applying Ito's Lemma or by simply expanding as a Taylor series, we find that:

$$d(B_s^2) = 2B_s dB_s + \tfrac{1}{2} \times 2(dB_s)^2 = 2B_s dB_s + ds$$

where, as usual, we have replaced the second order Brownian differential $(dB_s)^2$ with the time differential ds, using the 2×2 multiplication grid given.

To show this, let $G(B_s) = B_s^2$, *then* $\dfrac{\partial G}{\partial s} = 0$, $\dfrac{\partial G}{\partial B_s} = 2B_s$ *and* $\dfrac{\partial^2 G}{\partial B_s^2} = 2$, *and since we can write* $dB_s = 0\, ds + 1\, dB_s$, *the drift and volatility functions in the stochastic differential equation for* dB_s *are 0 and 1 respectively. Hence, using Ito's Lemma from page 46 in the Tables we have:*

$$dG = \left[0 \times 2B_s + ½ \times 1^2 \times 2 + 0 \right] ds + 1 \times 2B_s dB_s$$

ie $\quad d(B_s^2) = 2B_s dB_s + ds$

Alternatively, using a Taylor Series expansion we have:

$$dG = \frac{\partial G}{\partial s} ds + \frac{\partial G}{\partial B_s} dB_s + ½\frac{\partial^2 G}{\partial B_s^2}(dB_s)^2$$

$$= 0 \times ds + 2B_s dB_s + ½ \times 2 \times (dB_s)^2$$

$$= 2B_s dB_s + ds$$

If we now integrate this equation between time 0 and time t, we get:

$$\int_0^t d(B_s^2) = 2\int_0^t B_s dB_s + \int_0^t ds$$

ie $$\left[B_s^2\right]_0^t = 2\int_0^t B_s dB_s + \left[s\right]_0^t$$

ie $$B_t^2 - B_0^2 = 2\int_0^t B_s dB_s + t \quad \Rightarrow \int_0^t B_s dB_s = \tfrac{1}{2}(B_t^2 - t)$$

Note that the value of this integral is a random variable.

(ii) ***What is the expected value at time 0?***

At time 0 the future values of B_t are unknown, and we have:

$$E_0\left[\int_0^t B_s dB_s\right] = E_0\left[\tfrac{1}{2}(B_t^2 - t)\right] = \tfrac{1}{2}[E_0(B_t^2) - t]$$

But we know that $B_t - B_0 \sim N(0,t)$.

So: $$E_0(B_t^2) = \text{var}_0(B_t) + \left[E_0(B_t)\right]^2 = t + 0^2 = t$$

and $$E_0\left[\int_0^t B_s dB_s\right] = \tfrac{1}{2}[t - t] = 0$$

(iii) ***What is the expected value at time u?***

At time u (where $0 < u < t$) the values of B_t up to time u are known, but the values after time u are unknown. If we split the integral into two parts, we have:

$$\int_0^t B_s dB_s = \int_0^u B_s dB_s + \int_u^t B_s dB_s$$

The first integral on the RHS is known at time u and, from part (i), it is equal to:

$$\int_0^u B_s dB_s = \tfrac{1}{2}(B_u^2 - u)$$

The second integral on the RHS is a random quantity. Using the same method as in part (i), it is equal to:

$$\int_u^t B_s dB_s = \left[\tfrac{1}{2}(B_s^2 - s)\right]_u^t = \tfrac{1}{2}(B_t^2 - t) - \tfrac{1}{2}(B_u^2 - u)$$

The only random quantity in this expression is B_t^2 and we need to find the expected value of this based on the information we have at time u. We know that $B_t - B_u \sim N(0, t-u)$ or, since we know the value of B_u, $B_t \sim N(B_u, t-u)$. This tells us that:

$$E_u(B_t^2) = \text{var}_u(B_t) + [E_u(B_t)]^2 = t - u + B_u^2$$

So the expected value of our second integral is:

$$E_u\left[\int_u^t B_s dB_s\right] = 0$$

Combining the two parts of the integral, we get:

$$E_u\left[\int_0^t B_s dB_s\right] = \tfrac{1}{2}(B_u^2 - u) + 0 = \tfrac{1}{2}(B_u^2 - u)$$

Alternatively, we could use the fact that $B_t^2 - t$ is a martingale with respect to B_t to show this.

(iv) ***What can you say about this process?***

From part (i), we know that $\int_0^t B_s dB_s$ is always equal to $\tfrac{1}{2}(B_t^2 - t)$. So, in particular, $\tfrac{1}{2}(B_u^2 - u) = I_u$.

So we can write the result in (iii) as:

$$E_u[I_t] = I_u \quad \text{whenever } 0 < u < t$$

This tells us that the process I_t is a martingale with respect to B_t.

It is a general property that Ito integrals are martingales.

10.3 This integral can be thought of as the (limiting) sum of the small elements $t dB_t$.

The dB_t's are random quantities with mean 0 and variance dt. So the expected value of each element is 0 and the variance is $t^2 dt$.

So: $$E(I) = E\left(\int_0^1 t\, dB_t\right) = \int_0^1 t\, E(dB_t) = 0$$

Also:

$$Var(I) = \text{var}\left(\int_0^1 t\,dB_t\right) = \int_0^1 Var(t\,dB_t)$$

since the increments are independent. So:

$$Var(I) = \int_0^1 t^2\,dt = \left[\tfrac{1}{3}t^3\right]_0^1 = \tfrac{1}{3}$$

10.4 Since X_t is a function of standard Brownian motion, W_t, when applying Ito's Lemma, we note that the stochastic differential equation for the underlying stochastic process (standard Brownian motion) is:

$$dW_t = 1 \times dW_t + 0 \times dt$$

Let $G(W_t) = X_t = \alpha W_t^2 + \beta$, then:

$$\frac{\partial G}{\partial t} = 0$$

$$\frac{\partial G}{\partial W_t} = 2\alpha W_t$$

$$\frac{\partial^2 G}{\partial W_t^2} = 2\alpha$$

Hence, using Ito's Lemma from page 46 in the *Tables* we have:

$$dG = \left[0 \times 2\alpha W_t + ½ \times 1^2 \times 2\alpha + 0\right]dt + 1 \times 2\alpha W_t dW_t$$

ie $\quad dX_t = 2\alpha W_t dW_t + \alpha dt$

Alternatively, using a Taylor Series expansion we have:

$$dG = \frac{\partial G}{\partial t}dt + \frac{\partial G}{\partial W_t}dW_t + ½\frac{\partial^2 G}{\partial W_t^2}(dW_t)^2$$

$$= 0 \times dt + 2\alpha W_t dW_t + ½ \times 2\alpha \times (dW_t)^2$$

$$= 2\alpha W_t dW_t + \alpha dt$$

Remembering that $(dW_t)^2 = dt$ *from the multiplication table.*

10.5 (i) ***PDE for a martingale***

Taylor's theorem (ignoring higher order terms) is given by

$$df(X_t,t)=\frac{\partial f}{\partial t}dt+\frac{\partial f}{\partial X_t}dX_t+\frac{1}{2}\frac{\partial^2 f}{\partial X_t{}^2}(dX_t)^2$$

Substituting in for dX_t gives:

$$df(X_t,t)=\left[\frac{\partial f}{\partial t}+\frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f}{\partial X_t{}^2}+\mu(X_t,t)\frac{\partial f}{\partial X_t}\right]dt+\sigma(X_t,t)\frac{\partial f}{\partial X_t}dB_t$$

Note that in the function f we have ***explicit*** *time dependence. This gives the extra term* $\frac{\partial f}{\partial t}dt$ *in the Taylor expansion.*

For a martingale we require zero drift and hence

$$\frac{\partial f}{\partial t}+\frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2 f}{\partial X_t{}^2}+\mu(X_t,t)\frac{\partial f}{\partial X_t}=0$$

(ii) ***Verify that the equation holds***

Here $X_t=B_t$ and we are looking at a function of standard Brownian motion.

Note that $dB_t=0\times dt+1\times dB_t$, so $\mu=0$ and $\sigma=1$, and the derivatives are:

$$\frac{\partial f}{\partial t}=-1,\ \frac{\partial f}{\partial X_t}=2B_t,\ \frac{\partial^2 f}{\partial X_t^2}=2$$

Since the terms on the left-hand side of the equation in part (i) sum to zero, the equation holds.

(iii) ***Find g(t)***

Again $\mu=0$ and $\sigma=1$. Using the equation from part (i), we require:

$$g'(t)B_t+½\times 6B_t=0$$

So it follows that $g'(t)=-3$ and hence $g(t)=-3t$ will do, *ie* $B_t^3-3tB_t$ is a martingale.

We could also have used $g(t)=-3t+c$, *where* c *is any constant.*

10.6 (i) ***Solution for X_t***

The process X_t is an Ornstein-Uhlenbeck process so that:

$$X_t = X_0 e^{-\gamma t} + \sigma \int_0^t e^{-\gamma(t-s)} dB_s$$ [1]

(ii) ***Solution for R_t***

The process $R_t - 4$ is an Ornstein-Uhlenbeck process so that:

$$R_t - 4 = (R_0 - 4) e^{-\gamma t} + \sigma \int_0^t e^{-\gamma(t-s)} dB_s$$ [½]

which in this case becomes:

$$R_t = 4 + e^{-0.8t} + \int_0^t e^{-0.8(t-s)} dB_s$$ [½]

[Total 1]

(iii) ***Distribution of R_t***

Since $dB_s \sim N(0, ds)$, and these increments are independent, it follows that:

$$R_t \sim N\left(4 + e^{-0.8t}, \int_0^t e^{-1.6(t-s)} ds\right) = N\left(4 + e^{-0.8t}, \frac{1 - e^{-1.6t}}{1.6}\right)$$ [2]

In the long-term $\lim_{t \to \infty} R_t \sim N(4, 0.625)$. [1]

[Total 3]

10.7 Starting with a Taylor Series expansion for dY_t:

$$dY_t = d\ln\left(\frac{1}{X_t} - 1\right) = f'(X_t) dX_t + 0.5 f''(X_t)(dX_t)^2 \qquad (1)$$ [1]

where $f(x) = \ln\left(\frac{1}{x} - 1\right)$. We have:

$$f'(x) = \frac{-1/x^2}{\left(\frac{1}{x} - 1\right)} = \frac{1}{x^2 - x} \quad \text{and} \quad f''(x) = -\frac{2x - 1}{(x^2 - x)^2} = \frac{1 - 2x}{x^2 (1 - x)^2}$$ [2]

We also have:

$$(dX_t)^2 = X_t^2(1-X_t)^2\,dt \qquad [1]$$

by using the 2×2 multiplication grid for increments, which tells us that $(dt)^2 = 0$ and $dB_t \times dt = 0$. Substituting this into (1), along with the derivatives of f gives:

$$dY_t = \left[\frac{-1}{X_t(1-X_t)}\right]\left[0.5X_t(1-X_t)(1-2X_t)dt - X_t(1-X_t)dB_t\right]$$

$$+0.5\left[\frac{1-2X_t}{X_t^{\,2}(1-X_t)^2}\right]\left[X_t^2(1-X_t)^2\,dt\right]$$

$$= dB_t \qquad [1]$$

Alternatively, substituting the same partial derivatives as above into Ito's Lemma on page 46 in the Tables, together with the drift and volatility functions:

$$a(X_t) = 0.5X_t(1-X_t)(1-2X_t)$$

$$b(X_t) = -X_t(1-X_t)$$

gives:

$$dY_t = \left[0.5X_t(1-X_t)(1-2X_t)\times\frac{1}{X_t^2-X_t} + ½X_t^2(1-X_t)^2\frac{1-2X_t}{X_t^{\,2}(1-X_t)^2}\right]dt$$

$$-X_t(1-X_t)\times\frac{1}{X_t^2-X_t}dB_t$$

$$= dB_t \qquad [2]$$

It follows by integrating and taking the initial condition $X_0 = 0.5$ (which implies $Y_0 = 0$) into account that:

$$Y_t = B_t \quad ie: \quad \ln\left(\frac{1}{X_t}-1\right) = B_t \qquad [1]$$

Rearranging gives: $X_t = \dfrac{1}{1+e^{B_t}}$ [1]

[Total 7]

10.8 (i) ***Stochastic differential dS_t***

Rearranging the relationship given, we get:

$$S_t = S_0 e^{\mu t + \sigma B_t} \qquad [½]$$

Since S_t is a function of standard Brownian motion, B_t, when applying Ito's Lemma, we note that the stochastic differential equation for the underlying stochastic process (standard Brownian motion) is:

$$dB_t = 0 \times dt + 1 \times dB_t$$ [½]

Let $G(t, B_t) = S_t = S_0 e^{\mu t + \sigma B_t}$, then:

$$\frac{\partial G}{\partial t} = \mu S_0 e^{\mu t + \sigma B_t} = \mu S_t$$ [½]

$$\frac{\partial G}{\partial B_t} = \sigma S_0 e^{\mu t + \sigma B_t} = \sigma S_t$$ [½]

$$\frac{\partial^2 G}{\partial B_t^2} = \sigma^2 S_0 e^{\mu t + \sigma B_t} = \sigma^2 S_t$$ [½]

Hence, using Ito's Lemma from page 46 in the *Tables* we have:

$$dG = \left[0 \times \sigma S_t + ½ \times 1^2 \times \sigma^2 S_t + \mu S_t\right] dt + 1 \times \sigma S_t dB_t$$ [½]

ie $$dS_t = \left(\mu + ½\sigma^2\right) S_t dt + \sigma S_t dB_t$$ [½]

Alternatively, using a Taylor Series expansion we have:

$$dG = \frac{\partial G}{\partial t} dt + \frac{\partial G}{\partial B_t} dB_t + ½ \frac{\partial^2 G}{\partial B_t^2} (dB_t)^2$$

$$= \mu S_t dt + \sigma S_t dB_t + ½\sigma^2 S_t \times (dB_t)^2$$

$$= \left(\mu + ½\sigma^2\right) S_t dt + \sigma S_t dB_t$$

Remember that $(dB_t)^2 = dt$ *from the multiplication table.*

This can be written as:

$$\frac{dS_t}{S_t} = \sigma dB_t + \left(\mu + \tfrac{1}{2}\sigma^2\right) dt$$

So: $c_1 = \sigma$ and $c_2 = \mu + \tfrac{1}{2}\sigma^2$ [1]

[Maximum 4]

(ii) ***Expressions for the mean and variance***

The expected value of S_t is:

$$E[S_t] = E[S_0 e^{\mu t + \sigma B_t}] = S_0 e^{\mu t} E[e^{\sigma B_t}]$$

Since $B_t \sim N(0,t)$, its MGF is $E[e^{\theta B_t}] = e^{½\theta^2 t}$.

So: $E[S_t] = S_0 e^{\mu t} \times e^{½\sigma^2 t} = S_0 e^{\mu t + ½\sigma^2 t}$ [2]

The variance of S_t is:

$$Var(S_t) = E[S_t^2] - (E[S_t])^2$$

$$= E[S_0^2 e^{2\mu t + 2\sigma B_t}] - (S_0 e^{\mu t + ½\sigma^2 t})^2$$

$$= S_0^2 e^{2\mu t} E[e^{2\sigma B_t}] - S_0^2 e^{2\mu t + \sigma^2 t}$$

$$= S_0^2 e^{2\mu t + 2\sigma^2 t} - S_0^2 e^{2\mu t + \sigma^2 t} = S_0^2 e^{2\mu t}(e^{2\sigma^2 t} - e^{\sigma^2 t})$$ [2]

An alternative approach is to use the formulae for the mean and variance of the lognormal distribution.

[Total 4]

(iii) ***Expressions for the covariance and conditional expectation***

The covariance of S_{t_1} and S_{t_2} is:

$$Cov\left(S_{t_1}, S_{t_2}\right) = E[S_{t_1} S_{t_2}] - E[S_{t_1}]E[S_{t_2}]$$

From above:

$$E[S_{t_1}] = S_0 e^{\mu t_1 + ½\sigma^2 t_1} \quad \text{and} \quad E[S_{t_2}] = S_0 e^{\mu t_2 + ½\sigma^2 t_2}$$

The expected value of the product is:

$$E[S_{t_1} S_{t_2}] = E[S_0 \exp(\mu t_1 + \sigma B_{t_1}) S_0 \exp(\mu t_2 + \sigma B_{t_2})]$$

$$= S_0^2 e^{\mu(t_1 + t_2)} E[\exp(\sigma B_{t_1} + \sigma B_{t_2})]$$

To evaluate this we need to split B_{t_2} into two independent components:

$$B_{t_2} = B_{t_1} + (B_{t_2} - B_{t_1}) \quad \text{where} \quad B_{t_2} - B_{t_1} \sim N(0, t_2 - t_1)$$

We then get:

$$E[S_{t_1}S_{t_2}] = S_0^2 e^{\mu(t_1+t_2)} E[\exp(\sigma B_{t_1} + \sigma\{B_{t_1} + (B_{t_2} - B_{t_1})\})]$$

$$= S_0^2 e^{\mu(t_1+t_2)} E[\exp(2\sigma B_{t_1} + \sigma\{B_{t_2} - B_{t_1}\})]$$

$$= S_0^2 e^{\mu(t_1+t_2)} E[\exp(2\sigma B_{t_1})] E[\exp(\sigma\{B_{t_2} - B_{t_1}\})]$$

$$= S_0^2 e^{\mu(t_1+t_2)} \exp(2\sigma^2 t_1) \exp[½\sigma^2(t_2 - t_1)]$$

$$= S_0^2 e^{\mu(t_1+t_2)} e^{\frac{3}{2}\sigma^2 t_1 + \frac{1}{2}\sigma^2 t_2}$$

Putting these together gives:

$$Cov\left(S_{t_1}, S_{t_2}\right) = S_0^2 e^{\mu(t_1+t_2)} e^{\frac{3}{2}\sigma^2 t_1 + \frac{1}{2}\sigma^2 t_2} - S_0 e^{\mu t_1 + ½\sigma^2 t_1} . S_0 e^{\mu t_2 + ½\sigma^2 t_2}$$

$$= S_0^2 e^{\mu(t_1+t_2)} \left(e^{\frac{3}{2}\sigma^2 t_1} - e^{\frac{1}{2}\sigma^2 t_1} \right) e^{\frac{1}{2}\sigma^2 t_2}$$

[4]

The conditional expectation is:

$$E[S_{t_2} | S_{t_1}] = E[S_0 \exp(\mu t_2 + \sigma B_{t_2}) | B_{t_1}]$$

$$= S_0 e^{\mu t_2} E[\exp(\sigma\{B_{t_1} + (B_{t_2} - B_{t_1})\}) | B_{t_1}]$$

$$= S_0 e^{\mu t_1 + \mu(t_2 - t_1)} \exp(\sigma B_{t_1}) E[\exp(\sigma\{B_{t_2} - B_{t_1}\}) | B_{t_1}]$$

$$= S_{t_1} e^{\mu(t_2 - t_1)} e^{½\sigma^2(t_2 - t_1)} = S_{t_1} e^{(\mu + ½\sigma^2)(t_2 - t_1)}$$

[2]

[Total 6]

(iv) ***Martingale***

We can rearrange the last result in the form:

$$E[e^{-(\mu+½\sigma^2)t_2} S_{t_2} | S_{t_1}] = e^{-(\mu+½\sigma^2)t_1} S_{t_1}$$

which shows that, subject to the convergence criterion, the process $e^{-(\mu+½\sigma^2)t} S_t$ is a martingale with respect to $\{S_t\}$. [1]

This is an example of a 'discounted' security price process and it is a martingale. Such processes are very important in the theory of derivative pricing.

10.9 (i) ***Stochastic differential equation***

We know that:

$$X_t = \log S_t = \mu t + \sigma W_t$$

So X_t satisfies the SDE:

$$dX_t = \mu dt + \sigma dW_t$$ [1]

(ii) ***Apply Ito's Lemma***

We can now apply Ito's Lemma to find the SDE for S_t, which is a function of X_t, namely $S_t = \exp(X_t)$.

Now:

$$\frac{\partial S_t}{\partial X_t} = \exp(X_t)$$

$$\frac{\partial^2 S_t}{\partial X_t^2} = \exp(X_t)$$

and from (i) above, $dX_t = \mu dt + \sigma dW_t$, so the drift and volatility functions in the stochastic differential equation for X_t are μ and σ respectively.

Hence, Ito's Lemma gives:

$$\begin{aligned} dS_t &= [\mu \exp(X_t) + \tfrac{1}{2}\sigma^2 \exp(X_t)]\, dt + \sigma \exp(X_t)\, dW_t \\ &= S_t[(\mu + \tfrac{1}{2}\sigma^2)dt + \sigma dW_t] \end{aligned}$$

'Otherwise'

The 'otherwise' approach uses a Taylor Series expansion:

$$\begin{aligned} dS_t &= \frac{\partial S_t}{\partial X_t} dX_t + \tfrac{1}{2}\frac{\partial^2 S_t}{\partial X_t^2}(dX_t)^2 \\ &= \exp(X_t)\, dX_t + \tfrac{1}{2}\exp(X_t)\,(dX_t)^2 \\ &= \exp(X_t)[\mu dt + \sigma dW_t] + \tfrac{1}{2}\exp(X_t)[\mu dt + \sigma dW_t]^2 \\ &= \exp(X_t)[\mu dt + \sigma dW_t] + \tfrac{1}{2}\exp(X_t)[\mu^2 (dt)^2 + 2\mu\sigma\, dt\, dW_t + \sigma^2 (dW_t)^2] \end{aligned}$$

We can then disregard the second order terms, except for $(dW_t)^2$, which we replace with dt. This gives:

$$\begin{aligned} dS_t &= \exp(X_t)[\mu dt + \sigma dW_t] + \tfrac{1}{2}\exp(X_t)\sigma^2 dt \\ &= S_t[(\mu + \tfrac{1}{2}\sigma^2)dt + \sigma dW_t] \end{aligned}$$ [3]

(iii) ***Find the probability***

The probability that the share price will fall during a given year – say from time $t-1$ to time t – is:

$$P(S_t < S_{t-1}) = P\left[\frac{S_t}{S_{t-1}} < 1\right]$$ [½]

From the equation given in the question, we know that:

$$\frac{S_t}{S_{t-1}} = \frac{\exp[\mu t + \sigma W_t]}{\exp[\mu(t-1) + \sigma W_{t-1}]} = \exp\left[\mu + \sigma(W_t - W_{t-1})\right]$$ [½]

Using the fact that $W_t - W_{t-1} \sim N(0,1)$, we then find:

$$\begin{aligned}P(S_t < S_{t-1}) &= P\left[\exp\left[\mu + \sigma(W_t - W_{t-1})\right] < 1\right] \\ &= P\left[\mu + \sigma(W_t - W_{t-1}) < 0\right] \\ &= P\left[W_t - W_{t-1} < -\frac{\mu}{\sigma}\right] \\ &= \Phi\left(-\frac{0.06}{0.25}\right) = \Phi(-0.24) = 0.405\end{aligned}$$ [2]

[Total 3]

11

Stochastic models of security prices

Syllabus objectives

4.4 Stochastic models for security prices

4.4.1 Discuss the continuous-time lognormal model of security prices and the empirical evidence for or against the model.

0 Introduction

Stochastic models of security prices are used to simulate future investment returns as an important part of processes such as asset-liability modelling and pricing derivatives using the Black-Scholes framework.

Several basic types of models are available, each making different assumptions about the processes that generate investment returns. Each will, in general, yield different results. Consequently, the results obtained should be analysed with regard to both the nature of the model and the values of the model parameters employed.

This chapter focuses on one such model of security prices that is widely used in financial economics: the *continuous-time lognormal model*. This model assumes that the log of security prices follows a continuous-time random walk with drift, or equivalently, the log of security prices follows a Brownian motion with drift. The share prices themselves will therefore follow geometric Brownian motion as introduced earlier in the course.

We will look at the empirical evidence both for and against this model.

The Core Reading in this chapter is adapted from course notes written by Timothy Johnson.

1 Are stochastic processes good models for asset prices?

Econometrics is concerned with identifying relationships between economic data: is there a relationship between market indices and interest rates; are long-term bonds preferable to short-term Treasury bills? It involves inference – identifying a model – from discrete statistical data.

In mathematics it is generally easier to work with continuous objects rather than discrete objects, a good example is that the normal distribution is used more widely than the binomial model, on which it is based. Hence, modelling of complex financial entities is usually done using what are generally referred to as stochastic models.

Many econometric models are 'stochastic', in that they involve randomness, however the approach in econometrics is to see this randomness as a consequence of 'noise' or 'uncertainty'; stochastic models represent randomness as the key driving force.

The advantage of using rigorously defined mathematical objects is their definitions can be tested against intuition. For example, the Wiener Process is defined as:

(i) starting at zero

(ii) is continuous

(iii) has independent increments

(iv) which are normally distributed.

Clearly asset prices do not start at zero (they represent capital value so start at a value determined by initial capital).

Asset prices are not continuous. However fast algorithmic trading gets, trades are made at discrete points in time.

The evidence that asset prices are Markovian tend to refer to 'the long run' and over shorter timeframes this assumption appears to fail. Similarly, even if asset prices had stationary increments, it would be practically impossible to identify them, and they certainly don't exhibit normally distributed increments.

This analysis suggests that a Wiener process is not a good model for asset prices. To correct for these short-comings, Ito diffusions are used to model asset prices, principally geometric Brownian motion (GBM). GBM has lognormally distributed increments; this is an improvement on normally distributed increments in that the process does not hit zero.

However, overriding these practical limitations, stochastic calculus enables sophisticated analysis to be undertaken that can result in closed form expressions for asset prices to be deduced.

2 The continuous-time lognormal model

2.1 Definition

The *continuous-time lognormal model* is another name for geometric Brownian motion. Although the basic (non-geometric) Brownian motion model may be good at describing the movement of market indices in the short run, it is not very good in the long run for two reasons:

1. Brownian motion without a positive drift is certain to become negative eventually. Even with a positive drift, there is the possibility of negative security prices in the future, which isn't very realistic.
2. The Brownian motion model predicts that daily movements of size 10, say, would occur just as frequently when the process is at a level of 500 as when it is at a level of 5,000.

We can remove these two problems by working with logs.

The conventional continuous-time lognormal model of security prices assumes that log prices form a random walk. If S_t denotes the market price of an investment, then the model states that, for $u > t$, log returns are given by:

$$\log(S_u) - \log(S_t) \sim N\left[\eta(u-t), \sigma^2(u-t)\right]$$

where η is the *drift*, and σ is the *volatility*.

The parameter σ is also known as the *diffusion coefficient*.

Note that the η that appears in the lognormal model refers to the drift in the *log* price. This is not quite the same as the average rate of drift of the price itself, which is $\eta + \tfrac{1}{2}\sigma^2$. The corresponding stochastic differential equations for the log price and the price itself are $d(\log S_t) = \eta dt + \sigma dW_t$ and $dS_t = (\eta + \tfrac{1}{2}\sigma^2)S_t dt + \sigma S_t dW_t$.

Note that the parameterisation here is slightly different to that found in the previous chapter where we had:

$$\log(S_u) - \log(S_t) \sim N\left(\mu(u-t) - \tfrac{1}{2}\sigma^2(u-t), \sigma^2(u-t)\right) \text{ and } dS_t = \mu S_t dt + \sigma S_t dW_t$$

This simply means that the values for the drift terms are different in each case, *ie* $\eta(u-t) = \mu(u-t) - \tfrac{1}{2}\sigma^2(u-t)$. These parameters are specific to the investment considered.

By definition:

$$\log(S_u) - \log(S_t) = \log\left(\frac{S_u}{S_t}\right) \sim N\left(\eta(u-t), \sigma^2(u-t)\right) \quad \Rightarrow \quad \frac{S_u}{S_t} \sim LogN\left(\eta(u-t), \sigma^2(u-t)\right)$$

therefore the proportionate change in the share price from time t to time u (*ie* the percentage return) is lognormally distributed, and does not depend on S_t.

2.2 Properties

The lognormal model has the following properties:

- The mean and variance of the log returns are proportional to the length of the interval considered $(u-t)$, and so the standard deviation of the log returns (often taken to be a measure of volatility) increases with the square root of the interval.
- The dependence on the length of the interval means that the mean, variance and standard deviation tend to infinity as the length of the interval increases; the exceptions are the simplified cases where there is no drift in the log price $(\eta = 0)$ or where there is no volatility $(\sigma = 0)$.

In other words, the mean will tend to infinity unless the log security price doesn't drift and the standard deviation will tend to infinity unless there is no volatility.

Assuming zero volatility would be an extreme simplification because it renders the model deterministic. It means that there is nothing random about movements in the security price process, and in a no-arbitrage world the return on the security should equal the risk-free rate of interest because the security is itself risk-free.

The lognormal model has often been used to model ordinary share prices, in which case $\eta > 0$ represents the upward drift of log share prices due to growth in company profits, which is linked to other economic factors.

The fact that the mean and standard deviation tend to infinity as the length of the time period increases isn't necessarily as unrealistic as it may first appear. Since we generally expect share prices to grow over time, in the very long run it is reasonable to expect that share prices will be very large indeed. The increase in the uncertainty surrounding share prices over long future time periods is also consistent with the standard deviation tending to infinity.

Although the mean and standard deviation of a security price will tend to infinity over time under this model, the mean and standard deviation of the annual changes in the log of the security price are constant.

Question

Explain why the mean and standard deviation of the annual changes in the log of the security price are constant under this model.

Solution

The change in the log of the security price is $\log(S_u) - \log(S_t)$. This will always have the same distribution for any time period of length $u-t$.

So the expected value and standard deviation of annual changes in the log of the security price are constant.

- It is assumed that returns over non-overlapping intervals are independent of each other.

  This is because the normal variables generating the random variation in the log of the security price are assumed to be independent.

- We can write the value of the investment at time u as:

  $$S_u = S_t \exp(X_{u-t})$$

  where $X_{u-t} \sim N\left(\eta(u-t), \sigma^2(u-t)\right)$

  This format should be familiar from earlier chapters.

- Hence, we know that S_u is lognormally distributed and we can write:

  $$E[S_u] = S_t \exp\left(\eta(u-t) + \tfrac{1}{2}\sigma^2(u-t)\right)$$

  $$Var(S_u) = S_t^2 \exp\left(2\eta(u-t) + \sigma^2(u-t)\right)\left(\exp\left(\sigma^2(u-t)\right) - 1\right)$$

  The mean and variance of the lognormal distribution for a random variable X with parameters μ and σ are given on page 14 of the *Tables*:

  $$E[X] = \exp\left(\mu + \sigma^2/2\right)$$

  and $$Var(X) = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2) - 1\right)$$

2.3 Empirical tests of the lognormal model of security prices

Empirical results for testing the lognormal model are mixed.

As the model incorporates independent returns over disjoint intervals, it is impossible to use past history to deduce that prices are cheap or dear at any time. This implies weak form market efficiency, and is consistent with empirical observations that technical analysis does not lead to excess performance.

Recall from the material on the Efficient Markets Hypothesis that *technical analysis* involves using past data on asset prices to predict future price movements. If the market is weak-form efficient, there is no advantage to be gained from technical analysis, as the current share price reflects all information contained in the price history of an investment.

Before we move on to look at the evidence against the continuous-time lognormal model for share prices, it should be pointed out that this model is widely used. For example, it is the assumed underlying process for share prices that we will use to price share options later on in the course. This is largely due to the (relative) mathematical simplicity of the model.

This model also has the advantage that share prices cannot become negative. As we saw earlier:

$$S_u = S_t \exp(X_{u-t}) \quad \text{where} \quad X_{u-t} \sim N\left(\eta(u-t), \sigma^2(u-t)\right)$$

and we know that the exponential function does not take negative values.

We now consider a number of weaknesses in the lognormal model in more detail.

Volatility, σ

The most obvious weakness is that estimates of σ vary widely according to what time period is considered, and how frequently the samples are taken.

For example, the volatility has been found to be greater in recessions and periods of financial crisis. Also σ can be estimated based on daily, monthly or annual share price records. Even if the data used covers the same time period, this usually leads to different numerical estimates of σ.

We can also take some evidence from option prices. Later in this course, we will discuss the Black-Scholes formula. This formula expresses the price of an option as a function of several variables, one of which is σ. Hence, given the actual price of an option in the marketplace, together with values for all the other variables in the model (which we can actually observe), we can work backwards to derive the implied value for σ that is consistent with that observed price. In other words, the price of an option tells us implicitly what the market believes the volatility of the security price to be.

Examination of historical option prices suggests that volatility expectations fluctuate markedly over time.

Drift parameter, η

A more contentious area relates to whether the drift parameter η is constant over time. There are good theoretical reasons to suppose that η should vary over time. It is reasonable to suppose that investors will require a risk premium on equities relative to bonds.

The risk premium compensates the investor for the extra risk taken – both default risk (if compared to government bonds) and volatility of the share price.

So, if the expected return on bonds is currently high, investors will require a correspondingly higher expected return on equities in order to make it worthwhile to hold them, instead of bonds. If this is not the case, then investors will sell equities and buy bonds until the expected returns are again brought back into line.

Mean reversion

One unsettled empirical question is whether markets are *mean-reverting*, or not. A mean-reverting market is one where rises are more likely following a market fall, and falls are more likely following a rise.

Hence, if returns have recently been above the long-run average level, then we might expect them to be lower than average over the next few periods, so that average returns revert back towards their long-run trend level.

There appears to be some evidence for this, but the evidence rests heavily on the aftermath of a small number of dramatic crashes.

After a major crash, we might well expect the market to revert to its former level after sufficient time.

In the continuous-time lognormal model, we know that returns over non-overlapping intervals are independent. This implies that what happens over the next time period is unaffected by past experience, and is therefore inconsistent with evidence of mean-reversion.

Momentum effects

Furthermore, there also appears to be some evidence of momentum effects, which imply that a rise one day is more likely to be followed by another rise the next day.

For example, if returns increase, then everyone may jump on the bandwagon and drive prices even higher.

As above, this is inconsistent with the independent returns of the lognormal model.

Normality assumption

A further strand of empirical research questions the use of the normality assumption in market returns. Actual returns tend to have many more extreme events, both on the upside and downside, than is consistent with such a model. In particular, market crashes appear more often than one would expect from a normal (or lognormal) distribution. Furthermore, days with no change, or very small change, also happen more often than the normal distribution suggests.

While the lognormal model produces continuous price paths, jumps or discontinuities seem to be an important feature of real markets.

Overall, the distribution of actual market returns appears to be more peaked (relating to days of little or no change) and with fatter tails (relating to extreme upside and downside events) than is consistent with strict normality.

However, whilst a non-normal distribution can provide an improved description of the actual returns observed (in particular the greater frequency of more extreme events than would be the case under the lognormal model), the improved fit to empirical data comes at the cost of losing the tractability of working with normal (and lognormal) distributions.

Question

The shares of Abingdon Life can be modelled using a lognormal model in which $\eta = 0.104$ *pa* and $\sigma = 0.40$ *pa*. If the current share price is 2.00, derive a 95% confidence interval for the share price in one week's time, assuming that there are exactly 52 weeks in a year.

Solution

Letting S_t denote the current share price and S_u denote the share price in one week's time, we have:

$$\eta = 0.104 \quad \sigma = 0.40 \quad S_t = 2.00 \quad u - t = \tfrac{1}{52}$$

So a 95% confidence interval for $\log(S_u)$ is given by:

$$\log(S_u) = \log(S_t) + \eta(u-t) \pm 1.96 \times \sigma\sqrt{u-t}$$

Substituting in the relevant values gives:

$$\begin{aligned}\log(S_u) &= \log 2 + \frac{0.104}{52} \pm 1.96 \times 0.40\sqrt{\tfrac{1}{52}} \\ &= \log 2 + 0.002 \pm 0.1087 \\ &= (0.5864, 0.8039)\end{aligned}$$

Therefore, a 95% confidence interval for next week's share price is $(e^{0.5864}, e^{0.8039}) = (1.80, 2.23)$.

Summary

To summarise the points made in this section, the continuous-time lognormal model may be inappropriate for modelling investment returns because:

- The volatility parameter σ may not be constant over time. Estimates of volatility from past data are critically dependent on the time period chosen for the data and how often the estimate is re-parameterised.
- The drift parameter η may not be constant over time. In particular, bond yields will influence the drift.
- There is evidence in real markets of mean-reverting behaviour, which is inconsistent with the independent increments assumption.
- There is evidence in real markets of momentum effects, which is inconsistent with the independent increments assumption.
- The distribution of security returns $\log(S_u/S_t)$ has a taller peak in reality than that implied by the normal distribution. This is because there are more days of little or no movement in financial markets.
- The distribution of security returns $\log(S_u/S_t)$ has fatter tails in reality than that implied by the normal distribution. This is because there are more extreme movements in security prices.
- The sample paths of security prices are not continuous, but instead appear to jump occasionally.

2.4 Market efficiency

It is important to appreciate that many of the empirical deviations from the lognormal model do not imply market inefficiency.

If we believe in market efficiency, then we must be able to explain how an efficient market can be consistent with the evidence we've given above.

For example, periods of high and low volatility could easily arise if new information sometimes arrived in large measure and sometimes in small. Market jumps are consistent with the arrival of information in packets rather than continuously. Even mean reversion can be consistent with efficient markets. After a crash, many investors may have lost a significant proportion of their total wealth. It is not irrational for them to be more averse to the risk of losing what remains. As a result, the prospective equity risk premium could be expected to rise.

Consequently, the hypothesis of market efficiency can be difficult to disprove.

Many orthodox statistical tests are based around assumptions of normal distributions. If we reject normality, we will also have to retest various hypotheses. In particular, the evidence for time-varying mean and volatility is greatly weakened.

Chapter 11 Summary

The continuous-time lognormal model

$\ln S_u - \ln S_t \sim N\left(\eta(u-t), \sigma^2(u-t)\right)$ where S_t is the share price at time t, η is the *drift parameter*, σ is the *volatility parameter* and $u > t$.

- $E[S_u] = S_t \exp\left(\eta(u-t) + \frac{1}{2}\sigma^2(u-t)\right)$
- $Var(S_u) = S_t^2 \exp\left(2\eta(u-t) + \sigma^2(u-t)\right)\left(\exp\left(\sigma^2(u-t)\right) - 1\right)$
- The mean and variance of the log returns are proportional to the interval $u-t$.
- $\ln S_t$ has independent and stationary increments.
- $\ln S_t$ has continuous sample paths.

The continuous-time lognormal model may be inappropriate for modelling investment returns because:

- The volatility parameter σ may not be constant over time. Estimates of volatility from past data are critically dependent on the time period chosen for the data and how often the estimate is re-parameterised.
- The drift parameter η may not be constant over time. In particular, bond yields will influence the drift.
- There is evidence in real markets of mean-reverting behaviour, which is inconsistent with the independent increments assumption.
- There is evidence in real markets of momentum effects, which is inconsistent with the independent increments assumption.
- The distribution of security returns $\log\left(S_u / S_t\right)$ has a taller peak in reality than that implied by the normal distribution. This is because there are more days of little or no movement in financial markets.
- The distribution of security returns $\log\left(S_u / S_t\right)$ has fatter tails in reality than that implied by the normal distribution. This is because there are more extreme movements in security prices.
- The sample paths of security prices are not continuous, but instead appear to jump occasionally.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 11 Practice Questions

11.1 Explain what is meant by the continuous-time lognormal model of security prices.

11.2 When valuing derivatives it is often assumed that the price of the underlying security follows a geometric Brownian motion with stochastic differential equation:

$$dS_t = S_t(\mu dt + \sigma dZ_t)$$

where Z_t represents a standard Brownian motion.

List the advantages and disadvantages of this assumption.

11.3 An investor has decided to model PPB plc shares using the continuous-time lognormal model. Using historical data, the investor has estimated the annual drift and volatility parameters to be 6% and 25% respectively. PPB's current share price is $2.

Exam style

(i) Calculate the mean and variance of PPB's share price in one year's time. [3]

(ii) Calculate the probability that:

(a) PPB's shares fall in value over the next year.

(b) PPB's shares yield a return of greater than 30% over the next year.

Assume that no dividends are to be paid over the next year. [4]

[Total 7]

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 11 Solutions

11.1 The continuous-time lognormal model of security prices assumes that log prices form a random walk in continuous time.

If S_t denotes the market price of an investment at time t, then the model states that, for $u > t$, log returns are given by:

$$\log(S_u) - \log(S_t) \sim N\left(\eta(u-t), \sigma^2(u-t)\right)$$

where η is the parameter associated with the drift and σ is the parameter associated with the volatility.

The values of η and σ are specific to the investment considered.

Under the continuous-time lognormal model, the proportional change in the price is lognormally distributed, so that returns over any interval do not depend on the initial value of the investment, S_t.

The mean and variance of the log returns are proportional to the length of the interval considered $(u-t)$.

It is assumed that returns over non-overlapping intervals are independent of each other.

11.2 Geometric Brownian motion gives rise to a lognormal model for share prices.

Advantages

- This assumption makes the mathematics more tractable than other more complex models.
- Share prices cannot become negative in this model.
- The variance of the returns in a particular period is proportional to the length of that period, which seems intuitively reasonable.
- A lognormal model for share prices with independent returns in non-overlapping time intervals is consistent with the assumption that markets are efficient (at least, in weak form).
- The return and risk characteristics of the underlying share (given by μ and σ) are expressed as a proportion of the current share price, rather than in absolute terms.

Disadvantages

- In reality, share price movements may not be consistent with such a process, which generates continuous sample paths. For example, prices may 'jump', *ie* change suddenly by a significant amount.
- Also, in reality, μ and σ are not constant. For instance, the drift parameter might depend on interest rates, and studies of historical option prices suggest that investors' expectations of volatility fluctuate markedly over time.
- Historical evidence suggests that large movements in prices are more common than the lognormal distribution implied would suggest.
- Historical evidence also suggests that days of little or no movement in prices are more common than the lognormal distribution implied would suggest.
- The assumption of an efficient market may be invalid.
- There is evidence that share prices exhibit momentum effects in the short-term and mean-reversion in the long-term. This contradicts the property of independent increments.

11.3 (i) ***Mean and variance of PPB's share price in one year's time***

Based on the information given in the question:

$$\log S_1 - \log S_0 \sim N\left(\eta, \sigma^2\right) \qquad [1]$$

So, using the formulae for the mean and variance of the lognormal distribution from page 14 in the *Tables*, we have:

$$E\left[S_1 \middle| S_0\right] = S_0 \exp\left(\eta + ½\sigma^2\right)$$

$$Var\left(S_1 \middle| S_0\right) = S_0^2 \exp\left(2\eta + \sigma^2\right)\left(\exp\left(\sigma^2\right) - 1\right)$$

So, with $\eta = 0.06, \sigma = 0.25$ and $S_0 = 2$ we have:

$$E\left[S_1 \middle| S_0\right] = 2 \times \exp\left(0.06 + ½ \times 0.25^2\right) = 2.1911 \qquad [1]$$

$$Var\left(S_1 \middle| S_0\right) = 2^2 \times \exp\left(2 \times 0.06 + 0.25^2\right)\left(\exp\left(0.25^2\right) - 1\right) = 0.30963 \qquad [1]$$

[Total 3]

(ii)(a) ***Probability that PPB's shares fall in value over the next year***

Here we want the probability that $S_1 < S_0$.

$$P[S_1 < S_0] = P\left[\frac{S_1}{S_0} < 1\right]$$

$$= P\left[\log\left(\frac{S_1}{S_0}\right) < \log(1)\right]$$

$$= P\left[\frac{\log\left(\frac{S_1}{S_0}\right) - 0.06}{0.25} < \frac{0 - 0.06}{0.25}\right] \qquad [1]$$

Therefore:

$$P[S_1 < S_0] = P[Z < -0.24] = 1 - \Phi(0.24) = 1 - 0.59483 = 0.40517 \qquad [1]$$

So, there is a probability of almost 41% that the share price will fall over the next year.

(ii)(b) ***Probability that PPB's shares yield more than 30% over the next year***

Here we want the probability that $S_1 > 1.30 S_0$.

$$P[S_1 > 1.30 S_0] = P\left[\frac{S_1}{S_0} > 1.30\right]$$

$$= P\left[\log\left(\frac{S_1}{S_0}\right) > \log(1.30)\right]$$

$$= P\left[\frac{\log\left(\frac{S_1}{S_0}\right) - 0.06}{0.25} > \frac{\log(1.30) - 0.06}{0.25}\right] \qquad [1]$$

Therefore:

$$P[S_1 > 1.30 S_0] = P[Z > 0.80946] = 1 - \Phi(0.80946) = 1 - 0.79087 = 0.20913 \qquad [1]$$

So, there is an approximately 21% chance that the share will yield a return of 30% or more over the next year. [Total 4]

The relatively high probabilities in both (a) and (b) reflect the volatility parameter of 25%, which isn't unrealistic for an individual share.

12

Characteristics of derivative securities

Syllabus objectives

6.1 Option pricing and valuations

6.1.1 State what is meant by arbitrage.

6.1.2 Outline the factors that affect option prices.

6.1.3 Derive specific results for options which are not model dependent:

- Show how to value a forward contract.
- Develop upper and lower bounds for European and American call and put options.
- Explain what is meant by put-call parity.

0 Introduction

A *derivative* is a security or contract which promises to make a payment at a specified time in the future, the amount of which depends upon the behaviour of some *underlying* security up to and including the time of the payment.

The value of the derivative contract at any time is therefore *derived* from that of the underlying security – which is itself often referred to simply as the *underlying*. The underlying asset could be a share, bonds, index, interest rate, a currency, or a commodity such as gold or wheat.

Much of the remainder of this course focuses on the mathematics underlying the valuation of derivatives. After this introductory chapter, Part 3 outlines the more detailed analysis and procedures used to value derivatives and, in particular, options.

An *option* gives an investor the right, but not the obligation, to buy or sell a specified asset on a specified future date.

Options are contracts agreed between investors to trade in an underlying security at a given date at a set price. The holder of the option is not obliged to trade – hence the name 'option' – and will only do so if it is profitable. The other party, known as the 'writer', *is* obliged to trade if the holder of the option wants to. The writer of an option collects a premium from the holder for giving the holder the right to exercise (or not) the option.

There are two basic types of option:

A *call option* gives the right, but not the obligation, to buy a specified asset on a set date in the future for a specified price.

A *put option* gives the right, but not the obligation, to sell a specified asset on a set date in the future for a specified price.

From these two types of option, there are four positions that an investor could take:

1. holding (*ie* buying) a call option
2. holding (*ie* buying) a put option
3. writing (*ie* selling) a call option
4. writing (*ie* selling) a put option.

Question

Explain why buying a put is not the same as selling a call.

Solution

The difference is between *right* and *obligation*.

Buying a put costs money and allows the buyer to *choose* whether or not to sell the underlying asset. The buyer of a put option is likely to choose to sell the underlying asset if the market price *is less than* the exercise (or strike) price.

Selling a call means that the seller receives money and *must* sell the underlying asset if, and only if, the holder of the call option wants to. The writer of a call option is likely to be forced to sell the underlying asset if the market price *exceeds* the exercise (or strike) price.

Timing

Some options can be exercised only on the specified 'expiry' or 'exercise' date; others can be exercised on any working day prior to expiry. These are respectively known as 'European' and 'American' options.

A *European* option is an option that can only be exercised at expiry.

An *American* option is one that can be exercised on any date before its expiry.

Terminology

A *long* position on a contract is when the contract has been purchased, while a *short* position is when the contract is sold.

So, the buyer/holder of an option has a long position in the option, and the seller/writer of an option has a short position.

Therefore, a long position on a European option gives the holder the right but not the obligation to exercise the option. The holder of the short position will then be obliged to exercise it.

'Long' and 'short' do not only apply to options – they can be applied to any asset. So, for example, 'a long position in a share' means 'owning the share'.

Question

An investor has a short position in an American put option with an exercise price of 20p and an October expiry date.

Describe the nature of the contract that the investor has entered.

Solution

The investor has given another person the right, but not the obligation, to sell the underlying asset to the investor at a price of 20p at any time up to a date in the following October.

In return, the investor will have received a premium.

1 Arbitrage

One of the central concepts in this area of financial economics is that of *arbitrage*.

1.1 Definition

Put in simple terms, an *arbitrage opportunity* is a situation where we can make a certain profit with no risk. This is sometimes described as a *free lunch*. Put more precisely, an arbitrage opportunity means that:

(a) we can start at time 0 with a portfolio that has a net value of zero (implying that we are long in some assets and short in others).

This is usually called a zero-cost portfolio.

(b) at some future time *T*:

- **the probability of a loss is 0**
- **the probability that we make a strictly positive profit is greater than 0.**

If such an opportunity existed then we could multiply up this portfolio as much as we wanted to make as large a profit as we desired. The problem with this is that all of the active participants in the market would do the same and the market prices of the assets in the portfolio would quickly change to remove the arbitrage opportunity.

So, if we can find a strategy, or an investment portfolio, that gives an arbitrage profit, then we can simply repeat this strategy, or buy this investment portfolio, as many times as possible before asset prices change to 'close out' this opportunity for arbitrage profit.

1.2 The principle of no arbitrage

The *principle of no arbitrage* states simply that arbitrage opportunities do not exist.

This principle is vital to the pricing of derivative securities. Essentially, any two assets that behave in exactly the same way must have the same price. If this were not true, we could buy the 'cheap' one and sell the 'expensive' one as many times as we liked, making an unlimited arbitrage profit!

If we assume that there are no arbitrage opportunities in a market, then it follows that any two securities or combinations of securities that give exactly the same payments must have the same price. This is sometimes called the 'law of one price'. The ideas are demonstrated in the following examples.

Example 1

Consider a very simple securities market, consisting of two securities, A and B. At time $t=0$ the prices of the securities are P_0^A and P_0^B respectively. The term of both the securities is 1 year. At $t=1$ there are two possible outcomes. Either the 'market' goes up, in which case security A pays $P_1^A(u)$ and B pays $P_1^B(u)$, or it goes down, with payments $P_1^A(d)$ and $P_1^B(d)$ respectively.

Investors can buy securities, in which case they pay the time 0 price and receive the time 1 income, or they can sell securities, in which case they receive the time 0 price and must pay the time 1 outgo.

Now, assume that we have the following payment table:

Security	**Time 0 price** P_0 **£**	**Market goes up** $P_1(u)$ **£**	**Market goes down** $P_1(d)$ **£**
A	**6**	**7**	**5**
B	**11**	**14**	**10**

There is an arbitrage opportunity here. An investor could buy one unit of security B and sell two units of security A. This would give income at time 0 of £12 from the sale of security A and an outgo of £11 from the purchase of security B – which gives a net income at time 0 of £1. At time 1 the outgo due on the portfolio of 2 units of security A exactly matches the income due from security B, whether the market moves up or down. Thus, the investor makes a profit at time 0, with no risk of a future loss.

It is clear that investment A is unattractive compared with investment B. This will cause pressure to reduce the price of A and to increase the price of B, as there will be no demand for A and an excessive demand for B. Ultimately we would achieve balance, when $P_0^A = P_0^B / 2$**, when the arbitrage opportunity is eliminated, and the prices are consistent.**

The situation where $P_0^A = P_0^B / 2$ satisfies the 'law of one price' introduced earlier.

Example 2

Now, consider the following table:

Security	**Time 0 price** P_0 **£**	**Market goes up** $P_1(u)$ **£**	**Market goes down** $P_1(d)$ **£**
A	**6**	**7**	**5**
B	**6**	**7**	**4**

An arbitrage opportunity exists, as an investor could buy one unit of A and sell one unit of B. The net income at time 0 is £0, as the income from the sale of B matches the outgo on the purchase of A. At time 1 the net income is £0 if the market goes up, and £1 if the market goes down. So, for a zero investment, the investor has a possibility of making a profit (assuming the probability that the market goes down is not zero) and no possibility of making a loss.

With these prices, investors will naturally choose to buy investment A and will want to sell investment B. This will put pressure on the price of A to increase, and on the price of B to decrease. The arbitrage opportunity is eliminated when $P_0^A > P_0^B$**.**

From this point on, when considering derivative pricing, we will work with the assumption that no arbitrage opportunities exist.

2 Preliminary concepts

2.1 Notation

We will now consider options in more detail, starting by introducing the notation we will need:

t **is the current time**

S_t **is the underlying share price at time t**

K **is the strike or exercise price**

T **is the option expiry date**

c_t **is the price at time t of a European call option**

p_t **is the price at time t of a European put option**

C_t **is the price at time t of an American call option**

P_t **is the price at time t of an American put option**

r **is the risk-free continuously compounding rate of interest (assumed constant)**

Alternatively, we could describe r is the risk-free force of interest.

2.2 European call options

As we have already seen, the buyer of a European call option has the right, but not the obligation, to buy the share from the seller of the option on a set date, usually referred to as the exercise date (or expiry date). Conversely, the seller or *writer* of the option, who has no choice, is obliged to deliver the share should the holder of the option exercise the option.

If at time T the share price S_T is less than the strike price, then the holder would lose money by exercising the option to buy. So, in fact, the option will not be exercised if $S_T < K$. This is because the holder of the option would be silly to pay K for the share when it could be purchased for S_T ($< K$) in the open market. In this case, the holder of the option simply walks away and 'loses' the premium originally paid for the option itself. Conversely, the writer of the option gets to keep the premium received at the outset and therefore makes an overall profit on the deal equal in value to that premium.

If $S_T > K$, then the holder can buy the share at the strike price and sell immediately at the market price, S_T, receiving a positive cashflow at expiry of $S_T - K$.

Here the option should be exercised because it generates a positive cashflow for the holder.

Question

Determine the cashflow at expiry for the writer of the call option in this latter case.

Solution

The writer of the option experiences a cashflow of $K - S_T < 0$. This is because the writer is obliged to sell the share at K when it is in fact worth $S_T > K$.

The holder's net cashflow at the expiry date is called the 'payoff'. The payoff on a European call option is therefore:

$$f(S_T) = \max\{S_T - K, 0\} \text{ at time } T.$$

To purchase a traded option the buyer must pay the seller an option *premium* when the contract is made.

As always, the writer of the option makes a profit or loss that is equal and opposite to that of the buyer, so that the total profit or loss to the two parties sums to zero.

The writer of the option keeps the premium regardless of whether or not the option is ultimately exercised. It is paid by the buyer to the writer in order to enjoy the *choice* conferred by holding the option. Remember that the writer has no such choice, but must trade if the buyer wishes to do so. Later in this chapter, we will discuss the factors that influence the value of the option, *ie* how much the premium should be.

A call option is described as:

- *in-the-money* if the current price, S_t, is greater than the strike price, K
- *out-of-the-money* if $S_t < K$
- *at-the-money* if $S_t = K$.

Hence, an in-the-money call option is one that would result in an immediate profit – ignoring the premium originally paid – if it could be exercised now, whereas an out-of-the-money option would produce an immediate loss. Recall that if the option is European then it can be exercised only at the expiry date.

Example

Suppose that the current price of Share X is 115 and that a European call option is available on Share X with an exercise price of 110.

The option is currently *in-the-money*, as the share price exceeds the exercise price.

If the share price remained unchanged until the exercise date, then the holder of the option would exercise it, to generate a positive cashflow of 5.

If instead the share price fell to 105 at the exercise date, the option would not be exercised, as the share would then be worth less than the exercise price, and the option holder would be able to buy the share more cheaply on the open market.

2.3 European put options

The holder of a European put option has the right to *sell* one share at time T to the writer at the strike price K.

Again the choice, in this case to sell the share on the specified exercise date, lies with the buyer of the option, who is not obliged to sell. As before, the writer of the put is obliged to buy the share from the option holder should the latter exercise the option to sell.

The option will only be exercised if $S_T < K$ and the payoff is, therefore:

$$f(S_T) = \max\{K - S_T, 0\} \text{ at time } T.$$

Question

Explain why a European put option is exercised only if $S_T < K$.

Solution

A European put option should be exercised if $S_T < K$ because the option holder would then be able to sell the share for K when it is worth only S_T. The holder's cashflow at expiry would be $K - S_T > 0$ in this case, and as this is positive, it is worthwhile exercising the option.

Conversely, if $S_T > K$, the holder would be silly to sell the share to the writer of the put option for K, when it could instead be sold for S_T in the open market.

A put option is described as:

- *in-the-money* if $S_t < K$
- *out-of-the-money* if $S_t > K$
- *at-the-money* if $S_t = K$.

As before, an in-the-money option is one that would result in an immediate profit – ignoring the premium originally paid – if it could be exercised now, whereas an out-of-the-money option would produce an immediate loss.

2.4 American options

Recall that the only difference between an American and a European option is that with an American option the holder can exercise the option before the expiry date, not just *on* the expiry date, as is the case for a European option. As always, the writer of the option is obliged to trade should the holder wish to do so.

Note that the names *European option* and *American option* have arisen for historical reasons. There is no longer any direct link with the place where the contracts are traded.

2.5 Other terminology

Intrinsic value

The *intrinsic value* of a derivative is the value assuming expiry of the contract immediately rather than at some time in the future. For a call option, for example, the intrinsic value at time *t* is simply:

$$\max\{S_t - K, 0\}$$

Question

Write down the intrinsic value of a put option at time *t*.

Solution

The intrinsic value of a put option at time *t* is either:

- $K - S_t$ if the exercise price exceeds the share price – in which case it will be exercised, or
- 0 if the exercise price is less than the share price – in which case it will not be exercised.

So, overall, the intrinsic value of a put option is $\max\{K - S_t, 0\}$.

Thus, the intrinsic value of an option is:

- positive, if it is in-the-money
- equal to zero, if it is at-the-money
- equal to zero, if it is out-of-the-money.

Time value

The *time value* or *option value* of a derivative is defined as the excess (or the shortfall) of an option's current price over its intrinsic value.

It primarily represents the value of the *choice* that the option provides to its holder. The more valuable the choice is to the holder, perhaps because of the greater uncertainty there is about future share price movements for example, the greater is the time value.

3 Factors affecting option prices

3.1 Introduction

A number of mathematical models are used to value options. One of the more widely used is the *Black-Scholes* model. This uses five parameters to value an option on a non-dividend-paying share. The five parameters are:

- the underlying share price, S_t
- the strike price, K
- the time to expiry, $T-t$
- the volatility of the underlying share, σ
- the risk-free interest rate, r.

In the case of a dividend-paying share, we can consider dividends to be a sixth factor.

Here the price of an option means the size of the option premium paid at the outset.

3.2 Underlying share price

The first parameter is the initial price of the underlying share at time t, S_t.

The effect of the price of the underlying share on a typical call option is shown in Figure 12.1.

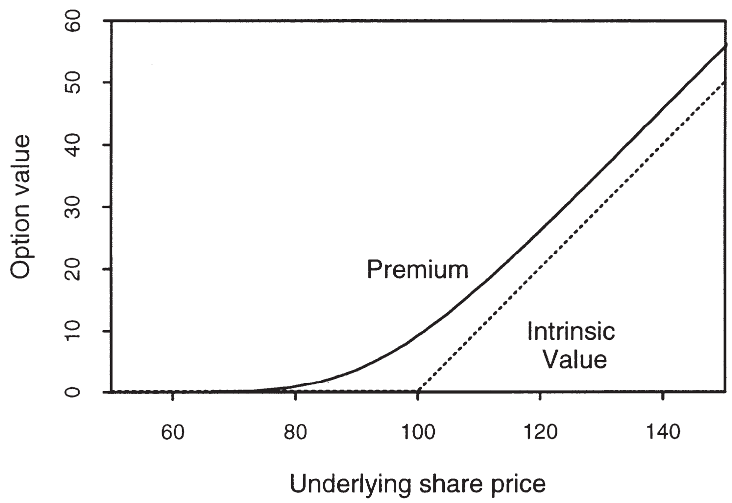


Figure 12.1: Call option (with strike price 100) premium and intrinsic value as a function of the current share price, S_t

The dotted line in the graph represents the intrinsic value. The time value is therefore the vertical distance between the actual price, *ie* premium, and the intrinsic value. For a call option on a non-dividend-paying share this is always positive.

Note that the price for a call option is always greater than the intrinsic value. This follows on from the lower bound derived below, namely that:

$$c_t \geq S_t - Ke^{-r(T-t)} > S_t - K$$

ie $$c_t \geq \max\left[S_t - Ke^{-r(T-t)}, 0\right] \geq \max\left[S_t - K, 0\right] = \text{intrinsic value}$$

The share price affects the option price (or premium) differently for call options and put options.

Question

Suppose that the price of Share X is 112 and that a put option on Share X with an exercise price of 110 is currently priced at 5.

Calculate the intrinsic value and time value of the option.

Solution

If $S_t = 112$, $K = 110$ and $p_t = 5$, then:

- Intrinsic value $=$ $\max\left[K - S_t, 0\right]$ $=$ $\max\left[110 - 112, 0\right]$ $= 0$
- Time value $=$ total value $-$ intrinsic value $=$ $5 - 0$ $=$ 5

Call option

In the case of a call option, a higher share price means a higher intrinsic value (or, where the intrinsic value is currently zero, a greater chance that the option is *in-the-money* at maturity). A higher intrinsic value means a higher premium.

Put option

For a put option, a higher share price will mean a lower intrinsic value and a lower premium.

In each case, the change in the value of the option will not match precisely the change in the intrinsic value because of the later timing of the option payoff.

So, the time value is not constant with respect to share price. The graph shows that the time value is greatest when there is more uncertainty in the outcome. When the share price is well above or below the exercise price, the question of whether exercise will take place is more certain, so the time value is smaller. The uncertainty is greatest when the share price is around the exercise price.

3.3 Strike price

In the case of a call option, a higher strike price means a lower intrinsic value. A lower intrinsic value means a lower premium. For a put option, a higher strike price will mean a higher intrinsic value and a higher premium. In each case the change in the value of the option will not match precisely the change in the intrinsic value because of the later timing of the option payoff.

3.4 Time to expiry, $T - t$

The longer the time to expiry, the greater the chance that the underlying share price can move significantly in favour of the holder of the option before expiry. So the value of an option will increase with term to maturity. This increase is moderated slightly by the change in the time value of money.

Question

We can also note that the longer the time to expiry, the greater the chance that the underlying share price can move *against* the holder of the option before expiry.

Justify, therefore, why the value of a call option *increases* with term to maturity.

Solution

Consider the holder of a call option. Although a longer time to expiry does mean that there is a greater chance that the share price can fall a long way, this does not fully offset the value of the chance of greater profits if the price goes up. This is because once the share price goes below the exercise price K, the option will not be exercised and the holder simply loses the premium paid at outset.

Now, regardless of how far the share price goes down, the maximum loss that the holder can make in this instance is the premium paid at the outset, whereas the maximum profit if the price goes up is unlimited. Thus, the holder of the option faces an asymmetric risk – *ie* lots of upside profit potential and limited downside risk. Consequently, the more chance there is of the share price moving a long way, the more valuable the option.

Whereas this is true for *all* the simple call options we consider in this course, it may not be true for a deeply in-the-money European put option. This is because the *guaranteed* gain from being able to exercise the option and receive money sooner could be worth more than the *possible* gain from the share price moving in your favour over a longer period to expiry.

3.5 Volatility of the underlying share

Within this context, volatility refers to the general level of variability in the market price of the underlying share.

The higher the volatility of the underlying share, the greater the chance that the underlying share price can move significantly in favour of the holder of the option before expiry. So the value of an option will increase with the volatility of the underlying share.

The argument here is the same as for the time to expiry. Note that the holder of an option will therefore like volatility or risk, whereas with most other assets we dislike risk and consequently place a lower value on a riskier asset.

3.6 Interest rates

An increase in the risk-free rate of interest will result in a higher value for a call option because the money saved by purchasing the option rather than the underlying share can be invested at this higher rate of interest, thus increasing the value of the option.

Buying a call option and later exercising it can be broadly compared with buying the share directly. By using the call option the buyer is deferring the payment of the bulk of the purchase price and can earn extra interest during the period until exercising the option (although the buyer will then miss out on any dividends).

For a put option, higher interest means a lower value.

This is because put options can be purchased as a way of deferring the sale of a share. Comparing this strategy with an immediate sale of the share, we see that the investor's money is tied up in the share for longer, and so is not benefiting from the higher interest rate.

The basic Black-Scholes model can be adapted to allow for a sixth factor determining the value of an option:

3.7 Income received on the underlying security

In many cases the underlying security might provide a flow of, say, dividend income. Normally such income is not passed onto the holder of an option. Then the higher the level of income received, the lower is the value of a call option, because by buying the option instead of the underlying share the investor foregoes this income. The reverse is true for a put.

Question

Consider an American put option on a non-dividend-paying share.

List the five factors that determine the price of this option and, for each factor, state whether an increase in its value produces an increase or a decrease in the value of the option.

Solution

The five factors and the effect of an increase in each of them on an American put option are:

Factor	*Value of American put option will ...*
share price	decrease
exercise price	increase
time to expiry	increase
volatility of share price	increase
risk-free rate of interest	decrease

3.8 The Greeks and risk management

So far, we have discussed the factors that affect the price of a derivative contract and suggested what happens to the price when you increase or decrease each factor. In addition, we can try to *quantify* what the change in the price will be in response to a change in one of the factors, while holding the others constant. In other words, we can calculate the (partial) derivative of the derivative price with respect to each of the above factors. These (mathematical) derivatives are given Greek letters and are collectively known as 'the Greeks'. The Greeks are covered in more detail in a later chapter.

In the same way that we can consider the price of a single derivative contract, we can also work out the effect of the various factors on a whole portfolio of shares and derivatives – we simply add up the Greeks for each constituent. These Greeks can then tell us about the exposure to risk of our portfolio, *eg* what will happen to the value of our portfolio if share prices fall by 10%, or the risk-free interest rate increases by 1%, and so on. The Greeks can therefore be used to manage risk.

4 Pricing forward contracts

4.1 Introduction

We now consider a different type of derivative to a share option – the forward contract.

A forward contract is the most simple form of derivative contract. It is also the most simple to price in the sense that the forward price can be established without reference to a model for the underlying share price.

The 'forward price' is the price the holder agrees to pay for the asset at the expiry date. So it corresponds to the exercise price for an option.

When pricing call options and put options later in the course using the Black-Scholes framework, we will assume that the underlying share price follows geometric Brownian motion. In order to calculate a forward price, however, no such assumption is needed.

A *forward contract* is an agreement made at some time $t = 0$, say, between two parties under which one agrees to buy from the other a specified amount of an asset (denoted *S*) at a specified price on a specified future date. The investor agreeing to sell the asset is said to hold a *short forward position* in the asset, and the buyer is said to hold a *long forward position*.

The underlying asset for a forward contract could be one of several types, *eg* a commodity (such as gold) or a financial security (such as a share). Here we will consider primarily shares.

Suppose that, besides the underlying share, we can invest in a cash account that earns interest at the continuously compounding rate of *r* per annum.

The forward price *K* should be set at a level such that the value of the contract at time 0 is zero (that is, no money changes hands at time 0).

In other words, there is no initial premium to pay when entering a forward contract.

Example

Investor A agrees to sell 1,500 of Company X's shares in six months' time to Investor B at a price of £1.40 per share. So, £1.40 is the forward price of one share.

Suppose that six months later Company X's share price is £1.35. Investor A then has to sell 1,500 shares, with a current market value of £2,025, for £2,100, and so makes a profit from the forward contract of £75. Similarly, Investor B has to buy the shares for a loss of £75.

4.2 Calculating the forward price for a security with no income

We now move on to determining the fair price for a forward contract.

We start off with the case where the underlying asset pays no income over the duration of the forward contract, and we demonstrate two proofs of the formula for the forward price in this case.

Proposition

The fair or economic forward price is $K = S_0 e^{rT}$.

Proof (a)

Before showing the actual proof, we will first demonstrate that when the forward is priced correctly, everything works well.

Suppose, first, that we have set the forward price at $K = S_0 e^{rT}$.

We also suppose that we take a short forward position, *ie* agree to sell the asset at time T.

At the same time we can borrow an amount S_0 in cash (subject to interest at rate *r*) and buy one share. The net cost at time 0 is then zero.

The forward contract costs nothing and the share costs S_0, which we have borrowed.

At time *T* we will have:

- **one share worth S_T on the open market**
- **a cash debt of $S_0 e^{rT}$**
- **a contract to sell the share at the forward price K.**

Therefore we hand over the one share to the holder of the forward contract and receive *K*. At the same time we repay the loan: an amount $S_0 e^{rT}$.

Since $K = S_0 e^{rT}$ we have made a profit of exactly 0. There is no chance of losing money on this transaction, nor is there any chance of making a positive profit. It is a risk-free trading strategy.

So $K = S_0 e^{rT}$ appears to be the correct forward price. The actual proof is one of contradiction. We will follow the same procedure as above in the case where $K \neq S_0 e^{rT}$ and demonstrate that this would allow an arbitrage profit.

Now suppose instead that $K > S_0 e^{rT}$. We can issue one forward contract and, at the same time, borrow an amount S_0 in cash (subject to interest at rate *r*) and buy one share. The net cost at time 0 is zero.

At time *T* we will have:

- **one share worth S_T on the open market**
- **a cash debt of $S_0 e^{rT}$**
- **a contract to sell the share at the forward price K.**

Therefore we hand over the one share to the holder of the forward contract and receive *K*. At the same time we repay the loan: an amount $S_0 e^{rT}$.

Since $K > S_0e^{rT}$ we have made a guaranteed profit, having made no outlay at time 0.

This is an example of arbitrage: that is, for a net outlay of zero at time 0 we have a probability of 0 of losing money and a strictly positive probability (in this case equal to 1) of making a profit greater than zero.

Instead of issuing one contract at this price, why not issue lots of them and we will make a fortune?

In practice a flood of sellers would come in immediately, pushing down the forward price to something less than or equal to S_0e^{rT}. In other words, the arbitrage possibility could exist briefly, but it would disappear very quickly before any substantial arbitrage profits could be made.

Having shown that $K > S_0e^{rT}$ leads to arbitrage opportunities, we complete the proof by contradiction by showing that $K < S_0e^{rT}$ also leads to arbitrage opportunities.

Now suppose that $K < S_0e^{rT}$.

We follow the same principles. At time 0:

- **buy one forward contract**
- **sell one share at a price S_0**
- **invest an amount S_0 in cash.**

The net value at time 0 is zero.

Above, 'buy one forward contract' means the same as 'take a long position in the forward contract, *ie* agree to buy the asset'.

At time *T*:

- **we have cash of S_0e^{rT}**
- **we pay K $\left(K < S_0e^{rT}\right)$ for one share, after which our net holding of shares is zero**
- **the shareholding has zero value and we have $S_0e^{rT} - K > 0$ cash.**

Again this is an example of arbitrage, meaning that we should not, in practice, find that $K < S_0e^{rT}$.

Since both $K > S_0e^{rT}$ and $K < S_0e^{rT}$ lead to arbitrage opportunities the only possibility for the fair price is $K = S_0e^{rT}$.

Proof (b)

Let K be the forward price. Now compare the setting up of the following portfolios at time 0:

- **A: one long forward contract**
- **B: borrow Ke^{-rT} cash and buy one share at S_0.**

If we hold both of these portfolios up to time T then both have a value of $S_T - K$ at T.

By the principle of no arbitrage, these portfolios must have the same value at all times before T. In particular, at time 0, portfolio B has value $S_0 - Ke^{-rT}$ which must equal the value of the forward contract.

This can only be zero (the value of the forward contract at $t = 0$) if:

$$K = S_0 e^{rT}$$

Question

A three-month forward contract exists on a zero-coupon corporate bond with a current price per £100 nominal of £42.60. The yield available on three-month government securities is 6% *pa* effective.

Calculate the forward price.

Solution

Being careful to note that the risk-free yield given is quoted as an annual effective rate, rather than a force:

$$K = 42.6e^{\frac{3}{12}r} = 42.6 \times 1.06^{\frac{3}{12}} = £43.23$$

4.3 Calculating the forward price for a security with fixed cash income

We now consider more complicated cases where the underlying asset pays income during the term of the forward contract.

In this section and the following one, the derivation of the formula for the forward price follows reasoning similar to Proof (b) above, although the methodology of Proof (a) would also achieve the same answer.

Assume now that at some time t_1, $0 \le t_1 < T$, the security underlying the forward contract provides a fixed amount c to the holder. For example, if the security is a government bond, there will be fixed coupon payments due every six months.

Now consider the following two portfolios:

Portfolio A: **Enter a forward contract to buy one unit of an asset *S*, with forward price *K*, maturing at time *T*; simultaneously invest an amount $Ke^{-rT} + ce^{-rt_1}$ in the risk-free investment.**

We are now assuming that the risk-free force of interest is constant throughout the term of the contract so that r is appropriate for investments of term T and t_1. If this was not the case, we would simply use two different forces of interest corresponding to the appropriate time periods.

Portfolio B: **Buy one unit of the asset, at the current price S_0. At time t_1 invest the income of *c* in the risk-free investment.**

At time $t = 0$ the price of Portfolio A is $Ke^{-rT} + ce^{-rt_1}$ for the risk-free investment, and zero for the forward contract.

The price of Portfolio B is S_0.

At time $t = T$ the payout from Portfolio A is: income of $K + ce^{r(T-t_1)}$ from the risk-free investment; outgo of *K* on the forward contract. Receive 1 unit of asset, value S_T. The net portfolio at *T* is one unit of the asset *S* plus $ce^{r(T-t_1)}$ units of the risk-free security.

The payout from Portfolio B is one unit of the asset, value S_T, plus $ce^{r(T-t_1)}$ units of the risk-free security, from the invested coupon payment.

The net cashflows of Portfolio A at time *T* are identical to those of Portfolio B – both give a net portfolio of one unit of the underlying asset *S* plus $ce^{r(T-t_1)}$ units of the risk-free security. Using the no arbitrage assumption the prices must also be the same – that is:

$$Ke^{-rT} + ce^{-rt_1} = S_0 \quad \Rightarrow \quad K = S_0e^{rT} - ce^{r(T-t_1)}$$

So the forward price is equal to the accumulated value at time *T* of the current price, less the accumulated value at time *T* of the income payment, which will not be received by the buyer of the asset under the forward contract.

Question

A fixed-interest security pays coupons of 8% *pa* half-yearly in arrear and is redeemable at 110%. Two months before the next coupon is due, an investor negotiates a forward contract to buy £60,000 nominal of the security in six months' time. The current price of the security is £80.40 per £100 nominal and the risk-free force of interest is 5% *pa*.

Calculate the forward price.

Solution

Only one coupon will be received during the 6-month term of the forward contract. This will be received in 2 months' time and will be for amount:

$$0.5 \times 0.08 \times 60,000 = £2,400$$

So, using the formula derived above, the forward price is equal to:

$$K = 80.4 \times \frac{60{,}000}{100} e^{0.05 \times 6/12} - 2{,}400 e^{0.05 \times (6/12 - 2/12)} = £47{,}021$$

For a long forward contract on a fixed-interest security there may be more than one coupon payment. It is easy to adapt the above method to allow for this. If we let *I* denote the present value at time $t = 0$ of the fixed income payments due during the term of the forward contract, then the forward price at time $t = 0$ per unit of security *S* is

$$K = (S_0 - I)e^{rT}$$

The word 'long' here is being used to say that the term is long, rather than being a reference to a 'long forward position' (*ie* agreeing to purchase the asset under the forward contract).

So, in this case, the forward price is equal to the current price, less the present value at time 0 of the income payments during the term of the contract, accumulated to time T. This formula appears on page 45 of the *Tables*, where the forward price is denoted by F rather than K.

Question

Consider the scenario in the previous question.

On the same day, a different investor negotiates a forward contract to purchase £50,000 nominal of the security in ten months' time.

Calculate the forward price of this contract.

Solution

Two coupons will be received during the 10-month term of this forward contract: one in 2 months' time and another in 8 months' time. The amount of each coupon will be:

$$0.5 \times 0.08 \times 50{,}000 = £2{,}000.$$

So, using the formula given above:

$$K = \left(80.4 \times \frac{50{,}000}{100} - 2{,}000\left(e^{-0.05 \times 2/12} + e^{-0.05 \times 8/12}\right)\right)e^{0.05 \times 10/12} = £37{,}826$$

As well as being used to calculate the forward price for contracts based on fixed-interest securities (which have known coupon payments), the formula above can also be used to calculate the forward price for contracts based on shares where the dividend payments to be received during the term of the contract are known in monetary terms. In this case, I would be equal to the present value of the dividends received during the term of the contract.

4.4 Calculating the forward price for a security with known dividend yield

We will now consider shares where the dividend income is specified in terms of a dividend yield, rather than in monetary terms. The dividend yield for an equity is defined to be:

$$\text{Dividend yield} = \frac{\text{Dividend per share}}{\text{Price per share}}$$

Let D be the known dividend yield per annum. We assume that dividends are received continuously, and are immediately reinvested in the security of S.

This has the effect that the number of shares will increase by a constant force, *D*. This rate of increase will not be affected by changes in the share price. If the share price increases then the value of dividends will increase by the same proportion.

If we start with one unit of the security at time $t = 0$, the accumulated holding at time T would be e^{DT} units of the security. This is because the number of units owned is continuously compounding at rate D per annum for T years. If instead of 1 unit at time $t = 0$ we hold e^{-DT} units, reinvesting the dividend income, at time T we would hold $e^{DT}e^{-DT} = 1$ unit of the security.

Now consider the following two portfolios:

Portfolio A: Enter a forward contract to buy one unit of an asset S, with forward price K, maturing at time T; simultaneously invest an amount Ke^{-rT} in the risk-free investment.

Portfolio B: Buy e^{-DT} units of the asset S, at the current price S_0. Reinvest dividend income in the security S immediately it is received.

At time $t = 0$ the price of Portfolio A is Ke^{-rT} for the risk-free investment, and zero for the forward contract.

The price of Portfolio B is $e^{-DT}S_0$.

At time $t = T$ the cashflows of Portfolio A are: an amount K is received from the risk-free investment. Outgo K is paid on the forward contract. Receive 1 unit of asset S. The net portfolio at T is one unit of the asset S.

The payout from Portfolio B is one unit of the asset S.

The net cashflows of Portfolio A at time T are identical to those of Portfolio B – both give a net portfolio of one unit of the underlying asset S. Using the no arbitrage assumption the prices must also be the same – that is:

$$Ke^{-rT} = S_0e^{-DT} \quad \Rightarrow \quad K = S_0e^{(r-D)T}$$

This formula also appears on page 45 of the *Tables*, where the forward price is denoted by F rather than K, and the continuously compounded dividend yield is denoted by q rather than D.

Question

The current price of a share is £200. The share pays dividends continuously to provide a fixed dividend yield, and the current dividend is £5 *pa*.

Calculate the forward price of a five-year contract on one share if the risk-free force of interest is 5% *pa*.

Solution

The dividend yield is $\frac{5}{200} = 2.5\%$. The forward price is therefore:

$$K = 200e^{(0.05-0.025)\times 5} = £226.63$$

It is simple to adjust the portfolios to get the forward price if the dividends are paid discretely.

The important principle for this case and the known income case is that, if the income is proportional to the underlying security, *S*, we assume the income is reinvested in the security. If the income is a fixed amount regardless of the price of the security at the payment date, then we assume it is invested in the risk-free security.

This is because when the payment is proportional to the stock price (*eg* dividends) we know how many units of stock they will purchase, but we do not know how much cash is paid (as the stock price is unknown). So we can predict the amount of stock held at the end if we assume reinvestment in the stock.

With a cash payment on the other hand, we would not know how much stock could be bought, but we do know how much the cash would accumulate to at the risk-free force of interest. Assuming dividends are reinvested in the security, but cash is invested at the risk-free (and known) force of interest enables us to predict the final portfolio without requiring any information about the price of the asset *S* during the course of the contract.

5 Bounds for option prices

In this section we deduce upper and lower limits for call and put option prices based on general reasoning. Remember that throughout this chapter we are considering options based on non-dividend-paying shares.

In this section, we assume that the current time is t and each option considered expires at time T.

5.1 Lower bounds on option prices

European calls

Consider a portfolio, A, consisting of a European call on a non-dividend-paying share and a sum of money equal to $Ke^{-r(T-t)}$.

At time *T*, portfolio *A* has a value which is equal to the value of the underlying share, provided the share price S_T is greater than *K*.

This is because the sum of money will grow with interest to be worth exactly *K* at time *T* (since $Ke^{-r(T-t)}$ is the present value at time *t* of a payment of *K* at time *T*). If $S_T > K$, the call option will be exercised, using the accumulated amount of money, *K*, to purchase the share (leaving zero cash). The payoff is thus S_T at time *T*.

If S_T is less than *K* then the payoff from portfolio A is greater than that from the share.

In this instance, the payoff from portfolio A is *K* – the accumulated amount of cash at the exercise date – because the option would not be exercised, since the share is only worth $S_T(<K)$.

Since the option plus cash produces a payoff that is at least as great as that from the share, it must have a value greater than or equal to S_t.

This follows from the no arbitrage assumption.

This gives us a lower bound for c_t:

$$c_t + Ke^{-r(T-t)} \geq S_t$$

ie $$c_t \geq S_t - Ke^{-r(T-t)}$$

Question

Suppose that the exercise price of a 3-month European call option on Share X is 100 and the continuously compounded risk-free rate of return is 12% *pa*.

Calculate the lower bound for the option's price if the current price of Share X is:

(i) 115

(ii) 125.

Solution

(i) In this case, the lower bound for the European call option price is:

$$c_t \geq 115 - 100\, e^{-0.12 \times 0.25}$$

$$c_t \geq 17.96$$

(ii) If instead the share price were 10 higher at 125, then the lower bound would be:

$$c_t \geq 125 - 100\, e^{-0.12 \times 0.25}$$

$$c_t \geq 27.96$$

ie it would also be 10 higher.

European puts

A similar argument can be used for put options: portfolio B contains a European put option and a share. Compare this with the alternative of cash, currently worth $Ke^{-r(T-t)}$. At time T, portfolio B will be worth at least as much as the cash alternative.

At time T the cash will be worth K.

Portfolio B (the share plus the put option) will be worth:

- K if $S_T < K$ (because the option will be exercised by selling the share, leaving K)
- S_T if $S_T > K$ (because the option will not be exercised).

Thus portfolio B is always worth at least as much as the cash deposit at time T.

Thus:

$$\boldsymbol{p_t + S_t \geq Ke^{-r(T-t)}}$$

$$\Rightarrow \quad \boldsymbol{p_t \geq Ke^{-r(T-t)} - S_t}$$

If the lower bound derived above does not hold, *ie* if:

$$p_t < Ke^{-r(T-t)} - S_t$$

then an arbitrage opportunity exists.

An investor could borrow an amount $p_t + S_t$ in the cash market to purchase the share and put option, thus creating a portfolio of a share, the put option and a cash loan. At expiry, this portfolio is worth:

$$\begin{aligned}\max\{K - S_T, 0\} + S_T - (p_t + S_t)e^{r(T-t)} &= \max\{K, S_T\} - (p_t + S_t)e^{r(T-t)} \\ &> \max\{K, S_T\} - Ke^{-r(T-t)}e^{r(T-t)} \\ &= \max\{0, S_T - K\} \\ &\geq 0\end{aligned}$$

Overall, for zero initial outlay, the investor has made a strictly positive profit, *ie* the investor has taken advantage of an arbitrage opportunity.

Question

Calculate the lower bound for a 3-month European put option on Share X if the current share price is 95, the exercise price is 100 and the continuously compounded risk-free rate is 12% *pa*.

Solution

In this case, the lower bound for the European put option on Share X is:

$$p_t \geq 100\, e^{-0.12 \times 0.25} - 95$$

$$p_t \geq 2.04$$

American calls

Recall that, unlike its European counterpart, an American option can be exercised at any date up to and including the expiry date.

A surprising result, however, is that it is never optimal to exercise an American call on a non-dividend-paying share early (*ie* before its expiry date). Hence the above relationship for European calls also holds for American calls, *ie*:

$$C_t \geq S_t - Ke^{-r(T-t)}$$

The reason for this surprising result is as follows:

If we exercise an American call option early then we receive $S_t - K$. Since an American option provides the same rights as a European option (*ie* exercise at expiry) and more besides (*ie* exercise before expiry), the value of the American call option must be at least that of the equivalent European call option (*ie* $C_t \geq c_t$). However, we already know that the European call option is worth more than its intrinsic value:

$$c_t \geq S_t - Ke^{-r(T-t)} > S_t - K$$

So:

$$C_t > S_t - K$$

and because we always receive more by selling the American call option than we do by exercising it, the option to exercise early is effectively worthless.

It is important to realise that this surprising result is only true for a *non-dividend paying share* (or a share with no dividends between time *t* and expiry). The crucial point is that if there is a dividend between time *t* and expiry then it may be beneficial to exercise the option early in order to receive this dividend.

American puts

The lower bound for an American put option can be increased above that derived above for a European put option. Since early exercise is always possible, we have:

$$\boldsymbol{P_t \geq K - S_t}$$

So, the intrinsic value is a lower bound for the price. This condition, which is stronger than for European puts, holds because early exercise may be sensible for an American put in order to receive the exercise money earlier. By receiving the cash before the last possible expiry date, the holder of the option then benefits from receiving interest on that cash for the remaining term.

Summary

Finally note that if any of the lower bounds given above are negative we can get a tighter bound from the fact that any option has a non-negative value to the holder – ignoring the premium already paid. This means that the following bounds can be given:

$$c_t \geq \max\left\{S_t - Ke^{-r(T-t)}, 0\right\}$$

$$C_t \geq \max\left\{S_t - Ke^{-r(T-t)}, 0\right\}$$

$$p_t \geq \max\left\{Ke^{-r(T-t)} - S_t, 0\right\}$$

$$P_t \geq \max\left\{K - S_t, 0\right\}$$

Question

Comment on the time value of each of the above options.

Solution

Recall that the time value is defined as the option price minus the intrinsic value.

The European and American call options have the same lower bound:

$$c_t \geq \max\left\{S_t - Ke^{-r(T-t)}, 0\right\} \quad \text{and} \quad C_t \geq \max\left\{S_t - Ke^{-r(T-t)}, 0\right\}$$

The intrinsic value of these call options is $\max\{S_t - K, 0\}$. Since:

$$\max\{S_t - K, 0\} \leq \max\left\{S_t - Ke^{-r(T-t)}, 0\right\}$$

the value of both call options is greater than the intrinsic value. The time value is therefore positive.

For an American put we can say that the time value is positive:

$$P_t - \max\{K - S_t, 0\} \geq \max\{K - S_t, 0\} - \max\{K - S_t, 0\} = 0$$

However, for a European put option the time value could be negative. The analysis above gives:

$$p_t - \max\{K - S_t, 0\} \geq \max\left\{Ke^{-r(T-t)} - S_t, 0\right\} - \max\{K - S_t, 0\}$$

A look at the right-hand side shows that this will be negative if $K > S_t > Ke^{-r(T-t)}$, since then $K - S_t > 0$ and $Ke^{-r(T-t)} - S_t < K - S_t$.

The reason that the time value of a European put can be negative is that by holding the option, rather than selling the share, an investor has money tied up that is not earning interest. This doesn't occur with the American put because of the possibility of early exercise.

It also doesn't happen with the call options because they work the other way round, *ie* holding the call means that an investor can invest money. If the share paid dividends, however, something like this could happen with the call, since then holding the option means that an investor is missing out on the dividends.

5.2 Upper bounds on option prices

European calls

A call option gives the holder the right to buy the underlying share for a certain price. The payoff $\max\{S_T - K, 0\}$ is always less than the value of the share at time T, S_T. Therefore the value of the call option must be less than or equal to the value of the share:

$$c_t \le S_t$$

This is obvious from the fact that an investor could just buy the share anyway, without needing to have the call option.

European puts

For a European put option the maximum value obtainable at expiry is the strike price *K*. Therefore the current value must satisfy:

$$p_t \le Ke^{-r(T-t)}$$

ie it cannot exceed the discounted value of the sum received on exercise – which it will equal if the share price falls to zero.

Note that these bounds make no assumptions about the behaviour of the share price.

For certain types of stochastic model for S_t we find that we are able to write down explicit formulae for the prices of European call and put options.

For example, if we assume that the process determining the share price is described by geometric Brownian motion, then we can use the Black-Scholes analysis and formulae discussed later in the course.

American calls

We can see that the call option inequality applies also to an American call option on a non-dividend-paying stock.

ie $\quad C_t \le S_t$

American puts

On the other hand, the possibility of early exercise of an American put option presents us with much more complexity. There are no simple rules for deciding upon the time to exercise. Partly as a result of this, there is no explicit formula for the price of an American put option.

For American puts the upper bound for a European put may not hold. For example, if the share price is zero, an American put is worth exactly *K*, not the discounted value of *K*. All we can be sure of is that a put will never be worth more than *K*.

ie $\quad P_t \le K$

Summary

The upper bounds derived in this section are:

$$c_t \leq S_t$$

$$C_t \leq S_t$$

$$p_t \leq Ke^{-r(T-t)}$$

$$P_t \leq K$$

6 Put-call parity

Consider the argument we used to derive the lower bounds for European call and put options on a non-dividend-paying stock. This used two portfolios:

- **A: one call plus cash of $Ke^{-r(T-t)}$**
- **B: one put plus one share.**

Recall that both portfolios included only *European* options on *non-dividend-paying* shares – this is an important condition underpinning the arguments that follow.

Both portfolios have a payoff at the time of expiry of the options of $\max\{K, S_T\}$.

We can see this as follows.

Portfolio A

First consider portfolio A, consisting of a European call plus cash of $Ke^{-r(T-t)}$. The value of portfolio A at the expiry date is given by:

	$S_T - K + K = S_T$	if $S_T > K$	(*ie* the call option is exercised)
and	$0 + K = K$	if $S_T \le K$	(*ie* the call expires worthless)

Portfolio B

Now consider portfolio B, consisting of the underlying share plus a European put with the same expiry date and exercise price as the call. On expiry the value of portfolio B is:

	$0 + S_T = S_T$	if $S_T > K$	(*ie* the put expires worthless)
and	$K - S_T + S_T = K$	if $S_T \le K$	(*ie* the put option is exercised)

Thus, the values at expiry are the same for both portfolios regardless of the share price at that time, namely $\max\{K, S_T\}$.

Since they have the same value at expiry and since the options cannot be exercised before then they should have the same value at any time $t < T$.

ie $$c_t + Ke^{-r(T-t)} = p_t + S_t$$

This relationship is known as *put-call parity*.

If the result was not true then this would give rise to the possibility of *arbitrage*. That is, for a net outlay of zero at time *t* we have a probability of 0 of losing money and a strictly positive probability (in this case 1) of making a profit greater than zero. In this case, the failure of put-call parity would allow an investor to sell calls and take a cash position and buy puts and shares with a net cost of zero at time *t* and a certain profit at time *T*.

One consequence of put-call parity is that, having found the value of a European call on a non-dividend-paying share (*eg* from the Black-Scholes formula discussed later in the course), we can easily find the value of the corresponding put.

In contrast to forward pricing, put-call parity does not tell us *what* c_t and p_t are individually: only the relationship between the two. To calculate values for c_t and p_t we require a model.

In all of these sections, the pricing of derivatives is based upon the principle of *no arbitrage*.

Note that we have made very few assumptions in arriving at these results. No model has been assumed for stock prices. All we have assumed is that we will make use of buy-and-hold investment strategies. Any model that we propose for pricing derivatives must, therefore, satisfy both put-call parity and the forward-pricing formula. If a model fails one of these simple tests then it is not arbitrage-free.

Question

Explain why the put-call parity relationship above does not hold in the case of:

(i) American options on non-dividend-paying shares.

(ii) European options on dividend-paying shares.

Solution

(i) This is because it can be worthwhile to exercise an American put early – in which case the cash will not have accumulated fully and so the payoffs do not work out to be the same. This means that Portfolio B is worth more than Portfolio A.

(ii) Dividends will be received on Portfolio B, but not on Portfolio A. Again we see that Portfolio B is then worth more than Portfolio A.

In the case of dividend-paying securities, the put-call parity relationship is:

$$c_t + Ke^{-r(T-t)} = p_t + S_t e^{-q(T-t)}$$

This formula can be found on page 47 of the *Tables*. Here q is the continuously compounded dividend rate, so we are assuming that all dividends are reinvested immediately in the same share.

Question

Company X issues 3-month European call options on its own shares with a strike price of 120p. The call options are currently priced at 30p per share. The current share price is 123p and the current risk-free force of interest is $r = 6\% pa$.

(i) If dividends are payable continuously at a rate of $q = 12\%\ pa$, calculate the fair price for put options on the share with the same strike price.

(ii) Explain the strategy for arbitrage profit if, instead, the price of the put options is 25p.

Solution

(i) ***Put option value***

Using put-call parity, the value of the put option should be:

$$p_t = c_t + Ke^{-r(T-t)} - S_t e^{-q(T-t)} = 30 + 120e^{-0.06\times0.25} - 123e^{-0.12\times0.25} = 28.8\text{p}$$

(ii) ***Arbitrage profit***

If the put options are only 25p, then they are cheap. If things are cheap, then we should buy them in order to generate arbitrage profits. So, looking at the put-call parity relationship, we 'buy the cheap side and sell the expensive side', *ie* we buy put options and shares and sell call options and cash.

For example:

- sell 1 call option 30p
- buy 1 put option (25p)
- buy 1 share (123p)
- sell (borrow) cash 118p

This is a zero-cost portfolio and, because put-call parity does not hold, we know it will make an arbitrage profit, which we can check as follows.

In 3 months' time, repaying the cash will cost us:

$$118e^{0.06\times0.25} = 119.78$$

We will also have received dividends totalling *d*, say, on the share.

1. If the share price is above 120 in 3 months' time, then the other party will exercise their call option and we will have to sell them the share. They will pay 120 for it and our profit is:

$$120 - 119.78 + d = 0.22 + d$$

(In this case, the put option will expire worthless.)

2. If the share price is below 120 in 3 months' time, then we will exercise our put option and sell the share for 120. Our profit is:

$$120 - 119.78 + d = 0.22 + d$$

(In this case, the call option will expire worthless.)

So we see that in either case, this zero-cost portfolio generates positive future profits.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 12 Summary

Derivatives

A *derivative* is a security or contract that promises to make a payment at a specified time in the future, the amount of which depends upon the behaviour of some underlying security up to and including the time of the payment.

Arbitrage

An *arbitrage opportunity* is a situation where we can make a certain profit with no risk.

The *principle of no arbitrage* states that arbitrage opportunities do not exist.

The *law of one price* says that any two portfolios that behave in exactly the same way must have the same price. If this were not true, we could buy the 'cheap' one and sell the 'expensive' one to make an arbitrage (risk-free) profit.

Intrinsic value and time value

The *intrinsic value* of a derivative is the value assuming expiry of the contract immediately rather than at some time in the future.

The *time value* of a derivative is the excess (or the shortfall) of the total value of an option over its intrinsic value.

Options

A *European call option* gives its holder the right, but not the obligation, to buy one share from the issuer of the contract at time T at the strike price K.

A *European put option* gives its holder the right, but not the obligation, to sell one share to the issuer of the contract at time T at the strike price K.

American options are similar to their European equivalents, except that they can be exercised at any time t up to expiry T.

Factors affecting option prices

- underlying share price, S_t
- strike price, K
- time to expiry, $T-t$
- volatility of the underlying share, σ
- risk-free interest rate, r
- dividend rate, q

Forwards

A forward contract is an agreement between two parties under which one agrees to buy from the other a specified amount of an asset at a specified price on a specified future date.

For a forward contract on an underlying asset that pays no income, because both $K > S_0e^{rT}$ and $K < S_0e^{rT}$ lead to an arbitrage opportunity, the fair forward price is given by:

$$K = S_0 e^{rT}$$

For a forward contract on an underlying asset that pays fixed cash income, the fair forward price is:

$$K = (S_0 - I)e^{rT}$$

where I is the present value of the income paid by the asset during the term of the forward contact.

For a forward contract on an underlying asset that pays dividends at a continuously compounded dividend yield of D, the fair forward price is:

$$K = S_0 e^{(r-D)T}$$

Bounds for option prices

Option prices lie within the following ranges:

$$S_t \geq c_t \geq \max\left\{S_t - Ke^{-r(T-t)}, 0\right\}$$

$$S_t \geq C_t \geq \max\left\{S_t - Ke^{-r(T-t)}, 0\right\}$$

$$Ke^{-r(T-t)} \geq p_t \geq \max\left\{Ke^{-r(T-t)} - S_t, 0\right\}$$

$$K \geq P_t \geq \max\left\{K - S_t, 0\right\}$$

Put-call parity

$$c_t + Ke^{-r(T-t)} = p_t + S_t e^{-q(T-t)}$$

Chapter 12 Practice Questions

12.1 Write down expressions for the payoff functions for the option holder for each of the following derivative contracts in terms of the current time t, the expiry time T, the price S_u of the underlying at time u and the exercise price K, indicating clearly the time(s) at which exercise may occur:

(a) a European call option

(b) a European put option

(c) an American call option

(d) an American put option

12.2 Identify the profit or loss for the investors in each of the following scenarios:

(a) Investor A purchases a European call option with an exercise price of 480p for a premium of 37p. The price of the underlying share is 495p at the expiry date.

(b) Investor B purchases a European put option with an exercise price of 180p for a premium of 12p. The price of the underlying share is 150p at the expiry date.

(c) Investor C issues a European put option with an exercise price of 250p for a premium of 22p. The price of the underlying share is 272p at the expiry date.

12.3 A speculator has a portfolio consisting of one short position on a European call option on a share.

Explain what this means and sketch the position diagram (a diagram of the overall profit/loss at expiry against the security price at expiry) for this portfolio assuming that no dividends are payable, and that the initial option premium was c.

12.4 A speculator has a portfolio consisting of one European call option and one European put option on the same underlying security. The two options have the same expiry date and the same strike price K. The prices paid for the options were c and p, respectively.

(i) Sketch the position diagram (a diagram of the overall profit/loss at expiry against the security price at expiry) for this portfolio, marking the coordinates of the key points on your graph.

(ii) Explain what this portfolio implies about the speculator's opinion concerning the future share price.

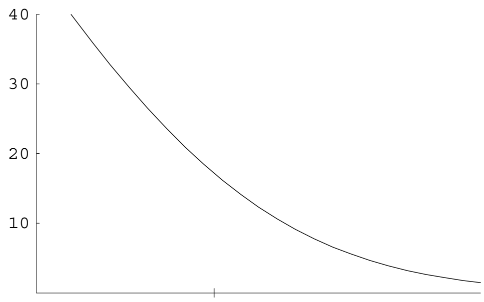
12.5 A call option on a stock that does not pay dividends has the following parameter values (in the usual notation):

$$S = 240\,,\ K = 250\,,\ T = 0.5\,,\ r = 0.06\,,\ \sigma = 0.2$$

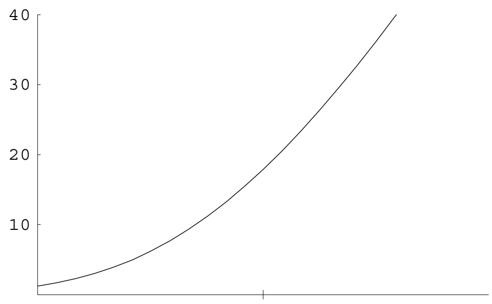
The graphs below show the theoretical price of this option at time $t = 0$ when each of the parameters S, K, T and r is varied without changing the values of the other parameters.

Identify which parameter has been plotted along the x-axis of each graph.

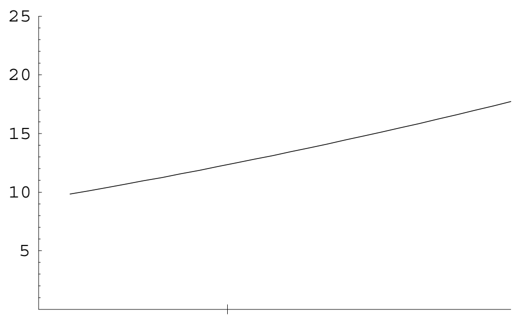
Graph 1



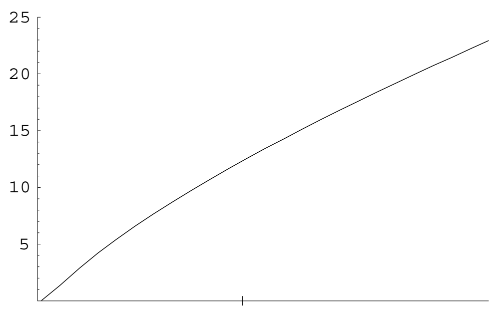
Graph 2



Graph 3



Graph 4



12.6 A 9-month forward contract is issued on a share that has a current price of £7. Dividends of 50p per share are expected in 2 months' time and 8 months' time.

Assuming a risk-free effective rate of interest of 6% per annum and no arbitrage, calculate the forward price.

12.7 Exam style The table below shows the closing prices (represented by letters) on a particular day for a series of European call options with different strike prices and expiry dates on a particular non-dividend-paying security.

	Strike price	
	125	150
3 months	W	Y
6 months	X	Z

(i) Write down, with reasons, the strictest inequalities that can be deduced for the relative values of W, X, Y, Z, assuming that the market is arbitrage-free. (Your inequalities should not involve any other quantities.) [4]

(ii) Calculate numerical values for a lower and an upper bound for X, given that the current share price is 120 and the continuously compounded risk-free interest rate is 6% *pa*. [2]

[Total 6]

12.8 Exam style Let p_t be the price of a European put option exercisable at time T with a strike price K on an underlying non-dividend-paying share with price S_t at time t.

(i) By considering a suitably chosen notional portfolio or portfolios (which should be specified carefully), show that the price p_t satisfies the inequality:

$$p_t \geq Ke^{-r(T-t)} - S_t$$ [4]

(ii) Explain how you would modify your inequality if you knew that holders of the share on the day before the option expires are entitled to receive a cash dividend of $0.02S_T$ payable at time T. [2]

[Total 6]

12.9 Exam style Consider an asset with price S_t at time t, paying a dividend at a constant dividend yield, D. Dividends are paid at the end of each year and are immediately reinvested in the asset. The continuously compounded risk-free rate of interest is r *pa.*

Derive the forward price, K, of a contract issued at time t, with maturity at time T, to trade one unit of the asset, where $T-t$ is an integer number of years. State any assumptions you make. [6]

12.10 Exam style By constructing two portfolios with identical payoffs at the exercise date of the options, derive an expression for the put-call parity of European options on a share that has a dividend of known amount d payable prior to the exercise date. [6]

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 12 Solutions

12.1 (a) Payoff $= \max(S_T - K, 0)$, exercise at time T only.

(b) Payoff $= \max(K - S_T, 0)$, exercise at time T only.

(c) Payoff $= \max(S_u - K, 0)$, exercise at any time $t \le u \le T$.

(d) Payoff $= \max(K - S_u, 0)$, exercise at any time $t \le u \le T$.

12.2 (a) Profit $= -37 + (495 - 480) = -22$, *ie* a loss of 22p.

(b) Profit $= -12 + (180 - 150) = +18$, *ie* a profit of 18p.

(c) Profit $= +22 + 0 = +22$, *ie* a profit of 22p.

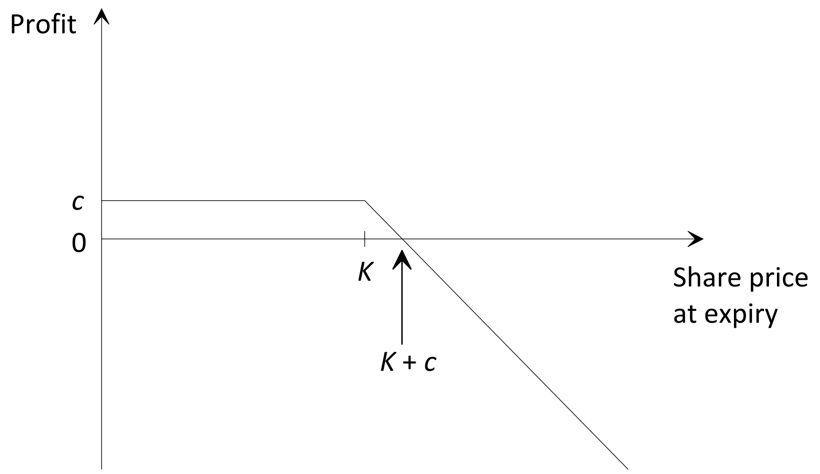
*In (c), Investor C has **collected** the premium and the option has expired worthless.*

12.3 Having a short position on one call option means that the speculator has issued (or written) a call option. The speculator will therefore have received the initial premium, c, and will have to deliver one share to the other party in exchange for a payment equal to the strike price K, say, on the expiry date if the other party elects to exercise the option.

The speculator's profit is given by:

$$c - \max(S_T - K, 0)$$

and the position diagram is:



12.4 (i) Position diagram

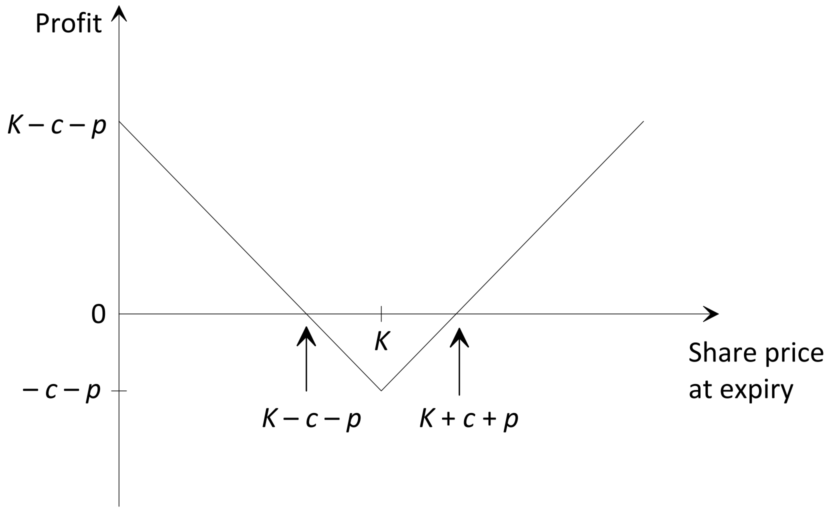
The profit for the call option is $\max(S_T - K, 0) - c$.

The profit for the put option is $\max(K - S_T, 0) - p$.

The profit for a portfolio containing both of these options is equal to the sum of these, which can be written as:

$$\max(S_T - K, 0) + \max(K - S_T, 0) - c - p = \left| S_T - K \right| - c - p$$

So the position diagram for the holder of this portfolio looks like this:

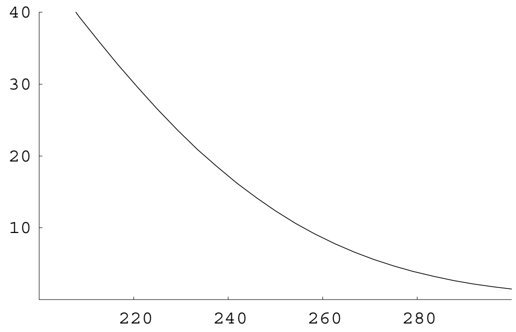


(ii) ***Speculator's opinion***

The speculator would have set up this portfolio in the belief that the share price at the expiry date would have moved a long way from the strike price K, where this movement could be in either direction. This situation can arise when the market is waiting for news (*eg* an announcement of the company's latest results or a court ruling) that will have a definite effect on the share price, but no one knows which way it will go.

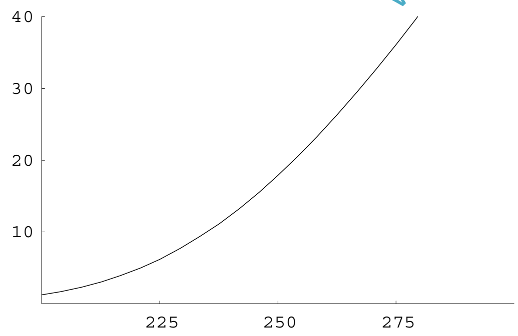
12.5 The graphs and associated parameters are shown below.

Graph 1



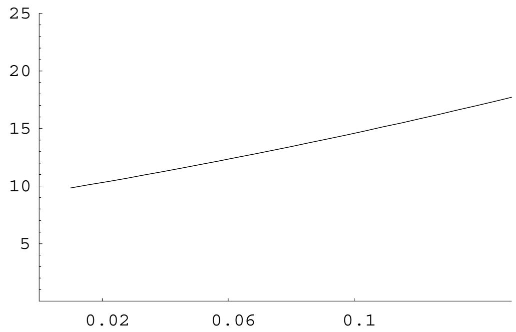
This is a graph of the option price against the strike price K.

Graph 2



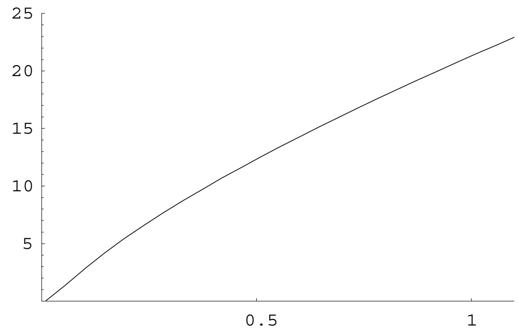
This is a graph of the option price against the current share price S.

Graph 3



This is a graph of the option price against the risk-free interest rate r.

Graph 4



This is a graph of the option price against the time to expiry T.

The clues to look for were:

- The value of a call option decreases as the strike price increases, and Graph 1 is the only graph that goes down.
- Since the payoff for a call option depends on the difference $S-K$, the graph against S should be similar to the graph against K, but reflected in the vertical axis (because increasing K by 10 would have a similar effect to reducing S by 10). So this is Graph 2.
- If $T=0$, this call option would be at its maturity date and it would be out-of-the-money with no value (since $S<K$). So the graph against T must go through the origin, as in Graph 4.
- The value of a call option increases with the interest rate, but would not have zero value when $r=0$. This is consistent with Graph 3.

12.6 The present value of the dividends, I, is:

$$I = 0.5\left(1.06^{-2/12} + 1.06^{-8/12}\right) = 0.9761$$

The forward price is given by:

$$K = (S_0 - I)(1+r)^T = (7 - 0.9761) \times 1.06^{0.75} = £6.29$$

12.7 (i) ***Inequalities***

The value of an option is greater if the remaining life is longer.

So: $X > W$ and $Z > Y$ [1]

The value of a call option is smaller if the exercise price is greater.

So: $Y < W$ and $Z < X$ [1]

Also, the value of an option will be strictly positive. [1]

Combining these results, we have:

$$0 < Y < (W, Z) < X$$

ie W and Z must have values between Y and X. [1]

[Total 4]

Based on the information given in the question we can't determine the order of W and Z.

(ii) ***Lower and upper bounds***

The lower bound for a European call option is given by the inequality:

$$c_t \geq S_t - Ke^{-r(T-t)}$$

Using the parameter values for X, this gives:

$$c_t \geq 120 - 125e^{-0.06 \times 0.5} = -1.31$$

Since this is negative, we take $c_t \geq 0$ instead. [1]

The upper bound for a call option is given by the inequality:

$$c_t \leq S_t \quad ie \quad c_t \leq 120$$ [1]

[Total 2]

12.8 (i) ***Show that the price satisfies the inequality***

Consider the following portfolios set up at the current time t :

Portfolio A: Cash of $Ke^{-r(T-t)}$ [½]

Portfolio B: 1 European put option and 1 share [½]

Portfolio A will be worth K at the expiry time T, in any event. [½]

Now consider the value of Portfolio B at the expiry time T. If the share price is below the strike price, we can exercise the put (selling the share in the process) and obtain an amount K. If the share is above the strike price, we have a share worth more than K.

So, either way, Portfolio B is worth at least K. [1]

We can show this mathematically by writing down the portfolio payoff, which is:

$$\max(K-S_T,0)+S_T=\max(K,S_T)\geq K$$

Hence, at the expiry date, Portfolio B is worth at least as much as Portfolio A. [½]

It follows (from the principle of no-arbitrage) that this must also be true at all earlier times t. In symbols:

$$p_t+S_t\geq Ke^{-r(T-t)} \quad ie \quad p_t\geq Ke^{-r(T-t)}-S_t \qquad [1]$$

[Total 4]

(ii) ***Modifying the inequality***

Consider an alternative Portfolio B* consisting of 1 European put option and $\frac{1}{1.02}$ shares.

The (cash) dividend payable on these shares at time T will be worth $0.02\times\frac{1}{1.02}$ times the share price at that time, which would allow us to buy an extra $\frac{0.02}{1.02}$ shares. So we would end up with $\frac{1}{1.02}+\frac{0.02}{1.02}=1$ share and 1 put option in our portfolio. [1]

By the same argument as before, this is worth at least as much as Portfolio A. So we now have the inequality:

$$p_t+\frac{1}{1.02}S_t\geq Ke^{-r(T-t)} \quad ie \quad p_t\geq Ke^{-r(T-t)}-\frac{1}{1.02}S_t \qquad [1]$$

[Total 2]

12.9 Consider the following two portfolios, set up at time t :

Portfolio A: A forward contract to buy one unit of the asset at time T for forward price K; simultaneously invest an amount $Ke^{-r(T-t)}$ in the risk-free investment. [1]

Portfolio B: Buy $(1+D)^{-(T-t)}$ units of the asset, reinvesting the dividend income in the asset immediately as it is received. [1]

At time T, the risk-free investment in Portfolio A has grown to amount K, which is used to buy one unit of the asset using the forward contract.

At time T, the amount of the asset in Portfolio B has grown with the reinvested dividend income to one unit.

So, the outcome of both of these portfolios is that one unit of the underlying asset is held at time T. [1]

Assuming no arbitrage, the value of these portfolios must therefore also be equal at time t. [1]

The cost of setting up Portfolio A is $Ke^{-r(T-t)}$. [½]

The cost of setting up Portfolio B is $S_t(1+D)^{-(T-t)}$ [½]

Equating these:

$$Ke^{-r(T-t)} = S_t(1+D)^{-(T-t)} \quad \Rightarrow \quad K = S_t e^{r(T-t)}(1+D)^{-(T-t)} \qquad [1]$$

[Total 6]

12.10 Suppose that the dividend d is payable at some date $t < t_1 < T$. Consider two portfolios at time t as follows:

1. Portfolio A – which consists of one European call option plus cash equal in amount to the discounted value of the strike price plus the present value of the dividends to be paid at time t_1 – *ie* a cash amount of $d\,e^{-r(t_1-t)} + K\,e^{-r(T-t)}$. [1]

2. Portfolio B – which consists of one European put plus one dividend-paying share. [1]

Then the value of portfolio A at the exercise date T is given by:

$$S_T - K + d\,e^{r(T-t_1)} + K \;=\; S_T + d\,e^{r(T-t_1)} \qquad \text{if } S_T > K \qquad [½]$$

(*ie* the call option is exercised leaving the investor with the share plus the accumulated value of the dividend received), and:

$$0 + d\,e^{r(T-t_1)} + K \;=\; d\,e^{r(T-t_1)} + K \qquad \text{if } S_T \le K \qquad [½]$$

(*ie* the call expires worthless and the investor is left with cash equal to the exercise price plus the accumulated value of the dividend).

Similarly, the value of portfolio B is:

$$0 + S_T + d\,e^{r(T-t_1)} = S_T + d\,e^{r(T-t_1)} \qquad \text{if } S_T > K \qquad [½]$$

(*ie* the put expires worthless and the investor is left with the share plus the accumulated value of the dividend received), and:

$$K - S_T + S_T + d\,e^{r(T-t_1)} = K + d\,e^{r(T-t_1)} \qquad \text{if } S_T \le K \qquad [½]$$

(*ie* the put option is exercised and the investor is left with cash equal to the exercise price plus the accumulated value of the dividend).

Thus, the payoffs at expiry are the same for both portfolios regardless of the share price at that time. Since they have the same value at expiry and since the options cannot be exercised before then, in an arbitrage-free market they should have the same value at any time $t < T$. [1]

Therefore:

$$c_t + d\,e^{-r(t_1-t)} + Ke^{-r(T-t)} = p_t + S_t \qquad [1]$$

[Total 6]

End of Part 2

What next?

1. Briefly **review** the key areas of Part 2 and/or re-read the **summaries** at the end of Chapters 7 to 12.
2. Ensure you have attempted some of the **Practice Questions** at the end of each chapter in Part 2. If you don't have time to do them all, you could save the remainder for use as part of your revision.
3. Attempt **Assignment X2**.

13

The Greeks

Syllabus objectives

6.1 Option pricing and valuations

6.1.12 Demonstrate an awareness of the commonly used terminology for the first, and where appropriate second, partial derivatives (the Greeks) of an option price.

0 Introduction

Later in the course we will consider how to price European options. We will first do this using a discrete-time model (the Binomial model) and then using a continuous-time model (the Black-Scholes model). The important point to realise is that no matter how complicated the model used to value a given derivative, it cannot conflict with the simpler properties that we already know from the introductory chapter on derivatives.

For example, no matter how ingenious a model for valuing a European call option might be, if it contradicts the fact that $c_T = \max\{S_T - K, 0\}$ then it must be ruled out. Similarly, we know that, regardless of the pricing model used, put-call parity must hold, *ie*:

$$c_t + Ke^{-r(T-t)} = p_t + S_t e^{-q(T-t)}$$

So, whatever we find out about the Greeks for call and put options by studying this chapter must remain true no matter how complicated the model used to value the derivatives is.

1 Portfolio risk management

1.1 Introduction

Call options allow exposure to be gained to upside movements in the price of the underlying asset. Put options allow the downside risks to be removed. In each case, however, because of the effect of gearing, both call and put options, on their own, are more risky than the underlying asset.

Question

Explain what is meant by gearing in this context.

Solution

When purchasing a call or a put option, the option premium paid is typically much smaller than the price of the underlying asset itself. Purchasing the option nevertheless enables us to obtain exposure to most of the variation in the price of the underlying asset. Consequently, the *percentage* returns obtained by purchasing the option will be much greater (or much less) than those obtained by purchasing the underlying asset instead. This effect is referred to as *gearing*.

We will look at a numerical example of gearing later in the chapter.

In more general terms, combinations of various derivatives and the underlying asset in a single portfolio allow us to modify our exposure to risk.

In particular, we can use derivatives to reduce the exposure of our portfolio to the risk of adverse movements in the market price of the underlying assets. If, for example, we are concerned about falls in the investment market, we might buy put options. By guaranteeing the price at which we can sell our assets, this reduces the risk associated with market falls. We will still, however, benefit from the resulting profits should the market instead go up.

Derivative contracts therefore give us more control over the market risks that we face, thereby increasing our opportunity set of possible risk and return combinations. Moreover, if we hold suitable derivatives and the underlying assets in appropriate combinations then we can sometimes eliminate almost all of the market risk facing our portfolio – though other risks such as lack of marketability or credit risk will remain. The strategy of reducing market risk in this way is known as *hedging*.

For example, in the proof of the Black-Scholes PDE we will take a mixed portfolio of a derivative and the underlying asset to create an instantaneously risk-free portfolio. This is called *delta hedging*.

Delta hedging involves the construction of a portfolio whose overall 'delta' is equal to zero. We will discuss this, and the Black-Scholes partial differential equation (PDE), later.

Delta is just one of what are called *the Greeks*. The Greeks are a group of *mathematical* derivatives that can be used to help us to manage or understand the risks in our portfolio.

Let $f(t, S_t)$ be the value at time t of a derivative when the price of the underlying asset at t is S_t.

We now introduce a set of six Greeks.

1.2 Delta

The delta for an individual derivative is:

$$\Delta = \frac{\partial f}{\partial S_t} \equiv \frac{\partial f}{\partial S_t}(t, S_t)$$

The notation here works as follows. We are thinking of the derivative price $f(t,S_t)$ as a function of time t and share price S_t. The partial derivative of $f(t,S_t)$ with respect to S_t, estimated at time 2 when the share price equals 100 is then written as:

$$\frac{\partial f(2,100)}{\partial S_2}$$

For the underlying asset, whose value is S_t, $\Delta = 1$.

Question

Explain briefly why the delta of the underlying asset is equal to 1.

Solution

The delta of the underlying asset is equal to the derivative of the underlying asset price with respect to itself. By definition, this must be equal to 1, *ie*:

$$\Delta = \frac{\partial S_t}{\partial S_t} = 1$$

Delta is defined as the rate of change of the derivative price with respect to changes in the underlying asset price (assuming all other parameters remain unchanged). It therefore tells us the (approximate) change in the derivative price when the underlying asset price changes.

Question

Investor A has £10,000 invested in a portfolio consisting of 1,000 shares in Company X. Investor B has £10,000 invested in a portfolio of 5,000 call options on shares in Company X and the delta of each call option is 0.5.

Calculate the percentage change in the value of each portfolio if the share price increases by 10%.

Solution

Each Company X share is worth £10, and each call option is worth £2.

If the share price increases by 10% to £11, Investor A's portfolio will be worth £11,000 – a 10% increase in value.

Investor B's portfolio consists entirely of call options. We can use the value of delta to estimate the change in value of a call option when the underlying share price changes. Since delta is equal to 0.5, when the share price increases by £1, the call option price will increase by (approximately) $0.5 \times £1 = 50p$, from £2 to £2.50. So Investor B's portfolio will be worth $5{,}000 \times 2.5 = £12{,}500$ – a 25% increase in value.

So we see that the percentage change in value is greater for the portfolio of options. This is an example of the effect of gearing mentioned earlier.

When we consider delta hedging, we add up the deltas for the individual assets and derivatives (taking account, of course, of the number of units held of each). If this sum is zero and if the underlying asset prices follow a diffusion then the portfolio is instantaneously risk-free.

A portfolio for which the weighted sum of the deltas of the individual assets is equal to zero is described as delta-hedged or *delta-neutral*.

Instantaneously risk-free means that if we know the value of the portfolio at time t, then we can predict its value at time $t + dt$ with complete certainty. In other words, there is no risk or uncertainty concerning the change in the value of the portfolio over the instantaneous time interval $(t, t + dt]$.

Later we will meet the following 'risk-free' portfolio that we will use to construct the Black-Scholes partial differential equation. It is:

- minus one derivative
- plus $\dfrac{\partial f}{\partial S_t} = \Delta$ shares.

Thus, the total delta of this portfolio is equal to:

$$(-1) \times \Delta \;+\; \Delta \times (+1) \;=\; 0$$

Consequently, the 'risk-free' portfolio must be instantaneously risk-free.

Question

By applying Ito's Lemma, show that a delta-hedged portfolio with value $V(t, S_t)$ is instantaneously risk-free if the underlying process S_t is a diffusion.

Solution

Consider the change in the value of the portfolio $dV(t,S_t)$. Ito's Lemma tells us that:

$$dV(t,S_t)=\frac{\partial V}{\partial t}dt+\frac{\partial V}{\partial S_t}dS_t+½\frac{\partial^2 V}{\partial S_t^{\,2}}(dS_t)^2$$

Now since S_t is a diffusion we have $dS_t=\mu(t,S_t)dt+\sigma(t,S_t)dZ_t$, so $(dS_t)^2=\sigma^2(t,S_t)dt$ giving:

$$\begin{aligned}dV&=\frac{\partial V}{\partial t}dt+\frac{\partial V}{\partial S_t}dS_t+½\sigma^2(t,S_t)\frac{\partial^2 V}{\partial S_t^{\,2}}dt\\&=\frac{\partial V}{\partial t}dt+\frac{\partial V}{\partial S_t}\left(\mu(t,S_t)dt+\sigma(t,S_t)dZ_t\right)+½\sigma^2(t,S_t)\frac{\partial^2 V}{\partial S_t^{\,2}}dt\\&=\left(\frac{\partial V}{\partial t}+\mu(t,S_t)\frac{\partial V}{\partial S_t}+½\sigma^2(t,S_t)\frac{\partial^2 V}{\partial S_t^{\,2}}\right)dt+\sigma(t,S_t)\frac{\partial V}{\partial S_t}dZ_t\end{aligned}$$

It follows that the change in the portfolio value *V* over the next instant will be deterministic if the stochastic dZ_t term vanishes, *ie* if the delta of the portfolio $\Delta=\frac{\partial V}{\partial S_t}=0$.

If it is intended that the sum of the deltas should remain close to zero (this is what is called *delta hedging*) then normally it will be necessary to rebalance the portfolio on a regular basis. The extent of this rebalancing depends primarily on *gamma*.

Within this context, we can distinguish between *dynamic hedging* and *static hedging*.

The process of simply constructing an initial portfolio with a total delta of zero, at time 0 say, and not rebalancing to reflect the subsequent changes in delta, is known as *static* delta hedging.

Note, however, that as the share price S_t varies with time, so does:

- the price of the derivative, $f(t,S_t)$
- $\Delta=\dfrac{\partial f(t,S_t)}{\partial S_t}$.

Hence, in order to ensure that the total portfolio delta remains equal to zero over time we need to 'rebalance' the constituents of the portfolio – so as to offset the changes in delta. Strictly speaking, since delta changes continuously through time, this rebalancing process must itself be continuous.

The process of continuously rebalancing the portfolio in this way in order to maintain a constant total portfolio delta of zero is known as *dynamic* delta hedging.

1.3 Gamma

$$\Gamma = \frac{\partial^2 f}{\partial S_t^2}$$

For the underlying asset, whose value is S_t, $\Gamma = 0$.

Question

Explain why the gamma of the underlying asset is equal to zero.

Solution

Gamma is the second derivative of f with respect to the underlying share price, *ie* the first derivative of delta with respect to the underlying share price. We saw previously that the delta of the underlying asset is equal to one and hence constant. The derivative of a constant is zero and so the gamma of the underlying asset must be zero, *ie*:

$$\Gamma = \frac{\partial \Delta}{\partial S_t} = \frac{\partial}{\partial S_t}(1) = 0$$

Gamma is the rate of change of Δ with the price of the underlying asset.

It therefore measures the *curvature* or convexity of the relationship between the derivative price and the price of the underlying asset.

Suppose a portfolio is following a delta hedging strategy. If the portfolio has a high value of Γ then it will require more frequent rebalancing or larger trades than one with a low value of gamma.

This is because a high value of Γ means that Δ is more sensitive to changes in the share price S_t. Consequently, a given change in S_t will produce a greater change in Δ, which means that a greater amount of rebalancing will be required in order to ensure that the overall portfolio delta remains equal to zero. Conversely, if Γ is small, then Δ will change less when the share price changes and so the adjustments needed to keep a portfolio delta-neutral will be minimal.

It is recognised that continuous rebalancing of the portfolio is not feasible and that frequent rebalancing increases transaction costs. The need for rebalancing can, therefore, be minimised by keeping gamma close to zero.

Hence, for practical hedging purposes, a portfolio with a low Γ is preferable, as costly rebalancing will be required less frequently in order to keep the portfolio approximately delta-hedged. Note that it may also be possible to construct a *gamma-neutral* portfolio – *ie* one with an overall gamma equal to zero.

1.4 Vega

$$\nu = \frac{\partial f}{\partial \sigma}$$

For the underlying asset, S_t, $\nu = 0$.

(Unlike the other names, *vega* is not a Greek letter.)

If you have met the Greeks before, possibly at university, then you may have seen this derivative called kappa (a real Greek letter). Derivative traders, however, call it vega.

This is the rate of change of the price of the derivative with respect to a change in the assumed level of volatility of S_t.

We refer here to the 'assumed' level of volatility. As we saw in the previous chapter, the value of an option depends on the volatility of the underlying share price. However, unlike the other parameters that affect option prices, the volatility cannot be observed directly in the market, and so its value must be estimated or assumed.

The value of a portfolio with a low value of vega will be relatively insensitive to changes in volatility. Put another way, it is less important to have an accurate estimate of σ if vega is low. Since σ is not directly observable, a low value of vega is important as a risk management tool. Furthermore, it is recognised that σ can vary over time. Since many derivative pricing models assume that σ is constant through time, the resulting approximation will be better if vega is small.

1.5 Rho

$$\rho = \frac{\partial f}{\partial r}$$

Rho tells us about the sensitivity of the derivative price to changes in the risk-free rate of interest. The risk-free rate of interest can be determined with a reasonable degree of certainty, but it can vary by a small amount over the (usually) short term of a derivative contract. As a result, a low value of ρ reduces risk relative to uncertainty in the risk-free rate of interest.

1.6 Lambda

$$\lambda = \frac{\partial f}{\partial q}$$

where q is the assumed, continuous dividend yield on the underlying security.

Note that in the definitions of ν, ρ and λ we are assuming that σ, r and q take constant values throughout the life of the derivative contract, but that these 'constant' values could change.

Question

Explain why the lambda of a long forward position in a dividend-paying share is negative.

Solution

If we enter a forward contract to purchase a dividend-paying share in the future (*ie* take a long forward position), instead of buying the share now, then we do not receive the dividends paid by the share over the lifetime of the forward contract. The greater the value of these dividends, the more income we miss out on by holding the forward and the lower the value of the forward. Consequently, lambda is negative for a forward contract.

1.7 Theta

$$\Theta = \frac{\partial f}{\partial t}$$

Since time is a variable which advances with certainty, it does not make sense to hedge against changes in t in the same way as we do for *unexpected* changes in the price of the underlying asset.

Note that:

- theta is usually written as a capital letter Θ
- t here is the time since the start of the contract, not the remaining life, which is $T-t$.

We will discuss theta again later in the context of the Black-Scholes PDE when its intuitive interpretation will become clearer.

Question

For each of the Greeks Δ, ν, Θ, ρ and λ, discuss whether its value will be positive or negative in the case of:

- a call option
- a put option.

Solution

Delta

An increase in the share price would either push a call option that is currently out-of-the-money towards being in-the-money or push one that is already in-the-money further into-the-money. Therefore delta is positive for a call option.

A decrease in the share price would either push a put option that is currently out-of-the-money towards being in-the-money or push one that is already in-the-money further into-the-money. Therefore delta is negative for a put option.

Vega

If the underlying security becomes more volatile then there is a greater chance of the price moving in favour of the option holder. Although there is also an increased chance of it moving against the holder, the downside loss is capped. Therefore vega should be positive for both a call option and a put option.

Theta

The intuitive argument is very similar to before. The greater the time to expiry, the more chance that the share price will move in the holder's favour, with the downside loss again being capped. Thus, because time t works in the opposite direction to time to expiry $T-t$, theta is usually negative for both a call and a put option.

There are circumstances where theta may be positive but these are considered beyond the scope of this subject.

Rho

We can think of holding a call option as having cash in the bank waiting to buy the share. If interest rates rise then the holder of a call option will benefit in the meantime. The holder of a put option may already own a share and is waiting to sell it for cash. So if interest rates rise then the holder of the put will lose out on that interest in the meantime. So rho is positive for a call option and negative for a put option.

Lambda

Again, we can think of holding a call option as having cash in the bank waiting to buy the share. If the dividend rate rises then the holder of a call option will lose out in the meantime. The holder of a put option may already own a share and is waiting to sell it for cash. So if dividend rates rise then the holder of the put will benefit from the extra dividends in the meantime. So lambda is negative for a call option and positive for a put option.

Chapter 13 Summary

The Greeks

Delta	$\Delta = \dfrac{\partial f}{\partial S_t}$	the change of the derivative price with the share price
Gamma	$\Gamma = \dfrac{\partial^2 f}{\partial S_t^2}$	the change of delta with the share price
Theta	$\Theta = \dfrac{\partial f}{\partial t}$	the change of the derivative price with time
Vega	$\nu = \dfrac{\partial f}{\partial \sigma}$	the change of the derivative price with volatility
Rho	$\rho = \dfrac{\partial f}{\partial r}$	the change of the derivative price with the risk-free rate
Lambda	$\lambda = \dfrac{\partial f}{\partial q}$	the change of the derivative price with the dividend rate

Delta hedging

A portfolio for which the weighted sum of the deltas of the individual assets is equal to zero is described as *delta-hedged*.

The practice questions start on the next page so that you can keep the chapter summaries together for revision purposes.

Chapter 13 Practice Questions

13.1 An investor claims to be able to value an unusual derivative on a non-dividend-paying share using the pricing formula:

$$V_t = S_t^2 e^{-4S_t}$$

where S_t denotes the price of the share at time t.

(i) Derive formulae for the delta and gamma of the derivative, based on the pricing formula above.

(ii) For each of the following scenarios, calculate the number of shares that must be purchased or sold along with a short holding in one derivative, in order to achieve a delta-hedged portfolio:

(a) the current share price is 1

(b) the current share price is 3.

(iii) Explain which of the scenarios in (ii) is likely to involve more portfolio management in the near future if the investor is determined to maintain a delta-hedged portfolio.

13.2 (i) Define the delta, gamma and theta of an option. [3]

Exam style

(ii) Describe, using a numerical example, the concept of delta hedging. [6]

[Total 9]

13.3 Give definitions of the 'Greeks' that could be used as an aid to management in each of the following situations. State also the desired ranges for their numerical values and define any notation you use.

Exam style

(a) A hedge fund manager wishes to establish a delta-neutral position that would not need frequent rebalancing.

(b) A derivatives trader is concerned that a change in the distribution of the daily price movements of particular shares might affect the values of the options held on those shares.

(c) The trustee of a pension fund that purchased a large number of options last year as a means of hedging is concerned about changes in the value of the fund as the options approach their expiry date. [6]

13.4 A call option has a price of 20.15p and a delta of 0.558 at time t. Determine the hedging portfolio of shares and cash for this option at time t, given that the price of the underlying share, $S_t = 240\text{p}$.

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 13 Solutions

13.1 (i) ***Delta and gamma***

Using the product rule, the delta of the derivative is:

$$\Delta = \frac{\partial V_t}{\partial S_t} = 2S_t e^{-4S_t} - 4S_t^2 e^{-4S_t}$$

$$= 2S_t e^{-4S_t}(1-2S_t)$$

The gamma of the derivative is:

$$\Gamma = \frac{\partial^2 V_t}{\partial S_t^2} = \frac{\partial \Delta}{\partial S_t} = \frac{\partial}{\partial S_t}\left(2S_t e^{-4S_t} - 4S_t^2 e^{-4S_t}\right)$$

Using the product rule on each term in the brackets gives:

$$\Gamma = 2e^{-4S_t} - 8S_t e^{-4S_t} - 8S_t e^{-4S_t} + 16S_t^2 e^{-4S_t}$$

$$= 2e^{-4S_t}\left(1 - 8S_t + 8S_t^2\right)$$

(ii)(a) ***Current share price is 1***

When the share price is 1, the delta of the derivative is:

$$\Delta = 2\times 1\times e^{-4\times 1}(1-2\times 1) = -0.03663$$

So a short holding of one derivative requires *selling* 0.03663 shares to achieve a delta-hedged portfolio.

(ii)(b) ***Current share price is 3***

When the share price is 3, the delta of the derivative is:

$$\Delta = 2\times 3\times e^{-4\times 3}(1-2\times 3) = -0.0001843$$

So a short holding of one derivative requires *selling* 0.0001843 shares to achieve a delta-hedged portfolio.

(iii) ***Portfolio management***

The amount of rebalancing required depends on the value of gamma. If the current share price is 1, then the gamma is:

$$\Gamma = 2e^{-4\times 1}\left(1 - 8\times 1 + 8\times 1^2\right) = 0.03663$$

If the current share price is 3, then the gamma is:

$$\Gamma = 2e^{-4\times 3}\left(1-8\times 3+8\times 3^2\right)=0.000602$$

So, in the first scenario, we have a higher value of gamma and so the delta of the portfolio is more sensitive to changes in the share price. This means it is more likely to involve more rebalancing of the portfolio in the near future if the investor is to maintain a delta-hedged position.

13.2 (i) ***Define delta, gamma and theta***

Δ measures the sensitivity of the price f of a derivative to changes in the price S_t of the underlying asset:

$$\Delta = \frac{\partial f}{\partial S_t}$$ [1]

Γ measures the sensitivity of the Δ of a derivative to changes in the price S_t of the underlying asset:

$$\Gamma = \frac{\partial \Delta}{\partial S_t} = \frac{\partial^2 f}{\partial S_t{}^2}$$ [1]

Θ measures the sensitivity of the price f of a derivative to changes in time:

$$\Theta = \frac{\partial f}{\partial t}$$ [1]

[Total 3]

(ii) ***Describe delta hedging and give a numerical example***

Delta hedging involves creating a portfolio consisting of a holding of a derivative and the underlying asset, so that the delta for the portfolio is zero. [1]

This means that the value of the portfolio will not change if the price of the underlying changes by a small amount (all other factors remaining unchanged). [1]

The delta of a portfolio can be calculated as a weighted sum of the deltas of the constituents of the portfolio. [1]

Suppose, for example, that an institution has sold 1,000,000 call options on a share, each with a delta of 0.5. The delta for this portfolio would be $-0.5\times 1{,}000{,}000 = -500{,}000$, and so this portfolio is not delta-hedged. [1]

If, however, the institution also purchased 500,000 shares, the portfolio would now be delta-hedged, because the shares themselves have a delta of 1, so that the delta for the portfolio would be $1{,}000{,}000 \times (-0.5) + 500{,}000 \times 1 = 0$. A small change in the value of the shares would now make no difference to the value of the portfolio as a whole. [2]

[Total 6]

13.3 Let f denote the value of the part of the portfolio containing the relevant shares and derivatives on those shares. [½]

Let S_t denote the share price. [½]

Let σ denote the market volatility of the share price. [½]

Let t denote calendar time. [½]

(a) ***Hedge fund manager***

For a delta-neutral position, the hedge fund manager will want to have an overall delta of zero:

$$\Delta = \frac{\partial f}{\partial S_t} = 0$$ [1]

To minimise the need for rebalancing to maintain a delta-neutral position, the manager will also want to have a low gamma:

$$\Gamma = \frac{\partial^2 f}{\partial S_t^{\,2}} \approx 0$$ [1]

(b) ***Derivatives trader***

The derivatives trader will be primarily concerned about the volatility. In order for changes in the volatility not to affect the value of the options, the trader will want to have a vega close to zero:

$$\mathcal{V} = \frac{\partial f}{\partial \sigma} \approx 0$$ [1]

Note that the question refers to a change in the <u>distribution</u> of the price movements, not the values themselves.

(c) ***Pension fund trustee***

The pension fund trustee will be primarily concerned about the effect of calendar time on the value of the options. To avoid the fund value falling, the trustee will prefer to have a non-negative theta:

$$\Theta = \frac{\partial f}{\partial t} \geq 0$$ [1]

[Total 6]

13.4 Let there be x shares and y units of cash in the hedging portfolio at time t.

The value of the hedging portfolio must be equal to the value of the call option at time t. This gives:

$$240x + y = 20.15 \qquad (1)$$

The delta of the hedging portfolio must be equal to the delta of the call option at time t. This gives:

$$x \times \Delta_{share} + y \times \Delta_{cash} = 0.558$$

Now:

$$\Delta_{share} = \frac{\partial S_t}{\partial S_t} = 1 \quad \text{and} \quad \Delta_{cash} = 0$$

giving $x = 0.558$.

Using equation (1), this gives $y = -113.77$.

So the hedging portfolio consists of 0.558 shares and cash of -113.77p.

14

The binomial model

Syllabus objectives

6.1 Option pricing and valuations

6.1.4 Show how to use binomial trees and lattices in valuing options and solve simple examples.

6.1.5 Derive the risk-neutral pricing measure for a binomial lattice and describe the risk-neutral pricing approach to the pricing of equity options.

6.1.6 Explain the difference between the real-world measure and the risk-neutral measure. Explain why the risk-neutral pricing approach is seen as a computational tool (rather than a realistic representation of price dynamics in the real world).

6.1.7 State the alternative names for the risk-neutral and state price deflator approaches to pricing.

6.1.11 Describe and apply in simple models, including the binomial model and the Black-Scholes model, the approach to pricing using deflators and demonstrate its equivalence to the risk-neutral pricing approach.

0 Introduction

In this chapter, we start to develop simple models that can be used to value derivatives. In particular, we use *binomial trees* or *lattices* to find the value at time 0 of a derivative contract that provides a payoff at a future date based on the value of a non-dividend-paying share at that future date.

The analysis throughout this and the subsequent chapters applies the no-arbitrage principle in order to value derivative contracts. The basic idea is that if we can construct a portfolio that replicates the payoff from the derivative under every possible circumstance, then that portfolio must have the same value as the derivative. So, by valuing the replicating portfolio we can value the derivative.

The models discussed in this chapter represent the underlying share price as a stochastic process in discrete time and with a discrete state space – in fact a geometric random walk. In subsequent chapters, we will discuss the continuous-time and continuous-state space analogue, geometric Brownian motion (or the lognormal model), which can be interpreted as the limiting case of the binomial model as the size of the time steps tends to zero.

1 Background

Here we will consider a model for stock prices in discrete time; the first point at which we introduce a model. The model will seem very simple and naïve at first sight, but:

(a) it introduces the key concepts of financial economic pricing and

(b) it leads us to the celebrated Black-Scholes model as a limiting case.

1.1 Assumptions

In the binomial model it is assumed that:

- **there are no trading costs**
- there are no taxes
- **there are no minimum or maximum units of trading**
- **stock and bonds can only be bought and sold at discrete times 1, 2, ...**
- the principle of no arbitrage applies.

As such the model appears to be quite unrealistic. However, it does provide us with good insight into the theory behind more realistic models. Furthermore, it provides us with an effective *computational* tool for derivatives pricing.

As well as being used to determine the derivative price, the principle of no arbitrage leads to a constraint on the parameters used in the binomial model.

1.2 Definitions

The share price process

We will use S_t to represent the price of a non-dividend-paying stock at discrete time intervals $t\ (t = 0,1,2,...)$. For $t > 0$, S_t is random.

Note that in this instance 'stock' specifically means a share or equity as opposed to a bond. For the time being we ignore the possibility of dividends, which would otherwise unnecessarily complicate matters.

The stock price S_t is assumed to be a random or stochastic process. Over any discrete time interval from $t-1$ to t, we assume that S_t either goes up or goes down. We also assume that we cannot predict beforehand which it will be and so future values of S_t cannot be predicted with certainty. We will, however, be able to attach probabilities to each possibility and we also assume that the sizes of the jumps up or down are known.

The cash process

Besides the stock, we can also invest in a bond or a cash account which has value B_t at time t per unit invested at time 0. This account is assumed to be risk-free and we will assume that it earns interest at the constant risk-free continuously compounding rate of r per annum. Thus, $B_t = e^{rt}$.

We usually assume that $r > 0$, so that:

- $B_t > B_0 = 1$ and $B_t > B_{t-1}$ for $t > 0$
- B_t increases in an entirely predictable manner as we move through time, *ie* at the continuously compounding rate of r per time period.

At all points in time there are no constraints (positive or negative) on how much we can hold in stock or cash.

2 The one-period model

2.1 Basic structure

The aim in this chapter is to find the value at an arbitrary time 0 – *ie now* – of a derivative that provides a payoff at some future date based on the value of the stock at that future date. As a starting point, we consider a one-period binomial model. In this model, we start at time $t = 0$, when the stock price is equal to S_0. Over the one-period time interval to $t = 1$, the stock price will do one of two things. It will either:

1. jump upwards or
2. jump downwards.

Here we are trying to find the value of a derivative that pays out an amount that depends directly on the value of the stock price at time 1, *ie* on S_1.

We have two possibilities for the price at time 1:

$$S_1 = \begin{cases} S_0 u & \text{if the price goes up} \\ S_0 d & \text{if the price goes down} \end{cases}$$

Here u is a fixed number bigger than 1 and d is a fixed number less than 1.

This is represented in Figure 14.1 (part of the Core Reading).

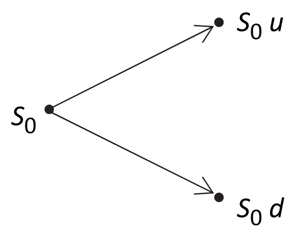


Figure 14.1: One-period binomial model for stock prices

This figure shows a one-period binomial model for the stock price. *One-period* because we consider only the time interval from time 0 to time 1, *binomial* because there are only two possible ways in which the stock price can move, *ie* up to $S_0 u$ or down to $S_0 d$.

Such a model is referred to as a *binomial tree* and each of the paths from S_0 to $S_0 u$ and from S_0 to $S_0 d$ is referred to as a *branch*. The points at each end of the branches are sometimes known as *nodes*.

We can now see the implication of there being no arbitrage in this model.

In order to avoid arbitrage we must have $d < e^r < u$. Suppose that this is not the case.

For example, if $e^r < d < u$, then we could borrow £1 of cash and buy £1 of stock. At time 0 this would have a net cost of £0. At time 1 our portfolio would be worth:

- $d - e^r$ **or**
- $u - e^r$

both of which are greater than 0. This is an example of arbitrage.

Or more generally, if one asset (*eg* the share) is certain to earn more over the one step than the other (*eg* the cash), then we simply borrow the lower-yielding asset to invest in the higher-yielding asset as many times over as we like, thus making an arbitrage profit. The only way to avoid this scenario is if the share performs either 'better' or 'worse' than the cash, *ie* $d < e^r < u$.

2.2 Determination of the derivative price at time 0

Suppose that we have a derivative which pays c_u if the price of the underlying stock goes up and c_d if the price of the underlying stock goes down.

So, the value of the payment made by the derivative at time 1, which we can denote by the random variable C_1, depends on the underlying stock price at time 1.

Question

Write down an expression for C_1.

Solution

C_1 is the derivative payoff at time 1, which is equal to:

$$C_1 = \begin{cases} c_u & \text{if the stock price went up (to } S_0 u) \\ c_d & \text{if the stock price went down (to } S_0 d) \end{cases}$$

At what price should this derivative trade at time 0?

At time 0, suppose that we hold ϕ units of stock and ψ units of cash. The value of this portfolio at time 0 is V_0.

So: $V_0 = \phi S_0 + \psi$

The portfolio is sometimes represented as an ordered pair (ϕ, ψ). ϕ and ψ are the Greek letters 'phi' and 'psi'.

At time 1, the same portfolio has the value:

$$V_1 = \begin{cases} \phi S_0 u + \psi e^r & \text{if the stock price went up} \\ \phi S_0 d + \psi e^r & \text{if the stock price went down} \end{cases}$$

eg if the stock price went up, then the value of one unit of the stock has increased to $S_0 u$ and the value of an initial unit in the cash account has increased to e^r.

So (ϕ, ψ) could be any values and this portfolio could consist of an amount of shares and an amount of cash in whatever quantities we choose. We are going to choose ϕ and ψ in order that the portfolio replicates the payoff of the derivative, no matter what the outcome of the share price process.

Let us choose ϕ and ψ so that $V_1 = c_u$ if the stock price goes up and $V_1 = c_d$ if the stock price goes down. Then:

$$\phi S_0 u + \psi e^r = c_u$$

and:

$$\phi S_0 d + \psi e^r = c_d$$

This choice of ϕ and ψ therefore ensures that the value of the portfolio (ϕ, ψ) at time 1 is equal to the derivative payoff whether the stock price goes up or down. Hence, by the no-arbitrage principle, the value of this portfolio at time 0, V_0, must also be the value of the derivative contract at that time. So, if we can solve these simultaneous equations to find ϕ and ψ, then we can determine V_0, which must be equal to the value of the derivative at time 0.

So, we have two linear equations in two unknowns, ϕ and ψ.

The easiest way to solve these equations is to subtract them. This enables us to find ϕ. We can then substitute into either equation to find ψ.

We solve this system of equations and find that:

$$\phi = \frac{c_u - c_d}{S_0(u - d)}$$

$$\psi = e^{-r}\left(\frac{c_d u - c_u d}{u - d}\right)$$

$$\Rightarrow \qquad V_0 = \phi S_0 + \psi$$

We now notice that all the terms in the numerators for ϕ and ψ involve either c_u or c_d. Therefore it is possible to rearrange V_0 so that it is a sum of a multiple of c_u and a multiple of c_d.

$$
\begin{aligned}
V_0 &= \frac{c_u - c_d}{u-d} + \frac{e^{-r}(c_d u - c_u d)}{u-d} \\
&= \frac{1}{u-d}\left[(1-de^{-r})c_u + (ue^{-r}-1)c_d\right] \\
&= \frac{e^{-r}}{u-d}\left[(e^r - d)c_u + (u-e^r)c_d\right] \\
&= e^{-r}\left[\frac{e^r-d}{u-d}c_u + \frac{u-e^r}{u-d}c_d\right]
\end{aligned}
$$

ie $\quad \boldsymbol{V_0 = e^{-r}\left(qc_u + (1-q)c_d\right)}$

where $\boldsymbol{q = \dfrac{e^r - d}{u-d}}$ **and** $\boldsymbol{1-q = \dfrac{u-e^r}{u-d}}$**.**

Note that the no-arbitrage condition $d < e^r < u$ ensures that $0 < q < 1$.

Recall that the formula for V_0 must, by the no-arbitrage principle, be equal to the price of the derivative at time 0. Since $0 < q < 1$, let's just 'pretend' for now that q is a probability. If we do this then we can see that the right-hand side simplifies to:

$$V_0 = e^{-r}E\left[C_1\right]$$

ie, the discounted value of the expected derivative payoff at time 1. However, we must remember we are just 'pretending' that q is a probability when in actual fact it isn't a real-world probability, so we add a Q to the notation to remind ourselves:

$$V_0 = e^{-r}E_Q\left[C_1\right]$$

We will see that, although it doesn't reflect reality, this 'pretending' idea is a good way of thinking about it.

If we denote the payoff of the derivative at $t = 1$ by the random variable C_1, we can write:

$$\boldsymbol{V_0 = e^{-r}E_Q\left[C_1\right]}$$

where Q is an artificial probability measure which gives probability q to an upward move in prices and $1-q$ to a downward move. We can see that q depends only upon u, d and r and not upon the potential derivative payoffs c_u and c_d. Note, by convention $E_Q\left[Y \middle| F_0\right]$ is sometimes written as $E_Q\left[Y\right]$.

We can think of a probability measure as a set of probabilities (more about this later). Note also that Q doesn't depend in any way on the real-world probabilities.

This is a surprising result. We might have thought at the start that the price of the derivative should be some sort of discounted value of the expected payoff. The above result is telling us that this is the case, but the probabilities used to calculate the expectation are not the real-world probabilities of the two possible derivative values. Instead we calculate the expectation using the 'probabilities' q and $1-q$.

2.3 Probability measures

General definition

Within the context of a binomial tree, a *probability measure* is simply a function that assigns a real number in the interval [0,1] to each of the branches in our binomial tree such that at any node they sum to 1. We can thus interpret each number as a probability.

Note carefully that the probabilities assigned by a probability measure do *not* need to correspond to the real-world probabilities that the stock price *actually* moves up or down between times 0 and 1. In fact, any function that assigns a value $q \in [0,1]$ to the up-branch and $1-q$ to the down-branch is a possible probability measure. Q is therefore the particular measure, amongst many possible measures, that happens to assign the probability $q = \dfrac{e^r - d}{u-d}$ to the up-branch between times 0 and 1, and $1-q = \dfrac{u-e^r}{u-d}$ to the down-branch.

$E_Q(C_1)$ therefore represents the expectation of the derivative payoff C_1 with respect to the probability measure Q – *ie* the expectation of C_1 based on an up-probability of q and a down-probability of $1-q$.

We could equally evaluate the expectation of C_1 based on a different probability measure, P say, which assigns a different probability $0 \le p \le 1$ to the up-branch. In that case, $E_P(C_1)$ would be the expectation of C_1 with respect to P:

ie $\qquad E_P(C_1) = p c_u + (1-p) c_d$

We could in fact use any probability measure we like to calculate an expectation – including the real-world probabilities – with different measures producing different expectations. The probability measure Q is, however, especially useful as it enables us to calculate the derivative price.

In fact Q is called the risk-neutral probability measure as we will see later.

Replicating portfolio

The portfolio (ϕ, ψ) is called a *replicating portfolio* because it replicates, precisely, the payoff at time 1 on the derivative without any risk.

In other words, $V_1 = c_u$ if the stock price goes up and $V_1 = c_d$ if the stock price goes down. So, there is *never* any risk or possibility that $V_1 \neq C_1$. Hence, in a world that is free of arbitrage, the values of the derivative and the replicating portfolio (ϕ,ψ) at time 0 must be equal – *ie* they are both equal to $V_0 = e^{-r}E_Q(C_1)$.

It is also a simple example of a *hedging strategy*: that is, an investment strategy which reduces the amount of risk carried by the issuer of the contract. In this respect not all hedging strategies are replicating strategies.

A replicating portfolio will always precisely reproduce the relevant payoff or cashflow. A hedging portfolio aims to reduce the amount of risk relating to a derivative strategy, but is not guaranteed to reproduce the payoff or cashflow precisely. Furthermore, a replicating portfolio is only a hedging portfolio if the position taken in it is opposite to that of the payoff or cashflow which it aims to reproduce.

So, if we hold the portfolio (ϕ,ψ) and sell the derivative, then the total value of the resulting portfolio at time 1 will be zero, however the stock price moves over the interval to time 1. An immediate consequence of this is that in an arbitrage-free world the value of the combined portfolio is zero at time 0. Hence, the value of the combined portfolio does not change from zero as we move from $t = 0$ to $t = 1$ and so we are said to have a perfectly-hedged position. This will also be true if we instead sell the portfolio (ϕ,ψ) and hold the derivative.

It is also possible to have an intermediate position where the risk is reduced but not eliminated. This would be the case, for example, if we sold the derivative and held the portfolio $(\tfrac{1}{2}\phi,\tfrac{1}{2}\psi)$. This portfolio is *not* a replicating strategy.

The real-world probability measure, *P*

Up until now we have not mentioned the *real-world* probabilities of up and down moves in prices. Let these be p and $1-p$ where $0 < p < 1$, defining a probability measure *P*.

So *P* is a set of probabilities that assigns the *actual* or *real-world* probability p of an upward jump in the stock price to an up-branch of the binomial tree and $1-p$ to a down-branch.

Other than by total coincidence, p will not be equal to q.

This is because p is the actual probability of the stock price moving upwards, whereas q is simply a number defined as:

$$q = \frac{e^r - d}{u - d}$$

So, q depends upon u, d and r, but not p.

When we wish to emphasise that q is *not* a real-world probability, it is often referred to as a *synthetic probability*. The important thing to note is that the real-world probability p (if indeed this can be determined) is irrelevant to our calculation of the derivative price, which is based solely on the synthetic probability q.

2.4 The risk-neutral probability measure, *Q*

Let us consider the expected stock price at time 1. Under *P* this is:

$$E_P\left[S_1\right] = S_0\left(pu + (1-p)d\right)$$

This is the expectation of the stock price at time 1 with respect to the real-world probability measure *P*.

Under *Q* it is:

$$E_Q\left[S_1\right] = S_0\left(qu + (1-q)d\right) = S_0\left(\frac{u(e^r - d)}{u-d} + \frac{d(u-e^r)}{u-d}\right) = S_0 e^r$$

Under *Q* we see that the expected return on the *risky* stock is the same as that on a risk-free investment in cash.

This is because the last equation shows that the expected stock price $E_Q\left[S_1\right]$ is equal to the accumulation of the initial stock price S_0 at the risk-free rate of return *r*. (So, the expected rate of return on the stock must be equal to the risk-free rate.)

In other words, under the probability measure *Q*, investors are neutral with regard to risk: they require no additional return for taking on more risk.

This is because both risk-free cash and the risky stock are priced so as to yield the same expected return with respect to the measure *Q*. So if we use the measure *Q* this is equivalent to assuming that investors do not require an additional risk premium to compensate them for the additional risk that they incur when investing in the risky stock, *ie* they are risk-neutral.

Note carefully that we are *not* saying that investors in the real world are risk-neutral – or equivalently that investors are risk-neutral under the real-world probability measure *P*. We are simply saying that they can be assumed to be risk-neutral under the non-real-world probability measure *Q*.

This is why *Q* is sometimes referred to as a *risk-neutral probability measure.*

It is the probability measure with respect to which any asset, whether risky or risk-free, offers the same expected return to investors, namely, the risk-free rate of return.

Under the real-world measure *P*, the expected return on the stock will not normally be equal to the return on risk-free cash. Under normal circumstances, investors demand higher expected returns in return for accepting the risk in the stock price. So we would normally find that $p > q$. However, this makes no difference to our analysis.

Question

Show why we would normally find that $p > q$.

Solution

The probability measure Q is the set of probabilities such that all assets, whether risky or risk-free, give the same expected return, namely the risk-free rate.

Under normal circumstances, investors demand higher expected returns in return for accepting the risk inherent in the random stock price. So, the expected return calculated with respect to the real-world probability measure P should be higher than that based on the risk-neutral probability measure Q, so as to compensate investors for the additional risk inherent in the stock. In order to generate this additional expected return, the expected stock price at time 1 must be higher under P than under Q, *ie*:

$$E_P[S_1] > E_Q[S_1]$$

$$\Leftrightarrow \quad S_0\left(pu+(1-p)d\right) > S_0\left(qu+(1-q)d\right)$$

$$\Leftrightarrow \quad pu-pd > qu-qd$$

$$\Leftrightarrow \quad (p-q)(u-d) > 0$$

Given that $u-d>0$ by assumption, we would therefore normally find that $p>q$.

2.5 Numerical example

The above computations will become clearer if we consider a simple numerical example.

Let's consider a one-period binomial model of a stock whose current price is 40. Suppose that:

- over the single period under consideration, the stock price can either move up to 60 or down to 30
- the actual probability of an up-movement is equal to ½
- the continuously compounded risk-free rate of return is 5% per time period
- we wish to find the current value of a one-period European call option, V_0, that has an exercise price of 45.

The binomial tree in respect of the stock price is therefore as follows:

$$\begin{tikzcd}
 & S_1=60 \\
S_0=40 \arrow[ur] \arrow[dr] & \\
 & S_1=30
\end{tikzcd}$$

We can likewise consider a corresponding binomial tree in respect of C_1, the payoff provided by the call option at time 1 – *ie* the profit paid at exercise:

$$C_0 = ? \quad \begin{array}{l} \nearrow \; C_1 = c_u = 60 - 45 = 15 \\ \\ \searrow \; C_1 = c_d = 0 \end{array}$$

Recall that V_0, the value of the replicating portfolio (ϕ, ψ) at time 0, must be equal to C_0, the derivative value at time 0.

Question

Explain why $c_d = 0$.

Solution

$c_d = 0$ because if the stock price goes down over the single period under consideration, then it ends up at 30. As this is less than the exercise price of 45, the call option will not be exercised and will expire worthless. Hence, the option payoff is then equal to 0.

In order to find V_0, and hence C_0, we can calculate the risk-neutral probabilities, q and $1 - q$, and then use the result that:

$$V_0 = e^{-r} E_Q(C_1)$$

ie
$$V_0 = e^{-r} \left[q\, c_u + (1-q) c_d \right]$$

Recall from above that the risk-neutral probability q is equal to:

$$q = \frac{e^r - d}{u - d}$$

In this instance:

$$d = \frac{30}{40} = 0.75$$

and:
$$u = \frac{60}{40} = 1.5$$

So:
$$q = \frac{e^{0.05} - 0.75}{1.5 - 0.75} = 0.40169$$

and:
$$1 - q = 0.59831$$

Hence:

$$V_0 = e^{-0.05}\left[0.40169 \times 15 + 0.59831 \times 0\right] = 5.732$$

ie $C_0 = 5.732$

Question

Find the constituents of the replicating portfolio (ϕ, ψ) and show that it costs 5.732 to set up this portfolio.

Solution

Recall from above that:

$$\phi = \frac{c_u - c_d}{S_0(u-d)}$$

and $$\psi = e^{-r}\left(\frac{c_d u - c_u d}{u-d}\right)$$

So, using the relevant values gives:

$$\phi = \frac{15-0}{40(1.5-0.75)} = 0.5$$

$$\psi = e^{-0.05}\left(\frac{0 \times 1.5 - 15 \times 0.75}{1.5 - 0.75}\right) = -14.268$$

Hence, the replicating portfolio consists of:

- a positive holding of half a share
- a negative holding of 14.268 of cash – *ie* we borrow cash.

Alternatively, we could have solved the simultaneous equations:

$$60\phi + \psi e^{0.05} = 15 \text{ and } 30\phi + \psi e^{0.05} = 0$$

Finally, the cost of the replicating portfolio is equal to:

$$V_0 = \phi S_0 + \psi$$

and using the relevant values from above gives:

$$V_0 = 0.5 \times 40 - 14.268 = 5.732$$

as required.

Three final points to note are that:

1. We did not use the actual or real-world probability of $p = ½$ to find V_0, *ie* V_0 is independent of p.

2. $p = ½ > 0.40169 = q$, as noted above.

3. V_0 differs from the discounted value of the expected derivative payoff based on the actual probability of $p = ½$, which is equal to:

$$V_0' = e^{-0.05}\left[0.5\times 15+0.5\times 0\right] = 7.134$$

Note that $V_0' > V_0$. This is because $p > q$, and so the real-world probability measure places greater weight upon the good outcome when the share price increases to 60.

3 Two-period binomial tree

3.1 Basic structure

The one-period binomial model is a good starting point for our analysis of derivative pricing. The stochastic processes underlying stock price movements will, however, typically be more complex than can be represented by the one-period model. We therefore need to extend the previous model into a multi-period context. The obvious next step is to develop a two-period binomial model. It turns out that the ideas discussed above carry over quite naturally to the two-period model.

To keep things simple, we assume that the periods we are dealing with are years.

Warning: The notation used in this section makes the calculations look more complicated than they really are. When carrying out numerical calculations using a diagram, it's actually quite a straightforward process.

We now look at a two-period binomial model. The most general model is as follows (see Figure 14.2, part of the Core Reading):

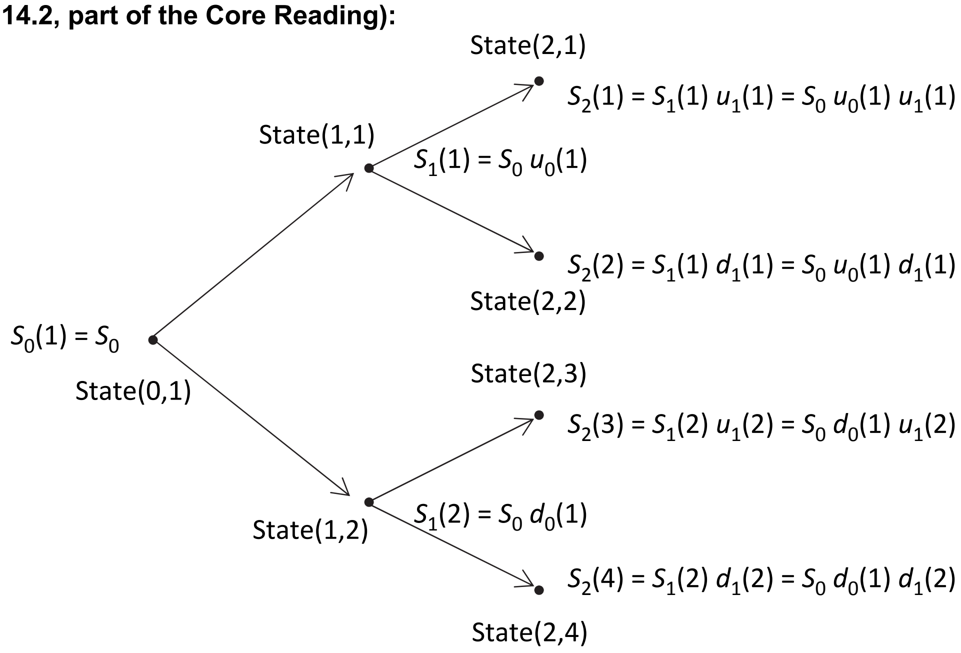


Figure 14.2: Two-period binomial model. Where the price at a particular node is denoted $S_t(j)$, this means the price in state (t, j).

The subscripts here denote time and the arguments in brackets denote the vertical position, counting from the top of the tree.

Period 1:

$$S_0 \rightarrow \begin{cases} S_0\ u_0(1) & \text{if up, to state (1,1)} \\ S_0\ d_0(1) & \text{if down, to state (1,2)} \end{cases}$$

The first period is essentially the same as our previous one-period model. The stock price therefore starts at S_0 at time 0 and either goes up or down over the period from $t=0$ to $t=1$. We do, however, need to be more careful with our notation once we move into a multi-period model.

Accordingly, the stock price is assumed to start from an initial state or node that is referred to as State (0,1) – *ie State 1 at time 0*. There is, of course, only one possible state at time 0, defined by the stock price S_0. The stock price then changes over the year from time 0 to time 1. It either moves along the up-branch to State (1,1), the first state (or up-node) at time 1, or down to State (1,2), the second state (or down-node) at time 1.

The stock price at State (1,1) following an up-movement is therefore $S_0\, u_0(1)$, where $u_t(j)$ denotes the proportionate increase in the stock price in an up-movement over the time interval from time t to time $t+1$, starting from State (t, j). Similarly, should the stock price move down to State (1,2), then it will be equal to $S_0\, d_0(1)$, where $d_t(j)$ is defined similarly to $u_t(j)$ but instead denotes the proportionate decrease in the stock price.

Period 2:

From state $(1,1)$, *ie* following a price *increase* in the first time interval:

$$S_0 u_0(1) \rightarrow \begin{cases} S_0\ u_0(1)\, u_1(1) & \text{if up, to state (2,1)} \\ S_0\ u_0(1)\, d_1(1) & \text{if down, to state (2,2)} \end{cases}$$

So, for example, if the stock price falls in the second period, then it moves from State (1,1) to State (2,2) via the down-branch $d_1(1)$.

From state $(1,2)$, *ie* following a price *decrease* in the first time interval:

$$S_0 d_0(1) \rightarrow \begin{cases} S_0\ d_0(1)\, u_1(2) & \text{if up, to state (2,3)} \\ S_0\ d_0(1)\, d_1(2) & \text{if down, to state (2,4)} \end{cases}$$

Recall that $u_1(2)$ denotes the up-branch price ratio from time 1 to time 2, starting at State (1,2).

3.2 Determination of the derivative price at time 0

Suppose that the derivative gives a payoff at time 2 of $c_2(j)$ if the price of the stock at time 2 is in state $(2, j)$.

How do we calculate the price of the derivative at time 0?

The approach used turns out to be the natural extension of that used in the one-period model. Here we first find the intermediate price at time 1 and then use this price to find the price at time 0.

We do this by working backwards from time 2. So we calculate the value of the contract at time 1 for each of the possible states at time 1.

These two states are State (1,1) following an upward movement in the first period and State (1,2) following a price fall.

Let $V_1(j)$ be the value of the contract if we are in state *j* at time 1. Then, by analogy with the one-period model:

$$V_1(1) = e^{-r}\left(q_1(1)c_2(1) + \left(1 - q_1(1)\right)c_2(2)\right)$$

$$V_1(2) = e^{-r}\left(q_1(2)c_2(3) + \left(1 - q_1(2)\right)c_2(4)\right)$$

where: $q_1(1) = \dfrac{e^r - d_1(1)}{u_1(1) - d_1(1)}$ **and:** $q_1(2) = \dfrac{e^r - d_1(2)}{u_1(2) - d_1(2)}$

So, the value of the derivative at time 1 is equal to the expectation at time 1 of the derivative payoff at time 2, calculated with respect to the risk-neutral probability measure *Q* and discounted at the risk-free rate of return.

No-arbitrage conditions imply that $d_t(j) < e^r < u_t(j)$ (and hence $0 < q_t(j) < 1$) for all *t* and *j*.

Question

Explain why no-arbitrage conditions imply that $d_t(j) < e^r < u_t(j)$.

Solution

This is the same as the restriction described earlier. Suppose that this is not the case. Say that at time 1 we are in State $(1, j)$ and $e^r < d_1(j) < u_1(j)$. Then we can buy a share for $S_1(j)$ and borrow the amount of cash needed to pay for this. At time 1, this would have a net cost of £0. At time 2, our portfolio will be worth either $(d_1(j) - e^r)S_1(j)$ or $(u_1(j) - e^r)S_1(j)$ both of which are greater than 0 – according to our assumption. So, we have a violation of the no-arbitrage condition.

The price at time 0 is found by treating the values $V_1(1)$ and $V_1(2)$ in the same way as derivative payoffs at time 1.

So:

$$V_0(1) = e^{-r}\left[q_0(1)V_1(1) + \left(1 - q_0(1)\right)V_1(2)\right]$$

Combining the two steps above we can then conclude that the value of the derivative at time 0 is equal to the expectation at time 0 of the derivative payoff *at time 2*, calculated with respect to the risk-neutral probability measure Q and discounted at the risk-free rate of return over two time periods. We return to this below.

Let C_2 be the random derivative payoff at time 2 (that is, it takes one of the values $c_2(j)$ for $j = 1,2,3,4$).

Let V_t be the random value of the contract at time *t*.

Let F_t be the history of the process up to and including time *t* (that is, the sigma-algebra generated by the sample paths up to and including time *t*) and let *F* be the sigma-algebra generated by all sample paths (up to the final time considered by the model).

'Sigma-algebra' is a technical term that is not important here. Recall that F_t, which we called the *filtration* in the chapter on Brownian motion and martingales, is basically the information known about the process S_t by time *t*. *F* is often used to denote the information known by the final time of the model, in this case the payoff date of the derivative at time 2.

Amongst the information that F_t gives us is:

- our current position in the binomial tree, State (t, j)
- the current stock price $S_t(j)$.

So, $V_t(j)$, the value of the derivative contract if we are in State *j* at time *t*, must depend on F_t.

Let *Q* be the probability measure generated by the probabilities $q_0(1)$, $q_1(1)$ and $q_1(2)$.

That is, let *Q* be the risk-neutral probability measure.

Then:

$$V_1(1) = e^{-r}E_Q\left[C_2 \mid \text{up in year } 1\right]$$

$$V_1(2) = e^{-r}E_Q\left[C_2 \mid \text{down in year } 1\right]$$

or $$V_1 = e^{-r}E_Q\left[C_2 \mid F_1\right]$$

The formula:

$$V_1 = e^{-r} E_Q[C_2 | F_1]$$

says that the no-arbitrage value of the derivative at time 1 is equal to:

- the expectation of the derivative payoff at time 2,
- calculated with respect to both the risk-neutral probability measure *Q* and the information set F_1, generated by the history of the stock price movements up to and including time 1,
- discounted at the continuously compounded risk-free rate of return, *r*.

Likewise:

$$\begin{aligned} \boldsymbol{V_0} &= \boldsymbol{e^{-r} E_Q\left[V_1 | F_0\right]} \\ &= \boldsymbol{e^{-r} E_Q\left[e^{-r} E_Q\left\{C_2 | F_1\right\} | F_0\right]} \\ &= \boldsymbol{e^{-2r} E_Q\left[C_2 | F_0\right]} \end{aligned}$$

So, the no-arbitrage value of the derivative at time 0 is equal to:

- the expectation of the derivative payoff at time 2
- calculated with respect to both the risk-neutral probability measure *Q* and the information set F_0, generated by the history of the stock price movements up to and including time 0,
- discounted at the continuously compounded risk-free rate of return, *r*.

Note that the final expression above is derived by applying the so-called *tower property* of conditional expectations. This states that for any random variable *X*, probability measure *P* and sigma-algebra F_i:

$$E_P[E_P\{X|F_j\}|F_i] = E_P[X|F_i]$$

for $i \leq j$. This is a generalisation of the formula $E[E[A|B]] = E[A]$.

Numerical example

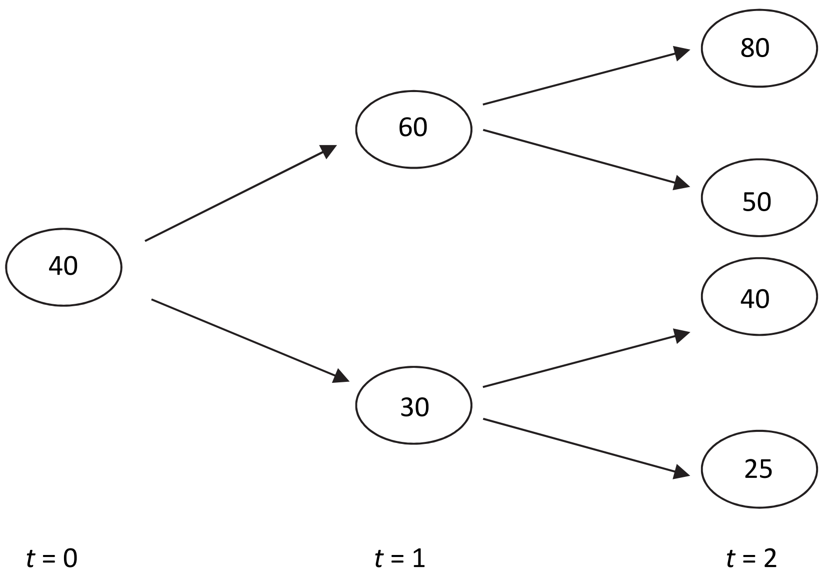
Again a numerical example should help to clarify matters. This example extends the previous one-period example to a two-period scenario.

Let's now consider a two-period binomial model of a stock whose current price is 40 (as before). Suppose that:

- Over the first period, the stock price can either move up to 60 or down to 30 (as before).
- Following an up-movement in the first period, the stock price can either move up to 80 or down to 50.
- Following a down-movement in the first period, the stock price can either move up to 40 or down to 25.
- The real-world probability of an up-movement is always equal to ½ (as before).
- The continuously compounded risk-free rate of return is 5% per time period.
- We wish to find V_0, the current value of a two-period European call option, that has an exercise price of 45.

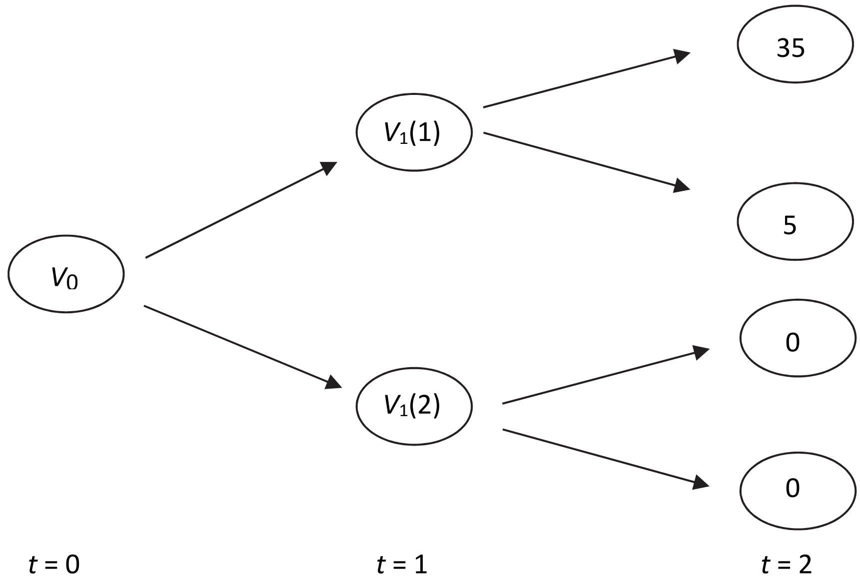
The binomial tree in respect of the *stock price* is therefore as follows:

Share prices



We can also draw a corresponding binomial tree showing the derivative value at each state or node:

Derivative values



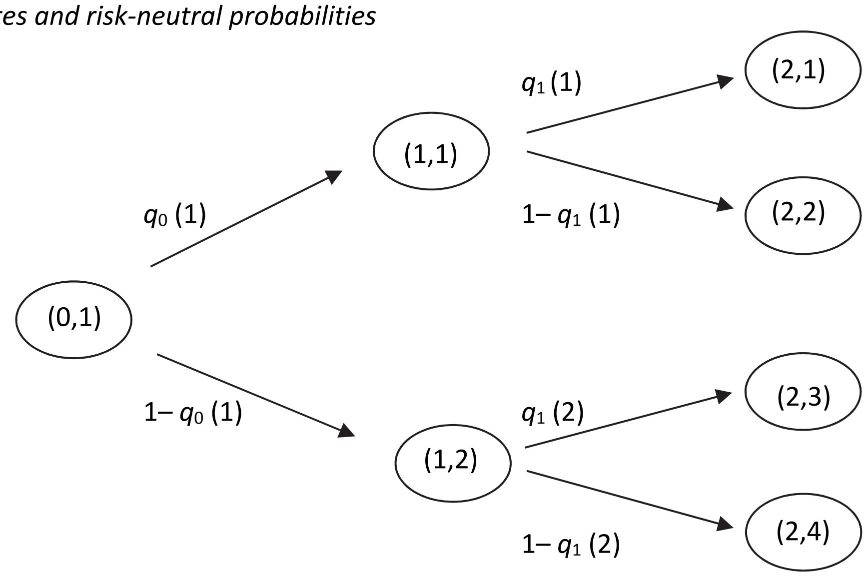
Note that the derivative value at each final node is always equal to the payoff at that node – which is certain once that particular node has been reached.

Question

Draw a further binomial tree corresponding to the above trees and annotate it with the states (t, j) and the risk-neutral probabilities.

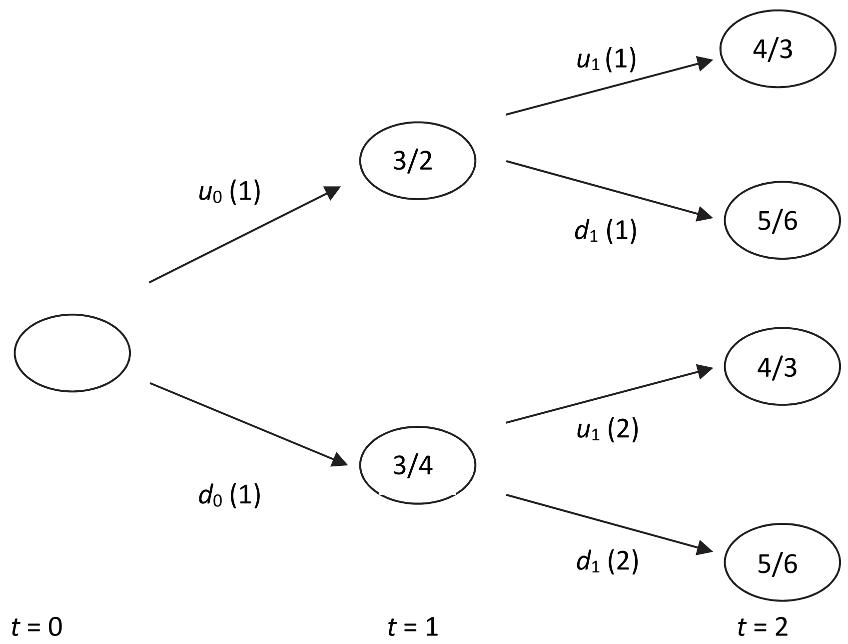
Solution

States and risk-neutral probabilities



Finally, we can also annotate the branches of the binomial tree with the price movement factors, *u* and *d* – and show the relevant values in the nodes:

Share price movements



To find V_0, the value of the derivative at time 0, we need to work backwards through the tree. In doing so, we can essentially consider the binomial tree as three distinct one-period trees – two in the second period and one in the first. For each subtree we apply the one-period approach of the previous section and then combine our results to find the two-period derivative price. We therefore proceed in three steps as follows:

1. We first determine $V_1(1)$, the derivative's value at time 1, assuming that we are then at State (1,1). This is found in exactly the same way as in the one-period model, but using the second-period risk-neutral probability $q_1(1)$.

2. We repeat the first step to find $V_1(2)$, the derivative's value at time 1, assuming that we are then at State (1,2), using the second-period risk-neutral probability $q_1(2)$.

3. We then repeat the procedure to find V_0, the derivative's value at time 0, as the discounted value of the expectation of V_1 using the first-period risk-neutral probability $q_0(1)$.

Step 1: Starting at State (1,1)

The price movement factors are $u_1(1) = 4/3$ and $d_1(1) = 5/6$. So, the risk-neutral probability is:

$$q_1(1) = \frac{e^r - d_1(1)}{u_1(1) - d_1(1)} = \frac{e^{0.05} - 5/6}{4/3 - 5/6} = 0.43588$$

Starting from State (1,1), the possible derivative payoffs at time 2 are $c_2(1) = 35$ and $c_2(2) = 5$.

Hence:

$$V_1(1) = e^{-r}\left(q_1(1)c_2(1) + (1 - q_1(1))c_2(2)\right)$$

ie $V_1(1) = e^{-0.05}\left(0.43588 \times 35 + (1 - 0.43588) \times 5\right) = 17.195$

Step 2: Starting at State (1,2)

The price movement factors are $u_1(2) = 4/3$ and $d_1(2) = 5/6$.

Starting from State (1,2), the possible derivative payoffs at time 2 are $c_2(3) = 0$ and $c_2(4) = 0$.

So, the risk-neutral probability is:

$$q_1(2) = \frac{e^r - d_1(2)}{u_1(2) - d_1(2)} = \frac{e^{0.05} - 5/6}{4/3 - 5/6} = 0.43588$$

Note that this answer is the same as in the first step, because the *u* and *d* price movement factors happen to be the same. This will not always be the case.

Finally, we have that:

$$V_1(2) = e^{-r}\left(q_1(2)c_2(3) + (1 - q_1(2))c_2(4)\right)$$

ie $V_1(2) = e^{-0.05}\left(0.43588 \times 0 + (1 - 0.43588) \times 0\right) = 0$

This is equal to zero as both of the final possible stock prices are below the exercise price and so the derivative payoff at time 2 must be zero, given that the stock price fell in the first period.

So, the calculations in Step 2 were pointless! We have given them for completeness but in the exam you should always be aware of a chance to simplify the calculations like this.

Step 3: Starting at State (0,1)

Question

Repeat the previous steps for the first time period to show that:

(i) the risk-neutral probability is $q_0(1) = 0.40169$

(ii) the derivative price at time 0 is $V_0 = 6.570$.

Solution

(i) ***Risk-neutral probability***

For the first time period, the price movement factors are $u_0(1) = 3/2$ and $d_0(1) = 3/4$.

So, the risk-neutral probability is:

$$q_0(1) = \frac{e^r - d_0(1)}{u_0(1) - d_0(1)} = \frac{e^{0.05} - 3/4}{3/2 - 3/4} = 0.40169$$

which is, of course, the same as in the one-period model in the previous section.

(ii) ***Derivative price at time 0***

Starting from State (0,1), the possible derivative values at time 1 are $V_1(1) = 17.195$ and $V_1(2) = 0$.

Hence:

$$V_0 = e^{-r}\left(q_0(1)V_1(1) + (1 - q_0(1))V_1(2)\right)$$

ie $$V_0 = e^{-0.05}\left(0.40169 \times 17.195 + (1 - 0.40169) \times 0\right) = 6.570$$

4 *n*-period binomial tree

The final step in the development of the binomial model approach to derivative pricing is to extend the model to a more general, multi-period or *n*-period context. Once again, the ideas discussed above carry over quite naturally into the general multi-period model.

4.1 Basic structure

We can generalise all the preceding results to *n* periods.

At time t there are 2^t possible states $(t,1),(t,2),\ldots,(t,2^t)$.

Recall that in the two-period model:

- at time 0, there is $2^0 = 1$ state
- at time 1, there are $2^1 = 2$ states
- at time 2, there are $2^2 = 4$ states.

In state (t,j) the price of the underlying stock is $S_t(j)$.

From this state the price can:

- **go up to $S_{t+1}(2j-1) = S_t(j)u_t(j)$ and state $(t+1,2j-1)$**
- **or down to $S_{t+1}(2j) = S_t(j)d_t(j)$ and state $(t+1,2j)$.**

The $2j$'s come in because of the way the states are numbered in each column. For example, if we are currently in State 3 (say) then we'll move to State 5 in the next column after an up-movement or State 6 after a down-movement.

4.2 Risk-neutral probability measure

The risk-neutral probabilities

If the risk-free rate of interest, *r*, is constant with $d_t(j) < e^r < u_t(j)$ then this induces the probabilities:

$$q_t(j) = \frac{e^r - d_t(j)}{u_t(j) - d_t(j)}$$

(as in the one-period and two-period models) of an up-move from state (t,j).

The probability measure *Q*

Putting the sample paths and the *q*–probabilities together gives us the probability measure *Q*.

We have noted before that $E_Q[S_1] = S_0 e^r$, giving rise to the use of the name *risk-neutral measure* for *Q*.

This follows from the fact that the time 0 expectation of the time 1 stock price with respect to the measure Q is equal to the accumulated value of the initial stock price at the risk-free rate.

Similarly in the *n*-period model we have:

$$E_Q\left[S_{t+1}|F_t\right]=S_t e^r$$

Question

Prove that $E_Q[S_{t+1}|F_t] = S_t e^r$ for $t=0,1,2,...$

Solution

We know that $q=\dfrac{e^r-d}{u-d}$. It follows that:

$$E_Q\left[S_{t+1}|F_t\right]=qS_t u+(1-q)S_t d=S_t\left[\frac{e^r-d}{u-d}(u-d)+d\right]=S_t e^r$$

Now applying the tower law, *ie* $E_Q[E_Q\{X|F_t\}|F_s] = E_Q[X|F_s]$ for $s\le t$:

$$E_Q[S_{t+1}|F_0] = E_Q\left[E_Q\left[S_{t+1}|F_t\right]|F_0\right]=E_Q\left[S_t e^r|F_0\right]=e^r E_Q\left[S_t|F_0\right]$$

It follows that:

$$E_Q\left[S_{t+1}\right]=E_Q\left[S_{t+1}|F_0\right]=S_0 e^{(t+1)r}$$

by induction. So the use of the expression *risk-neutral measure* for Q is still valid.

This is because the expectation with respect to Q of the stock price in $t+1$ periods' time is equal to the initial stock price accumulated at the risk-free rate of return over those $t+1$ periods.

4.3 Finding the derivative price at time 0

We can write $C_n(j)$ for the payoff under a derivative maturing at time *n* in state (n,j) and $V_t(j)$ for the price of the derivative at time *t* in state (t,j). The corresponding random variables are denoted by C_n and V_t respectively. As was shown in the two-period model, we calculate prices by starting at the maturity date and working backwards.

Thus:

$$V_n(j)=C_n(j)$$

and for $t<n$:

$$V_t=e^{-r}E_Q\left[V_{t+1}|F_t\right]$$

So, moving backwards through the $n-t$ stages of the tree from time n to time t we get:

$$= e^{-r(n-t)} E_Q\left[C_n \middle| F_t\right]$$

As we work our way backwards through the binomial tree, by analogy with the one-period model we can also construct a replicating strategy $\left(\phi_{t+1}, \psi_{t+1}\right)$ where:

$\phi_{t+1}(j)$ **= number of units of stock when in state (t, j) at time t**

$$= \frac{V_{t+1}(2j-1) - V_{t+1}(2j)}{S_t(j)\,(u_t(j) - d_t(j))}$$

and $\psi_{t+1}(j)$ **= amount held in cash when in state (t, j) at time t**

$$= e^{-r}\left(\frac{V_{t+1}(2j)\,u_t(j) - V_{t+1}(2j-1)\,d_t(j)}{u_t(j) - d_t(j)}\right)$$

$\phi_{t+1}(j)$ and $\psi_{t+1}(j)$ denote the appropriate holdings of shares and cash respectively for the replicating portfolio over the time interval from t to t+1.

Note here that the holding of $\phi_{t+1}(j)$ shares means that the difference between the value of the holding in shares when prices go up and when prices go down precisely matches the difference between the two possible values of the option.

ie $\phi_{t+1}(j)\times\left[S_t(j)\left(u_t(j)-d_t(j)\right)\right] = V_{t+1}(2j-1)-V_{t+1}(2j)$

This means that all risk has been removed.

Question

Which of the following statements would be true if we wanted to have a replicating portfolio at all times in a general n-period binomial tree?

1. Once we set up the initial replicating portfolio, we would leave it unchanged throughout.
2. There is a particular replicating strategy required for each time interval and we would have to adjust our portfolio at each step.
3. A different replicating strategy is required for every node in the tree and we would have to adjust our portfolio at each step depending on which node we are at.

Solution

Statement 3 is true. We would usually have to rebalance our replicating portfolio in a different way at each node visited.

This model, as well as those in the preceding sections, allows share prices to move in a way which is clearly much simpler than reality. However, models are always approximations to reality and the quality of a model can be gauged by how closely answers provided by the model resemble reality.

All models are, by definition, simplified characterisations of reality. They attempt to capture the important features of a particular situation in order to help us understand that situation. The strengths, weaknesses and assumptions of the model are as important as the results the model produces, and their explicit consideration often provides additional insight into the situation being modelled.

In this case, the binomial model is recognised as an effective model (provided the time to maturity is broken up into a suitable number of sub-periods) for pricing and valuing derivative contracts. In this respect, we might describe the binomial model as a good *computational tool*.

5 Recombining binomial trees

5.1 Ease of computation

The model in the previous subsection is very flexible, given that it allows for different levels of volatility when in different states.

Question

How does the previous model allow for different levels of volatility in different states?

Solution

The previous model allows for different levels of volatility in different states by allowing for different up and down price factors in different states, *ie* $u_t(j)$ and $d_t(j)$ vary with t and j.

However, the usefulness of the model is severely limited by the number of states which exist even for relatively low numbers of time periods up to maturity (that is, 2^n states), since computation times even for simple derivative securities are at best proportional to the number of states.

One solution to this problem is to assume instead that the volatility is the same in all states, so that the price ratios for the up-steps and down-steps are the same size, irrespective of where they appear in the binomial tree. This may not be an unreasonable assumption given that the steps are expressed in proportionate (or percentage), rather than absolute, terms.

Note that this now means that 2 up-steps and 1 down-step (say) over 3 time periods will take us to the same share price whatever order the steps occur in.

Suppose that we assume that the sizes of the up-steps and down-steps are the same in all states. That is:

$$u_t(j) = u \text{ and } d_t(j) = d$$

$$\Rightarrow \qquad q_t(j) = q$$

for all t, j with $d < e^r < u$ and $0 < q < 1$.

Note that the constancy of the risk-neutral probability q, follows directly from that of the step sizes, given that q is a direct function of them:

$$q_t(j) = \frac{e^r - d_t(j)}{u_t(j) - d_t(j)}$$

ie $$q_t(j) = \frac{e^r - d}{u - d} = q \text{, constant}$$

Then we have:

$$S_t = S_0 u^{N_t} d^{t-N_t}$$

where N_t is the number of up-steps between time 0 and time *t*. This means that we have $n+1$ possible states at time *n* instead of 2^n.

So, in a ten-period model, for example, the number of possible states at time 10 is reduced from 1,024 to just 11.

Consequently, computing times are substantially reduced, provided the payoff on the derivative is not path-dependent: that is, it depends upon the number of up-steps and down-steps but not their order. For such non-path-dependent derivatives we have $C_n = f(S_n)$ for some function *f*. For example, for a European call option we have $f(x) = \max\{x-K, 0\}$, where *K* is the strike price.

For a European put option, $f(x) = \max\{K - x, 0\}$ where K is the strike price.

Question

Would the following derivatives satisfy the non-path-dependent assumption mentioned here?

- Derivative A pays a cash amount in one year's time equal to the highest value the share price reached during the year.
- Derivative B pays a cash amount in one year's time equal to the average of the share price at the start and end of the year.

Solution

Derivative A is path-dependent since the highest value would be different if, for example, we had the sequences *uudd* and *dduu*.

Derivative B is not path-dependent, since an average calculated based only on the initial and final values does not depend on the particular path taken in between.

This special form for the *n*-period model allows us to call it a *recombining* binomial tree or a *binomial lattice* (see Figure 14.3).

The term *recombinant* is also used in this context.

A further implication of this model is that, unlike the non-recombining model discussed in the previous sections, there will usually be more than one route from the initial node to any particular final node.

5.2 Binomial distribution

Under this model the q-probabilities are, as stated above, equal, and all steps are made independent of one another. So the number of up-steps up to time t, N_t, has a binomial distribution with parameters t and q. Furthermore, for $0 < t < n$:

- **N_t is independent of $N_n - N_t$**

  ie the number of up-steps (and hence down-steps) in non-overlapping time intervals is independent,

- **and $N_n - N_t$ has a binomial distribution with parameters $n-t$ and q.**

The price at time t of the derivative is:

$$V_t = e^{-r(n-t)} \sum_{k=0}^{n-t} f\left(S_t u^k d^{n-t-k}\right) \frac{(n-t)!}{k!(n-t-k)!} q^k (1-q)^{n-t-k}$$

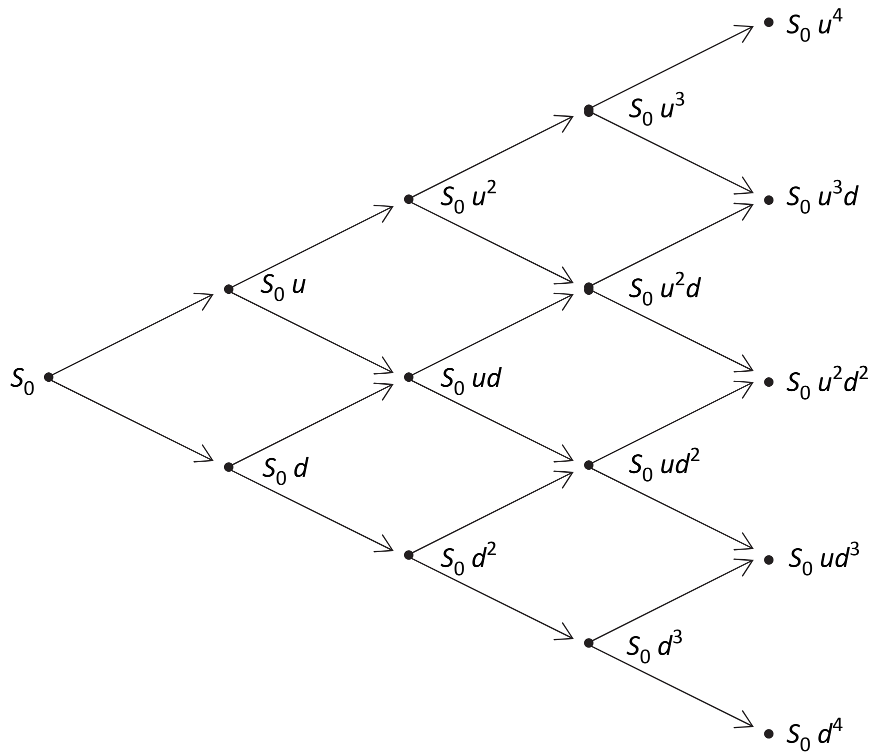


Figure 14.3: Recombining binomial tree or binomial lattice

This figure is part of the Core Reading.

To derive the above expression for V_t, we note that the general expression for the value of the derivative contract at time *t* is (irrespective of whether the tree is recombining or not):

$$V_t = e^{-r(n-t)} E_Q[C_n | F_t]$$
$$= e^{-r(n-t)} E_Q[f(S_n) | F_t]$$
$$= e^{-r(n-t)} \sum_s f(s) \times P_Q[S_n = s | F_t]$$

where the summation is over all the possible values that S_n can take and the probabilities are the risk-neutral probabilities.

If the tree *is* recombining then:

$$S_n = S_t u^{N_n - N_t} d^{(n-t)-(N_n - N_t)}$$

where $N_n - N_t$ is the number of up-steps between time t and time n. So $N_n - N_t$ can take values from 0 to $n-t$ and $N_n - N_t \sim Bin(n-t, q)$. Therefore:

$$V_t = e^{-r(n-t)} \sum_{k=0}^{n-t} f\left(S_t u^k d^{n-t-k}\right) \times P_Q[S_t u^k d^{n-t-k} | F_t]$$

or $$V_t = e^{-r(n-t)} \sum_{k=0}^{n-t} f\left(S_t u^k d^{n-t-k}\right) \frac{(n-t)!}{k!(n-t-k)!} q^k (1-q)^{n-t-k}$$

Question

Consider a binomial lattice model for a 2-month call option with an exercise price of 200. Suppose that the share price either goes up by 4% or down by 3% each month, the risk-free continuously compounded rate is ½% per month and the current share price is also 200.

Use the formula above to estimate the value of the option.

Solution

Here:

$$u = 1.04 \qquad d = 0.97 \qquad r = 0.005 \qquad n - t = 2$$

So, the risk-neutral probability is equal to:

$$q = \frac{e^r - d}{u - d} = \frac{e^{0.005} - 0.97}{1.04 - 0.97} = 0.500179$$

Hence:

Share price at exercise	Call option payoff	Probability of payoff
$200 \times 1.04^2 = 216.32$	16.32	$q^2 = 0.250179$
$200 \times 1.04 \times 0.97 = 201.76$	1.76	$2q(1-q) = 0.500000$
$200 \times 0.97^2 = 188.18$	0	$(1-q)^2 = 0.249821$

The current call option price is therefore:

$$V_0 = e^{-0.005\times 2}\left[16.32\times 0.250179 + 1.76\times 0.500000 + 0\times 0.249821\right]$$

$$= 4.914$$

6 Calibrating binomial models

It is often convenient when calibrating the binomial model to have the mean and variance implied by the binomial model correspond to the mean and variance of a lognormal distribution. The reasoning will become clearer when considering continuous-time versions in later chapters.

For recombining binomial models, an additional condition that leads to a unique solution is:

$$u = \frac{1}{d}$$

ie, an up-step and a down-step would mean the share price, after two steps, is the same as it is at time 0.

Recall the solution to the SDE for geometric Brownian motion. If $dS_t = \alpha S_t\, dt + \sigma S_t\, dB_t$, then S_t / S_0 has a lognormal distribution with parameters $\left(\alpha - \tfrac{1}{2}\sigma^2\right)t$ and $\sigma^2 t$.

Question

If S_t / S_0 has a lognormal distribution with parameters $\left(\alpha - \tfrac{1}{2}\sigma^2\right)t$ and $\sigma^2 t$, then give formulae for the mean and variance of S_t / S_0.

Solution

The formulae on page 14 of the *Tables* give us the expectation and variance of S_t / S_0:

$$E\left(\frac{S_t}{S_0}\right) = \exp\left(\left(\alpha - \frac{1}{2}\sigma^2\right)t + \frac{1}{2}\sigma^2 t\right) = e^{\alpha t}, \quad \text{var}\left(\frac{S_t}{S_0}\right) = e^{2\alpha t}(e^{\sigma^2 t} - 1)$$

If we parameterise the lognormal distribution (under the risk-neutral law) so that:

$$\ln\{S(t)/S(t_0)\} \sim N\left[\left(r - \frac{\sigma^2}{2}\right)(t - t_0), \sigma^2 (t - t_0)\right]$$

then the conditions that must be met are:

$$E\left[S(t+\delta t)/S(t)\right] = \exp(r\delta t) \qquad \textbf{(1)}$$

and $$\text{var}\left[\ln\{S(t+\delta t)/S(t)\}\right] = \sigma^2 \delta t \qquad \textbf{(2)}$$

Here:

- δt **is the time interval of each step in the binomial model**
- $S(t)$ **denotes the price of the asset at time** t **.**

Noting that:

$$E\left[S(t+\delta t)/S(t)\right]=qu+(1-q)d$$

it follows from equation (1) that:

$$qu+(1-q)d=\exp(r\delta t)$$

Rearranging to make q the subject, we get:

$$q=\frac{\exp(r\delta t)-d}{u-d}$$

Using the equation (2) and the assumption that $u=\dfrac{1}{d}$ **:**

$$\begin{aligned}\operatorname{var}\left[\ln\{S(t+\delta t)/S(t)\}\right]&=q(\ln u)^2+(1-q)(-\ln u)^2-E\left[\ln\{S(t+\delta t)/S(t)\}\right]^2\\&=(\ln u)^2-E\left[\ln\{S(t+\delta t)/S(t)\}\right]^2\end{aligned}$$

The last term involves terms of higher order than δt **.**

Using the lognormal distribution given:

$$\ln\{S(t+\delta t)/S(t)\}\sim N\left[\left(r-\frac{\sigma^2}{2}\right)\delta t,\sigma^2\delta t\right]$$

we have:

$$E\left[\ln\{S(t+\delta t)/S(t)\}\right]^2=\left(r-\frac{\sigma^2}{2}\right)^2\delta t^2$$

which tends to 0 as $\delta t\to 0$.

So, if we ignore the $E\left[\ln\{S(t+\delta t)/S(t)\}\right]^2$ **terms and equate the expression to** $\sigma^2\delta t$ **, we get:**

$$\operatorname{var}\left[\ln\{S(t+\delta t)/S(t)\}\right]=(\ln u)^2=\sigma^2\delta t$$

Then, we solve to get:

$$u = \exp(\sigma\sqrt{\delta t})$$

and hence also:

$$d = \exp(-\sigma\sqrt{\delta t})$$

When a (continuously payable) dividend is paid on the underlying asset, it is convenient and conventional to adjust the up-steps and down-steps to be:

$$u = \exp(\sigma\sqrt{\delta t} + \nu\delta t)$$

$$d = \exp(-\sigma\sqrt{\delta t} + \nu\delta t)$$

where ν is the continuously payable dividend rate.

These formulae can be found on page 45 of the *Tables* (using the letter q instead of ν). In some exam questions, explicit values for u and d are not given. In this case, we assume that $ud = 1$ and use these formulae to calculate u and d.

Question

A non-dividend-paying share has volatility $\sigma = 20\%$ *pa.* Calculate the values of u and d for the share price movements over one month.

Solution

Applying the formula for u, we get:

$$\begin{aligned} u &= \exp\left(\sigma\sqrt{\delta t} + \nu\delta t\right) \\ &= \exp\left(0.2\sqrt{1/12} + 0/12\right) \\ &= 1.0594 \end{aligned}$$

Since the dividends are zero, we have:

$$d = \frac{1}{u} = 0.9439$$

7 The state price deflator approach

In this section we will present a different, but equivalent, approach to pricing.

7.1 One-period case

Recall the one-period binomial model where:

$$V_1 = \begin{cases} c_u & \text{if } S_1 = S_0 u \\ c_d & \text{if } S_1 = S_0 d \end{cases}$$

Then:

$$\begin{aligned} V_0 &= e^{-r} E_Q[V_1] \\ &= e^{-r}\left[q c_u + (1-q) c_d \right] \end{aligned}$$

So far we have been pricing derivatives on a risk-neutral basis. Here the fair price for the derivative is the discounted value of:

'probability' share price goes up $\times$ derivative value if share price goes up

+

'probability' share price goes down $\times$ derivative value if share price goes down

We can re-express this value in terms of the real-world probability p :

$$\begin{aligned} V_0 &= e^{-r}\left[p \frac{q}{p} c_u + (1-p) \frac{(1-q)}{(1-p)} c_d \right] \\ &= E_P[A_1 V_1] \end{aligned}$$

where A_1 is a random variable with:

$$A_1 = \begin{cases} e^{-r} \dfrac{q}{p} & \text{if } S_1 = S_0 u \\ e^{-r} \dfrac{1-q}{1-p} & \text{if } S_1 = S_0 d \end{cases}$$

The expression for the value of the derivative takes the same form as before. It is the discounted value of:

'probability' share price goes up $\times$ derivative value if share price goes up

+

'probability' share price goes down $\times$ derivative value if share price goes down

However, in this case we have real-world probabilities and a different discount factor. The discount factor A_1 depends on whether the share price goes up or goes down. This means it is random and so we call it a *stochastic discount factor*.

A_1 is called a *state price deflator*. It also has a variety of other names:

- **deflator**
- **state price density**
- **pricing kernel**
- **stochastic discount factor.**

Question

Let the value of a share at time 0 be $S_0 = 100$ and let the continuously compounded risk-free rate be 3% per time period. In one time period's time the share price will either have gone up to 120 or down to 85. The real-world probability that the share price goes up is 0.6. Calculate the possible values of the state price deflator A_1.

Solution

We can calculate the risk-neutral probability q in the same way as earlier in the chapter:

$$u = \frac{120}{100} = 1.2 \text{ and } d = \frac{85}{100} = 0.85$$

$$\Rightarrow \qquad q = \frac{e^r - d}{u - d} = \frac{e^{0.03} - 0.85}{1.2 - 0.85} = 0.51558$$

We can then use the formula for the state price deflator A_1 and the real-world probability, $p = 0.6$, to calculate the value of A_1:

$$A_1 = \begin{cases} e^{-r}\dfrac{q}{p} & \text{if } S_1 = S_0 u \\ e^{-r}\dfrac{1-q}{1-p} & \text{if } S_1 = S_0 d \end{cases}$$

$$= \begin{cases} e^{-0.03}\dfrac{0.51558}{0.6} & \text{if } S_1 = 120 \\ e^{-0.03}\dfrac{0.48442}{0.4} & \text{if } S_1 = 85 \end{cases}$$

$$= \begin{cases} 0.8339 & \text{if } S_1 = 120 \\ 1.1752 & \text{if } S_1 = 85 \end{cases}$$

Note that:

(a) if $V_1 = 1$ then:

$$V_0 = E_P[A_1 \times 1] = e^{-r}$$

This is not surprising. If a derivative pays 1 in one period's time regardless of the outcome of the share price, then, by no-arbitrage, the fair price to pay for this derivative is just the discounted value of 1.

(b) if $V_1 = S_1$ then:

$$V_0 = E_P[A_1 \times S_1] = S_0$$

Again, this is not surprising. If a derivative pays the value of the share in one period's time regardless of what this turns out to be, then, by no-arbitrage, the fair price to pay for this derivative is the value of the share now.

7.2 *n*-period case, binomial lattice

We now extend the theory to the *n*-period case.

Recall that, under the risk-neutral approach to pricing, we have:

$$V_n = f\left(S_n\right)$$

Now:

$$S_n = S_0 u^i d^{n-i}$$

where i is the number of up-steps.

Over n steps the share price will go up i times and go down $n-i$ times. Each time it goes up, its price is multiplied by u and each time it goes down its price is multiplied by d.

So let us define:

$$V_n(i) = f\left(S_0 u^i d^{n-i}\right)$$

Then we have:

$$V_0 = e^{-rn} E_Q[V_n]$$

$$= e^{-rn} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k} f(S_0 u^k d^{n-k})$$

Although the algebra is heavier, this is the same form as before and is an expression in terms of the risk-neutral probability q. In the same way as the one-period case, we now re-express this in terms of the real-world probability p.

$$V_0 = e^{-rn} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \left(\frac{q}{p}\right)^k \left(\frac{1-q}{1-p}\right)^{n-k} V_n(k)$$

$$= \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} A_n(k)\, V_n(k)$$

$$= E_P[A_n V_n]$$

where $A_n = e^{-rn} \left(\frac{q}{p}\right)^{N_n} \left(\frac{1-q}{1-p}\right)^{n-N_n}$

and N_n is the number of up-steps up to time n.

So, again, the discount factor A_n is random and so we call it a stochastic discount factor.

A_n is again called the state price deflator. An important property of A_n is that for all n = 1, 2, … we have:

$$A_n = A_{n-1} \times e^{-r} \left(\frac{q}{p}\right)^{I_n} \left(\frac{1-q}{1-p}\right)^{1-I_n}$$

where $I_n = \begin{cases} 1 & \text{if } S_n = S_{n-1}u \\ 0 & \text{if } S_n = S_{n-1}d \end{cases}$

It follows that:

- $S_n = S_{n-1} u^{I_n} d^{1-I_n}$ **and**
- $N_n = \sum_{k=0}^{n} I_k$

A very important point to note is that, for this model, the risk-neutral and the state price deflator approaches give the same price V_0. Theoretically, they are the same; they only differ in the way that they present the calculation of a derivative price.

One presents it using normal discount factors and the risk-neutral probability q and the other presents it using stochastic (random) discount factors and the real-world probability p.

Finally, note that:

$$E_P[A_n] = e^{-rn}$$

If a derivative pays 1 in n periods' time regardless of the outcome of the share price, then, by no-arbitrage, the fair price to pay for this derivative is just the discounted value of 1.

We also have:

$$E_P[A_n S_n] = S_0$$

If a derivative pays the value of the share in n periods' time regardless of what this turns out to be, then, by no-arbitrage, the fair price to pay for this derivative is the value of the share now.

The state price deflator approach can be adapted to price a derivative at any time *t* and it is straightforward to show that:

$$V_t = \frac{E_P[A_T V_T]}{A_t}$$

In words, the formula above is saying that the fair price for the derivative is the expected value of the derivative payoff at time *T*, discounted back to time *t*. However, since we are taking that expectation with respect to the real-world probability P, the discount factor we use is the state price deflator A.

$\frac{A_T}{A_t}$ is analogous to the deterministic discount factor $\frac{v^T}{v^t} = v^{T-t}$, *ie* the 'present value' at time *t* of a payment at time *T*.

Chapter 14 Summary

Binomial model assumptions

- Assets may be bought and sold at integer times $t = 0,1,2,3,...$
- Assets may be held in any amount.
- There are no taxes or transaction costs.
- There are no arbitrage opportunities $\Leftrightarrow d < e^r < u$.

One-period model: replicating portfolio and risk-neutral valuation

In the one-period model:

$$\begin{aligned} V_0 &= \phi S_0 + \psi \\ &= e^{-r}\left(q\,c_u + (1-q)c_d\right) \\ &= e^{-r}E_Q[C_1] \end{aligned}$$

where:

$$\phi = \frac{c_u - c_d}{S_0(u-d)} \qquad \psi = e^{-r}\left(\frac{c_d u - c_u d}{u-d}\right)$$

$$q = \frac{e^r - d}{u-d} \qquad 1-q = \frac{u-e^r}{u-d}$$

The portfolio consisting of ϕ shares and ψ cash is a *replicating portfolio*.

Q is the *risk-neutral probability measure*, which gives the risk-neutral probability q to an upward move in prices and $1-q$ to a downward move.

The risk-neutral probabilities ensure that the underlying security yields an expected return equal to the risk-free rate.

Finding the derivative price using the risk-neutral probabilities is referred to as *risk-neutral valuation*.

Two-period model: risk-neutral valuation

Here the value of the derivative is:

$$V_0 = e^{-2r} E_Q[C_2 | F_0]$$

So, the no-arbitrage value of the derivative at time 0, is equal to:

- the expectation of the derivative payoff at time 2
- calculated with respect to both the risk-neutral probability measure Q and the information set F_0, generated by the history of the stock price movements up to and including time 0 and
- discounted at the continuously compounded risk-free rate of return.

n-period model

The results of the one-period and two-period binomial models generalise to the multi-period or n-period context. In this case:

- the risk-neutral up-step probability from State (t, j) is:

$$q_t(j) = \frac{e^r - d_t(j)}{u_t(j) - d_t(j)}$$

- the expectation of the stock price in n periods' time, calculated with respect to the risk-neutral measure Q, is equal to the current stock price, accumulated at the continuously compounded risk-free rate of return over those n periods, *ie* $E_Q[S_{t+n} | F_t] = S_t e^{rn}$
- the derivative price at time t is $V_t = e^{-r(n-t)} E_Q[C_n | F_t]$
- the number of units of stock in the replicating portfolio when in State (t, j) at time t is:

$$\phi_{t+1}(j) = \frac{V_{t+1}(2j-1) - V_{t+1}(2j)}{S_t(j)\big(u_t(j) - d_t(j)\big)}$$

- the amount held in cash in the replicating portfolio when in State (t, j) at time t is:

$$\psi_{t+1}(j) = e^{-r}\left(\frac{V_{t+1}(2j)u_t(j) - V_{t+1}(2j-1)d_t(j)}{u_t(j) - d_t(j)}\right)$$

Recombining binomial trees

A recombining binomial tree (or binomial lattice) is one in which values of u and d, and consequently the risk-neutral probabilities, are the same in all states.

With such models:

- the volume of computation required is greatly reduced
- N_t, the number of up-steps up to time t, has a binomial distribution with parameters t and q
- $S_t = S_0 u^{N_t} d^{t-N_t}$

Calibrating binomial models

If we assume that $ud = 1$, then a binomial tree model with steps of length δt, and a continuous dividend rate ν, can be calibrated to have the same mean and variance as a continuous-time model. In this instance, we need:

$$q = \frac{\exp(r\delta t) - d}{u-d} \qquad u = \exp\left(\sigma\sqrt{\delta t} + \nu\delta t\right) \qquad d = \exp\left(-\sigma\sqrt{\delta t} + \nu\delta t\right)$$

The state price deflator approach in the one-period binomial tree

The state price deflator is:

$$A_1 = \begin{cases} e^{-r}\dfrac{q}{p} & \text{if } S_1 = S_0 u \\ e^{-r}\dfrac{1-q}{1-p} & \text{if } S_1 = S_0 d \end{cases}$$

Then the fair price for the derivative is:

$$V_0 = E_P[A_1 V_1]$$

The state price deflator approach in the *n*-period binomial tree

The state price deflator is:

$$A_n = e^{-rn}\left(\frac{q}{p}\right)^{N_n}\left(\frac{1-q}{1-p}\right)^{n-N_n}$$

where N_n is the number of up-steps up to time n. The fair price for the derivative is then:

$$V_0 = E_P[A_n V_n]$$

The practice questions start on the next page so that you can keep the chapter summaries together for revision purposes.

Chapter 14 Practice Questions

14.1 Exam style

The market price of a security can be modelled by assuming that it will either increase by 12% or decrease by 15% each month, independently of the price movement in other months. No dividends are payable during the next two months. The continuously compounded monthly risk-free rate of interest is 1%. The current market price of the security is 127.

(i) Use the binomial model to calculate the value of a two-month European put option on the security with a strike price of 125. [3]

(ii) Calculate the value of a two-month American put option on the same security with the same strike price. [3]

(iii) Calculate the value of a two-month American call option on the same security with the same strike price. [2]

[Total 8]

14.2 Exam style

A company share price is to be modelled using a 5-step recombining binomial tree, with each step in the tree representing one day. Each day, it is assumed that the share price:

- increases by 2%, or
- decreases by 1%.

Assume that the risk-free force of interest is $\delta = 5.5\%\ pa$ and that there are 365 days in a year. No dividends are to be paid over the next five days.

(i) Calculate the risk-neutral probability of an up-step on any given day. [2]

(ii) Calculate the fair price of a 5-day at-the-money call option on £10,000 worth of shares in this company. [5]

A special option is available where the payoff after 5 days is:

$$\max\left\{S_5^* - K, 0\right\}$$

where S_5^* is the arithmetic average share price recorded at the end of each of the 5 days and K is the strike price.

(iii) Calculate the fair price of the special option (strike price $K = 1.06S_0$) on £10,000 worth of shares in this company. [4]

(iv) Explain whether an at-the-money special option is likely to have a higher value of vega than a standard call option. [3]

[Total 14]

14.3 In a one-step binomial tree model it is assumed that the initial share price of 260 will either increase to 285 or decrease to 250 at the end of one year. Assume that the annual risk-free force of interest is 0.05 and that no dividends are payable.

(i) Calculate the price of a one-year European call option with a strike price of 275, using each of the following:

(a) a replicating portfolio method

(b) risk-neutral valuation

(c) a risk-free portfolio method.

(ii) Repeat your calculations in (i) for a one-year European put option with a strike price of 275.

(iii) Verify numerically that the put-call parity relationship holds in this case.

14.4 The market price of a non-dividend-paying security with current market price S is being modelled using a one-step binomial tree in which the proportionate changes in the security price following an up- and a down-movement are denoted by u and d. The risk-free force of interest over the period is r.

Show that if an option on this security has a payoff of z_u following an up-movement and a payoff of z_d following a down-movement, then the option can be replicated exactly using a portfolio consisting of Δ securities, where $\Delta S = \dfrac{z_u - z_d}{u-d}$, and an amount of cash, ψ, which you should specify.

14.5 The increase in the price of a share over the next year is believed to have a mean of 10% and a standard deviation of 10%.

(i) Determine the values of u and d for a one-step binomial tree model that are consistent with the mean and standard deviation of the return on the underlying share, assuming that the share price is twice as likely to go up than to go down.

(ii) Hence calculate the value of each of the following options, given that the current share price is 250, the risk-free force of interest is 7½% per annum and dividends can be ignored:

(a) a one-year European call option with a strike price of 275

(b) a one-year European put option with a strike price of 300.

14.6 (i) Show that in a one-step binomial tree model of the price of a non-dividend-paying share, the risk-neutral probability q of an up movement is given by:

$$q = \frac{e^{r\delta t} - d}{u - d}$$

where d, u, r and δt are quantities you should define.

(ii) Explain briefly why it must be assumed that $d < e^{r\delta t} < u$.

(iii) Write down a formula for θ, the expected one-step rate of return on the share based on the real-world probability p of an up-movement.

(iv) Show that the real-world variance, σ^2, of the one-step rate of return on the share is $p(1-p)(u-d)^2$.

(v) Show that $p > q$ if and only if $1 + \theta > e^{r\delta t}$ and interpret this result.

14.7 Exam style

The movement of a share price over the next two months is to be modelled using a two-period recombining binomial model. Over each month, it is assumed that the share price will either increase or decrease by 10%.

(i) Over each month, the risk-neutral probability of an up-step is $q = 0.55$. Calculate the monthly risk-free force of interest r that has been used to arrive at this figure. [1]

(ii) The current share price is 1. The annualised expected force of return on the share is $\mu = 30\%$. Calculate the state-price deflators in each of the three possible final states of the share price. [4]

(iii) Calculate the value of each of the following two-month derivatives:

(a) a derivative with payoff profile $(1,0,0)$

(b) a derivative with payoff profile $(0,1,0)$

(c) a derivative with payoff profile $(0,0,1)$

(d) a European call option with a strike price of $K = 0.95$

(e) a European put option with a strike price of $K = 1.05$

(f) a derivative whose payoff is $2 \times |S - 0.98|$, where S is the share price at the end of the two months. [5]

A payoff profile of (x,y,z) means that the derivative returns x if the share price goes up twice, y if the share price goes up once and down once, and z if the share price goes down twice.

[Total 10]

14.8 (i) Explain the difference between a recombining and a non-recombining binomial tree. [2]

Exam style

(ii) A researcher is using a two-step binomial tree to determine the value of a 6-month European put option on a non-dividend-paying share. The put option has a strike price of 450.

During the first 3 months it is assumed that the share price of 400 will either increase by 10% or decrease by 5% and that the continuously compounded risk-free rate (per 3 months) is 0.01. During the following 3 months it is assumed that the share price will either increase by 20% or decrease by 10% and that the continuously compounded risk-free rate is 0.015 (per 3 months).

Calculate the value of the put option. [6]

(iii) The researcher is considering subdividing the option term into months. Explain the advantages and disadvantages of this modification of the model and suggest an alternative model based on months that might be more efficient numerically. [5]

[Total 13]

Chapter 14 Solutions

14.1 (i) ***Calculate the value of the European put option***

We are given $u = 1.12$, $d = 0.85$ and $r = 0.01$ in the question. The risk-neutral probability under the binomial model is:

$$q = \frac{e^r - d}{u - d} = \frac{e^{0.01} - 0.85}{1.12 - 0.85} = 0.5928$$ [½]

The possible values for the security price after two months are:

$$S_{uu} = 1.12^2 \times 127 = 159.3088$$

$$S_{ud} = S_{du} = 1.12 \times 0.85 \times 127 = 120.904$$

$$S_{dd} = 0.85^2 \times 127 = 91.7575$$ [1]

Therefore the possible payoffs of the put option after two months are:

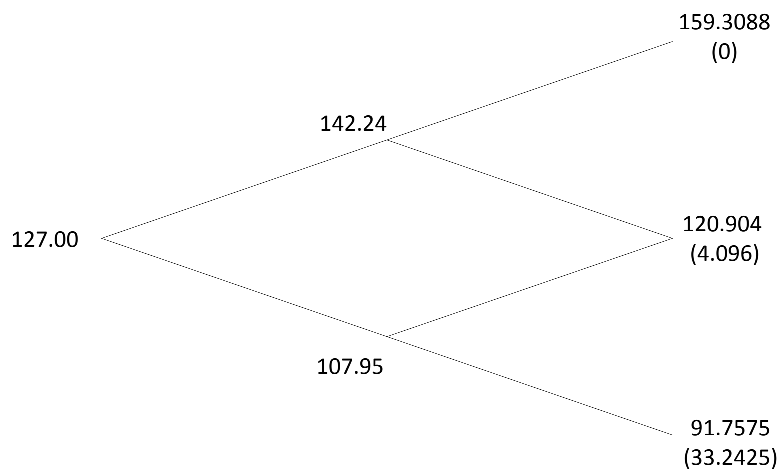
$$p_{uu} = 0$$

$$p_{ud} = 125 - 120.904 = 4.096$$

$$p_{dd} = 125 - 91.7575 = 33.2425$$ [½]

The up and down steps are the same over each month, so this binomial tree is recombining.

The binomial tree of security prices looks like this, with the final put payoffs in italics.



We can discount the expected value of the payoff under the risk-neutral probability to find the initial value of the European put option:

$$e^{-2r}\left(q^2 p_{uu} + 2q(1-q)p_{ud} + (1-q)^2 p_{dd}\right)$$

$$= e^{-0.02}\left(0 + 2\times 0.5928\times 0.4072\times 4.096 + 0.4072^2 \times 33.2425\right)$$

$$= 7.342$$ [1]

[Total 3]

(ii) ***Calculate the value of the American put option***

The only difference from part (i) is that with an American option, it may be beneficial to exercise the option early, after one month. [½]

If the security price rises over the first month then there would be no point in exercising the put option because the payoff would be zero. However, if the security price falls during the first month, the payoff from exercising the option early is:

$$125 - 127\times 0.85 = 17.05$$ [½]

Using the risk-neutral probability we can calculate the value of holding on to the put option:

$$V_1(2) = e^{-0.01}\{q\times 4.096 + (1-q)\times 33.2425\} = 15.8062$$ [½]

Since this is less than 17.05, if the security price falls, it would be advantageous to exercise after one month. The extra value at time 0 of having this option after one month is:

$$e^{-0.01}(1-q)\times(17.05 - 15.8062) = 0.5015$$ [½]

Finally the value of the American option is the value of the European option plus the extra value of the option to exercise after one month:

$$V_0 = 7.342 + 0.5015 = 7.843$$ [1]

[Total 3]

(iii) ***Calculate the value of the American call option***

It is never optimal to exercise an American call option early on a non-dividend-paying security, so the value of the American call option is the same as that of a European call option. [1]

Using put-call parity we can derive this value from the value of the European put option in part (i):

$$c_0 = p_0 + S_0 - Ke^{-rT} = 7.342 + 127 - 125e^{-0.01\times 2} = 11.817$$ [1]

[Total 2]

Alternative methods are valid, eg:

$$c_0 = e^{-0.02}q^2 \times (159.3088 - 125) = 11.817$$

14.2 (i) ***Risk-neutral probability***

Using the formula for the risk-neutral up-step probability on page 45 of the *Tables*, we have:

$$q = \frac{e^{0.055/365} - 0.99}{1.02 - 0.99} = 0.338357$$ [2]

(ii) ***Call option value***

For simplicity, we can assume that we are valuing one call option on a share worth £10,000. We can use the up and down steps to calculate the six possible final share prices in this binomial tree:

$$S_5 = £10,000 \times 1.02^5 \times 0.99^0 = £11,040.81 \quad \text{for 5 up jumps}$$
$$S_5 = £10,000 \times 1.02^4 \times 0.99^1 = £10,716.08 \quad \text{for 4 up jumps}$$
$$S_5 = £10,000 \times 1.02^3 \times 0.99^2 = £10,400.90 \quad \text{for 3 up jumps}$$
$$S_5 = £10,000 \times 1.02^2 \times 0.99^3 = £10,094.99 \quad \text{for 2 up jumps}$$
$$S_5 = £10,000 \times 1.02^1 \times 0.99^4 = £9,798.08 \quad \text{for 1 up jump}$$
$$S_5 = £10,000 \times 1.02^0 \times 0.99^5 = £9,509.90 \quad \text{for 0 up jumps}$$ [2]

Since the option is at-the-money, $K = £10,000$, so the payoffs from the call option are:

$$\begin{array}{ll} £1,040.81 & \text{for 5 up jumps} \\ £716.08 & \text{for 4 up jumps} \\ £400.90 & \text{for 3 up jumps} \\ £94.99 & \text{for 2 up jumps} \\ £0 & \text{for 1 up jump} \\ £0 & \text{for 0 up jumps} \end{array}$$ [1]

So, the fair price of the call option is:

$$V_0 = e^{-5\times 0.055/365} \left\{ \begin{array}{l} 1,040.81q^5 + 716.08 \times 5q^4(1-q) \\ +400.90 \times 10q^3(1-q)^2 + 94.99 \times 10q^2(1-q)^3 \end{array} \right\}$$ [1]

Substituting in the value of q, this is:

$$V_0 = 0.999247\{4.62 + 31.05 + 67.98 + 31.50\}$$
$$= £135.05$$ [1]

[Total 5]

(iii) ***Special option value***

Again, for simplicity, we can assume that we are valuing one special option on a share worth £10,000. By investigation, we find that the average share price only exceeds £10,600 if the share price goes up on all 5 days. If we try 4 up jumps and 1 down jump, we get:

$$\frac{1.02+1.02^2+1.02^3+1.02^4+1.02^4\times 0.99}{5}=1.055<1.06$$ [1]

The average share price over the 5 days, if it goes up every day, is:

$$£10,000\times\frac{1.02+1.02^2+1.02^3+1.02^4+1.02^5}{5}=£10,616.24$$ [1]

The payoff of the special option in this case is:

$$\max\{10,616.24-10,600,0\}=£16.24$$ [1]

So, the fair price of the special option is:

$$V_0=e^{-5\times 0.055/365}\times 16.24q^5$$
$$=£0.07$$ [1]

[Total 4]

(iv) ***Value of vega***

Vega is the rate of change of an option value f with respect to the volatility σ of the underlying asset, *ie* the shares in the company:

$$\nu=\frac{\partial f}{\partial \sigma}$$ [1]

The special option's payoff is dependent on the average share price, which is a much smoother process than the current share price. So, because the special option's final payoff is less dependent on the current share price, its value will vary less with changes in the volatility of the share price. [1]

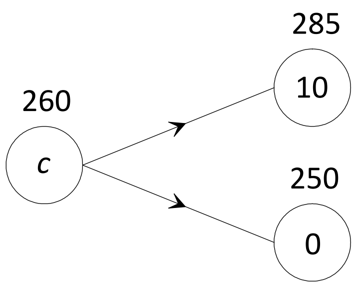
So, the special option will have a lower value of vega. [1]

[Total 3]

This observation is reinforced by noting that the payoff from the standard call option in part (ii) with strike £10,600 lies in the range $(0,440.81)$ according to this model, whereas the payoff for the special option lies in the much narrower range $(0,16.24)$.

14.3 (i)(a) ***Price of call option using a replicating portfolio method***

The diagram for this call option (with the option values / payoffs inside the circles and the share prices on the top) is:



A replicating portfolio, consisting of ϕ_1 shares and ψ_1 invested in cash, is set up at $t=0$ to give the same payoff as the derivative at $t=1$, irrespective of whether the share price increases or decreases over the year.

If the share price increases:

$$285\phi_1 + \psi_1 e^{0.05} = 10 \qquad (1)$$

and if the share price decreases:

$$250\phi_1 + \psi_1 e^{0.05} = 0 \qquad (2)$$

Solving equations (1) and (2) simultaneously gives:

$$\phi_1 = \tfrac{2}{7} \quad \text{and} \quad \psi_1 = -250 \times \tfrac{2}{7} \times e^{-0.05} = -67.945$$

This portfolio gives the same payoff as the call option at $t=1$, so assuming no arbitrage, the value of the call option at $t=0$, c, must equal the value of the portfolio at $t=0$. So:

$$c = 260\phi_1 + \psi_1 = 6.34$$

(i)(b) ***Price of call option using a risk-neutral valuation***

The risk-neutral probability of an up-movement is:

$$q = \frac{e^{0.05} - 250/260}{285/260 - 250/260} = 0.66659$$

So the value of the call option is:

$$\begin{aligned} c &= \left[10 \times 0.66659 + 0 \times (1 - 0.66659)\right] e^{-0.05} \\ &= 6.34 \end{aligned}$$

(i)(c) ***Price of call option using a risk-free portfolio method***

A risk-free portfolio, consisting of 1 call option and x shares, is set up at $t=0$ so that the value of the portfolio at $t=1$ will be the same irrespective of whether the share price increases or decreases over the year.

Considering the possibilities at $t=1$, this means that:

$$10+285x=0+250x \quad \Rightarrow \quad x=-\tfrac{2}{7}$$

The value of this portfolio at $t=1$ is:

$$10+285\left(-\tfrac{2}{7}\right)=0+250\left(-\tfrac{2}{7}\right)=-\tfrac{500}{7}$$

This portfolio has been set up to be risk-free (it has the same value at $t=1$ no matter what happens to the share price), so we can use the risk-free force of interest to calculate its value at $t=0$:

$$-\tfrac{500}{7}e^{-0.05}$$

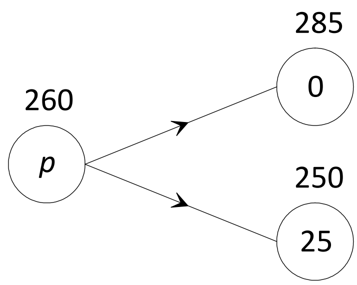
Since this is the cost of setting up the risk-free portfolio at $t=0$:

$$c+260\left(-\tfrac{2}{7}\right)=-\tfrac{500}{7}e^{-0.05} \quad \Rightarrow \quad c=6.34$$

Note that we obtain the same value of the call option, whichever approach we take.

(ii)(a) ***Price of put option using a replicating portfolio method***

The diagram for this put option (with the option values / payoffs inside the circles and the share prices on the top) is:



A replicating portfolio, consisting of ϕ_2 shares and ψ_2 invested in cash, is set up at $t=0$ to give the same payoff as the derivative at $t=1$, irrespective of whether the share price increases or decreases over the year.

If the share price increases:

$$285\phi_2 + \psi_2 e^{0.05} = 0 \qquad (3)$$

and if the share price decreases:

$$250\phi_2 + \psi_2 e^{0.05} = 25 \qquad (4)$$

Solving equations (3) and (4) simultaneously gives:

$$\phi_2 = -\tfrac{5}{7} \quad \text{and} \quad \psi_2 = 285 \times \tfrac{5}{7} \times e^{-0.05} = 193.643$$

This portfolio gives the same payoff as the put option at $t=1$, so assuming no arbitrage, the value of the put option at $t=0$, p, must equal the value of the portfolio at $t=0$. So:

$$p = 260\phi_2 + \psi_2 = 7.93$$

(ii)(b) ***Price of put option using a risk-neutral valuation***

The risk-neutral probability of an up-movement is:

$$q = \frac{e^{0.05} - 250/260}{285/260 - 250/260} = 0.66659$$

This is the same as for the call option in (i).

So the value of the put option is:

$$p = \left[0 \times 0.66659 + 25 \times (1 - 0.66659)\right] e^{-0.05}$$
$$= 7.93$$

(ii)(c) ***Price of put option using a risk-free portfolio method***

A risk-free portfolio, consisting of 1 put option and y shares, is set up at $t=0$ so that the value of the portfolio at $t=1$ will be the same irrespective of whether the share price increases or decreases over the year.

Considering the possibilities at $t=1$, this means that:

$$0 + 285y = 25 + 250y \quad \Rightarrow \quad y = \tfrac{5}{7}$$

The value of this portfolio at $t=1$ is:

$$0 + 285\left(\tfrac{5}{7}\right) = 25 + 250\left(\tfrac{5}{7}\right) = \tfrac{1425}{7}$$

This portfolio has been set up to be risk-free (it has the same value at $t=1$ no matter what happens to the share price), so we can use the risk-free force of interest to calculate its value at $t=0$:

$$\frac{1425}{7}e^{-0.05}$$

Since this is the cost of setting up the risk-free portfolio at $t=0$:

$$p+260\left(\frac{5}{7}\right)=\frac{1425}{7}e^{-0.05} \quad \Rightarrow \quad p=7.93$$

Note that we obtain the same value of the put option, whichever approach we take.

(iii) ***Verify put-call parity***

The put-call parity relationship states that:

value of put + share price = value of call + discounted strike price

The LHS is: $7.93+260=267.93$

The RHS is: $6.34+275e^{-0.05}=6.34+261.59=267.93$

Since these are equal, the put-call parity relationship holds in this case.

14.4 Suppose the portfolio consists of Δ securities and an initial amount of cash ψ.

If the security price moves up, the value of the portfolio will be:

$$\psi e^r+\Delta Su = \psi e^r+\frac{z_u-z_d}{u-d}u$$

We want this to equal the payoff following an upward movement in the security price:

ie $$\psi e^r+\frac{z_u-z_d}{u-d}u = z_u$$

So: $$\psi = e^{-r}\left(z_u-\frac{z_u-z_d}{u-d}u\right) = e^{-r}\left(\frac{uz_d-dz_u}{u-d}\right)$$

If the security price moves down, the value of the portfolio will be:

$$\psi e^r+\Delta Sd = \psi e^r+\frac{z_u-z_d}{u-d}d = \left(\frac{uz_d-dz_u}{u-d}\right)+\frac{z_u-z_d}{u-d}d = z_d$$

So, with an initial cash holding of $\psi=e^{-r}\left(\frac{uz_d-dz_u}{u-d}\right)$, the portfolio replicates the derivative payoff, irrespective of the actual price movement.

Note that Δ here is calculated as the proportionate change in the derivative price relative to the price of the underlying security. This corresponds to the definition of the 'Greek' delta, namely $\Delta = \dfrac{\partial V}{\partial S}$.

14.5 (i) ***Values of u and d***

Equating the mean and variance of the returns gives:

$$\tfrac{2}{3}u + \tfrac{1}{3}d = 1.1 \quad \Rightarrow d = 3.3 - 2u$$

and $$\tfrac{2}{3}u^2 + \tfrac{1}{3}d^2 - 1.1^2 = 0.1^2 \quad \Rightarrow 2u^2 + d^2 = 3.66$$

Eliminating d from these simultaneous equations gives:

$$2u^2 + (3.3 - 2u)^2 = 3.66$$

ie $$6u^2 - 13.2u + 7.23 = 0$$

Solving this using the quadratic formula gives:

$$u = \frac{13.2 \pm \sqrt{13.2^2 - 4(6)(7.23)}}{2(6)} = \frac{13.2 \pm \sqrt{0.72}}{12}$$

So:

$$u = 1.17071 \text{ and } d = 0.95858$$

or:

$$u = 1.02929 \text{ and } d = 1.24142$$

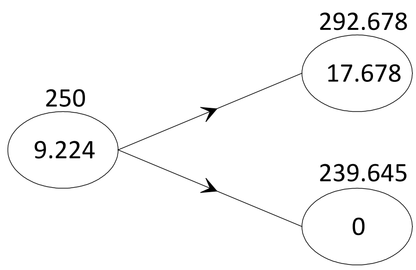
Since we need $u > d$, we can eliminate the second pair of values and conclude that the appropriate parameter values are $u = 1.17071$ and $d = 0.95858$.

(ii)(a) ***Call option***

The risk-neutral probability of an up-movement is:

$$q = \frac{e^{0.075} - 0.95858}{1.17071 - 0.95858} = 0.56241$$

The tree diagram for the call option looks like this:

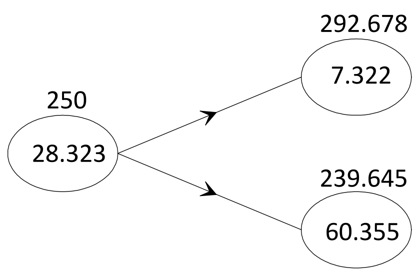


So the value of the call option is:

$$(0.56241 \times 17.678 + 0)e^{-0.075} = 9.224$$

(ii)(b) ***Put option***

The tree diagram for the put option looks like this:



So the value of the put option is:

$$(0.56241 \times 7.322 + 0.43759 \times 60.355)e^{-0.075} = 28.323$$

14.6 (i) ***Risk-neutral probability***

Let u and d be the assumed proportionate changes in the price of the underlying share if it goes up and down respectively, and let r be the risk-free interest rate (continuously compounded). δt is the length of the one-step time period.

Let S be the current price of the share.

If q and $1-q$ are the risk-neutral probabilities for the tree, the expected final value of the share should be the same as if it had been invested in risk-free cash.

So we need:

$$qSu + (1-q)Sd = Se^{r\delta t}$$

Cancelling the S's gives:

$$qu + (1-q)d = e^{r\delta t}$$

Rearranging gives:

$$q(u-d)+d=e^{r\delta t} \quad \Rightarrow q=\frac{e^{r\delta t}-d}{u-d}$$

(ii) ***Explain the inequality***

The condition $d<e^{r\delta t}<u$ is needed to ensure that the market is arbitrage-free. Otherwise we could make a guaranteed profit.

For example, if $d<u<e^{r\delta t}$, the cash investment would outperform the share in all circumstances. So we could make a guaranteed profit by selling the share at the start and investing the proceeds in cash. When we buy back the share at the end, we would have a positive profit of either $Se^{r\delta t}-Su$ or $Se^{r\delta t}-Sd$.

(iii) ***Formula for θ***

Let R be the one-step return on the share. Then

$$1+R=\begin{cases} u & \text{with probability } p \\ d & \text{with probability } 1-p \end{cases}$$

So:

$$E(1+R)=pu+(1-p)d \quad \Rightarrow \quad \theta=E(R)=pu+(1-p)d-1$$

(iv) ***Real-world variance, σ^2***

Similarly, the variance of the rate of return is:

$$\begin{aligned}\sigma^2 &= \text{var}(R)\\ &= \text{var}(1+R)\\ &= E\left[(1+R)^2\right]-\left(E[1+R]\right)^2\\ &= pu^2+(1-p)d^2-[pu+(1-p)d]^2\end{aligned}$$

If we collect the p's together, expand and simplify, we get:

$$\sigma^2 = p(u^2 - d^2) + d^2 - [p(u-d)+d]^2$$

$$= p(u^2 - d^2) - p^2(u-d)^2 - 2pd(u-d)$$

$$= p(u-d)[(u+d) - p(u-d) - 2d]$$

$$= p(u-d)[(u-d) - p(u-d)]$$

$$= p(u-d)^2[1-p]$$

$$= p(1-p)(u-d)^2$$

(v) ***Show the inequality***

Since $u > d$ by definition, the inequality $p > q$ is equivalent to:

$$p(u-d)+d > q(u-d)+d$$

ie $$1+\theta > \frac{e^{r\delta t}-d}{u-d}(u-d)+d \;=\; e^{r\delta t}-d+d \;=\; e^{r\delta t}$$

In words, this says that the real-world probability of an up movement is greater than the risk-neutral probability whenever the expected increase in the value of the underlying security exceeds the risk-free interest rate, *ie* whenever the underlying security is risky.

14.7 (i) ***The risk-free force of interest r***

Under the risk-neutral probability measure we have:

$$E_Q\left[S_1\right] = S_0 e^r$$

$$\Leftrightarrow \qquad q \times 1.1S_0 + (1-q)\times 0.9S_0 = S_0 e^r$$ [½]

Dividing through by S_0, substituting in the given value of q and solving for r gives:

$$r = \log\left(0.55\times 1.1 + 0.45\times 0.9\right)$$

$$= \log(1.01) = 0.995\%$$

[½]
[Total 1]

(ii) ***State price deflators***

The state price deflators for each of the three final possible states can be derived from the formulae:

$$A_2(1)=e^{-2r}\frac{q^2}{p^2} \qquad A_2(2)=e^{-2r}\frac{2q(1-q)}{2p(1-p)} \qquad A_2(3)=e^{-2r}\frac{(1-q)^2}{(1-p)^2} \qquad [1]$$

We can use the annualised expected force of return on the share, μ, to calculate the real-world probability of an up-step p:

$$E_P[S_1]=S_0e^{\mu/12}$$

$$\Leftrightarrow \qquad p\times 1.1+(1-p)\times 0.9=e^{\mu/12}=e^{0.3/12}$$

$$\Leftrightarrow \qquad p=\frac{e^{0.3/12}-0.9}{1.1-0.9}=0.62658 \qquad [1]$$

Using this value of p, and noting that, since $r=\log(1.01)$, the monthly effective risk-free interest rate is 1%, the state price deflators are:

$$A_2(1)=\frac{1}{1.01^2}\frac{0.55^2}{0.62658^2}=0.75533$$

$$A_2(2)=\frac{1}{1.01^2}\frac{2\times 0.55\times 0.45}{2\times 0.62658\times 0.37342}=1.03695$$

$$A_2(3)=\frac{1}{1.01^2}\frac{0.45^2}{0.37342^2}=1.42356 \qquad [2]$$

[Total 4]

(iii)(a) ***Payoff* $(1,0,0)$**

$$e^{-2r}q^2=\frac{0.55^2}{1.01^2}=0.2965 \qquad [½]$$

(iii)(b) ***Payoff* $(0,1,0)$**

$$e^{-2r}2q(1-q)=\frac{2\times 0.55\times 0.45}{1.01^2}=0.4852 \qquad [½]$$

(iii)(c) ***Payoff* $(0,0,1)$**

$$e^{-2r}(1-q)^2=\frac{0.45^2}{1.01^2}=0.1985 \qquad [½]$$

Note that the values calculated in parts (a), (b) and (c) are known as 'state prices'. We can use them to calculate the value of more complicated derivatives as we now do for the rest of part (iii).

We could alternatively have calculated the values in (a), (b) and (c) using:

$$A_2(1)\times p^2 \ , A_2(2)\times 2p(1-p) \ , A_2(3)\times(1-p)^2$$

(iii)(d) ***Call option, strike price*** $K = \mathbf{0.95}$

Depending on the final price of the share, the payoff for this option will be:

$$\max(1\times 1.1^2 - 0.95, 0) = 0.26 \,,$$

$$\max(1\times 1.1\times 0.9 - 0.95, 0) = 0.04 \text{ , or}$$

$$\max(1\times 0.9^2 - 0.95, 0) = 0 \,.$$

So, the value is:

$$0.26\times 0.2965 + 0.04\times 0.4852 = 0.0965 \qquad [1]$$

(iii)(e) ***Put option, strike price*** $K = \mathbf{1.05}$

Depending on the final price of the share, the payoff for this option will be:

$$\max(1.05 - 1\times 1.1^2, 0) = 0 \,,$$

$$\max(1.05 - 1\times 1.1\times 0.9, 0) = 0.06 \text{ , or}$$

$$\max(1.05 - 1\times 0.9^2, 0) = 0.24 \,.$$

So, the value is:

$$0.06\times 0.4852 + 0.24\times 0.1985 = 0.0768 \qquad [1]$$

(iii)(f) ***Payoff*** $\mathbf{2\times|S-0.98|}$

Depending on the final price of the share, the payoff for this option will be:

$$2\times\left|1\times 1.1^2 - 0.98\right| = 0.46 \,,$$

$$2\times\left|1\times 1.1\times 0.9 - 0.98\right| = 0.02 \text{ , or}$$

$$2\times\left|1\times 0.9^2 - 0.98\right| = 0.34 \,.$$

So, the value is:

$$0.46\times 0.2965 + 0.02\times 0.4852 + 0.34\times 0.1985 = 0.2136 \qquad [1½]$$

[Total 5]

14.8 (i) ***Explain the difference***

In a *recombining* binomial tree u and d, the proportionate increase and decrease in the underlying security price at each step, are assumed constant throughout the tree. As a result, the security price after a specified number of up- and down- movements is the same, irrespective of the order in which the movements occurred. [1]

In a *non-recombining* binomial tree the values of u and d can change at each stage. As a result, each node in the tree will in general generate two new nodes, making the tree much larger than a recombining tree. Consequently, an n-period tree will have 2^n, rather than just $n+1$, possible states at time n. [1]

[Total 2]

(ii) ***Calculate the value of the put option***

The risk-neutral probabilities of an up-movement at the first and second step are:

$$q_1 = \frac{e^{0.01} - 0.95}{1.10 - 0.95} = 0.40033 \quad \text{and} \quad q_2 = \frac{e^{0.015} - 0.90}{1.20 - 0.90} = 0.38371 \qquad [2]$$

The put option payoffs at each of the four possible states at expiry are 0, 54, 0 and 108. [1]

Working backwards through the tree, we can then find the option value $V_1(1)$ following an up-step over the first 3 months from:

$$V_1(1)e^{0.015} = 0.38371 \times 0 + (1 - 0.38371) \times 54$$

ie $V_1(1) = 32.784$ [1]

Similarly, the option value $V_1(2)$ following a down-step over the first 3 months is found from:

$$V_1(2)e^{0.015} = 0.38371 \times 0 + (1 - 0.38371) \times 108$$

ie $V_1(2) = 65.568$ [1]

Finally, the current value V_0 of the put option is found from:

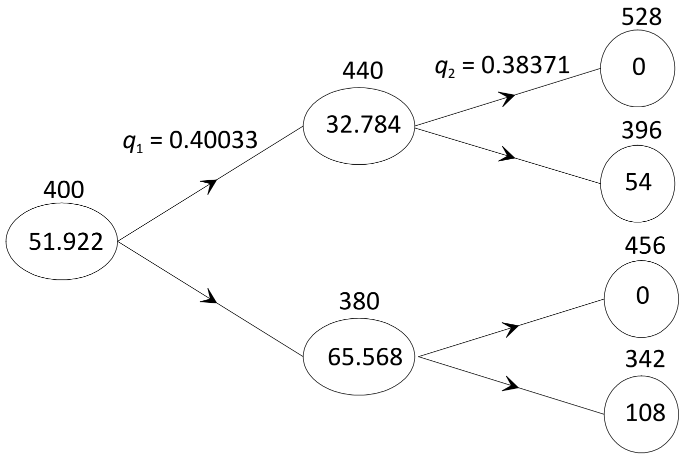
$$V_0 \, e^{0.01} = 0.40033 \times 32.784 + (1 - 0.40033) \times 65.568$$

ie $V_0 = 51.922$

So the value of the put option is 51.922. [1]

[Total 6]

In fact, the tree diagram looks like this:



(iii) ***Suggest an alternative model***

The researcher is proposing using a 6-step non-recombining tree.

This would result in a model that was much less crude than the two-step tree and should be capable of producing a more accurate valuation. [1]

However, there would be a lot more parameter values to specify (although some of these may be assumed to be equal). Appropriate values of u and d would be required for each branch of the tree and values of r for each month. [2]

The new tree would be big, having $2^6 = 64$ nodes in the expiry column. This would make the calculations prohibitive to do manually and would require more programming and calculation time on a computer. [1]

An alternative model that might be more efficient numerically would be a 6-step recombining tree (lattice), which would have only 7 nodes in the final column. [1]

[Total 5]

15

The Black-Scholes option pricing formula

Syllabus objectives

6.1 Option pricing and valuations

6.1.8 Demonstrate an understanding of the Black-Scholes derivative-pricing model:

- Derive the Black-Scholes partial differential equation both in its basic and Garman-Kohlhagen forms. (part)

6.1.9 Show how to use the Black-Scholes model in valuing options and solve simple examples.

6.1.10 Discuss the validity of the assumptions underlying the Black-Scholes model.

0 Introduction

In the previous chapter we developed a simple binomial model to price derivatives in discrete time with a discrete state space. The assumed share price process was a geometric random walk. We now extend our analysis to continuous time and a continuous state space, in an attempt to more realistically model the processes driving stock prices. Since a random walk becomes Brownian motion in the continuous-time limit, we assume our share price follows a geometric Brownian motion or lognormal model.

We will discuss two methods of pricing derivatives in the continuous-time case – these are basically alternative proofs of the same result, the *Black-Scholes* pricing formula. Both methods can be described as no-arbitrage approaches. In this chapter we focus primarily on the *risk-free* construction or partial differential equation (PDE) approach. The proof allows us to set up a PDE for the value of the derivative. It can then be shown that the Garman-Kohlhagen formula (see page 47 of the *Tables*) is a solution of this equation, and also satisfies the boundary conditions for a call or put option. In later chapters we turn our attention to the development of a derivative pricing approach based on the use of *replicating strategies,* as in the discrete-time case.

Section 2.4 sets up the PDE by applying Ito's Lemma, which you will recall from the chapter on stochastic calculus. The boundary conditions are also given. In principle, it is straightforward to differentiate the Garman-Kohlhagen formula and show that it does satisfy the PDE and boundary conditions. However, to do so we need to calculate the Greeks. We will return to our intuitive interpretation of delta, gamma and theta within the context of the Black-Scholes PDE.

The Black-Scholes analysis of option prices is underpinned by a number of key assumptions. We discuss these first in Section 1 and consider how realistic they are in practice. Even though the assumptions do not all hold in practice, this does not prevent the Black-Scholes model providing a good approximation to reality. The approach offers valuable insight into option pricing and is widely used in practice.

1 The assumptions underlying the Black-Scholes model

1.1 The assumptions

The assumptions underlying the Black-Scholes model are as follows:

1. **The price of the underlying share follows a geometric Brownian motion.**

   ie the share price changes *continuously* through time according to the stochastic differential equation:

   $$dS_t = S_t\left(\mu dt + \sigma dZ_t\right)$$

   This is the same as the lognormal model discussed in an earlier chapter and therefore has the same properties.

2. **There are no risk-free arbitrage opportunities.**

3. **The risk-free rate of interest is constant, the same for all maturities and the same for borrowing or lending.**

   In fact, this simplifying assumption is not critical and can be relaxed.

4. **Unlimited short selling (that is, negative holdings) is allowed.**

   So, we are allowed to sell unlimited amounts of securities that we do not own. This is necessary because, in order to hedge a derivative whose price is positively correlated with that of the underlying asset – *eg* a call option, which will have a positive delta – we need to hold a negative quantity of the underlying asset.

5. **There are no taxes or transaction costs.**

   This is important since we will need to continuously rebalance some risk-free portfolios.

6. **The underlying asset can be traded continuously and in infinitesimally small numbers of units.**

   Infinite divisibility of securities is necessary to ensure that perfect hedges can be achieved. Continuous trading requires that security markets are open 24 hours a day, every day, and this is necessary so that the hedging portfolio can be rebalanced *continuously*.

The key general implication of the underlying assumptions is that the market in the underlying share is complete: that is, all derivative securities have payoffs which can be replicated. This consequence is at odds with the real world and implies problems with the underlying assumptions.

1.2 How realistic are the assumptions?

Question

List the seven main defects of the lognormal model of security prices discussed earlier in the course.

Solution

The continuous-time lognormal model may be inappropriate for modelling investment returns because:

- The volatility parameter σ may not be constant over time. Estimates of volatility from past data are critically dependent on the time period chosen for the data and also how often the price history is sampled
- The long-term drift parameter μ may not be constant over time. In particular, interest rates will impact the drift.
- There is evidence in real markets of mean-reverting behaviour, which is inconsistent with the independent increments assumption.
- There is evidence in real markets of momentum effects, which is inconsistent with the independent increments assumption.
- The distribution of security returns $\log\left(S_u/S_t\right)$ has a taller peak in reality than that implied by the normal distribution. This is because there are more days of little or no movement in the share price.
- The distribution of security returns $\log\left(S_u/S_t\right)$ has fatter tails in reality than that implied by the normal distribution. This is because there are more extreme movements in security prices.
- The sample paths of security prices are not continuous, but instead appear to jump occasionally.

It is clear that each of these assumptions is unrealistic to some degree, for example:

- **Share prices can jump. This invalidates assumption 1 since geometric Brownian motion has continuous sample paths.**

  An important consequence of discontinuous share prices is that it is not possible to rebalance the risk-free portfolio at each moment so as to eliminate movements in the value of the portfolio. Hence, the portfolio is not entirely risk-free.

  However, hedging strategies can still be constructed which substantially reduce the level of risk.

- **The risk-free rate of interest does vary and in an unpredictable way.**

  We might, for example, assume that the risk-free rate is either the base rate set by the central bank or the yield on Treasury bills, both of which can vary over time.

  However, over the short term of a typical derivative, the assumption of a constant risk-free rate of interest is not far from reality. (More specifically the model can be adapted in a simple way to allow for a stochastic risk-free rate, provided this is a predictable process.)

  In addition, different rates may apply for borrowing and lending.

- **Unlimited short selling may not be allowed, except perhaps at penal rates of interest. These problems can be mitigated by holding mixtures of derivatives which reduce the need for short selling. This is part of a suitable risk management strategy as discussed in Section 2 below.**

  Individual investors probably have to pay a higher interest rate on an overdraft than they receive on savings. So too do financial institutions.

- **Shares can normally only be dealt in integer multiples of one unit, not continuously, and dealings attract transaction costs: invalidating assumptions 4, 5 and 6. Again we are still able to construct suitable hedging strategies which substantially reduce risk.**

  Transactions costs do arise in practice, their impact depending upon their size. Several extensions to the standard Black-Scholes model have been developed to allow for the effect of transactions costs on option prices.

- **Distributions of share returns tend to have fatter tails than suggested by the lognormal model, invalidating assumption 1.**

  The assumption that share prices follow a geometric Brownian motion of the form $dS_t = S_t(\mu dt + \sigma dZ_t)$ implies that the future share price S_T, $T > t$, is lognormally distributed. Actual share prices, however, experience large up and down movements more commonly than suggested by a lognormal distribution. A particular consequence of this is that large jumps make it more difficult to maintain a delta-neutral portfolio.

Despite all of the potential flaws in the model assumptions, analyses of market derivative prices indicate that the Black-Scholes model does give a very good approximation to the market.

It is worth stressing here that all models are only approximations to reality. It is always possible to take a model and show that its underlying assumptions do not hold in practice.

This does not mean that a model has no use. A model is useful if, for a specified problem, it provides answers which are a good approximation to reality or if it provides insight into underlying processes.

A model is a stylised representation of a more complex situation and as such aims to characterise the most important features of that situation in a way that enables it to be analysed. It thereby provides useful insight into that situation. More complex models often provide greater insight, but at the cost of greater complexity and perhaps reduced tractability.

In this respect the Black-Scholes model is a good model, since it gives us prices which are close to what we observe in the market (despite the fact that we can criticise quite easily the individual assumptions) and because it provides insight into the usefulness of dynamic hedging.

In addition, it is widely used by derivative traders, who can make adjustments to allow for its known deficiencies. The model can be calibrated to reproduce observable market prices and it is valuable in understanding the sensitivity of option prices to the various factors that affect them.

2 The Black-Scholes model

2.1 Introduction

In this section we will show how to derive the price of a European call or put option using a model under which share prices evolve in continuous time and are characterised at any point in time by a continuous distribution rather than a discrete distribution.

2.2 The underlying SDE

Suppose that we have a European call option on a non-dividend-paying share S_t which is governed by the stochastic differential equation (SDE):

$$dS_t = S_t(\mu dt + \sigma dZ_t)$$

where Z_t is a standard Brownian motion.

The share price process is therefore being modelled as a geometric Brownian motion or lognormal model, as discussed previously. The constants μ and σ are referred to as the *drift* and *volatility* parameters respectively.

Investors are allowed to invest positive or negative amounts in this share. Investors can also have holdings in a risk-free cash bond with price B_t at time t.

This is governed by the ordinary differential equation:

$$dB_t = rB_t dt$$

where r is the (assumed-to-be) constant risk-free rate of interest. Hence:

$$S_t = S_0 \exp\left[(\mu - \frac{1}{2}\sigma^2)t + \sigma Z_t\right]$$

$$B_t = B_0 \exp(rt)$$

To check this solution, we define the function $g(t, Z_t) = S_0 \exp[(\mu - \frac{1}{2}\sigma^2)t + \sigma Z_t]$ so that $S_t = g(t, Z_t)$. Note that this is a function of t and Z_t.

Recall that the Ito process for standard Brownian motion is:

$$dZ_t = 1\, dZ_t + 0\, dt$$

Now apply Ito's Lemma to $g(t, Z_t)$, which we write as g for notational ease. That is:

$$\begin{aligned} dg &= \sigma g dZ_t + \left[\left(\mu - \frac{1}{2}\sigma^2\right)g + 0 + \frac{1}{2}\sigma^2 g\right]dt \\ &= \mu g dt + \sigma g dZ_t \end{aligned}$$

Replacing $g = g(t, Z_t)$ with S_t, we have the original SDE and the check is complete:

$$dS_t = \mu S_t dt + \sigma S_t dZ_t$$

Since Z_t is normally distributed, S_t is lognormally distributed with all of the usual properties of that distribution.

From the formula given for S_t we can deduce that:

$$\log S_t = \log S_0 + \left(\mu - ½\sigma^2\right)t + \sigma Z_t \sim N\left[\log S_0 + \left(\mu - ½\sigma^2\right)t, \sigma^2 t\right]$$

So S_t has a lognormal distribution with parameters $\log S_0 + \left(\mu - ½\sigma^2\right)t$ and $\sigma^2 t$.

This price process is sometimes called a lognormal process, geometric Brownian motion or exponential Brownian motion.

2.3 The Black-Scholes formula

Let $f(t,s)$ be the price at time t of a call option given:

- **the current share price is $S_t = s$**
- **the time of maturity is $T > t$**
- **the exercise price is K.**

Proposition 15.1 (The Black-Scholes formula)

For such a call option:

$$f(t,S_t) = S_t \Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

where:

$$d_1 = \frac{\log \frac{S_t}{K} + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$$

and:

$$d_2 = d_1 - \sigma\sqrt{T-t}$$

and $\Phi(z)$ is the cumulative distribution function of the standard normal distribution.

This formula is the Garman-Kohlhagen formula found on page 47 of the *Tables*. Here the dividend rate q is equal to zero.

For a put option we also have $f(t,S_t) = Ke^{-r(T-t)}\Phi(-d_2) - S_t\Phi(-d_1)$ where d_1 and d_2 are as defined above.

Question

Starting from the formula for the price of a call option given in Proposition 15.1, use put-call parity to derive the formula just given for the price of a put option.

Solution

Using put-call parity and the result from Proposition 15.1, we have:

$$p_t = Ke^{-r(T-t)} - S_t + c_t = Ke^{-r(T-t)} - S_t + S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

$$= Ke^{-r(T-t)}(1-\Phi(d_2)) - S_t(1-\Phi(d_1))$$

$$= Ke^{-r(T-t)}\Phi(-d_2) - S_t\Phi(-d_1)$$

Notice that the bottom line answer is also given on page 47 of the *Tables*. Again $q=0$.

We will give two proofs of this result, one here using the partial differential equation (PDE) approach and the other in a later chapter using the martingale approach.

2.4 The PDE approach

Here we use Ito's Lemma to derive an expression for the price of the derivative as a function, f, of the underlying share price process S_t. Here S_t again refers to the share price excluding any dividends received. This method involves the construction of a risk-free portfolio, which in an arbitrage-free world must yield a return equal to the risk-free rate of return.

An expression for $df(t,S_t)$

We first use Ito's Lemma to write a stochastic differential equation (SDE) for the change in the derivative price as a function of the change in the share price. Here $df(t,S_t)$ means the change in the value of the derivative over a very small time period.

Given the Ito process:

$$dS_t = \mu S_t dt + \sigma S_t dZ_t$$

which is the SDE for geometric Brownian motion, with drift and volatility functions for S_t of μS_t and σS_t respectively,

then applying Ito's Lemma to the function $f(t,S_t)$, we have:

$$df(t,S_t) = \frac{\partial f}{\partial s}\sigma S_t dZ_t + \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial s}\mu S_t + \frac{1}{2}\frac{\partial^2 f}{\partial s^2}\sigma^2 S_t^2\right]dt$$

On grounds of notational compactness we have used the notation $\frac{\partial f}{\partial t}$ to mean $\frac{\partial f}{\partial t}(t,S_t)$, and $\frac{\partial f}{\partial s}$ to mean $\frac{\partial f}{\partial s}(t,S_t)$ *etc.* In some textbooks you will see the alternative form $\frac{\partial f}{\partial S_t}$ to represent $\frac{\partial f}{\partial s}(t,S_t)$. However, this is slightly too casual and can lead to confusion. The correct way to apply Ito's Lemma is thus to derive the partial derivatives of the deterministic function $f(t,s)$ and then evaluate these at the random point (t,S_t).

The risk-free portfolio

Suppose that at any time t, $0 \le t < T$, we hold the following portfolio:

- **minus one derivative**
- **plus $\frac{\partial f}{\partial s}(t, S_t)$ shares.**

Let $V(t, S_t)$ be the value of this portfolio. That is:

$$V(t, S_t) = -f(t, S_t) + \frac{\partial f}{\partial s} S_t$$

The pure investment gain over the period $(t, t+dt]$ is the change in the value of the minus one derivative plus the change in the value of the holding of $\partial f / \partial s$ units of the share. That is:

$$-df(t, S_t) + \frac{\partial f}{\partial s} dS_t$$

Note that $\frac{\partial f}{\partial s}$, which represents the number of shares held in the portfolio over the time interval $[t, t+dt)$, is constant. Therefore:

$$-df(t, S_t) + \frac{\partial f}{\partial s} dS_t = -\left\{ \frac{\partial f}{\partial s} \sigma S_t dZ_t + \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial s} \mu S_t + \frac{1}{2} \frac{\partial^2 f}{\partial s^2} \sigma^2 S_t^2 \right] dt \right\}$$
$$+ \frac{\partial f}{\partial s} \left[\mu S_t dt + \sigma S_t dZ_t \right]$$

After cancelling some terms on the right-hand side of the equation we are left with:

$$-df(t, S_t) + \frac{\partial f}{\partial s} dS_t = \left(-\frac{\partial f}{\partial t} - \frac{1}{2} \frac{\partial^2 f}{\partial s^2} \sigma^2 S_t^2 \right) dt$$

We are assuming here that there is no net investment into or out of the portfolio.

Note that this portfolio strategy is not self-financing. That is, the pure investment gain derived above is not equal to the instantaneous change in the value of the portfolio, $dV(t, S_t)$.

Self-financing portfolios are defined formally later in the course.

Note also that the right-hand side of the above expression, which represents the pure investment gain over the interval $[t, t+dt)$, involves the share price, which is random. However, the value of S_t is known at time t. Likewise, if we know the relevant functions, then the values of the derivatives $\frac{\partial f}{\partial t}$ and $\frac{\partial^2 f}{\partial s^2}$ are also known at time t. So, the expression on the right-hand side involves no terms whose values are unknown at time t.

Now note that the expression for $-df(t,S_t)+\frac{\partial f}{\partial s}dS_t$ involves dt but not dZ_t so that the instantaneous investment gain over the short interval t to $t+dt$ is risk-free.

Given that the market is assumed to be arbitrage-free, this rate of interest must be the same as the risk-free rate of interest on the cash bond. (If this was not true then arbitrage opportunities would arise by going long in cash and short in the portfolio (or vice versa) with zero cost initially and a certain, risk-free profit an instant later.)

Therefore we must have, for all t and $S_t > 0$, the alternative expression:

$$-df(t,S_t)+\frac{\partial f}{\partial s}dS_t = rV(t,S_t)dt$$

Question

Derive the right-hand side of this equation.

Solution

We can derive the right-hand side of the equation $(rV\,dt)$ as follows. The value of the portfolio must accumulate to $e^{rdt}V_t$ during the short time interval dt. If we expand the exponential as a series, and ignore second-order and higher-order terms (because dt is infinitesimal), then we see that the accumulated value is $V_t(1+rdt)=V_t+rV_t dt$. The change in the value must therefore be $rV_t dt$.

This is analogous to the equation for the cash bond, which was $dB_t = rB_t dt$.

Recall that $V(t,S_t)=-f(t,S_t)+\dfrac{\partial f}{\partial s}S_t$ is the value of the portfolio. We now have two different expressions for $-df(t,S_t)+\dfrac{\partial f}{\partial s}dS_t$. If we equate these, we get:

$$\left(-\frac{\partial f}{\partial t}-\frac{1}{2}\frac{\partial^2 f}{\partial s^2}\sigma^2 S_t^2\right)dt = r\left(-f+\frac{\partial f}{\partial s}S_t\right)dt$$

$$\Rightarrow \qquad \frac{\partial f}{\partial t}+rS_t\frac{\partial f}{\partial s}+\frac{1}{2}\sigma^2 S_t{}^2\frac{\partial^2 f}{\partial s^2}=rf$$

This is known as the Black-Scholes PDE.

So, we have a non-stochastic partial differential equation (PDE) that can be solved to determine the value of the derivative.

The value of the derivative is found by specifying appropriate boundary conditions and solving the PDE.

The boundary conditions are:

$$f(T, S_T) = \max\{S_T - K, 0\} \quad \text{for a call option}$$

$$f(T, S_T) = \max\{K - S_T, 0\} \quad \text{for a put option}$$

We have now constructed a partial differential equation that the value of any fairly priced derivative based on the underlying share must satisfy. This means that if a proposed model for the fair price of a derivative does not satisfy the PDE it is not an accurate model.

Finally, we can try out the solutions given in the proposition for the value of a call and a put option. We find that they satisfy the relevant boundary conditions and the PDE.

As an example we could check that the Black-Scholes formula for the fair price of a European call option:

- satisfies the Black-Scholes PDE and
- satisfies the boundary condition $f(T, S_T) = \max\{S_T - K, 0\}$

Intuitive interpretation of the PDE

We now return to the intuitive interpretation of the Greeks, this time within the context of the Black-Scholes PDE.

From the Black-Scholes PDE, we have:

$$\frac{\partial f}{\partial t} + rS_t \frac{\partial f}{\partial s} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 f}{\partial s^2} = rf$$

or $$\Theta + rS_t \Delta + \frac{1}{2}\sigma^2 S_t^2 \Gamma = rf$$

Recall that a portfolio for which the weighted sum of the deltas of the individual assets is equal to zero is sometimes described as *delta-neutral*. Also recall that if Γ is small, then Δ will change only slowly over time and so the adjustments needed to keep a portfolio delta-neutral will be minimal.

So, if the delta and gamma of a portfolio are both zero then Θ is the risk-free rate of growth of the portfolio.

Question

A forward contract is arranged where an investor agrees to buy a share at time T for an amount K. It is proposed that the fair price for this contract at time t is:

$$f(t, S_t) = S_t - Ke^{-r(T-t)}$$

Show that this:

(i) satisfies the boundary condition

(ii) satisfies the Black-Scholes PDE.

Solution

(i) ***Boundary condition***

At expiry:

$$f(T,S_T)=S_T-K$$

which is what we would expect. The investor will pay K and receive a share worth S_T.

(ii) ***Black-Scholes PDE***

We first differentiate $f(t,S_t)$ with respect to S_t and t to find the Greeks:

$$\Delta=\frac{\partial f}{\partial S_t}=1 \qquad \Gamma=\frac{\partial \Delta}{\partial S_t}=0 \qquad \Theta=\frac{\partial f}{\partial t}=-rKe^{-r(T-t)}$$

Using these Greeks we see that:

$$\Theta+\Delta r S_t+\frac{1}{2}\sigma^2\Gamma S_t^2$$

$$=-rKe^{-r(T-t)}+1\times rS_t+0$$

$$=r\left(S_t-Ke^{-r(T-t)}\right)$$

$$=rf(S_t,t)$$

and so the PDE is satisfied.

3 The Black-Scholes model for dividend-paying shares

3.1 Introduction

In previous sections we assumed that the underlying asset produced no income, so that the price of the underlying asset would give us the total return directly.

So far we have considered only non-dividend-paying shares and, in doing so, we have treated the share as a non-dividend-paying *pure asset*, *ie* an asset that provides an investment return only via capital growth. In practice, of course, most shares do pay a stream of dividends and so we must modify our previous analysis to allow for this possibility.

We now, essentially, repeat Section 2 in the case that the underlying security has income payable continuously at a rate q. This leads to the *Garman-Kohlhagen formula* for the value of a call option on a dividend-paying share in a continuous-time framework.

3.2 The underlying SDE with dividends

Suppose instead that dividends are payable continuously at the constant rate of q per annum per share. That is, the dividend payable over the interval $(t, t+dt]$ is:

$$qS_t dt$$

Note that the dividend amount is proportional to the value of the share at that time. Note also that this q has nothing to do with the risk-neutral probability measure discussed earlier in the course.

Suppose that S_t is subject to the same stochastic differential equation as before:

$$dS_t = S_t\left(\mu dt + \sigma dZ_t\right)$$

although μ may be different from the rate of growth on the non-dividend-paying asset described before.

Recall that the total return on a share is the sum of the growth rate and the dividend rate.

Modification due to dividends

Once we introduce dividends into our model, the problem we face is that the share price process S_t no longer represents the whole value of the asset. More specifically, if we buy the share for S_0 at time 0, then by time T the total value of what we have bought is equal to:

- S_T , the share price at time T
- *plus* the total of the accumulated dividends received to date.

Moreover, we have defined the dividend payable over the interval $(t, t + dt]$ to be $qS_t dt$ – *ie* it depends on S_t . So, the total of the accumulated dividends received to date must depend on the share price at every instant over the interval 0 to T.

We therefore need to construct a new process, $\tilde{S}_t$ say, that is closely related to S_t and that does represent the entire value of the asset we purchase by buying the share – *ie* the total of the past dividend income *and* the capital growth. This is achieved by considering the following investment.

Suppose that we start with one unit of the share worth S_0 at time 0. Subsequently, any dividends that we receive are assumed to be reinvested instantaneously, by purchasing additional units of the same share. Note that as the dividends are assumed to be received continuously, the reinvestment process will itself be continuous. Over any time interval $[t, t+dt)$ the dividend payout on one share is equal to $qS_t dt$, which can be reinvested at time $t+dt$ to purchase an additional qdt units of the share.

Question

How many units of the share will we hold at time *t*?

Solution

Given that we started with one unit of share at time 0, if we purchase additional shares continuously at the rate of *q* – *ie* we purchase qdt over the interval $(t, t+dt]$ for *each share that we already hold* – then by time *t* our total holding will be e^{qt}.

It is the value of this portfolio, allowing for the continuous reinvestment of dividends, that we define as the process of interest and that we denote by $\tilde{S}_t$. Many unit-linked savings funds offer a facility where investors can choose either to receive dividends as cash or to reinvest them in the fund. In the first case the value of the fund corresponds to S_t, in the second case $\tilde{S}_t$.

Let $\tilde{S}_t$ be the value of an investment of $\tilde{S}_0 = S_0$ at time 0 in the underlying asset, assuming that all dividends are reinvested in the same asset at the time of payment of the dividend. Sometimes $\tilde{S}_t / \tilde{S}_0$ is described as the *total return* on the asset from time 0 to time t.

It is important to note that $\tilde{S}_t$ is the tradable asset and not S_t in the following sense. If we pay S_0 at time 0 for the asset, then we are buying the right to future dividends as well as future growth of the capital. In other words, the value of the asset at time t should account for the accumulated value of the dividends as well as the value of the capital at time t.

It is straightforward to see that the stochastic differential equation for $\tilde{S}_t$ is:

$$d\tilde{S}_t = \tilde{S}_t\left((\mu + q)dt + \sigma dZ_t\right)$$

As the dividends are received continuously at the rate of *q per annum* and are reinvested immediately, so the growth rate or drift of the *total value* of the share asset at time *t* must be the drift in the share price alone, μ, *plus* the instantaneous income yield, *q*.

Solving this we find that:

$$\tilde{S}_t = \tilde{S}_0 \exp\left[(\mu + q - \frac{1}{2}\sigma^2)t + \sigma Z_t\right]$$

Question

Show that $d\tilde{S}_t = \tilde{S}_t\{(\mu+q)dt + \sigma dZ_t\}$ is the corresponding SDE to the above equation.

Solution

If:

$$\tilde{S}_t = \tilde{S}_0 \exp\left[\left(\mu+q-½\sigma^2\right)t + \sigma Z_t\right]$$

then the SDE for $\tilde{S}_t$ can be found by Ito's Lemma.

Let $\tilde{S}_t = f(X_t)$ where $X_t = \left(\mu+q-½\sigma^2\right)t + \sigma Z_t$ and $f(x) = \tilde{S}_0 e^x$. Then we have:

$$f(X_t) = f'(X_t) = f''(X_t) \text{ and } dX_t = \left(\mu+q-½\sigma^2\right)dt + \sigma dZ_t$$

It follows that:

$$\begin{aligned} d\tilde{S}_t = df(X_t) &= f'(X_t)dX_t + ½f''(X_t)(dX_t)^2 \\ &= \tilde{S}_t\left\{\left(\mu+q-½\sigma^2\right)dt + \sigma dZ_t + ½\sigma^2 dt\right\} \\ &= \tilde{S}_t\{(\mu+q)dt + \sigma dZ_t\} \end{aligned}$$

Alternatively, if we consider t as an explicit variable and think of $\tilde{S}_t = g(t, Z_t)$ then we need an extra $\frac{\partial \tilde{S}_t}{\partial t}$ term:

$$\begin{aligned} d\tilde{S}_t(t, Z_t) &= \frac{\partial \tilde{S}_t}{\partial t}dt + \frac{\partial \tilde{S}_t}{\partial z}dZ_t + \frac{1}{2}\frac{\partial^2 \tilde{S}_t}{\partial z^2}(dZ_t)^2 \\ &= \left[\mu+q-½\sigma^2\right]\tilde{S}_t\,dt + \sigma\tilde{S}_t\,dZ_t + ½\sigma^2\tilde{S}_t(dZ_t)^2 \\ &= \left[\mu+q-½\sigma^2\right]\tilde{S}_t\,dt + \sigma\tilde{S}_t\,dZ_t + ½\sigma^2\tilde{S}_t\,dt \\ &= \tilde{S}_t\left\{(\mu+q)dt + \sigma dZ_t\right\} \end{aligned}$$

as required.

Effectively, the tradable asset $\tilde{S}_t$ is just the share price process accumulated at the fixed rate q, where S_t follows the geometric Brownian motion as we assumed earlier.

Question

How are $\tilde{S}_t$ and S_t related?

Solution

We can see from the last Core Reading equation that:

$$\tilde{S}_t = S_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma Z_t\right]e^{qt}$$

As the tradable asset $\tilde{S}_t$ is just the share price process accumulated at the fixed rate *q,* it must be the case that:

$$\tilde{S}_t = e^{qt} S_t$$

Note that if the model is assumed to start at time zero then S_0 and $\tilde{S}_0$ are the same.

3.3 The Garman-Kohlhagen formula

Let us consider a European call option on the underlying asset S_t with strike price K and time of maturity T. The payoff on this option will be $\max\{S_t - K, 0\}$ as before. However, in valuing the option we must take account of the fact that dividends are payable on the underlying asset which do not feed through to the holder of the option. Let us denote the value of this option at time t by $f(t, S_t)$.

Proposition 15.2 (The Garman-Kohlhagen formula for a call option on a dividend-paying share)

For such a call option:

$$f(t, S_t) = S_t e^{-q(T-t)}\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

where:

$$d_1 = \frac{\log\frac{S_t}{K} + \left(r - q + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2 = d_1 - \sigma\sqrt{T-t}$$

and $\Phi(z)$ is the cumulative distribution function of the standard normal distribution.

This formula is given on page 47 of the *Tables*.

For a put option we also have $f(t, S_t) = Ke^{-r(T-t)}\Phi(-d_2) - S_t e^{-q(T-t)}\Phi(-d_1)$ where d_1 and d_2 are as defined above.

As before we will give two proofs of this result, one here using the partial differential equation (PDE) approach and the other in a later chapter using the martingale approach.

3.4 The PDE approach

An expression for $df(t,S_t)$

Given the Ito process:

$$dS_t = \mu S_t dt + \sigma S_t dZ_t$$

then applying Ito's Lemma to the function $f(t,S_t)$ we have:

$$df(t,S_t) = \frac{\partial f}{\partial s}\sigma S_t dZ_t + \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial s}\mu S_t + \frac{1}{2}\frac{\partial^2 f}{\partial s^2}\sigma^2 S_t^2\right)dt$$

The risk-free portfolio

Suppose that at any time t, $0 \le t < T$, we hold the following portfolio:

- **minus one derivative**
- **plus $\frac{\partial f}{\partial s}(t,S_t)$ shares.**

Let $V(t,S_t)$ be the value of this portfolio:

ie $$V(t,S_t) = -f(t,S_t) + \frac{\partial f}{\partial s}S_t$$

The pure investment gain over the period $(t,t+dt]$ is the change in the value of the minus one derivative plus the change in the value of the holding of $\partial f / \partial s$ units of the share including the dividend payment. That is:

$$\begin{aligned}-df(t,S_t) + \frac{\partial f}{\partial s}(dS_t + qS_t dt) &= -\left\{\frac{\partial f}{\partial s}\sigma S_t dZ_t + \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial s}\mu S_t + \frac{1}{2}\frac{\partial^2 f}{\partial s^2}\sigma^2 S_t^2\right)dt\right\} \\ &\quad + \frac{\partial f}{\partial s}\left((\mu S_t dt + \sigma S_t dZ_t) + qS_t dt\right) \\ &= \left(-\frac{\partial f}{\partial t} + \frac{\partial f}{\partial s}qS_t - \frac{1}{2}\frac{\partial^2 f}{\partial s^2}\sigma^2 S_t^2\right)dt\end{aligned}$$

So, the difference here from before is that the change in the value of the portfolio also includes the dividend income $qS_t dt$.

Now note that the expression for $-df(t,S_t) + \frac{\partial f}{\partial s}(dS_t + qS_t dt)$ involves dt but not dZ_t, so that the investment gain over the next instant is risk-free.

Given that the market is assumed to be arbitrage-free, this rate of interest must be the same as the risk-free rate of interest on the cash bond. Therefore we must have the alternative expression:

$$-df(t,S_t)+\frac{\partial f}{\partial s}(dS_t+qS_t dt)=rV(t,S_t)dt$$

So, equating the two expressions we have for pure investment gain and remembering that $V(t,S_t)=-f(t,S_t)+\frac{\partial f}{\partial s}S_t$ gives:

$$\left(-\frac{\partial f}{\partial t}+\frac{\partial f}{\partial s}qS_t-\frac{1}{2}\frac{\partial^2 f}{\partial s^2}\sigma^2 S_t^2\right)dt=r\left(-f+\frac{\partial f}{\partial s}S_t\right)dt$$

$$\Rightarrow \qquad \frac{\partial f}{\partial t}+(r-q)S_t\frac{\partial f}{\partial s}+\frac{1}{2}\sigma^2 S_t{}^2\frac{\partial^2 f}{\partial s^2}=rf$$

Notice that the only difference from before is that dividends are now included. The *r* factor on the LHS is replaced by $r-q$.

The value of the derivative is found by specifying appropriate boundary conditions and solving the PDE.

The boundary conditions are:

$$f(T,S_T)=\max\{S_T-K,0\} \quad \text{for a call option}$$

$$f(T,S_T)=\max\{K-S_T,0\} \quad \text{for a put option}$$

Question

Verify that the Garman-Kohlhagen formula for a European call option on a dividend-paying share satisfies the above boundary condition.

Solution

We need to check that:

$$\lim_{t\to T} f(t,S_t)=\max\left[S_T-K,0\right]$$

Because $d_2=d_1-\sigma\sqrt{T-t}$, the limit of d_1 and d_2 are equal:

$$\lim_{t\to T} d_1=\lim_{t\to T} d_2=\begin{cases}\dfrac{+ve+0}{0}=+\infty & \text{if } S_T>K\\[2ex] \dfrac{-ve+0}{0}=-\infty & \text{if } S_T<K\end{cases}$$

If $S_T = K$, and writing d_1 as:

$$d_1 = \frac{\ln(S_t / K)}{\sigma\sqrt{T-t}} + \frac{(r-q+\frac{1}{2}\sigma^2)}{\sigma}\sqrt{T-t}$$

we see that:

$$\lim_{t \to T} d_1 = 0 + 0 = 0$$

Finally, noting that $\Phi(\infty) = 1$, $\Phi(-\infty) = 0$ and $\Phi(0) = 0.5$, we have:

$$\lim_{t \to T} f(t, S_t) = \begin{cases} S_T - K & \text{if } S_T > K \\ 0 & \text{if } S_T < K \\ 0.5 \times (S_T - K) = 0 & \text{if } S_T = K \end{cases}$$

$$= \max[S_T - K, 0]$$

Finally, we can try out the solutions given in the proposition for the value of a call and a put option. We find that they satisfy the relevant boundary conditions and the PDE.

3.5 Summary

Note that the Black-Scholes formula is simply the Garman-Kohlhagen formula with $q = 0$ and so by verifying that the Garman-Kohlhagen formula satisfies the above PDE, we will implicitly verify that the Black-Scholes formula satisfies the same PDE with $q = 0$.

Delta

It is worth noting the following results relating to delta, Δ, as this may help with exam questions. The following results are derived by differentiating the Black-Scholes and Garman-Kohlhagen formulae with respect to S_t:

- For a European call option on a non-dividend-paying share, $\Delta = \Phi(d_1)$.
- For a European put option on a non-dividend-paying share, $\Delta = -\Phi(-d_1)$.
- For a European call option on a dividend-paying share, with a continuously-compounded dividend yield, q, $\Delta = e^{-q(T-t)}\Phi(d_1)$.
- For a European put option on a dividend-paying share, with a continuously-compounded dividend yield, q, $\Delta = -e^{-q(T-t)}\Phi(-d_1)$.

4 Implied volatility

Both the standard Black-Scholes formula and the Garman-Kohlhagen formula for a dividend-paying share require knowledge of σ, the volatility of the price of the underlying asset. Unlike the other parameters in these formulae, the volatility cannot be observed directly in the market. It must therefore be estimated if either formula is to be used to price a derivative in practice.

One way of doing this is to use an observed option price from the market and derive the volatility that is consistent with this price. This is possible because we can observe the values of the current price of the underlying asset, the risk-free rate of interest and the dividend yield on the underlying asset. This information is added to the strike price and the maturity date leaving the volatility as the only unknown quantity in the Black-Scholes formula. The resulting estimate is known as the *implied volatility*.

We observed in an earlier chapter that the higher the volatility of the underlying share price, the greater the chance that the underlying share price can move significantly in favour of the holder of the option before expiry. So higher volatility will be associated with higher option prices, and lower volatility with lower option prices.

Suppose we know that the price of an option is 6.87. It is not possible for us to write down a nice formula for σ in terms of the other parameters. However, it is straightforward to find the value of σ working backwards by trial and error. Since the option price is a strictly increasing function of σ, the solution to this problem is unique.

Suppose we find that with σ = 0.17, Black-Scholes implies that the corresponding option price would be 6.841 and with σ = 0.18, the price would be 7.006. Using linear interpolation we can then estimate the volatility as σ = 0.17176. This estimate is called the *implied volatility*.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 15 Summary

Assumptions of Black-Scholes

- The underlying share price follows geometric Brownian motion.
- The market is arbitrage-free.
- The risk-free rate *r* is constant, the same for all maturities and the same for borrowing and lending.
- Assets may be bought and sold at any time $t > 0$.
- Assets may be held in any amount, including negative holdings.
- There are no taxes or transaction costs.

An implication of these is that the market in the underlying share is complete. The validity of each assumption can be questioned.

Black Scholes PDE

$$rf(S_t) = \Theta + \Delta(r-q)S_t + \frac{\sigma^2}{2}\Gamma S_t^{\,2}$$

For any derivative to be fairly priced, it must:

- satisfy the boundary conditions, *ie* have the correct payoff at expiry and
- satisfy the above PDE.

Garman-Kohlhagen formulae for options on a dividend-paying share

European call

$$f(S_t) = S_t\Phi(d_1)e^{-q(T-t)} - Ke^{-r(T-t)}\Phi(d_2)$$

European put

$$f(S_t) = Ke^{-r(T-t)}\Phi(-d_2) - S_t\Phi(-d_1)e^{-q(T-t)}$$

where:

$$d_1 = \frac{\ln\left(S_t/K\right) + (r-q+½\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\ , \qquad d_2 = d_1 - \sigma\sqrt{T-t}$$

The Black-Scholes formula for a non-dividend-paying share is the same but with $q = 0$.

The practice questions start on the next page so that you can keep the chapter summaries together for revision purposes.

Chapter 15 Practice Questions

15.1 Exam style

A building society issues a one-year bond that entitles the holder to the return on a weighted-average share index (ABC500) up to a maximum level of 30% growth over the year. The bond has a guaranteed minimum level of return so that investors will receive at least $x\%$ of their initial investment back. Investors cannot redeem their bonds prior to the end of the year.

(i) Explain how the building society can use a combination of call and put options to prevent making a loss on these bonds. [4]

(ii) The volatility of the ABC500 index is 30% *pa* and the continuously compounded risk-free rate of return is 4% *pa*. Assuming no dividends, use the Black-Scholes pricing formulae to determine the value of x (to the nearest 1%) that the building society should choose to make neither a profit nor a loss. [6]

[Total 10]

15.2 Exam style

A company's directors have decided to provide senior managers with a performance bonus scheme. The bonus scheme entitles the managers to a cash payment of £10,000 should the company share price have increased by more than 20% at the end of the next 6 months. In addition, the managers will be entitled to 5,000 free shares each, should the share price have increased by more than 10% at the end of the next 6 months.

You are given the following data:

Current share price	£7.81
Risk-free rate	5% *pa* (continuously compounded)
Share price volatility	25% *pa*

No dividends to be paid over the next 6 months.

(i) By considering the terms of the Black-Scholes call option pricing formula, calculate the value of the bonus scheme to one manager. [6]

(ii) Explain the main disadvantages of this bonus scheme as an incentive for managers to perform. [2]

(iii) Some shareholders are concerned that this scheme might cause an undesirable distortion to the managers' behaviour. Suggest some modifications to the scheme that will ensure that the managers' aims coincide with the long-term objectives of the shareholders. [3]

[Total 11]

15.3 (i) One of the assumptions underlying the Black-Scholes model is that the price of the underlying follows a geometric Brownian motion. Explain briefly what this means and why this assumption may not be valid in practice.

(ii) State the other assumptions underlying the Black-Scholes model.

15.4 One form of the Black-Scholes partial differential equation is:

$$\theta + rs\Delta + \tfrac{1}{2}\sigma^2 s^2 \Gamma = rf$$

(i) State the context in which this formula applies, indicating what s and f represent. (You are not required to state the Black-Scholes assumptions.)

(ii) What do r and σ represent? What assumptions are made about these quantities in this equation?

(iii) State the names and give the mathematical definitions of the 'Greeks' that appear in this equation.

(iv) What boundary condition would you need to use in order to solve this equation when applied to a European call option with a strike price K?

15.5 Exam style

An investor buys, for a premium of 187.06, a call option on a non-dividend-paying stock whose current price is 5,000. The strike price of the call is 5,250 and the time to expiry is 6 months. The risk-free rate of return is 5% *pa* continuously compounded.

The Black-Scholes formula for the price of a call option on a non-dividend-paying share is assumed to hold.

(i) Calculate the price of a put option with the same time to maturity and strike price as the call. [2]

(ii) The investor buys a put option with strike price 4,750 with the same time to maturity. Calculate the price of the put option if the implied volatility were the same as that in (i).

[You need to estimate the implied volatility to within 1% *pa* of the correct value.] [7]
[Total 9]

15.6 Exam style

The solution to the Black-Scholes equation for the price V (assuming a risk-free force of interest r) of a European put option maturing u years from now with strike price K on a stock that pays dividends at force q whose current spot price is S is:

$$V = Ke^{-ru}\Phi(-d_2) - Se^{-qu}\Phi(-d_1)$$

where $d_1, d_2 = \dfrac{\log(S/K) + (r - q \pm ½\sigma^2)u}{\sigma\sqrt{u}}$.

(i) Show that the hedge ratio $\Delta = \dfrac{\partial V}{\partial S}$ is given by $\Delta = -e^{-qu}\Phi(-d_1)$. [8]

(ii) Hence find a formula for $\Gamma = \dfrac{\partial^2 V}{\partial S^2}$. [2]

[Total 10]

15.7 An investment bank has issued a special derivative security which provides a payoff in one year of:

Exam style

$S_1-(S_0-15)$	if	$S_0-15 \le S_1 \le S_0-5$
10	if	$S_0-5 \le S_1 \le S_0+5$
$(S_0+15)-S_1$	if	$S_0+5 \le S_1 \le S_0+15$
0	otherwise	

where S_t is the price of the underlying share at time t.

An investor purchases one of these special derivatives on a share with initial price £50.

(i) Write down the investor's payoff from this special derivative in one year's time. [1]

(ii) Explain how this payoff can be written in terms of two long and two short call options with different strike prices. [4]

(iii) Calculate the fair price for this special derivative paid by the investor, using the following basis:

- volatility of the share price, $\sigma = 15\%$ *pa*
- risk-free interest rate, $r = 3\%$ *pa* (continuously compounded)
- no dividends are paid on the underlying share. [5]

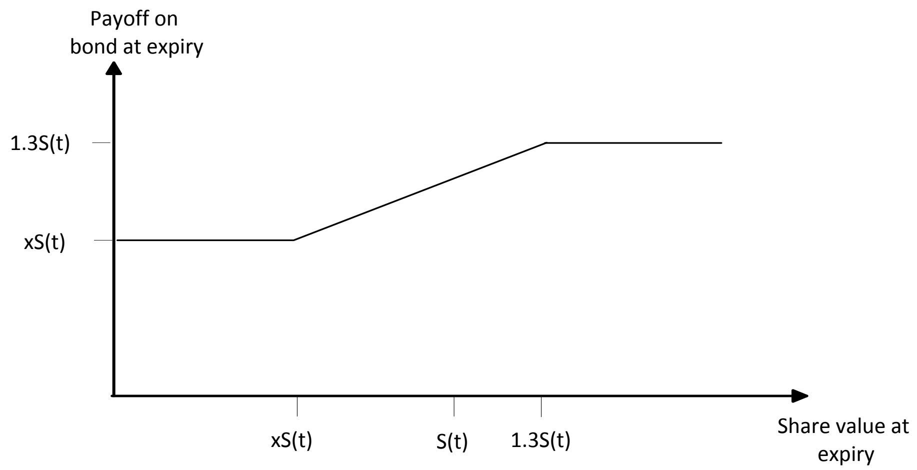
[Total 10]

The solutions start on the next page so that you can separate the questions and solutions.

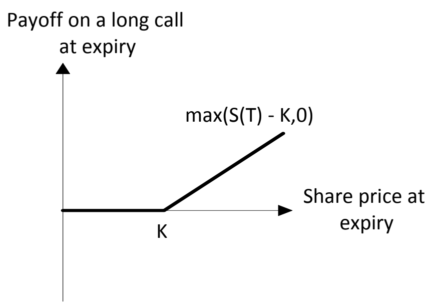
Chapter 15 Solutions

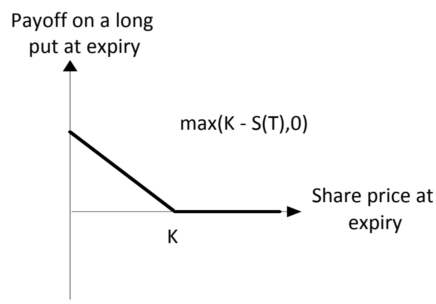
15.1 (i) ***Preventing a loss***

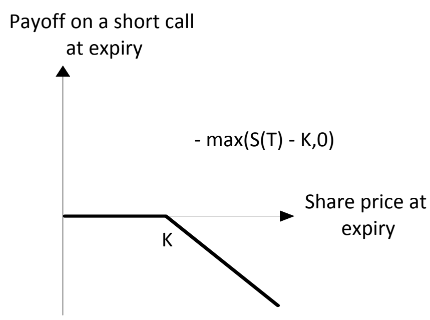
A helpful way of tackling tricky Black-Scholes questions such as this is to draw a graph comparing the payoff on the bond at expiry with the value of the underlying asset at expiry. In the diagram below, S(t) is the value of the initial investment.

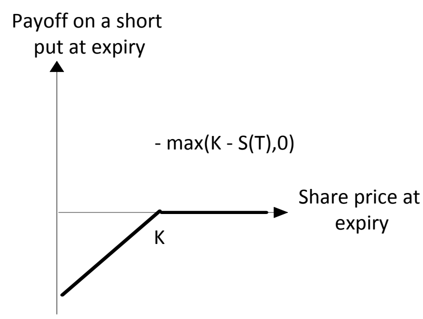


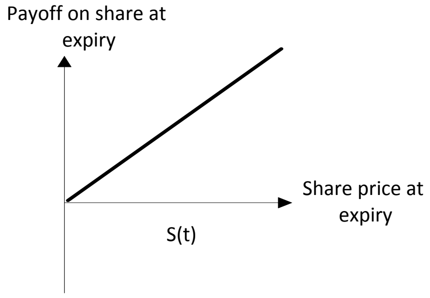
We can compare the shape of this graph against graphs for the payoffs on call and put options, and on the underlying shares.











The graphs demonstrate that a combination of the underlying shares, a long put with exercise price xS_t *and a short call with exercise price* $1.3S_t$ *will replicate the graph and hence the payoff on the bond.*

If an investor buys a bond the building society can invest the money in the ABC500 so that it is not exposed to movements in the ABC500 index. However, the building society is guaranteeing that investors will receive at least $x\%$ of their initial investment back. The building society can hedge this loss by buying a put option on the index with a strike price of x% of the current share price. This put option will cost money – let's say p. [2]

The building society is also limiting the investors' return to 130% of their initial investment. They can do this by selling call options with a strike price of 130% of the current share price. This call option will be priced at c, say. If $c \geq p$ then the building society will not make a loss. [2]

[Total 4]

(ii) ***No profit or loss***

If $c = p$ then the building society will not make a profit or a loss. So the problem requires us to work out the price of the call option c and then work out the value of x such that $c = p$.

We will be using the Black-Scholes formula to price the options and, because the numbers are all relative, we can assume that the initial index price is 100, say.

Using the formulae on page 47 of the *Tables*, we can calculate the price of a call option:

$$d_1 = \frac{\ln(S_t/K) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$= \frac{\ln\left(\frac{100}{130}\right) + \left(0.04 + \frac{0.3^2}{2}\right)\times 1}{0.3\sqrt{1}} = -0.5912$$

$$d_2 = d_1 - \sigma\sqrt{T-t} = -0.5912 - 0.3\sqrt{1} = -0.8912 \quad [1]$$

$$c = S_t\Phi(d_1) - K\Phi(d_2)e^{-r(T-t)}$$
$$= 100\times\Phi(-0.5912) - 130\times\Phi(-0.8912)e^{-0.04}$$
$$= 100\times 0.277 - 130e^{-0.04}\times 0.186$$
$$= 4.44 \qquad [1]$$

We now need to work out the value of x so that $p = 4.44$. We will try $K = 90$ to begin with:

$$d_1 = \frac{\ln\left(S_t/K\right) + \left(r + \dfrac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$
$$= \frac{\ln\left(\dfrac{100}{90}\right) + \left(0.04 + \dfrac{0.3^2}{2}\right)\times 1}{0.3\sqrt{1}} = 0.6345$$

$$d_2 = d_1 - \sigma\sqrt{T-t} = 0.6345 - 0.3\sqrt{1} = 0.3345 \qquad [1]$$

$$p = K\Phi(-d_2)e^{-r(T-t)} - S_t\Phi(-d_1)$$
$$= 90\times\Phi(-0.3345)e^{-0.04} - 100\times\Phi(-0.6345) \qquad [1]$$
$$= 90e^{-0.04}\times 0.369 - 100\times 0.263$$
$$= 5.62$$

This is higher than the required value, and so we try a lower value for the strike price, say $K = 80$:

$$d_1 = \frac{\ln\left(S_t/K\right) + \left(r + \dfrac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$
$$= \frac{\ln\left(\dfrac{100}{80}\right) + \left(0.04 + \dfrac{0.3^2}{2}\right)\times 1}{0.3\sqrt{1}} = 1.0271$$

$$d_2 = d_1 - \sigma\sqrt{T-t} = 1.0271 - 0.3\sqrt{1} = 0.7271$$

$$p = K\Phi(-d_2)e^{-r(T-t)} - S_t\Phi(-d_1)$$
$$= 80\times\Phi(-0.7271)e^{-0.04} - 100\times\Phi(-1.0271)$$
$$= 80e^{-0.04}\times 0.234 - 100\times 0.152$$
$$= 2.74 \qquad [1]$$

We require a put option with a premium of $p = 4.44$. So we linearly interpolate to find the value of K that will give us this:

$$K = 80 + (90-80) \times \frac{4.44-2.74}{5.62-2.74} = 85.90$$ [1]

So $x = 86\%$ approximately. So, the building society can use $x = 86\%$ and the discussed hedging portfolio to avoid making a loss.

In fact, the exact figure is 86.43%, which is still 86% to the nearest 1%.

[Total 6]

15.2 (i) ***Value of the bonus scheme***

First, consider the share options. The manager will receive 5,000 shares if the share price in 6 months' time is greater than:

$$£7.81 \times 1.1 = £8.59$$ [½]

This is like having 5,000 call options on the share with a strike price of £8.59, except that no payment is actually required. We can use the Black-Scholes formula for one call option, which is on page 47 of the *Tables* (the Garman-Kohlhagen formula with $q = 0$):

$$C_{[K=8.59]} = 7.81\Phi(d_1) - 8.59e^{-0.05\times 6/12}\Phi(d_2)$$ [1]

But the manager will not need to pay the £8.59 so the second term is not required. So, the value of each share option is:

$$\begin{aligned} 7.81\Phi(d_1) &= 7.81\Phi\left[\frac{\ln(S/K) + \left(r + \sigma^2/2\right)(T-t)}{\sigma\sqrt{T-t}}\right] \\ &= 7.81\Phi\left[\frac{\ln(1/1.1) + \left(0.05 + 0.25^2/2\right)\times 0.5}{0.25\sqrt{0.5}}\right] \\ &= 7.81\Phi(-0.30935) \\ &= 7.81[1 - 0.6215] \\ &= £2.96 \end{aligned}$$ [1]

So, the share options for each manager are worth $5{,}000 \times £2.96 = £14{,}800$. [½]

Next, consider the £10,000 cash bonus. The managers will receive this if the share price is greater than:

$$£7.81 \times 1.2 = £9.37$$ [½]

The value of a call option here would be:

$$C_{[K=9.37]} = 7.81\Phi(d_1) - 9.37e^{-0.05\times 6/12}\Phi(d_2)$$ [½]

The second term here corresponds to the value of the strike price £9.37 that would be paid if the share price is greater than £9.37 in 6 months' time.

The cash bonus is made in the same circumstances as this. However, the amount is £10,000 rather than £9.37. [½]

So the value of the cash bonus is:

$$10{,}000e^{-0.05\times\frac{6}{12}}\Phi(d_2) = 10{,}000\times 0.97531\Phi\left[\frac{\ln(S/K)+\left(r-\sigma^2/2\right)(T-t)}{\sigma\sqrt{T-t}}\right]$$

$$= 9{,}753.1\Phi\left[\frac{\ln(7.81/9.37)+\left(0.05-0.25^2/2\right)\times 0.5}{0.25\sqrt{0.5}}\right]$$

$$= 9{,}753.1\Phi(-0.97833)$$

$$= 9{,}753.1(1-0.8360)$$

$$= £1{,}600$$ [1]

Finally, the total value of the bonus scheme is:

$$£14{,}800 + £1{,}600 = £16{,}400$$ [½]

[Total 6]

(ii) ***Incentive for managers to perform?***

Both the £10,000 cash bonus and the share options do not give managers any incentive to help the company share price once the 6-month period is over. [½]

Managers may be able to sell their free shares, add the proceeds to the £10,000 and may have little interest in how the company subsequently performs. [½]

In addition, because the managers do not receive any bonus at all whether the share price increases by 9%, or decreases by 50%, say, they may be tempted to undertake a riskier business/investment strategy that is not in the best interests of the shareholders. [½]

Any increase above 20% is not further rewarded. [½]

[Total 2]

(iii) ***Improvements to the scheme***

The share options will provide managers more of a long-term incentive if they are restricted from selling the shares for a fixed time period after they are awarded, 3 years say. [½]

In addition, a condition may be imposed that the shares will only be awarded provided the manager continues to work for the company for a fixed time period, 3 years say. [½]

These will ensure the managers have an interest in how the company performs after the 6-month period is over. [½]

Instead of a cash bonus, the managers could be given the equivalent amount in more bonus shares, again with the restrictions mentioned above. [½]

The number of free shares issued could be made to depend more gradually on the company's share price performance, *eg* 100 free shares for each percentage point performance above a specified benchmark level. [½]

This may stop any manager being tempted to employ an all-or-nothing approach in their business/investment strategy. [½]

It may be possible to pay the managers' salaries almost entirely in shares so that their interests are the same as that of the shareholders. [½]

[Maximum 3]

15.3 (i) ***Explain why the assumption might not be valid***

This means that S_t, the share price at time t, can be considered to be a random variable that obeys the stochastic differential equation:

$$dS_t = S_t(\mu dt + \sigma dZ_t)$$

where Z_t represents a standard Brownian motion.

Another way of expressing this is to say that the distribution of $\log \frac{S_t}{S_0}$ is $N\left((\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t\right)$ and that movements in S_t in non-overlapping time intervals are statistically independent.

This will not be realistic in practice for any underlying asset that experiences sudden changes (discontinuous jumps) in price or whose drift or volatility is not constant.

(ii) ***Other assumptions***

The other Black-Scholes assumptions are:

- There are no risk-free arbitrage opportunities.
- The risk-free interest rate is constant, the same for all maturity dates and the same for borrowing or lending.
- Unlimited short selling (*ie* having a negative holding) is allowed.
- There are no taxes or transaction costs.
- The underlying asset can be traded continuously and in arbitrarily small amounts.

15.4 (i) ***State the context, and indicate what s and f represent***

This PDE describes the relationship between the theoretical price f of a derivative and the current price s of the underlying stock (which does not pay dividends) on whose value the derivative payoff depends.

(ii) ***What do r and σ represent?***

r represents the force of interest earned on a risk-free asset (usually expressed as an annualised rate).

σ represents the volatility of the underlying stock, *ie* the standard deviation of the log of the price ratio movements (again usually expressed as an annualised rate).

This PDE assumes that r and σ are constant.

(iii) ***Names of the Greeks***

Theta is defined as $\Theta = \dfrac{\partial f}{\partial t}$.

Delta is defined as $\Delta = \dfrac{\partial f}{\partial S_t}$.

Gamma is defined as $\Gamma = \dfrac{\partial^2 f}{\partial S_t^2}$.

(iv) ***Boundary condition***

The boundary condition would specify the payoff at the end of the contract (time T):

$$f(T, S_T) = \max(S_T - K, 0)$$

15.5 (i) ***Calculate the price of a put option with strike price 5,250***

The put-call parity relationship states that:

$$c_t + Ke^{-r(T-t)} = p_t + S_t \qquad [1]$$

Substituting in the values given:

$$187.06 + 5{,}250e^{-0.05\times ½} = p_t + 5{,}000$$

$$\Rightarrow \quad p_t = 307.44 \qquad [1]$$

[Total 2]

(ii) ***Calculate the price of a put option with strike price 4,750***

We first need to estimate the implied volatility of the stock using the information given in the question. This can be done by trial and improvement. We start by substituting the parameter values into the Black-Scholes formula:

$$c_t = 5{,}000\Phi(d_1) - 5{,}250e^{-0.05\times½}\Phi(d_2)$$

where $d_1 = \dfrac{\log\left(\dfrac{5{,}000}{5{,}250}\right) + \left(0.05 + ½\sigma^2\right)\times ½}{\sigma\sqrt{½}}$ and $d_2 = d_1 - \sigma\sqrt{½}$ [1]

In most instances, $\sigma = 0.2$ is a reasonable starting point for the interpolation. If $\sigma = 0.2$, then substituting both this value and the other parameter values into the Black-Scholes formula gives:

$$d_1 = -0.0975 \text{ and } d_2 = -0.2389$$ [½]

$$\Rightarrow \quad c_t = 5{,}000\underbrace{\Phi(-0.0975)}_{0.4612} - 5{,}250e^{-0.05\times½}\underbrace{\Phi(-0.2389)}_{0.4056} = 229.18$$ [1]

This is above the actual price of 187.06, so we need to try a lower value of σ. If we try $\sigma = 0.1$, then we obtain:

$$d_1 = -0.3011 \text{ and } d_2 = -0.3718$$ [½]

$$\Rightarrow \quad c_t = 5{,}000\underbrace{\Phi(-0.3011)}_{0.3817} - 5{,}250e^{-0.05\times½}\underbrace{\Phi(-0.3718)}_{0.3550} = 90.77$$ [1]

As the two call option prices straddle the actual price of 187.06, we can interpolate between the two values of σ to obtain an estimate for the implied volatility:

$$\frac{187.06 - 90.77}{229.18 - 90.77} \approx \frac{\sigma - 0.1}{0.2 - 0.1} \quad \Rightarrow \quad \sigma \simeq 17\%$$ [1]

We can now use this estimate of σ to determine the price of a put option with a strike price of 4,750:

$$d_1 = \frac{\log\left(\dfrac{5{,}000}{4{,}750}\right) + \left(0.05 + ½\times 0.17^2\right)\times ½}{0.17\times\sqrt{½}} = 0.6948$$ [½]

and:

$$d_2 = d_1 - 0.17\times\sqrt{½} = 0.5746$$ [½]

So the price of the 4,750 put option is:

$$p_t = 4{,}750e^{-0.05\times½}\Phi(-0.5746) - 5{,}000\Phi(-0.6948) = 92.11$$ [1]

[Total 7]

15.6 (i) ***Formula for delta***

Differentiating the formula given for V with respect to S, using the product rule for the second term, we get:

$$\Delta = Ke^{-ru} \frac{\partial}{\partial S} \Phi(-d_2) - e^{-qu} \Phi(-d_1) - Se^{-qu} \frac{\partial}{\partial S} \Phi(-d_1) \qquad [1]$$

Using the chain rule, this can be written as:

$$\Delta = -Ke^{-ru} \phi(-d_2) \frac{\partial d_2}{\partial S} - e^{-qu} \Phi(-d_1) + Se^{-qu} \phi(-d_1) \frac{\partial d_1}{\partial S} \qquad [1]$$

Here $\phi(\,)$ denotes the derivative of $\Phi(\,)$ *ie* it is the probability density function of the standard normal distribution, which is:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-½z^2} \qquad [½]$$

From the definitions of d_1 and d_2, we find that:

$$\frac{\partial d_1}{\partial S} = \frac{\partial d_2}{\partial S} = \frac{1}{S\sigma\sqrt{u}} \qquad [½]$$

Putting these together and regrouping gives:

$$\Delta = \frac{-Ke^{-ru} e^{-½d_2^2} + Se^{-qu} e^{-½d_1^2}}{\sqrt{2\pi} S\sigma\sqrt{u}} - e^{-qu} \Phi(-d_1) \qquad [1]$$

In fact, the first term is zero. To show this, we add and subtract the definitions of d_1 and d_2 to get:

$$d_1 + d_2 = \frac{2\log(S/K) + 2(r-q)u}{\sigma\sqrt{u}} \qquad [½]$$

and $$d_1 - d_2 = \frac{\sigma^2 u}{\sigma\sqrt{u}} = \sigma\sqrt{u} \qquad [½]$$

Multiplying these two equations gives us a difference of two squares:

$$d_1^2 - d_2^2 = 2\log(S/K) + 2(r-q)u \qquad [1]$$

Halving gives:

$$\tfrac{1}{2}d_1^2 - \tfrac{1}{2}d_2^2 = \log(S/K) + (r-q)u$$

Exponentiating gives:

$$e^{½d_1^2}e^{-½d_2^2}=\frac{S}{K}\times e^{ru}e^{-qu}$$

Rearranging then gives:

$$Ke^{-ru}e^{-½d_2^2}=Se^{-qu}e^{-½d_1^2}$$ [1]

ie the numerator in the expression for Δ is zero.

So: $\Delta=-e^{-qu}\Phi(-d_1)$ [1]

[Total 8]

It would not be acceptable here to differentiate the expression for V ignoring the fact that d_1 *and* d_2 *are functions of S. Although this happens to give the correct answer, it is not a valid method.*

(ii) ***Formula for gamma***

Finding the second derivative, starting from the formula in (i), is more straightforward:

$$\Gamma=\frac{\partial^2 V}{\partial S^2}=\frac{\partial\Delta}{\partial S}=\frac{\partial}{\partial S}\left[-e^{-qu}\Phi(-d_1)\right]$$ [½]

So:

$$\Gamma=-e^{-qu}\frac{\partial}{\partial S}\Phi(-d_1)=e^{-qu}\phi(-d_1)\frac{\partial d_1}{\partial S}$$

$$=e^{-qu}\frac{e^{-½d_1^2}}{\sqrt{2\pi}}\frac{1}{S\sigma\sqrt{u}}=\frac{e^{-½d_1^2-qu}}{S\sigma\sqrt{2\pi u}}$$ [1½]

[Total 2]

15.7 (i) ***Investor's payoff***

If $S_0=50$, then the payoff is:

S_1-35	if $35\le S_1\le 45$	
10	if $45\le S_1\le 55$	
$65-S_1$	if $55\le S_1\le 65$	
0	otherwise	[1]

Alternatively, this may be expressed:

0	if $S_1 \le 35$
$S_1 - 35$	if $35 \le S_1 \le 45$
10	if $45 \le S_1 \le 55$
$65 - S_1$	if $55 \le S_1 \le 65$
0	if $S_1 \ge 65$

(ii) ***Express the payoff in terms of call options***

The payoff from the special derivative can be replicated with the following combination of call options:

- a long call option with strike price £35 [1]
- a short call option with strike price £45 [1]
- a short call option with strike price £55 [1]
- a long call option with strike price £65 [1]

[Total 4]

With this combination of options the overall payoff is:

$$\max(S_1 - 35, 0) - \max(S_1 - 45, 0) - \max(S_1 - 55, 0) + \max(S_1 - 65, 0)$$

We can check that this combination of options does indeed give the same payoff as the special derivative, by considering the payoff for different ranges of the share price.

If $S_1 \le 35$*, no option is exercised, so the payoff is 0.*

If $35 \le S_1 \le 45$*, only the call option with strike price £35 is exercised, so the payoff is:*

$$S_1 - 35$$

If $45 \le S_1 \le 55$*, the call options with strike prices £35 and £45 will be exercised, giving a payoff of:*

$$S_1 - 35 - (S_1 - 45) = 10$$

If $55 \le S_1 \le 65$*, the call options with strike prices £35, £45 and £55 will be exercised, giving a payoff of:*

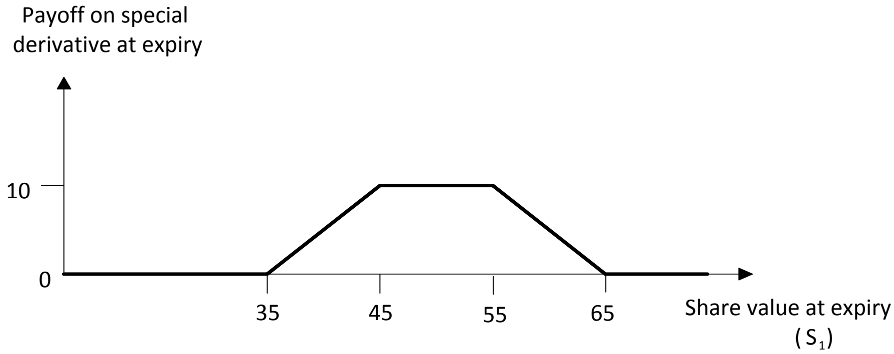
$$S_1 - 35 - (S_1 - 45) - (S_1 - 55) = 65 - S_1$$

If $S_1 \ge 65$*, all call options will be exercised, giving a payoff of:*

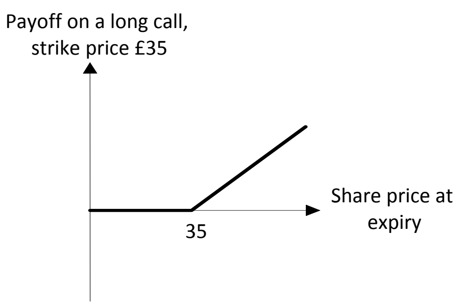
$$S_1 - 35 - (S_1 - 45) - (S_1 - 55) + S_1 - 65 = 0$$

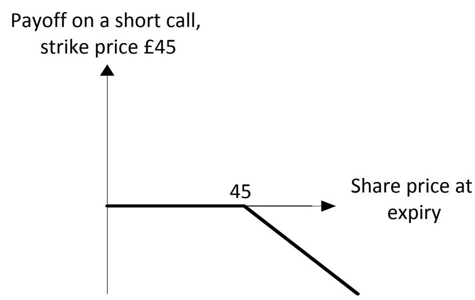
So, in all cases, this combination of call options gives the same payoff as the special derivative.

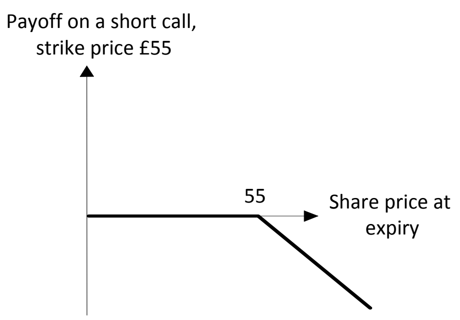
Alternatively, it may help to consider the graph of the payoff from the special derivative:

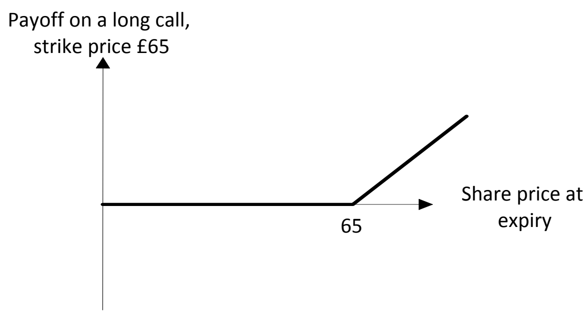


This can be replicated by superimposing the following four payoff graphs relating to the four call options:









(iii) ***Cost of special derivative***

To work out the cost of the special derivative, we can calculate the total cost of the combination of the four call options that replicate its payoff.

In this case we have $S_0 = 50, \sigma = 15\%, r = 3\%, T - t = 1$ and $q = 0$, so the Black-Scholes formula for the price of a call option is:

$$50\Phi(d_1) - Ke^{-0.03}\Phi(d_2)$$

where:

$$d_1 = \frac{\ln\left(\frac{50}{K}\right) + 0.03 + \frac{1}{2}(0.15)^2}{0.15} \quad \text{and} \quad d_2 = d_1 - 0.15$$

For the call option with strike price £35:

$$d_1 = 2.6528 \;\Rightarrow\; \Phi(d_1) = 0.99601$$

$$d_2 = 2.5028 \;\Rightarrow\; \Phi(d_2) = 0.99384$$

So, the first call option has value:

$$c_1 = 50 \times 0.99601 - 35e^{-0.03} \times 0.99384 = £16.04$$ [1]

For the call option with strike price £45:

$$d_1 = 0.9774 \;\Rightarrow\; \Phi(d_1) = 0.83582$$

$$d_2 = 0.8274 \;\Rightarrow\; \Phi(d_2) = 0.79600$$

So, the second call option has value:

$$c_2 = 50 \times 0.83582 - 45e^{-0.03} \times 0.79600 = £7.03$$ [1]

For the call option with strike price £55:

$$d_1 = -0.3604 \;\Rightarrow\; \Phi(d_1) = 0.35927$$

$$d_2 = -0.5104 \;\Rightarrow\; \Phi(d_2) = 0.30489$$

So, the third call option has value:

$$c_3 = 50 \times 0.35927 - 55e^{-0.03} \times 0.30489 = £1.69$$ [1]

For the call option with strike price £65:

$$d_1 = -1.4741 \;\Rightarrow\; \Phi(d_1) = 0.07023$$

$$d_2 = -1.6241 \;\Rightarrow\; \Phi(d_2) = 0.05218$$

So, the fourth call option has value:

$$c_4 = 50 \times 0.07023 - 65e^{-0.03} \times 0.05218 = £0.22$$ [1]

The overall cost of the special derivative is therefore:

$$16.04 - 7.03 - 1.69 + 0.22 = £7.54$$ [1]

[Total 5]

16

The 5-step method in discrete time

Syllabus objectives

6.1 Option pricing and valuations

6.1.8 Demonstrate an understanding of the Black-Scholes derivative-pricing model:

- Explain what is meant by a complete market.
- Explain what is meant by risk-neutral pricing and the equivalent martingale measure.
- Demonstrate how to price and hedge a simple derivative contract using the martingale approach. (part)

0 Introduction

This chapter centres on the development of a derivative pricing approach based on the use of *replicating strategies*. It is referred to here as the 5-step method. Before presenting the 5-step method in the continuous case, in this chapter we will demonstrate the structure of the proof in the simpler discrete case, *ie* for the binomial model.

The approach in the binomial case involves constructing a replicating portfolio. In other words, at a given time $t-1$ we need to be able to set up a portfolio of shares and cash that will evolve by time t to be worth the same as the derivative at that time, no matter how the share price moves. In the binomial case there are only two possibilities for the share price movement over each time step. This will allow us to calculate the correct amounts of shares ϕ_t and cash ψ_t simply by solving two simultaneous equations.

However, there is a problem in continuous time with such a replicating strategy. In the continuous case the number of possible share price movements is infinite, in fact they form a continuum, and so this simple method might seem doomed to failure.

The solution lies in recognising that the concepts underlying the replication method can be formulated in terms of *martingales.* A powerful theorem known as the *martingale representation theorem* then comes to the rescue. An application of this theorem confirms the existence of the ϕ and ψ we need for replication. This forms the basis of the *martingale approach* to pricing or *risk-neutral valuation*.

The formula we end up with is the general derivative pricing formula:

$$V_t = e^{-r(n-t)} E_Q[C_n \mid F_t]$$

We will see in a later chapter that this same formula can be used to derive the Black-Scholes and Garman-Kohlhagen formulae for the call and put options, which we met previously.

Some of the material in this chapter and the following chapter is quite technical.

1 Preliminary concepts

It is helpful to recognise that the basic set-up for the continuous case is analogous to the one used in the discrete-time framework outlined in the chapter on the binomial model. Consequently, we will on occasion refer back to the binomial model. First of all, however, we introduce some preliminary concepts.

Suppose that we are given some probability triple (Ω, F, P) where P is a *real-world* probability measure (sometimes also referred to as the *physical* or *natural* or *objective* measure).

Here Ω is the sample space, *ie* the set of all possible outcomes.

Question

What is represented by F?

Solution

F represents the set of all information concerning the stock price process S_t that could (eventually) be known.

1.1 Background

The share price process and the filtration

We will make use in this and subsequent sections of a stochastic process S_t for prices, where S_t is measurable with respect to F_t (that is, given F_t we know the value of S_u for all $u \le t$). Let F_t be the sigma-algebra generated by S_u (and B_u if this is stochastic) for $0 \le u \le t$; that is, F_t gives the history of the process up to time t.

B_u is the value of a risk-free bond, and is introduced below.

So, S_t will be used to denote the value of the share at time t. Because share prices are random, this is not known before time t, and so we model it with a stochastic process. As mentioned in the chapter on Brownian motion, where it was called a filtration, we think of F_t as meaning that we are at time t, and we know the history of the share value up to and including time t.

The real-world probability measure, P

The real-world probability measure P can be interpreted in the following way. Let A be some event contained in F (for example, suppose that A is the event that S_1 is greater than or equal to 100). Then $P(A)$ is the *actual* probability that the event A will occur.

On a more intuitive level, with m independent realisations of the future instead of one, we would find that the event A occurs on approximately a proportion $P(A)$ occasions (with the approximation getting better as m gets larger and larger).

So at time 0 we may think that the probability that the share price will be greater than or equal to 100 at time t is 0.2. Another way of saying this is that, if we could run the future up until time t lots of times, we would expect the share price to be greater than or equal to 100 about one fifth of the time.

In practice we can only ever *estimate* the real-world probabilities, using a mathematical model of a situation, since it is impossible for us to 'replay' the real world.

Cash

Suppose also that we have a risk-free cash bond which has a value at time t of B_t.

We will refer to this as either the *cash bond* or simply as *cash.*

Sometimes the risk-free rate of interest will itself follow a stochastic process, but in the sections which follow we will assume that the risk-free rate of interest is constant. That is, B_t is deterministic and equal to $B_0 e^{rt}$ for some constant r.

Assuming that $r > 0$, we have the following properties:

- $B_0 = 1$ is the value of cash at $t = 0$
- $B_t > B_0 = 1$ and $B_t > B_{t-1}$ for $t > 0$
- B_t increases in an entirely predictable manner as we move through time – *ie* at the continuously compounding rate of r per time period.

Be careful not to confuse this with when B_t is used as the notation for Brownian motion. It should be clear from the context in which B_t is being used but we will add comments in sections where the distinction is not so obvious.

1.2 Tradeable assets

In the basic form of the Black-Scholes model that we will see in the next section, we will assume that S_t represents the price of a non-dividend-paying share. This means that if we invest S_0 at time 0, S_t represents the *total return* on the investment up to time t, assuming that we hold onto the share until that time.

This, as well as the cash bond B_t, is an example of a *tradeable asset*. Such an asset is one where its price at time t is equal to the total return on that investment up to time t with no dividend income payable or inputs of cash required.

1.3 Self-financing strategies

We now consider the properties that are necessary for a replicating portfolio.

The portfolio

Suppose that at time t we hold the portfolio (ϕ_t, ψ_t), where:

- **ϕ_t represents the number of units of S_t held at time t**
- **ψ_t is the number of units of the cash bond held at time t.**

So a portfolio is an ordered pair of processes ϕ_t and ψ_t that describes the number of units of each security held at time *t*. Note that:

- we do not constrain the values of ϕ_t and ψ_t to be non-negative – *ie* short selling of securities is permitted.
- the choice of the portfolio (ϕ_t, ψ_t) at time *t* represents a *dynamic* strategy, as the values of ϕ_t and ψ_t can change continuously through time.

We assume that S_t is a tradeable asset as described above.

Previsible processes

The only significant requirement on (ϕ_t, ψ_t) is that they are *previsible*; that is, F_{t^-} -measurable (so ϕ_t and ψ_t are known based upon information up to *but not including* time t).

So, ϕ_t and ψ_t are dependent only on the history of stock prices up to *but not including* time *t*.

This is sometimes referred to as the history of stock prices up to time t^- and the corresponding filtration is written as F_{t^-}. In the binomial case, $F_{t^-} = F_{t-1}$.

The reason we require ϕ_t and ψ_t to be previsible is that we want to be able to replicate the portfolio in advance. Consider the binomial case again. At time $t-1$ we need to be able to set up a portfolio that replicates the value of the derivative at time t, no matter what happens to the share price. It is no use being able to do this in retrospect.

In continuous time we would need to continuously change the holdings ϕ_t and ψ_t. This is obviously not possible in practice, but we should be able to approximate the theoretical ideal by rebalancing the portfolio on a regular basis.

Changes in the value of the portfolio

Let $V(t)$ be the value at time t of this portfolio: that is:

$$V(t) = \phi_t S_t + \psi_t B_t$$

Now consider the instantaneous pure investment gain in the value of this portfolio over the period from time t up to $t+dt$: that is, assuming that there is no inflow or outflow of cash during the period $[t,t+dt)$.

Any increase or decrease in the value of the portfolio is therefore due entirely to capital gains or losses generated within the fund. We are assuming that the stock does not pay dividends. There is no net investment/new money injected into or withdrawn from the fund.

This instantaneous pure investment gain is equal to:

$$\phi_t dS_t + \psi_t dB_t$$

Question

Why is the pure investment gain equal to $\phi_t dS_t + \psi_t dB_t$?

Solution

In this equation, dB_t represents the infinitesimal change in the value of one unit of cash during the time interval $[t, t + dt)$. dS_t represents the corresponding change in the stock price, which is a random variable.

The absence of any new net investment into or out of the portfolio means that there are no new purchases or sales of either the stock or the cash bond. Hence, ϕ_t and ψ_t remain unchanged over this short time interval. Consequently, any changes to the value of the portfolio over the time interval $[t, t + dt)$ must arise solely through changes in the stock price dS_t and /or the increase in the value of cash dB_t .

Let's see what the change in the value would be if we did have inflows or outflows, *ie* if we were to buy $d\phi_t$ new shares and pay in $d\psi_t$ units of the cash bond.

The instantaneous change in the value of the portfolio, allowing for cash inflows and outflows, is given by:

$$dV_t = (\phi_t + d\phi_t)(S_t + dS_t) + (\psi_t + d\psi_t)(B_t + dB_t) - (\phi_t S_t + \psi_t B_t)$$

which simplifies to:

$$dV(t) \equiv V(t+dt) - V(t) = \phi_t dS_t + S_t d\phi_t + d\phi_t dS_t + B_t d\psi_t + \psi_t dB_t + d\psi_t dB_t$$

If $d\phi_t = d\psi_t = 0$, this simplifies to:

$$dV_t = \phi_t dS_t + \psi_t dB_t$$

Self-financing

The portfolio strategy is described as *self-financing* if $dV(t)$ is equal to $\phi_t dS_t + \psi_t dB_t$. That is, at $t+dt$ there is no inflow or outflow of money necessary to make the value of the portfolio back up to $V(t+dt)$.

ie $\quad (\phi_t, \psi_t)$ is self-financing $\Leftrightarrow$ $dV_t = \phi_t dS_t + \psi_t dB_t$

So, the required change in the value of a self-financing portfolio is equal to the pure investment gain. The significance of being self-financing will become clearer once we talk about replicating strategies.

1.4 Replicating strategies and complete markets

Let X be any derivative payment contingent upon F_U where U is the time of payment of X. That is, X is F_U-measurable and, therefore, depends upon the path of S_t up to time U. The time of payment U may be fixed or it may be a random stopping time.

'X is F_U-measurable' means that if we know the history of stock prices up to and including time U, then we know the value of X. In this particular instance, F_U tells us the value of S_U, the stock price at time U, and hence the value of X, the derivative payment contingent upon S_U.

Question

For which types of derivative is the time of payment U fixed? For which type is it not fixed?

Solution

The time of payment U is fixed for:

- a European option, which can be exercised only at the fixed exercise or expiry date
- a future or a forward, both of which have a fixed delivery date.

The time of payment is not fixed for an American option, where the holder can choose to exercise the option at any time on or before the exercise date.

Replicating strategy

A *replicating strategy* is a self-financing strategy (ϕ_t, ψ_t), defined for $0 \le t < U$, such that:

$$V(U) = \phi_U S_U + \psi_U B_U = X$$

In other words, for an initial investment of $V(0)$ at time 0, if we follow the self-financing portfolio strategy (ϕ_t, ψ_t) we will be able to reproduce the derivative payment without risk.

'Following the self-financing portfolio strategy (ϕ_t, ψ_t)' means that we rebalance our portfolio continuously so that we have ϕ_t units of stock and ψ_t units of cash at time *t*, and we do this without any inflow or outflow of funds.

Hence, the self-financing portfolio strategy (ϕ_t, ψ_t) replicates the derivative payment *X* at the time of payment *U*, irrespective of what S_U, the stock price at time *U*, and hence the value of *X*, actually turns out to be. Moreover, if we choose the initial portfolio (ϕ_0, ψ_0) at time 0, then we can replicate the derivative payment at time *U* without the need for any further cash injections into the portfolio.

This is because the change in the value of the portfolio (ϕ_t, ψ_t) over any time interval will be such that its value at the end of that time interval will exactly equal the cost of the portfolio that must then be purchased to maintain the replicating strategy over the next time interval.

Complete market

The market is *complete* if for any such contingent claims X there is a replicating strategy (ϕ_t, ψ_t).

This is important because it means that, in a complete market, we will always be able to price a contingent claim *X*, such as a derivative, based on the arbitrage-free approach and using a replicating strategy (ϕ_t, ψ_t).

We have already seen one example of a complete market: the binomial model. In that model we saw that we could replicate any derivative payment contingent on the history of the underlying asset price.

To replicate a call option or put option we needed shares and cash, which were both available assets. We will see shortly that to replicate these derivatives in continuous time we can use these same assets.

Another example of a complete market is the continuous-time lognormal model for share prices.

The form of the lognormal model used in the Black-Scholes framework is:

$$S_t = S_0 \exp\left[(\mu - \tfrac{1}{2}\sigma^2)t + \sigma Z_t\right]$$

where Z_t is a standard Brownian motion.

1.5 Equivalent measures

Two measures P and Q which apply to the same sigma-algebra F are said to be *equivalent* if for any event E in F:

$$P(E) > 0 \text{ if and only if } Q(E) > 0$$

where $P(E)$ and $Q(E)$ are the probabilities of E under P and Q respectively.

We have already seen equivalent measures when we looked at the binomial model.

We can think of this intuitively as meaning that any event E that is possible/impossible under probability measure P is also possible/impossible under probability measure Q. However, remember that when infinite sequences of events are involved, any particular sequence we specify will normally have probability zero.

Here we will be using P to denote the real-world probabilities and Q for the risk-neutral probabilities.

From the above definition of equivalence, the only constraint on the real-world measure P is that at any point in the binomial tree the probability of an up-move lies strictly between 0 and 1. The only constraint on Q is the same but this can be equated to the requirement that the risk-free return must lie strictly between the return on a down-move and the return on an up-move.

Question

Why is this?

Solution

As was the case in the binomial model chapter, this restriction is required in order to avoid any violation of the no-arbitrage condition – as we are pricing the derivative using an arbitrage-free approach. If arbitrage opportunities were present, then the value of our replicating portfolio might differ from that of the derivative claim that we are trying to price.

Recall that in order to avoid arbitrage we must have $d < e^r < u$, where u and d are the up-step and down-step factors respectively in our binomial tree. Recall from the binomial model chapter that:

$$q = \frac{e^r - d}{u - d}$$

If $0 < q < 1$, then:

$$0 < q = \frac{e^r - d}{u - d} < 1$$

Multiplying this through by $u - d$ gives:

$$0 < e^r - d < u - d$$

ie $\quad d < e^r < u$

This gives us considerable flexibility in the range of possible equivalent measures.

We could, if we wanted to, come up with a whole range of other probability measures, but we will mostly be interested in the real-world and the risk-neutral ones.

1.6 Sample paths problem

In the binomial model, up to a given finite time horizon each sample path has a probability greater than zero – so equivalence is straightforward to prove.

This is because there are only two paths that the share price process can take at each individual node. The share price goes up with probability $p \neq 0$ or down with probability $1-p \neq 0$.

In many continuous-time models (for example, Brownian motion) all individual sample paths have zero probability.

Question

Why do all the sample paths have zero probability?

Solution

The probability that a continuous random variable takes an exact value is zero, *ie* $P[X=x]=0$. We can think of a sample path as the realisation of a 'series of random outcomes up to a certain point', so the random variable logic applies equally to sample paths.

This makes equivalence more difficult to establish.

Remember that, for equivalence, the probabilities for any event (*ie* set of sample paths) have to either both equal zero or be strictly positive together.

Continuous-time example

Suppose that Z_t is a standard Brownian motion under P:

$$Z_t \sim N(0,t) \text{ for all } t$$

Let $X_t = \gamma t + \sigma Z_t$ be a Brownian motion with drift under P.

Is there a measure Q under which X_t is a standard Brownian motion and which is equivalent to P?

The answer is *yes* if $\sigma = 1$ but *no* if $\sigma \neq 1$ (but we do not give a proof here). So we can change the drift of the Brownian motion but not the volatility.

This example can be expressed more formally in the following theorem.

1.7 The Cameron-Martin-Girsanov theorem

Suppose that Z_t is a standard Brownian motion under P. Furthermore suppose that γ_t is a previsible process. Then there exists a measure Q equivalent to P and where:

$$\tilde{Z}_t = Z_t + \int_0^t \gamma_s ds$$

is a standard Brownian motion under Q.

Conversely, if Z_t is a standard Brownian motion under P and if Q is equivalent to P then there exists a previsible process γ_t such that:

$$\tilde{Z}_t = Z_t + \int_0^t \gamma_s ds$$

is a standard Brownian motion under Q.

Note that the converse of the Cameron-Martin-Girsanov theorem tells us that we can change the drift but not the volatility of the Brownian motion.

Geometric Brownian motion

As we saw in the chapter on stochastic calculus, the process $S_t = S_0 \exp[(\mu - \frac{1}{2}\sigma^2)t + \sigma Z_t]$ where Z_t is a standard Brownian motion under P is sometimes called *geometric Brownian motion*.

Let us consider the discounted asset price $e^{-rt}S_t$. We have:

$$E_P[e^{-rt}S_t] = S_0 e^{(\mu - r)t}$$

So $e^{-rt}S_t$ is not a martingale under P (unless $\mu = r$).

Question

Show that $E_P\left[e^{-rt}S_t\right] = S_0 e^{(\mu-r)t}$

Solution

We know that:

$$\log S_t = \log S_0 + (\mu - ½\sigma^2)t + \sigma Z_t$$

So $$\log S_t \sim N\left[\log S_0 + (\mu - ½\sigma^2)t, \sigma^2 t\right]$$

This means that S_t has a lognormal distribution with the same parameters. Using the formula in the *Tables*, the mean of S_t is:

$$E[S_t] = \exp\left\{\log S_0 + (\mu - ½\sigma^2)t + ½\sigma^2 t\right\} = S_0 e^{\mu t}$$

Since r is a constant, we then have:

$$E[e^{-rt}S_t] = e^{-rt}S_0 e^{\mu t} = S_0 e^{(\mu - r)t}$$

Now take $\gamma_t = \gamma = \dfrac{\mu - r}{\sigma}$, and define:

$$\tilde{Z}_t = Z_t + \int_0^t \gamma_s ds = Z_t + \frac{(\mu - r)}{\sigma} t$$

Then:

$$S_t = S_0 \exp\left[(\mu - ½\sigma^2)t + \sigma Z_t \right]$$

$$= S_0 \exp\left[(r - ½\sigma^2)t + \sigma Z_t + \mu t - rt \right]$$

$$= S_0 \exp\left[(r - ½\sigma^2)t + \sigma\left(Z_t + \frac{\mu - r}{\sigma} t \right) \right]$$

$$= S_0 \exp\left[(r - ½\sigma^2)t + \sigma \tilde{Z}_t \right]$$

Also by the Cameron-Martin-Girsanov theorem there exists a measure Q equivalent to P and where $\tilde{Z}_t$ is a standard Brownian motion under Q.

Furthermore, we find, for $u < t$, that:

$$E_Q[e^{-rt} S_t \mid F_u] = e^{-rt} S_u E_Q[\exp\{(r - ½\sigma^2)(t-u) + \sigma(\tilde{Z}_t - \tilde{Z}_u)\}]$$

$$= e^{-rt} S_u e^{(r - \frac{1}{2}\sigma^2)(t-u) + \frac{1}{2}\sigma^2 (t-u)}$$

$$= e^{-ru} S_u$$

So $e^{-rt} S_t$ is a martingale under Q.

2 The martingale representation theorem

Our ability to replicate the derivative payoff relies on setting up a self-financing, previsible portfolio. In the binomial model this was straightforward, but as mentioned above, in the continuous case it isn't. Here we need to use the *martingale representation theorem.*

2.1 Theorem

Suppose that X_t is a martingale with respect to a measure P. That is:

for any $t < s$, $E_P[X_s \mid F_t] = X_t$

Suppose also that Y_t is another martingale with respect to P. The *martingale representation theorem* states that there exists a unique previsible process ϕ_t such that (in continuous time):

$$Y_t = Y_0 + \int_0^t \phi_s dX_s$$

or $$dY_t = \phi_t dX_t$$

if and only if there is no other measure equivalent to P under which X_t is a martingale.

Recall that a measure P' is equivalent to P (written $P' \sim P$) if $P'(A) = 0 \Leftrightarrow P(A) = 0$.

The theorem also holds when X_t is a vector of martingales.

Note that the integral in the equation for Y_t above is a *stochastic integral*. The subsequent equation is the same relationship expressed as a stochastic differential equation.

It will turn out that in the continuous case, ϕ_t is the amount of stock to be held in the replicating portfolio. As mentioned before, it is therefore essential that ϕ_t be a previsible process. This is what the theorem guarantees.

The proof of the martingale representation theorem is beyond the syllabus, but we will show why such a result might hold by analogy with the discrete case. In the discrete case, we have finite differences, $\Delta X_t, \Delta Y_t$ so that the theorem essentially becomes:

- the process $\phi_t = \dfrac{\Delta Y_t}{\Delta X_t}$ is previsible.

Note that ϕ_t would not be well-defined here if ΔX_t was zero.

2.2 Proof of the martingale representation theorem in discrete time

We can illustrate the proof of the martingale representation theorem in the context of a discrete-time random walk. Note that this is slightly different to the binomial model because the process in the binomial model was assumed to be a geometric random walk, *ie* we multiplied by a random factor each time, rather than *adding* a random term.

If we apply this to the discrete-time binomial model we might have:

$$X_{t+1} = \begin{cases} X_t + u(t, X_t) & \text{with probability } q \\ X_t + d(t, X_t) & \text{with probability } 1-q \end{cases}$$

Note that $X_{t+1} = X_t + \Delta X_{t+1}$ where ΔX_{t+1} is a random variable which takes the value $u(t, X_t)$ with probability q and the value $d(t, X_t)$ with probability $1-q$. Here u is a positive quantity and d is a negative quantity.

If X_t is a martingale with respect to the implied measure Q (*ie* the probabilities q and $1-q$) then:

$$qu(t, X_t) + (1-q)d(t, X_t) = 0$$

$$\Rightarrow \quad q = \frac{-d}{u-d}$$

Notice that this uniquely specifies Q.

Question

Find an expression for $d(t, X_t)$ in terms of q and $u(t, X_t)$.

Solution

We have just seen that:

$$q = \frac{-d(t, X_t)}{u(t, X_t) - d(t, X_t)}$$

Rearranging this gives:

$$d(t, X_t) = \frac{-qu(t, X_t)}{1-q}$$

Now if Y_t is also a martingale with respect to Q (*ie* based on the same probabilities q and $1-q$) then, first, Y_t must also follow a binomial model with:

$$Y_{t+1} = \begin{cases} Y_t + \tilde{u}(t, Y_t) & \text{with probability } q \\ Y_t + \tilde{d}(t, Y_t) & \text{with probability } 1-q \end{cases}$$

since both X_t and Y_t must be measurable with respect to the same sigma-algebra F_t for all t.

It is important to realise that the fact that the processes are measurable with respect to the same filtration, implies that $\Delta X_{t+1} = u(t, X_t)$ if and only if $\Delta Y_{t+1} = \tilde{u}(t, Y_t)$.

Second, since Y_t is a martingale with respect to Q, we must have:

$$E_Q\left[Y_{t+1}|F_t\right] = Y_t$$

ie $$q\tilde{u}(t,Y_t)+(1-q)\tilde{d}(t,Y_t)=0$$

or $$\tilde{d}(t,Y_t)=\frac{-q\tilde{u}(t,Y_t)}{1-q}$$

so that the one-step martingale property is satisfied.

Now consider $\phi_{t+1}=\dfrac{\Delta Y_{t+1}}{\Delta X_{t+1}}$. Since the denominator and numerator are both random white noise terms, it would appear that ϕ_{t+1} will also be random, and not known until time $t+1$. However, consider the outcomes for ϕ_{t+1}.

One possibility is:

$$\phi_{t+1}=\frac{\Delta Y_{t+1}}{\Delta X_{t+1}}=\frac{\tilde{d}(t,Y_t)}{d(t,X_t)}$$

On the other hand, we might have:

$$\phi_{t+1}=\frac{\Delta Y_{t+1}}{\Delta X_{t+1}}=\frac{\tilde{u}(t,Y_t)}{u(t,X_t)}$$

Now we know that:

$$d(t,X_t) = \frac{-qu(t,X_t)}{1-q} \quad \text{and} \quad \tilde{d}(t,Y_t) = \frac{-q\tilde{u}(t,Y_t)}{1-q}$$

So dividing these gives:

$$\frac{\tilde{u}(t,Y_t)}{u(t,X_t)}=\frac{\tilde{d}(t,Y_t)}{d(t,X_t)}$$

So, at time *t* there is actually only *one* possible outcome. Hence ϕ_{t+1} is known at time *t* and therefore previsible.

Then if $\phi_{t+1}=\tilde{u}(t,Y_t)/u(t,X_t)$ (so ϕ_t is previsible: that is F_{t-} or F_{t-1} measurable) we have:

$$Y_t = Y_0+\sum_{s=1}^{t}\phi_s\Delta X_s$$

where $\Delta X_s = X_s - X_{s-1}$

or $\Delta Y_t = \phi_t \Delta X_t$

What we have done here is to calculate the total change in *Y* between time 0 and time *t*, which is found by adding up the steps:

$$Y_t = Y_0 + \Delta Y_1 + \Delta Y_2 + \ldots + \Delta Y_t$$

If *Y* experiences an up-movement at time *s* then $\Delta Y_s = \tilde{u}(s-1, Y_{s-1})$, which from the definition of ϕ_s is equal to $\phi_s u(s-1, X_{s-1})$. Since $u(s-1, X_{s-1})$ is the size of the up-movement in *X* at time *s*, this corresponds to $\phi_s \Delta X_s$ when there is an up-movement. (Down-movements are connected in a similar way.)

So $\Delta Y_t = \phi_t \Delta X_t$, and we can write:

$$Y_t = Y_0 + \phi_1 \Delta X_1 + \phi_2 \Delta X_2 + \ldots + \phi_t \Delta X_t$$

ie

$$Y_t = Y_0 + \sum_{s=1}^{t} \phi_s \Delta X_s$$

So, one *Q*-martingale, Y_t say, can be represented in terms of a different but equivalent *Q*-martingale, X_t.

More specifically, the change in Y_t over any interval [*t*–1, *t*] is equal to the corresponding change in X_t, ΔX_t, scaled up by the ratio of the relative sizes of the up-steps in the respective binomial trees.

So, the actual value of Y_t can be represented in terms of its initial value Y_0 and the ΔX_s's.

Question

Explain how ϕ_t is related to the two martingale processes X_t and Y_t.

Solution

ϕ_t is equal to the ratio of the changes in the two martingale processes, and it reflects the relative *volatilities* of the two processes.

2.3 Diffusion models

Our proof in continuous time will be based on a diffusion model for the share price.

Diffusion models are characterised by their stochastic differential equation:

$$dX_t = \mu_X(t, X_t)dt + \sigma_X(t, X_t)dZ_t$$

where Z_t is a standard Brownian motion with respect to a measure P.

Recall that μ represents the drift and σ the volatility.

If X_t is a martingale with respect to P then we must have:

$$\mu_X(t, X_t) = 0 \text{ for all } t\,,\ X_t$$

So, the process X_t has no drift.

If Y_t is also a martingale then we have:

$$dY_t = \sigma_Y(t, Y_t)dZ_t = \phi_t \sigma_X(t, X_t)dZ_t = \phi_t dX_t$$

where $\phi_t = \dfrac{\sigma_Y(t, Y_t)}{\sigma_X(t, X_t)}$ **provided** $\sigma_X(t, X_t) > 0$ **with probability 1.**

3 Another look at the binomial model

3.1 Introduction

We are now going to use the definitions and theorems we've built up to put together a proof of the formula $V_t = e^{-r(n-t)} E_Q\left[C_n | F_t\right]$ for the fair price for a derivative. First, we should give a worded description of how we are going to use various results. We start by working backwards from what we are trying to show, in order to give an intuitive idea of why the proof involves the steps that it does.

- The aim of the 5-step proof is to show that a certain portfolio replicates the derivative at all times.
- Looking at the definition of a replicating strategy we see that we first need a self-financing portfolio.
- Looking at the definition of a self-financing strategy we see that we need the holdings in the portfolio to be previsible.
- To show a process is previsible, we have the martingale representation theorem to help us.
- However, the martingale representation theorem requires that we have two martingales.
- The discounted share price process is a Q-martingale but we do need one more.
- The other martingale is constructed from the derivative payoff, discounting this all the way back (past the current time t) to time 0.

You should refer back to this section to help you understand the purpose of any of the steps in the 5-step method, whether the method is in the discrete-time case (as here), or for the continuous-time case.

3.2 The 5-step method

Before we look at the continuous-time model, let us apply some of these results to the binomial model.

The 5-step proof in the next chapter has exactly the same structure as the one given in this section.

We work through a series of steps which can be used to solve the problems of pricing and hedging of derivatives.

The idea of this section is to reformulate the risk-neutral pricing method of the binomial model in a way that can be applied to the continuous-time case. In so doing, we show that the fair price for a derivative contract is the discounted value of the expected payoff, where the expectation is taken with respect to the risk-neutral probability measure. This probability measure is defined as the assignment of probabilities that make the discounted share price process a martingale.

The proof relies on constructing a self-financing, replicating portfolio, made out of shares and cash. The martingale representation theorem can be used to guarantee that this portfolio is previsible.

When working through the 5 steps, bear in mind that we are reformulating the replicating proof from scratch. For example, we don't yet know what q is. We will help illustrate the proof by reference to the numerical example of a 2-period binomial tree discussed in the binomial model chapter.

Step 1

Establish the equivalent measure Q under which the discounted asset price process $D_t = e^{-rt}S_t$ is a martingale.

In fact, we found such a measure for the binomial model.

Question

Find the measure Q (according to this definition) when the share price process is a geometric random walk (*ie* it increases by a factor of u or decreases by a factor of d at each step), as in the binomial model.

Solution

Recall that the martingale condition requires that:

$$E_Q[D_{t+1}|F_t] = D_t$$

ie $$E_Q[e^{-r(t+1)}S_{t+1}|F_t] = e^{-rt}S_t$$

$\Leftrightarrow$ $$quS_t + (1-q)dS_t = S_t e^r$$

$\Leftrightarrow$ $$qu + (1-q)d = e^r$$

$\Leftrightarrow$ $$q = \frac{e^r - d}{u - d}$$

Note that this definition is equivalent to finding a measure with respect to which the expected share price evolves at the risk-free rate, *ie* it is a risk-neutral probability measure.

Question

Recall the numerical example of a 2-period binomial tree discussed in the binomial model chapter. In that example we used the tree to find (amongst other things) the value of the call option at time 1 $(V_1(1) = 17.195)$ when the share price was 60. In the example:

- the risk-free rate was 5% per period
- the risk-neutral probability for the up-step was $q = 0.43588$
- the share price either went up to 80 or down to 50 at time 2 (the exercise date)
- the call option had a strike price of 45.

Find the corresponding value of D_1, the discounted asset process, and show that the martingale property is satisfied.

Solution

Using the definition of D_1 we have:

$$D_1 = e^{-r\times 1}S_1 = e^{-0.05\times 1}\times 60 = 57.074$$

We can also find the two possible values of D_t at time 2. If the share price goes up to 80, then:

$$D_2 = e^{-r\times 2}S_2 = e^{-0.05\times 2}80 = 72.387$$

If the share price goes down to 50, then:

$$D_2 = e^{-r\times 2}S_2 = e^{-0.05\times 2}50 = 45.242$$

Thus, using the probability measure with $q = 0.43588$:

$$E\left[D_2 \mid F_1\right] = 72.387\times 0.43588 + 45.242(1-0.43588) = 57.074$$

ie the martingale property is satisfied.

Step 2 (proposition)

Define $V_t = e^{-r(n-t)}E_Q[C_n \mid F_t]$, where the random variable C_n is the derivative payoff at time n. It is proposed that this is the fair price of the derivative and we will prove this over the next few steps.

This is what the 5-step method is trying to show, *ie* that the fair price for the derivative is the discounted value of the expected payoff, where the expectation is taken with respect to Q found in Step 1. This formula is exactly the same as the one that we saw in the chapter on the binomial model.

Question

Recall again the numerical example in the binomial model chapter. Show that the value of the call option at time 1 when the share price is 60 can be calculated using the above formula.

Solution

Applying the formula in Step 2 of the proof gives:

$$V_1 = e^{-r(2-1)} E_Q[C_2|F_1] = e^{-0.05\times 1}[35\times 0.43588 + 5\times(1-0.43588)] = 17.195$$

This agrees with the value that we found in the earlier chapter.

The quantity V_t is the value at time *t* of the replicating portfolio we will be using.

Step 3

Let $E_t = B_n^{-1} E_Q[C_n | F_t] = e^{-rt} V_t$.

This is the definition of a new process E_t. It is the discounted value of the replicating portfolio, which will turn out to be the same as the discounted value of the derivative process.

B_t^{-1} is sometimes referred to as the discount process. It is equal to the inverse of the value of an initial unit of cash at time *t*. Hence:

$$B_t^{-1} = e^{-rt}$$

In particular $B_n^{-1} = e^{-rn}$ and so we have:

$$E_t = B_n^{-1} E_Q[C_n | F_t] = e^{-rn} E_Q[C_n | F_t] = e^{-rt} e^{-r(n-t)} E_Q[C_n | F_t] = e^{-rt} V_t$$

using the definition of V_t from Step 2.

Question

Recall again the numerical example in the binomial model chapter. Calculate the value of E_1 when the share price is 60 at time 1.

Solution

Applying the formula just given, we get:

$$E_1 = B_2^{-1} E_Q[C_2|F_1] = e^{-0.05\times 2}[35\times 0.43588 + 5\times(1-0.43588)] = 16.356$$

Alternatively, we can calculate it by discounting the value of V_1, which we found in the previous question:

$$E_1 = e^{-r} V_1 = e^{-0.05} \times 17.195 = 16.356$$

Under Q, E_t is a martingale. That is, for $s > 0$:

$$E_Q[E_{t+s} \mid F_t] = E_Q[B_n^{-1} E_Q\{C_n \mid F_{t+s}\} \mid F_t] = B_n^{-1} E_Q[C_n \mid F_t] = E_t$$

by the *tower property* of conditional expectation.

Step 4

Steps 1 and 3 have given us two martingales with respect to the equivalent martingale measure Q, namely D_t and E_t.

Since the measure Q is the unique martingale measure, by the martingale representation theorem there exists a previsible process ϕ_t (that is, ϕ_t is F_{t-1}-measurable) such that:

$$\begin{aligned} \Delta E_t &\equiv E_t - E_{t-1} \\ &= \phi_t (D_t - D_{t-1}) \\ &\equiv \phi_t \Delta D_t \end{aligned}$$

The next few equations use the binomial tree notation introduced in the binomial model chapter, where the number in brackets shows the position, counting from the top of the tree. It is assumed that the share price is in state j at time $t-1$.

Let us see if we can establish what ϕ_t is. Now:

$$\Delta E_t = \begin{cases} e^{-rt} V_t(2j-1) - e^{-r(t-1)} V_{t-1}(j) & \text{if up} \\ e^{-rt} V_t(2j) \quad - e^{-r(t-1)} V_{t-1}(j) & \text{if down} \end{cases}$$

and $$\Delta D_t = \begin{cases} e^{-rt} S_{t-1}(j) u_{t-1}(j) - e^{-r(t-1)} S_{t-1}(j) & \text{if up} \\ e^{-rt} S_{t-1}(j) d_{t-1}(j) - e^{-r(t-1)} S_{t-1}(j) & \text{if down} \end{cases}$$

In fact, we found the relationship between the derivative values in the binomial model chapter, which we derived using a replicating portfolio.

Recall that $V_{t-1}(j) = e^{-r}\left(q_{t-1}(j) V_t(2j-1) + (1 - q_{t-1}(j)) V_t(2j)\right)$.

We can substitute this expression for $V_{t-1}(j)$ and simplify.

Then we can see that:

$$\Delta E_t = \begin{cases} e^{-rt}\left(V_t(2j-1) - V_t(2j)\right)\left(1 - q_{t-1}(j)\right) \\ e^{-rt}\left(V_t(2j-1) - V_t(2j)\right)\left(-q_{t-1}(j)\right) \end{cases}$$

We also had a formula for the risk-neutral probability q.

Furthermore, since $q_{t-1}(j) = \dfrac{e^r - d_{t-1}(j)}{u_{t-1}(j) - d_{t-1}(j)}$, we can also see that:

$$\Delta D_t = \begin{cases} e^{-rt} S_{t-1}(j)\left(u_{t-1}(j) - d_{t-1}(j)\right)\left(1 - q_{t-1}(j)\right) \\ e^{-rt} S_{t-1}(j)\left(u_{t-1}(j) - d_{t-1}(j)\right)\left(-q_{t-1}(j)\right) \end{cases}$$

Therefore $\Delta E_t = \phi_t \Delta D_t$ where:

$$\begin{aligned}\phi_t(j) = \frac{\Delta E_t(j)}{\Delta D_t(j)} &= \frac{E_t(2j-1) - E_{t-1}(j)}{D_t(2j-1) - D_{t-1}(j)} \\ &= \frac{E_t(2j-1) - \left(q_{t-1}(j)E_t(2j-1) + \left(1 - q_{t-1}(j)\right)E_t(2j)\right)}{D_t(2j-1) - \left(q_{t-1}(j)D_t(2j-1) + \left(1 - q_{t-1}(j)\right)D_t(2j)\right)} \\ &= \frac{E_t(2j-1) - E_t(2j)}{D_t(2j-1) - D_t(2j)} \\ &= \frac{V_t(2j-1) - V_t(2j)}{S_{t-1}(j)\left[u_{t-1}(j) - d_{t-1}(j)\right]}\end{aligned}$$

ie $$\phi_t = \frac{V_t(2j-1) - V_t(2j)}{S_{t-1}(j)\left(u_{t-1}(j) - d_{t-1}(j)\right)}$$

Note that we've used $\phi_t(j)$ to emphasise the fact that ϕ_t does depend on j in general.

So we've shown that the martingale representation theorem does work in this case.

As we expected, this is what we found in the chapter on the binomial model.

Question

Recall again the numerical example in the binomial model chapter. Calculate the value of ϕ_2 when the share price is 60 at time 1.

Solution

Applying the formula in the Core Reading gives:

$$\phi_2 = \frac{35-5}{80-50} = 1$$

A value of 1 is obtained here because the option is in-the-money whether the share price goes up or down.

So, we define $\phi_t = \dfrac{\Delta E_t}{\Delta D_t} = \dfrac{E_t - E_{t-1}}{D_t - D_{t-1}}$ and the martingale representation theorem guarantees us that this is previsible. ϕ_t will be the number of shares in our replicating portfolio.

Step 5

Let $\psi_t = E_{t-1} - \phi_t D_{t-1}$.

Note that ψ_t, which is previsible, will be the number of units of the cash bond in our replicating portfolio.

In this final step we need to show that we can construct a previsible, self-financing, replicating portfolio.

Between times $t-1$ and t^- suppose that we hold the portfolio consisting of:

- **ϕ_t units of the underlying asset S_t**
- **ψ_t units of the cash account B_t.**

So we hold the portfolio $(\phi_t = \Delta E_t / \Delta D_t, \psi_t = E_{t-1} - \phi_t D_{t-1})$.

The value of this portfolio at time $t-1$ is:

$$\phi_t S_{t-1} + \psi_t B_{t-1} = e^{r(t-1)}(\phi_t D_{t-1} + \psi_t) = e^{r(t-1)} E_{t-1} = V_{t-1}$$

At time t, the value of the portfolio will be:

$$\phi_t S_t + \psi_t B_t = e^{rt}(\phi_t D_t + \psi_t)$$

Writing D_t as $D_{t-1} + \Delta D_t$ gives:

$$= e^{rt}(\phi_t D_{t-1} + \psi_t + \phi_t \Delta D_t)$$

Then using $\psi_t = E_{t-1} - \phi_t D_{t-1}$ and $\Delta E_t = \phi_t \Delta D_t$ gives:

$$= e^{rt}(E_{t-1} + \Delta E_t)$$

$$= e^{rt} E_t$$

$$= V_t$$

In other words, the accumulated value of this portfolio is exactly equal to the value, or cost, of the 'new' portfolio that will be purchased at time *t*.

Therefore the portfolio is self-financing.

Furthermore, $V_n = C_n$, the derivative payoff at time n.

The formula in Step 2 above gives, $V_n = e^{-r(n-n)} E_Q[C_n | F_n] = C_n$.

Therefore the hedging strategy (ϕ_t, ψ_t) is replicating, implying that:

$$V_t = e^{-r(n-t)} E_Q[C_n \mid F_t]$$

is the fair price at time t for this contract.

This completes the proof, *ie* that the formula proposed in Step 2 is correct, which therefore proves the result we derived using the binomial tree approach earlier.

Question

Describe what the above formula says about the fair price for the derivative at time *t*.

Solution

The expression is:

$$V_t = e^{-r(n-t)} E_Q\left[C_n | F_t\right]$$

It says that:

- the value of the derivative at time *t* is equal to the expectation of the derivative payoff at time *n*
- taken with respect to both the risk-neutral probability measure *Q* and the filtration F_t, generated by the history of the stock price movements up to and including time *t*
- discounted at the continuously compounded risk-free rate of return *r*.

Question

Recall again the numerical example in the binomial model chapter. Calculate the value of ψ_2 when the share price is 60 at time 1 and use it to calculate the corresponding option price.

Solution

Here:

$$\psi_2 = E_1 - \phi_2 D_1 = 16.356 - 1 \times 57.074 = -40.718$$

So the value of the call option at time 1 when the share price is 60 is given by:

$$\phi_2 S_1 + \psi_2 B_1 = 1 \times 60 - 40.718 \times e^{0.05} = 17.195$$

This value again agrees with the answer found earlier.

3.3 Summary

Let us summarise here what we have done, because we will repeat the steps when we look at the continuous-time model in the next chapter.

- **We have established the equivalent martingale measure Q.**
- **We have proposed a fair price, V_t, for a derivative and its discounted value $E_t = e^{-rt}V_t$.**
- **We have used the martingale representation theorem to construct a hedging strategy (ϕ_t, ψ_t).**
- **We have shown that this hedging strategy replicates the derivative payoff at time n.**
- **Therefore V_t is the fair value of the derivative at time t.**

Here 'hedging' is used as a synonym for 'replicating'.

Students should understand this proof, but they will not be expected to reproduce the algebra in the examinations.

Question

Have we assumed in this proof that the markets are arbitrage-free?

Solution

Yes. Otherwise, it would not necessarily follow that the replicating portfolio has the same value as the derivative.

Chapter 16 Summary

Preliminary concepts

ϕ_t is *previsible* if it is known based on information up to but not including time *t*.

A portfolio (ϕ_t, ψ_t) is *self-financing* if ϕ_t, ψ_t are previsible and $dV_t = \phi_t dS_t + \psi_t dB_t$,

ie the required change in the value of the portfolio over each instant of time is equal to the pure instantaneous investment gain.

Consider a derivative with random variable payoff X at time T. A self-financing portfolio V_t is a *replicating* portfolio for *X* if $V_T = X$.

So, for an initial investment of V_0 at time 0, if we follow the self-financing portfolio strategy we will be able to reproduce the derivative payment exactly and without risk.

An investment market is *complete* if for every derivative in that market, there exists a replicating strategy for that derivative.

X_t is a Q*-martingale* if $E_Q[X_u | F_t] = X_t$ whenever $t < u$.

The Cameron-Martin-Girsanov theorem

Suppose that Z_t is a standard Brownian motion under P. Furthermore suppose that γ_t is a previsible process. Then there exists a measure Q equivalent to P and where:

$$\tilde{Z}_t = Z_t + \int_0^t \gamma_s ds$$

is a standard Brownian motion under Q.

Conversely, if Z_t is a standard Brownian motion under P and if Q is equivalent to P then there exists a previsible process γ_t such that:

$$\tilde{Z}_t = Z_t + \int_0^t \gamma_s ds$$

is a standard Brownian motion under Q.

Martingale representation theorem (discrete-time version)

Let X_t be a P-martingale and let Y_t be a second P-martingale. Then there exists a unique previsible process ϕ_t such that:

$$Y_t = Y_0 + \sum_{s=1}^{t} \phi_s \Delta X_s \quad \text{where} \quad \Delta X_s = X_s - X_{s-1}$$

or $\quad \Delta Y_t = \phi_t \Delta X_t$

if and only if there is no other measure equivalent to P under which X_t is a martingale.

Martingale representation theorem (continuous-time version)

Let X_t be a P-martingale and let Y_t be a second P-martingale. Then there exists a unique previsible process ϕ_t such that:

$$Y_t = Y_0 + \int_0^t \phi_s dX_s$$

or $\quad dY_t = \phi_t dX_t$

if and only if there is no other measure equivalent to P under which X_t is a martingale.

The 5-step approach

The binomial model result can be proved using the martingale approach, which consists of five steps:

Step 1

Find the equivalent martingale measure Q under which $D_t = e^{-rt} S_t$ is a martingale.

Step 2

Let $V_t = e^{-r(n-t)} E_Q\left[C_n | F_t\right]$ where C_n is the derivative payoff at time n. This is proposed as the fair price of the derivative at time t.

Step 3

Let $E_t = e^{-rn} E_Q\left[C_n | F_t\right] = e^{-rt} V_t$. This is a martingale under Q.

Step 4

By the martingale representation theorem, there exists a previsible process ϕ_t such that $\Delta E_t = \phi_t \Delta D_t$.

Step 5

Let $\psi_t = E_{t-1} - \phi_t D_{t-1}$ and at time t hold the portfolio:

- ϕ_t units of the tradeable asset S_t
- ψ_t units of the cash account.

At time $t-1$ the value of this portfolio is equal to V_{t-1} . At time t the value of this portfolio is equal to V_t . Also $V_n = C_n$. Therefore, the hedging strategy (ϕ_t, ψ_t) is replicating and so, by no arbitrage, V_t is the fair price of the derivative at time t.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 16 Practice Questions

16.1 Exam style

(i) A risky asset has value A_t at time t. A probability measure R is defined so that $A_t v^t$ is an R-martingale, where v is calculated using the risk-free interest rate. Explain why R can be described as a risk-neutral probability measure. [2]

(ii) Let $X_t = v^T E_R\left[C \middle| F_t\right]$, where C is a discrete random variable occurring at time $T > t$. Prove that X_t is an R-martingale. [2]

Hint: if X is a discrete random variable and $\underline{Y}$ is a vector of random variables,

then $\mathrm{E}\left[\mathrm{E}\left[\mathrm{X} \middle| \underline{\mathrm{Y}}\right]\right] = \mathrm{E}\left[\mathrm{X}\right]$.

(iii) Stating any results that you use, deduce that θ_t is previsible, where $dX_t = \theta_t dD_t$ and $D_t = A_t v^t$. [2]

[Total 6]

16.2 (i) Explain what is meant by a 'complete market'.

(ii) Give two reasons why in practice financial markets may not be complete.

(iii) State the relevance of the concept of complete markets in derivative pricing.

16.3 S_t denotes the price of a security at time t. The discounted security process $e^{-rt}S_t$, where r denotes the continuously compounded risk-free interest rate, is a martingale under the risk-neutral measure Q.

(i) Express mathematically the fact that the discounted security process is a Q-martingale.

B_t denotes the accumulated value at time t of an initial investment of 1 unit of cash.

(ii) (a) Write down an expression for B_t.

(b) Show that the discounted cash process is also a Q-martingale.

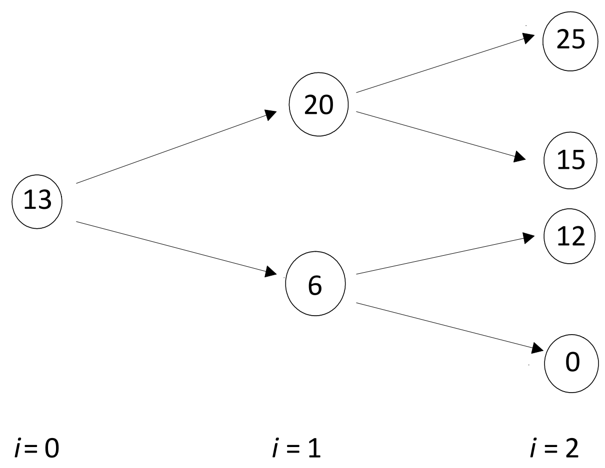
(c) Deduce that the discounted value of any self-financing portfolio (where transactions are made only by switching funds between the security and cash, with no injections or withdrawals of funds from the portfolio) will also be a Q-martingale.

V_t is a process defined by $V_t = e^{-r(T-t)}E_Q[X \mid F_t]$, where X is a function of S_T, T is a fixed time, and F_t denotes the filtration representing the history of the security price up to and including time t.

(iii) Show that the discounted process $e^{-rt}V_t$ is also a Q-martingale.

(iv) Explain the significance of these results in derivative pricing.

16.4 The diagram shows a two-step non-recombining binomial tree. The numerical values shown are X_i, the possible values of a particular derivative at times $i = 0,1,2$, based on a probability measure P that attributes equal probability to the two branches at each step. F_i denotes the filtration of the derivative value process at time i.



(i) (a) If $X_1 = 20$, what are the realised values of F_0 and F_1?

(b) What is the value of $E_P(X_2 \mid F_1)$ in this case?

(c) What is the value of $E_P(X_2 \mid F_1)$ in the case where $X_1 = 6$?

(d) Hence calculate $E_P[E_P(X_2 \mid F_1) \mid F_0]$.

(e) Calculate $E_P(X_2 \mid F_0)$ and comment on your answer.

(ii) The risk-neutral probability measure Q attributes probabilities of 0.4 and 0.6 to the up-paths and down-paths at each branch of this tree.

(a) Does your conclusion in (i)(e) still apply when the probability measure Q is used in place of P?

(b) State briefly why this type of result is useful.

(c) Calculate the value of the derivative at time 0, presenting your calculations in the form of a tree. Ignore interest.

Chapter 16 Solutions

16.1 (i) ***Risk-neutral probability measure***

If $A_t v^t$ is an R-martingale then we have:

$$E_R\left[A_t v^t \middle| F_s\right] = A_s v^s$$

$$\Leftrightarrow \qquad E_R\left[A_t \middle| F_s\right] = A_s v^{s-t} = A_s(1+i)^{t-s} \qquad [1]$$

So, using R as our probability measure we see that, statistically speaking, we expect A_t to 'act like cash', *ie* to increase at the risk-free rate. The riskiness within it has been 'neutralised' and therefore R is the risk-neutral probability measure. [1]

[Total 2]

(ii) ***Prove that X_t is a martingale***

For a martingale we need:

$$E_R\left[X_t \middle| F_s\right] = X_s \qquad [½]$$

The LHS is:

$$E_R\left[X_t \middle| F_s\right] = E_R\left[v^T E_R\left[C \middle| F_t\right] \middle| F_s\right] \qquad [½]$$

By analogy with the hint in the question, we have:

$$E_R\left[E_R\left[C \middle| F_t\right] \middle| F_s\right] = E_R\left[C \middle| F_s\right] \qquad [½]$$

ie the tower law. So, we can simplify the RHS of the previous equation to give:

$$v^T E_R\left[E_R\left[C \middle| F_t\right] \middle| F_s\right] = v^T E_R\left[C \middle| F_s\right] = X_s \qquad [½]$$

[Total 2]

(iii) ***Previsible θ_t***

We make use of the martingale representation theorem in the following form:

Let D_t and X_t be R-martingales. Then there exists a unique process θ_t such that $dX_t = \theta_t dD_t$. Furthermore, θ_t is previsible. [1]

Since θ_t satisfies this equation and the conditions for the theorem to apply, we conclude that θ_t is previsible. [1]

[Total 2]

16.2 (i) ***Explain what is meant by a 'complete market'***

A financial market is said to be complete if any contingent claim (*ie* derivative payoff) can be replicated using the underlying asset and cash.

For example, we can replicate the payoff from a call option on a share using a long holding in the underlying share and a short holding of cash.

(ii) ***Give two reasons why markets might not be complete in practice***

A replicating strategy will typically involve holding fractions of shares and fractions of units of cash. Since these are indivisible it will not be possible to replicate exactly.

A replicating strategy for a put option will require a short holding of the underlying security. If the financial markets do not allow short selling (or if there are restrictions on it), then replication may not be possible.

(iii) ***Relevance for derivative pricing***

Although it is usually just a technicality, the concept of a complete market is important because, without it, we could not be sure that we could replicate the payoff from a derivative. This means that the steps in the derivation of the risk-neutral pricing formula cannot be applied. So we could not use this formula to price the derivative.

16.3 (i) ***Express the martingale property mathematically***

The martingale property tells us that, whenever $t<T$:

$$E_Q\left[e^{-rT}S_T \,\middle|\, F_t\right]=e^{-rt}S_t$$

(ii)(a) ***Expression for B_t***

B_t is the accumulated value at time t of an initial investment of 1 unit of cash.

So: $B_t = e^{rt}$

(ii)(b) ***Show that the discounted cash process is a martingale***

The discounted cash process is therefore:

$$e^{-rt}B_t = e^{-rt}e^{rt} = 1$$

which trivially satisfies the martingale equation $E_Q\left[e^{-rT}B_T \,\middle|\, F_t\right]=e^{-rt}B_t$.

(ii)(c) ***Deduce that any self-financing portfolio is also a martingale***

We have established that the discounted values of both of the components (the shares and the cash) of such a portfolio are martingales. So any multiple of these will also be a martingale.

Also, if we 'rebalance' the portfolio by making switches from cash to shares or vice versa, this will not affect the martingale property.

A good intuitive way to think of martingales here is that, on average, they don't drift up or down. So, if our (discounted) cash and shares are not drifting up or down, neither will any combination of them.

(iii) ***Show that $e^{-rt}V_t$ is also a martingale***

To show that the process $e^{-rt}V_t$ is a martingale, we need to prove that, if $t_1 < t_2$ then:

$$E_Q\left[e^{-rt_2}V_{t_2}\middle|F_{t_1}\right] = e^{-rt_1}V_{t_1} \qquad (1)$$

We've used t_1 and t_2 here, rather than t and T, to avoid confusion, because there's already a T in the definition of V_t.

Substituting the definition of V_{t_2} into the left-hand side (*LHS*) of (1), we have:

$$LHS = E_Q\left[e^{-rt_2}e^{-r(T-t_2)}E_Q(X\,|\,F_{t_2})\middle|F_{t_1}\right]$$

$$= e^{-rT}E_Q\left[E_Q(X\,|\,F_{t_2})\middle|F_{t_1}\right]$$

We can simplify this nested expectation using the tower law to get:

$$LHS = e^{-rT}E_Q(X\,|\,F_{t_1})$$

From the definition of V_{t_1}, the right-hand side (*RHS*) of (1) equals:

$$RHS = e^{-rt_1}e^{-r(T-t_1)}E_Q(X\,|\,F_{t_1}) = e^{-rT}E_Q(X\,|\,F_{t_1})$$

So: $\quad LHS = RHS$

This shows that $e^{-rt}V_t$ is also a Q-martingale.

(iv) ***Explain the significance***

We have shown that, if we have a self-financing portfolio consisting of shares and cash, its discounted value will be a Q-martingale, *ie* it will have no drift.

If it is possible to rebalance such a portfolio so that it will always replicate over the next instant the value of a derivative based on the share, then the discounted value of the derivative must equal the discounted value of this portfolio.

It turns out that such a replicating strategy *is* possible. The Martingale Representation Theorem guarantees this.

But the discounted process $e^{-rt}V_t$ behaves in precisely this way, and gives the correct payoff X when $t=T$.

So V_t must equal the value of the derivative.

This gives us the derivative pricing formula $V_t = e^{-r(T-t)}E_Q[X \mid F_t]$.

16.4 (i)(a) The filtration F_i is just the set of values representing the path followed up to time i. So, in this case (where we are told what has happened up to time 1), we have:

$$F_0 = \{13\} \quad \text{and} \quad F_1 = \{13,20\}$$

(i)(b) If $F_1 = \{13,20\}$, there are two possible values for X_2, namely 25 and 15, which we are assuming are equally likely.

So: $E_P(X_2 \mid F_1) = \frac{1}{2}\times 25 + \frac{1}{2}\times 15 = 20$

This figure has already been written in on the tree in the question at time 1.

(i)(c) If $X_1 = 6$, then $F_1 = \{13,6\}$.

Here: $E_P(X_2 \mid F_1) = \frac{1}{2}\times 12 + \frac{1}{2}\times 0 = 6$

Again, this matches the figure shown in the tree at time 1.

(i)(d) We have established that, if we start at the '13' node (corresponding to the only possible value of F_0), the conditional expectation $E_P(X_2 \mid F_1)$ can take two possible values, and these are equally likely under the probability measure P.

So: $E_P[E_P(X_2 \mid F_1) \mid F_0] = \frac{1}{2}\times 20 + \frac{1}{2}\times 6 = 13$

This matches the figure shown in the tree at time 0.

(i)(e) If we start at the '13' node, X_2 can take four possible values (25, 15, 12 or 0), each with probability $\frac{1}{2}\times\frac{1}{2} = \frac{1}{4}$.

So: $E_P(X_2 \mid F_0) = \frac{1}{4}\times 25 + \frac{1}{4}\times 15 + \frac{1}{4}\times 12 + \frac{1}{4}\times 0 = 13$

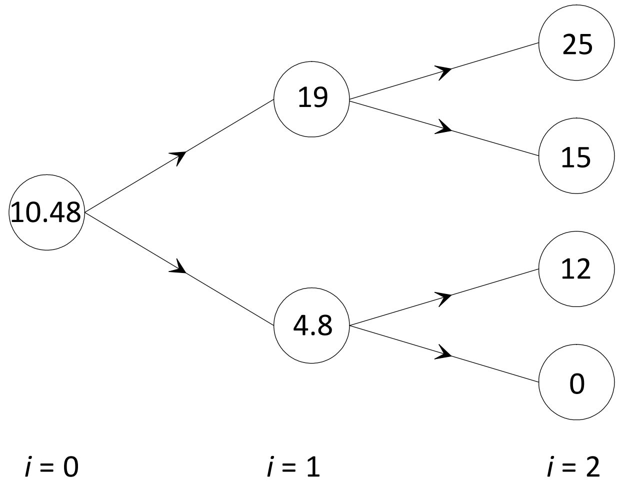
This expectation is the same as the expectation in (i)(d), *ie* $E_P[E_P(X_2 \mid F_1) \mid F_0] = E_P(X_2 \mid F_0)$. This is an example of the 'tower property' of conditional expectations.

(ii)(a) Yes. The tower property applies equally well in this case, *ie* it is also true that:

$$E_Q[E_Q(X_2 \mid F_1) \mid F_0] = E_Q(X_2 \mid F_0)$$

(ii)(b) Equalities of the form $E_Q(X_n \mid F_{i-1}) = E_Q[E_Q(X_n \mid F_i) \mid F_{i-1}]$ are useful because they enable us to work backwards through a binomial tree, calculating the value of the derivative at time $i-1$ from the values at time i.

(ii)(c) We need to calculate $E_Q(X_2 \mid F_0)$. Calculating the values of the conditional expectations by working backwards using the risk-neutral probability measure Q, leads to the following tree:



So the value of the derivative at time 0 is 10.48.

17

The 5-step method in continuous time

Syllabus objectives

6.1 Option pricing and valuations

1. Demonstrate an understanding of the Black-Scholes derivative-pricing model:

   - Derive the Black-Scholes partial differential equation both in its basic and Garman-Kohlhagen forms. (part)

   - Demonstrate how to price and hedge a simple derivative contract using the martingale approach (part).

11. Describe and apply in simple models, including the binomial model and the Black-Scholes model, the approach to pricing using deflators and demonstrate its equivalence to the risk-neutral pricing approach.

0 Introduction

In the last chapter we built up the theory required for an alternative proof, known as the 5-step method, of the derivative pricing formula. We also showed how the structure of this proof works by proving the formula for the binomial model once more. You should remember the main steps we performed.

Question

What were the five main steps?

Solution

- We established the equivalent martingale measure Q.
- We proposed a fair price, V_t, for a derivative and its discounted value $E_t = e^{-rt}V_t$.
- We used the martingale representation theorem to construct a hedging strategy (ϕ_t, ψ_t).
- We then showed that this hedging strategy replicates the derivative payoff at time n.
- So V_t was the fair value of the derivative at time t.

In this chapter we will use the 5-step method again to prove the formula $V_t = e^{-r(T-t)}E_Q[X \mid F_t]$. We will then use this to derive the Black-Scholes formula, which we met in an earlier chapter. Later in this chapter we will extend the theory to incorporate dividends and hence prove the Garman-Kohlhagen formula, which appears on page 47 of the *Tables*.

Although some of the theory may appear abstract and purely mathematical, in this chapter we also see how delta-hedging can give the martingale approach a more intuitive appeal. Specifically, we will see that the ϕ component of the replicating portfolio turns out to equal Δ. As a by-product we will deduce the Black-Scholes PDE by an alternative method.

We will also discuss the advantages and disadvantages of the 5-step method as compared to the PDE method from the Black-Scholes model chapter, and look briefly at the state price deflator approach.

1 The 5-step method in continuous time

1.1 Introduction

Having seen the structure of the 5-step method in the previous chapter, we now repeat the steps of the proof in continuous time. To make the theory easier to follow, you should recall, also from the previous chapter, the purpose of the steps in the proof, *ie* recall the worded description of what we are trying to achieve:

- The aim of the 5-step proof is to show that a certain portfolio replicates the derivative at all times.
- Looking at the definition of a replicating strategy we see that we first need a self-financing portfolio.
- Looking at the definition of a self-financing strategy we see that we need the holdings in the portfolio to be previsible.
- To show a process is previsible, we have the martingale representation theorem to help us.
- However, the martingale representation theorem requires that we have two martingales.
- The discounted share price process is a Q-martingale but we do need one more.
- The other martingale is constructed from the derivative payoff, discounting this all the way back (past the current time t) to time 0.

Remember that, in continuous time, the share price process is being modelled as a geometric Brownian motion or lognormal model, which we discussed previously.

Question

Fully describe what it means for a share price to follow geometric Brownian motion.

Solution

Geometric Brownian motion means that we are modelling the share price using the continuous-time lognormal model. Alternatively, we can express this in terms of the stochastic differential equation:

$$dS_t = \left[\mu\, dt + \sigma\, dZ_t\right] S_t$$

This can be seen on page 46 of the *Tables*.

1.2 The martingale approach (the 5-step method)

When we looked at the binomial model in the previous chapter we demonstrated how the value of a derivative could be expressed as:

$$V_t = e^{-r(T-t)} E_Q[X \mid F_t]$$

where X is the value of the derivative at maturity and Q is the equivalent martingale measure. This was shown for a discrete-time process with a finite state space. Here we are working in continuous time and with continuous state spaces (that is, S_t can take any value greater than zero).

At this stage we do not introduce dividends. So, we are looking at a non-dividend-paying share in continuous time.

However, the binomial result can be extended in the obvious way to give the following result:

Proposition

Let X be any derivative payment contingent on F_T, payable at some fixed future time T, where F_T is the sigma-algebra generated by S_u for $0 \le u \le T$.

So F_T is the history of S_t up to and including time *T*.

Then the value of this derivative payment at time $t < T$ is:

$$V_t = e^{-r(T-t)} E_Q[X \mid F_t].$$

Proof

We follow the same sequence of steps described in the previous chapter.

Step 1

Establish the unique equivalent measure Q under which the discounted asset price process $D_t = e^{-rt} S_t$ is a martingale.

It can be shown that this measure exists, is unique and that under Q:

$$D_t = D_0 \exp\left(\sigma \tilde{Z}_t - \frac{\sigma^2 t}{2}\right)$$

where $\tilde{Z}_t$ is a Brownian motion under Q.

This is the same idea as in the discrete case. We want to construct the measure that assigns probabilities to the possible asset price paths in such a way that the discounted asset price is a martingale. In discrete time this was easy because we only had to find a single probability for each branch of the tree. In the continuous case we have a whole continuum of possible paths and the problem is not so straightforward.

It is beyond the syllabus to actually construct the probability measure. We can say something about it, however, based on the result of the following question.

Question

Let Z_t be a standard Brownian motion. By considering the stochastic differential equation or otherwise, prove that $e^{\sigma Z_t - ½\sigma^2 t}$ is a martingale.

Solution

We've already seen that the process $X_t = e^{\sigma Z_t - ½\sigma^2 t}$ satisfies the stochastic differential equation:

$$dX_t = X_t\left[\left(-½\sigma^2 + ½\sigma^2\right)dt + \sigma dZ_t\right] = \sigma X_t dZ_t$$

It follows that the process X_t has no drift and hence must be a martingale.

Alternatively:

$$E\left[\exp\left(-½\sigma^2 t + \sigma Z_t\right) | F_s\right]$$

$$= E\left[\exp\left(-½\sigma^2 t + \sigma Z_s + \sigma(Z_t - Z_s)\right) | F_s\right]$$

$$= \exp\left(-½\sigma^2 t + \sigma Z_s\right) E\left[\exp\left(\sigma(Z_t - Z_s)\right) | F_s\right]$$

since $\exp\left(-½\sigma^2 t + \sigma Z_s\right)$ is a constant given F_s.

Now, $Z_t - Z_s \sim N(0, t-s)$, so we can use the moment generating function of a normal distribution to note that:

$$E\left[\exp\left(\sigma(Z_t - Z_s)\right) | F_s\right] = M_{N(0,t-s)}(\sigma) = \exp\left(½\sigma^2 (t-s)\right)$$

Together with the above, this gives:

$$E\left[\exp\left(-½\sigma^2 t + \sigma Z_t | F_s\right)\right] = \exp\left(-½\sigma^2 s + \sigma Z_s\right)$$

ie $E[X_t | F_s] = X_s$ and the process X_t satisfies the definition of a martingale.

Since:

$$S_t = S_0 \exp\left((\mu - ½\sigma^2)t + \sigma Z_t\right)$$

it follows that:

$$D_t = e^{-rt} S_t = S_0 \exp\left((\mu - r - ½\sigma^2)t + \sigma Z_t\right)$$

Now in order to be a martingale with respect to *Q*, we could assign probabilities to $\sigma\tilde{Z}_t = (\mu - r)t + \sigma Z_t$ that make $\tilde{Z}_t$ a standard Brownian motion. We saw in the previous chapter that the Cameron-Martin-Girsanov theorem can be used to achieve this. So, with respect to the real-world probabilities *P*, the random process $\tilde{Z}_t$ is a Brownian motion with drift. But, when we assign new probabilities, using the measure *Q,* we remove this drift.

We then have:

$$D_t = S_0 \exp\left[-½\sigma^2 t + \sigma\tilde{Z}_t\right]$$

where $\tilde{Z}_t$ is a standard Brownian motion under *Q*, so that D_t is a martingale with respect to *Q* by using the previous question.

Hence we can write:

$$S_t = S_0 \exp\left[\left(r - ½\sigma^2\right)t + \sigma\tilde{Z}_t\right]$$

where $\tilde{Z}_t = \left(\dfrac{\mu - r}{\sigma}\right)t + Z_t$

Step 2 (proposition)

Define:

$$\boldsymbol{V_t = e^{-r(T-t)} E_Q[X \mid F_t]}$$

We propose that this is the fair price of the derivative.

Step 3

Let:

$$\boldsymbol{E_t = e^{-rT} E_Q[X \mid F_t] = e^{-rt} V_t}$$

Under *Q*, E_t is a martingale.

Recall that, as in the previous chapter:

$$\begin{aligned} E_t &= B_n^{-1}E_Q\left[C_n \mid F_t\right] \\ &= e^{-rn}E_Q\left[C_n \mid F_t\right] \\ &= e^{-rt}e^{-r(n-t)}E_Q\left[C_n \mid F_t\right] \\ &= e^{-rt}V_t \end{aligned}$$

where C_n is the claim amount at time n, here denoted by X.

Question

Why is E_t a Q-martingale?

Solution

E_t is a martingale with respect to Q, since for $s > 0$:

$$E_Q\left[E_{t+s} \mid F_t\right] = E_Q\left[B_T^{-1}E_Q\left\{C_T \mid F_{t+s}\right\} \mid F_t\right] = B_T^{-1}E_Q\left[C_T \mid F_t\right] = E_t$$

Step 4

By the martingale representation theorem there exists a previsible process ϕ_t (that is ϕ_t is F_{t^-} -measurable) such that:

$$dE_t = \phi_t dD_t$$

As in the discrete case, this application of the martingale representation theorem guarantees that the stock process ϕ_t is previsible.

Step 5

Let:

$$\psi_t = E_t - \phi_t D_t$$

We will again see that this is just the right holding of the cash bond that makes the value of the portfolio held equal to the value of the derivative at that time.

Suppose that at time t we hold the portfolio:

- **ϕ_t units of the underlying asset S_t**
- **ψ_t units of the cash account B_t**

where $\phi_t dD_t = dE_t$ and $\psi_t = E_t - \phi_t D_t$.

Remembering that $e^{-rt}S_t = D_t$ and $E_t = e^{-rt}V_t$ we can examine the change in the value of this portfolio over the very short time period $[t, t+dt]$.

At time t, the portfolio has value:

$$\phi_t S_t + \psi_t B_t = e^{rt}(\phi_t D_t + \psi_t) = e^{rt}E_t = V_t$$

At time $t+dt$, the portfolio has value:

$$\begin{aligned}\phi_t S_{t+dt} + \psi_t B_{t+dt} &= e^{r(t+dt)}(\phi_t D_{t+dt} + \psi_t)\\ &= e^{r(t+dt)}(\phi_t D_t + \phi_t dD_t + \psi_t)\\ &= e^{r(t+dt)}(E_t + dE_t)\\ &= e^{r(t+dt)}E_{t+dt} = V_{t+dt}\end{aligned}$$

Therefore, the change in the value of the portfolio over t up to $t+dt$ is:

$$\begin{aligned}dV_t &= V_{t+dt} - V_t\\ &= \left(\phi_t S_{t+dt} + \psi_t B_{t+dt}\right) - \left(\phi_t S_t + \psi_t B_t\right)\\ &= \phi_t\left(S_{t+dt} - S_t\right) + \psi_t\left(B_{t+dt} - B_t\right)\\ &= \phi_t dS_t + \psi_t dB_t\end{aligned}$$

So the change in the value of the portfolio over the period t up to $t+dt$ is the pure investment gain:

$$\phi_t dS_t + \psi_t dB_t$$

Hence, the hedging strategy (ϕ_t, ψ_t) is self-financing.

We need to check that the portfolio has the correct value at the expiry date.

Furthermore:

$$V_T = E_Q[X \mid F_T] = X\,.$$

Therefore the hedging strategy is replicating, so that $V_t = e^{-r(T-t)}E_Q[X \mid F_t]$ is the fair price at time t for this derivative contract.

As before, V_t, the proposed no-arbitrage value of the derivative at time $t < T$, is equal to:

- the time-t *expectation* of the claim amount paid at time T
- calculated with respect to the *probability measure* Q and
- the *filtration* F_t generated by the history of the stock price up to and including time t and
- discounted at the continuously compounded risk-free rate of return, r.

The formula:

$$V_t = e^{-r(T-t)} E_Q\left[X|F_t\right]$$

is a general formula that applies to any derivative on a dividend-paying share. In order to find an expression for any specific derivative we would need to specify the derivative payoff X and then calculate the expectation in the above formula.

1.3 Delta hedging and the martingale approach

Recall an earlier chapter where we defined the delta of a derivative as one of the *Greeks*.

Question

(i) What is the definition of delta?

(ii) In what numerical range would you expect delta to be for:

(a) a call option

(b) a put option?

Solution

(i) ***Definition of delta***

Delta is the rate of change of the value of the derivative with respect to the share price:

$$\Delta = \frac{\partial f}{\partial S_t}$$

(ii)(a) ***Call option***

The delta of a call option can be written as $\Delta = \Phi(d_1)$, therefore $0 \le \Delta \le 1$.

(ii)(b) ***Put option***

The delta of a put option can be written as $\Delta = -\Phi(-d_1)$, therefore $-1 \le \Delta \le 0$.

It is important to mention delta hedging at this stage. In the martingale approach we showed that *there exists* a portfolio strategy (ϕ_t, ψ_t) which would replicate the derivative payoff.

We did not say what ϕ_t actually is or how we work it out. This is quite straightforward.

In fact it turns out to be delta.

First, we can evaluate directly the price of the derivative $V_t = e^{-r(T-t)} E_Q[X \mid F_t]$ either analytically (as in the Black-Scholes formula) or using numerical techniques.

In general, if S_t represents the price of a *tradeable* asset:

$$\phi_t = \frac{\partial V}{\partial s}(t, S_t) = \Delta$$

ϕ_t is usually called the *delta* Δ of the derivative.

Recall that ϕ_t is the change in the discounted derivative process relative to the change in the discounted share process. If we ignored the discount factors up to the current time then this would be the change in the derivative process relative to the change in the share process, *ie* delta.

The martingale approach tells us that provided we:

- **start at time 0 with V_0 invested in cash and shares**
- **follow a self-financing portfolio strategy**
- **continually rebalance the portfolio to hold exactly ϕ_t** (delta) **units of S_t with the rest in cash**

then we will precisely replicate the derivative payoff.

This replication is achieved without any risk, and is a form of delta-hedging.

1.4 Example: the Black-Scholes formula for a call option

The 5-step method has shown us that the fair price at time *t* for a derivative contract that pays a (random) amount X at time *T* is $V_t = e^{-r(T-t)} E_Q\left[X|F_t\right]$. We now want to evaluate this expression in the case where the derivative is a European call option on a non-dividend-paying share.

Question

What is the payoff function for a European call option?

Solution

$$f(S_T, T) = \max\{S_T - K, 0\}$$

Proposition

The Black-Scholes formula for a call option on a share with no dividends is:

$$V_t = S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

where $d_1 = \dfrac{\log\frac{S_t}{K} + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$ **and** $d_2 = d_1 - \sigma\sqrt{T-t}$

Proof

Using risk-neutral valuation, the fair price of the derivative with payoff function X_T is:

$$V_t = e^{-r(T-t)} E_Q\left[X_T \mid F_t\right]$$

This is the general risk-neutral pricing formula, which we derived using the 5-step method.

The payoff function for a European call option is $\max\left[S_T - K, 0\right]$.

We therefore need to substitute this payoff function into the general risk-neutral pricing formula and work out the resulting expression. To do this, we also need to know the risk-neutral probability of each possible value that this payoff function might take at the expiry date T, conditioned on the current share price S_t.

Question

Why do we not need to condition on the full past history of the share price process F_t ?

Solution

The assumption of *independent increments* for Brownian motion means that future values of the share price depend only on the current share price S_t and not the past history of how we arrived at it.

Hence, the share price at the maturity date T, S_T, and likewise the derivative payoff at time T, depends only on S_t.

So:

$$V_t = e^{-r(T-t)} E_Q\left[\max\left[S_T - K, 0\right] \mid S_t\right]$$

The process S_t is assumed to be a geometric Brownian motion and so has a continuous state space. This means we need to use integration to work out the expected value of the payoff function. Hence:

$$V_t = e^{-r(T-t)} \int_0^\infty \max\left[S_T - K, 0\right] f(S_T \mid S_t)\, dS_T$$

where $f(S_T \mid S_t)$ is the (conditional) probability density function for S_T, given S_t, and we are summing over all the possible values of S_T from zero to infinity.

This expression can be simplified to give:

$$V_t = e^{-r(T-t)} \int_K^{\infty} (S_T - K)\, f(S_T \mid S_t)\, dS_T$$

which can be written in terms of two integrals as follows:

$$V_t = e^{-r(T-t)} \int_K^{\infty} S_T f(S_T \mid S_t)\, dS_T - Ke^{-r(T-t)} \int_K^{\infty} 1 \times f(S_T \mid S_t)\, dS_T$$

Under the risk-neutral measure *Q*:

$$S_T = S_t \exp\left[(r - \tfrac{1}{2}\sigma^2)(T-t) + \sigma(\tilde{Z}_T - \tilde{Z}_t)\right]$$

The above expression for S_T in terms of S_t can be derived from the stochastic differential equation for the share price under *Q*, which is:

$$dS_t = rS_t dt + \sigma S_t d\tilde{Z}_t$$

where r is the risk-free rate and $\tilde{Z}_t$ is a standard Brownian motion process under *Q*.

Thus:

$$\log S_T \mid S_t \sim N\left[\log S_t + (r - \tfrac{1}{2}\sigma^2)(T-t), \sigma^2(T-t)\right]$$

which tells us the distribution of S_T given the current share price S_t, *ie* it has a lognormal distribution.

So, we can use the formula for the truncated moments of a lognormal distribution on page 18 in the *Tables*.

Recall that in general the share price can take any value from zero to infinity. However, in the above integrals, we are only summing from *K* to infinity, *ie* we need to evaluate truncated moments of the share price.

When applying the formula from the *Tables*, note that the power of S_T in the first integral is one, whereas that in the second integral is zero – as S_T to the power of zero is equal to one.

Thus:

$$V_t = e^{-r(T-t)}\left[e^{\log S_t + (r - \frac{1}{2}\sigma^2)(T-t) + \frac{1}{2}\sigma^2(T-t)}\right]\{\Phi(U_1) - \Phi(L_1)\}$$
$$-Ke^{-r(T-t)}\left[e^0\right]\{\Phi(U_0) - \Phi(L_0)\}$$

which after some cancelling becomes:

$$V_t = S_t\{\Phi(U_1) - \Phi(L_1)\} - Ke^{-r(T-t)}\{\Phi(U_0) - \Phi(L_0)\}$$

Finally, we just need to evaluate the $\Phi(U_1), \Phi(L_1), \Phi(U_0)$ and $\Phi(L_0)$ terms.

Now:

- $U_1 \approx \infty$
- $U_0 \approx \infty$
- $L_1 = \dfrac{\log K - \log S_t - (r - \frac{1}{2}\sigma^2)(T-t) - \sigma^2(T-t)}{\sigma\sqrt{T-t}} = -d_1$
- $L_0 = \dfrac{\log K - \log S_t - (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} = -d_2$

These are obtained using the formulae on page 18 in the *Tables*.

Question

Explain why $U_0 \approx \infty$.

Solution

Using the formula for U_k with $k=0$ gives:

$$U_0 = \frac{\log(\infty) - \log S_t - (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

As $x \to \infty$, $\log(x) \to \infty$. So, the log term is *much* bigger than the other terms in the numerator, which can therefore be ignored. In addition, we can argue that dividing infinity by a finite term, such as $\sigma\sqrt{T-t}$, still gives us infinity.

Hence:

$$U_0 \approx \infty$$

Note that a similar argument can be used for U_1, where the subtraction of $\sigma\sqrt{T-t}$ does not affect the end result.

Thus:

- $\Phi(U_1) = \Phi(U_0) \approx \Phi(\infty) = 1$

Hence:

$$\begin{aligned} V_t &= S_t\{1 - \Phi(-d_1)\} - Ke^{-r(T-t)}\{1 - \Phi(-d_2)\} \\ &= S_t\,\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \end{aligned}$$

This is the Black-Scholes formula for the value of a call option on a non-dividend-paying share.

1.5 Example: deriving the Black-Scholes PDE using the martingale approach

Let's now find an actual formula for ϕ_t and look at an alternative method of deriving the Black-Scholes PDE using martingales.

We define:

$$E_t = e^{-rt}V_t \quad \text{and} \quad D_t = e^{-rt}S_t$$

which are both martingales under Q. Then:

- $dD_t = \sigma D_t d\tilde{Z}_t$
- $dS_t = B_t(rD_t dt + dD_t)$ **and**
- $dE_t = -re^{-rt}V_t dt + e^{-rt}dV_t = e^{-rt}(-rV_t dt + dV_t)$.

Question

Derive these three relationships.

Solution

First equation

Since $S_t = S_0 e^{(r-½\sigma^2)t+\sigma\tilde{Z}_t}$, we have:

$$D_t = e^{-rt}S_t = e^{-rt}S_0 e^{(r-½\sigma^2)t+\sigma\tilde{Z}_t} = S_0 e^{-½\sigma^2 t+\sigma\tilde{Z}_t}$$

Let $X_t = -½\sigma^2 t + \sigma\tilde{Z}_t$, so that $dX_t = -½\sigma^2 dt + \sigma d\tilde{Z}_t$ and $D_t = S_0 e^{X_t}$.

Now apply Ito's Lemma:

$$dD_t = S_0\left\{-½\sigma^2 e^{X_t} + ½\sigma^2 e^{X_t}\right\}dt + \sigma S_0 e^{X_t} d\tilde{Z}_t = \sigma D_t d\tilde{Z}_t$$

Second equation

$$D_t = e^{-rt}S_t$$

So: $\quad S_t = D_t e^{rt}$

Provided that the processes X_t and Y_t are not both stochastic, the product rule:

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t$$

applies.

We can use this in conjunction with the result from ordinary calculus that $\frac{d}{dt}e^{rt} = re^{rt}$ (or $de^{rt} = re^{rt}dt$) to get:

$$\begin{aligned} dS_t = d(e^{rt}D_t) &= e^{rt}dD_t + D_t de^{rt} \\ &= e^{rt}dD_t + D_t re^{rt}dt \\ &= e^{rt}(dD_t + rD_t dt) \\ &= B_t(rD_t dt + dD_t) \end{aligned}$$

Third equation

$$dE_t = d(e^{-rt}V_t) = e^{-rt}dV_t - V_t re^{-rt}dt = e^{-rt}(-rV_t dt + dV_t)$$

Applying Ito's Lemma to the function $V(t, S_t)$:

$$\begin{aligned} dV_t &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial s}dS_t + \frac{1}{2}\frac{\partial^2 V}{\partial s^2}(dS_t)^2 \\ &= \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2}\right)dt + \frac{\partial V}{\partial s}B_t(rD_t dt + dD_t) \\ &= \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} + rS_t\frac{\partial V}{\partial s}\right)dt + \frac{\partial V}{\partial s}B_t dD_t \end{aligned}$$

Hence:

$$\begin{aligned} dE_t &= e^{-rt}(-rV_t dt + dV_t) \\ &= e^{-rt}\left[-rV_t dt + \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} + rS_t\frac{\partial V}{\partial s}\right)dt + \frac{\partial V}{\partial s}B_t dD_t\right] \\ &= e^{-rt}\left[-rV_t dt + \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} + rS_t\frac{\partial V}{\partial s}\right)dt\right] + \frac{\partial V}{\partial s}dD_t \end{aligned}$$

Now we know that E_t and D_t are both martingales under Q. Therefore by the martingale representation theorem there exists some previsible process ϕ_t such that:

$$dE_t = \phi_t\, dD_t = \sigma\, \phi_t\, D_t d\tilde{Z}_t$$

This can be written as:

$$dE_t = 0dt + \phi_t dD_t$$

We can compare this with the SDE derived above:

$$dE_t = e^{-rt}\left(-rV_t + \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} + rS_t \frac{\partial V}{\partial s}\right)dt + \frac{\partial V}{\partial s}dD_t$$

This means that:

$$\phi_t = \frac{\partial V}{\partial s}$$

and $$-rV_t + \frac{\partial V}{\partial t} + rS_t \frac{\partial V}{\partial s} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} = 0$$

$$\Leftrightarrow \quad rV_t = \frac{\partial V}{\partial t} + rS_t \frac{\partial V}{\partial s} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2}$$

(otherwise E_t would not be a martingale under Q).

We recognise the last equation as the Black-Scholes PDE.

This can be found on page 46 in the *Tables*.

It can be shown that:

$$\frac{\partial V}{\partial s} = \Phi(d_1)$$

So this martingale approach has provided an alternative derivation of the Black-Scholes PDE and it has also given us an explicit formula for ϕ_t, in case we wanted to set up a replicating portfolio in real life.

1.6 Advantages of the martingale approach

We have now seen two ways of deriving the Black-Scholes formula for a call option $c_t = S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$. One method involved evaluating the martingale formula $e^{-r(T-t)}E_Q[X \mid F_t]$. The other method involved 'solving' the PDE with the correct boundary condition.

The main advantage of the martingale approach is that it gives us much more clarity in the process of pricing derivatives. Under the PDE approach we derived a PDE and had to 'guess' the solution for a given set of boundary conditions.

Of course, we ourselves did not have to literally guess the solution, we just had to look it up on page 47 of the *Tables*!

Under the martingale approach we have an expectation which can be evaluated explicitly in some cases and in a straightforward numerical way in other cases.

So, the point is that, without knowing the formula on page 47 of the *Tables* there is no easy way we could work it out using the PDE approach, whereas it can be worked out using the martingale approach without knowing it beforehand.

Furthermore the martingale approach also gives us the replicating strategy for the derivative.

Question

What is the replicating strategy for a European call option?

Solution

Hold:

- $\phi_t = \dfrac{\partial V}{\partial S} = \Phi(d_1)$ shares and
- $\psi_t = E_t - \phi_t D_t = e^{-rt}(V_t - \phi_t S_t)$ units of the cash bond (*ie* an actual cash amount of $V_t - \phi_t S_t$).

Question

You are trying to replicate a 6-month European call option with strike price 500, which you purchased 4 months ago. If $r = 0.05$, $\sigma = 0.2$, and the current share price is 475, what portfolio should you be holding, assuming that no dividends are expected before the expiry date?

Solution

Here $T = \dfrac{6}{12}$, $t = \dfrac{4}{12}$, $r = 0.05$, $\sigma = 0.2$, $K = 500$ and $S_t = 475$. So you need:

$$\phi_t = \Phi(d_1) = \Phi\left(\frac{\ln(475/500) + (0.05 + 0.2^2/2) \times (6-4)/12}{0.2\sqrt{(6-4)/12}}\right) = 0.314 \text{ shares}$$

and:
$$V_t - \phi_t S_t = 475\Phi(d_1) - 500e^{-0.05\times 2/12}\Phi(d_1 - 0.2\sqrt{2/12}) - 0.314 \times 475$$
$$= -142 \text{ cash}$$

Finally, the martingale approach can be applied to any F_T-measurable derivative payment, including path-dependent options (for example, Asian options), whereas the PDE approach, in general, cannot.

An Asian option is one where the payoff depends on the average share price up to expiry.

Question

What would you say is the main disadvantage of the martingale approach as compared to the PDE approach?

Solution

The PDE approach is quicker and easier to describe, and more easily understood.

1.7 Risk-neutral pricing

This martingale approach is often referred to as *risk-neutral pricing*.

In this approach, the measure Q is commonly called the *risk-neutral* measure. However, Q is also referred to as the *equivalent martingale* measure because the discounted price processes $S_t e^{-rt}$ and $V_t e^{-rt}$ are both martingales under Q.

2 The state price deflator approach

Recall that we have:

$$dS_t = S_t[\mu dt + \sigma dZ_t] \text{ under } P$$

and:

$$dS_t = S_t[rdt + \sigma d\tilde{Z}_t] \text{ under } Q$$

where:

$$d\tilde{Z}_t = dZ_t + \gamma dt \text{ and } \gamma = \frac{\mu - r}{\sigma}$$

Corollary to the Cameron-Martin-Girsanov Theorem

There exists a process η_t such that, for any F_T-measurable derivative payoff X at time T:

$$E_Q[X \mid F_t] = E_P\left[\frac{\eta_T}{\eta_t} X \middle| F_t\right]$$

We do not prove this corollary in this subject.

In the present case, where $\tilde{Z}_t$ is a Q-Brownian motion, Z_t is a P-Brownian motion and $d\tilde{Z}_t = dZ_t + \gamma dt$, we have:

$$\eta_t = e^{-\gamma Z_t - \frac{1}{2}\gamma^2 t}$$

Question

What interesting property does η_t have?

Solution

It's a martingale.

Now if we further define:

$$A_t = e^{-rt}\eta_t$$

the price at time t for the derivative X payable at time T is then:

$$V_t = e^{-r(T-t)}E_Q[X \mid F_t]$$

under the martingale approach and:

$$V_t = e^{-r(T-t)}E_P\left[\frac{\eta_T}{\eta_t}X\middle|F_t\right]$$

$$= \frac{E_P[e^{-rT}\eta_T X \mid F_t]}{e^{-rt}\eta_t}$$

$$= \frac{E_P[A_T X \mid F_t]}{A_t}$$

under the new approach.

The process A_t is called a state price deflator (also *deflator, state price density, pricing kernel* or *stochastic discount factor*).

Note that A_t is defined in terms of η_t, which is a function of Z_t. So A_t is a stochastic process linked to the random behaviour of the share price.

Question

Use Ito's Lemma to show that the SDE for A_t is:

$$dA_t = -A_t\left(rdt + \gamma dZ_t\right)$$

Solution

We have that:

$$A_t = e^{-rt}\eta_t = e^{-\gamma Z_t - (r + ½\gamma^2)t}$$

If we let $X_t = -\gamma Z_t - (r + ½\gamma^2)t$, so that $dX_t = -\gamma dZ_t - (r + ½\gamma^2)dt$, then:

$$A_t = e^{X_t}$$

We can then apply Ito's Lemma to get:

$$dA_t = \left[-(r + ½\gamma^2)e^{X_t} + ½\gamma^2 e^{X_t}\right]dt - \gamma e^{X_t}dZ_t$$

$$= -rA_t dt - \gamma A_t dZ_t$$

$$= -A_t(rdt + \gamma dZ_t)$$

This shows that A_t is just a 'randomised' version of the ordinary discount factor e^{-rt}, for which $d(e^{-rt}) = -(e^{-rt})rdt$.

A very important point to note is that, for this model, the risk-neutral and the state price deflator approaches give the same price V_t. Theoretically they are the same. They only differ in the way that they present the calculation of a derivative price.

3 The 5-step approach with dividends

3.1 Introduction

We now extend the theory involved in the 5-step method so that it can deal with an underlying asset that pays dividends. It may be useful to review the Black-Scholes option pricing formula chapter at this stage, where we discussed how the share price process must be modified to incorporate dividends.

Recall that the solution to the modified SDE was:

$$\tilde{S}_t = S_0 \exp\left[\left(\mu + q - \tfrac{1}{2}\sigma^2\right)t + \sigma Z_t\right]$$

The cash process will stay the same:

$$B_t = e^{rt}$$

3.2 The martingale approach

We have already mentioned that for a continuous-dividend-paying asset S_t, the tradeable asset is:

$$\tilde{S}_t = \tilde{S}_0 \exp[(\mu + q - \tfrac{1}{2}\sigma^2)t + \sigma Z_t]$$

rather than just S_t.

To price a derivative contingent on this underlying asset we can repeat the steps which allow us to price and replicate the derivative.

Step 1

Find the unique equivalent martingale measure Q under which:

$$\tilde{D}_t = e^{-rt}\tilde{S}_t$$

is a martingale.

Again, it is beyond the syllabus to actually construct the probability measure.

ie $\tilde{D}_t = \tilde{S}_0 \exp[-\tfrac{1}{2}\sigma^2 t + \sigma \tilde{Z}_t]$

where $\tilde{Z}_t$ is a standard Brownian motion under Q.

Hence we can write:

$$S_t = S_0 \exp[(r - q - \tfrac{1}{2}\sigma^2)t + \sigma \tilde{Z}_t]$$

Step 2 (proposition)

Let:

$$V_t = e^{-r(T-t)} E_Q[X \mid F_t]$$

where X is the derivative payoff at time T. We propose that this is the fair price of the derivative at t.

Note that this is exactly the same formula as before, except that the risk-neutral measure used to calculate the expectation is now the one for a dividend-paying share.

Step 3

Let:

$$E_t = e^{-rT} E_Q[X \mid F_t] = e^{-rt} V_t$$

This is a martingale under Q.

Step 4

By the martingale representation theorem there exists a previsible process $\tilde{\phi}_t$ such that $dE_t = \tilde{\phi}_t d\tilde{D}_t$.

As before without dividends, this application of the martingale representation theorem guarantees that the stock process $\tilde{\phi}_t$ is previsible.

Step 5

Let:

$$\psi_t = E_t - \tilde{\phi}_t \tilde{D}_t$$

At time t we hold the portfolio:

- **$\tilde{\phi}_t$ units of the tradeable asset $\tilde{S}_t$**

  (This is equivalent to $\phi_t = e^{qt} \tilde{\phi}_t$ units of S_t.)

- **ψ_t units of the cash account**

where $\tilde{\phi}_t d\tilde{D}_t = dE_t$ and $\psi_t = E_t - \phi_t \tilde{D}_t$.

Remembering that $e^{-rt} \tilde{S}_t = \tilde{D}_t$ and $E_t = e^{-rt} V_t$ we examine the change in the value of this portfolio over the very small time period $[t, t+dt]$.

At time t, the value of this portfolio is equal to:

$$\tilde{\phi}_t \tilde{S}_t + \psi_t B_t = e^{rt} \tilde{\phi}_t \tilde{D}_t + e^{rt}\left(E_t - \tilde{\phi}_t \tilde{D}_t\right) = e^{rt} E_t = V_t$$

At time $t+dt$ the value of this portfolio is equal to:

$$\begin{aligned}\tilde{\phi}_t\tilde{S}_{t+dt}+\psi_t B_{t+dt} &= e^{r(t+dt)}(\tilde{\phi}_t\tilde{D}_{t+dt}+\psi_t)\\ &= e^{r(t+dt)}(\tilde{\phi}_t\tilde{D}_t+\tilde{\phi}_t d\tilde{D}_t+\psi_t)\\ &= e^{r(t+dt)}(E_t+dE_t)\\ &= e^{r(t+dt)}E_{t+dt}=V_{t+dt}\end{aligned}$$

So the change in the value of the portfolio over the same period is:

$$\begin{aligned}\boldsymbol{V_{t+dt}-V_t=dV_t} &\boldsymbol{= B_t dE_t+E_t dB_t}\\ &\boldsymbol{= B_t\left[\tilde{\phi}_t d\tilde{D}_t+rE_t dt\right]}\\ &\boldsymbol{= \tilde{\phi}_t d\tilde{S}_t+\psi_t dB_t}\end{aligned}$$

The pure investment gain over the period t up to $t+dt$ is:

$$\boldsymbol{\tilde{\phi}_t d\tilde{S}_t+\psi_t dB_t=B_t\left[\tilde{\phi}_t d\tilde{D}_t+r\left(\tilde{\phi}_t\tilde{D}_t+\psi_t\right)dt\right]}$$

So the change in the value of the portfolio is the same as the pure investment gain.

Hence the portfolio is self-financing.

Also, $V_T=X$.

So, the hedging strategy $(\tilde{\phi}_t,\psi_t)$ is replicating and V_t is the fair price at time t.

Yet again, V_t, the proposed no-arbitrage value of the derivative at time $t<T$, is equal to:

- the time-t *expectation* of the claim amount paid at time T
- calculated with respect to the *probability measure* Q and
- the *filtration* F_t generated by the history of the stock price up to and including time t and
- discounted at the continuously compounded risk-free rate of return r.

3.3 Example: the price of a European call option on a share with dividends

The idea in this section is that we can use the Black-Scholes formula we have already derived for a call option on a non-dividend-paying share to derive the corresponding (Garman-Kohlhagen) formula when there are dividends.

In the absence of dividends, we know from the derivative pricing formula and the Black-Scholes formula that:

$$c_t=e^{-r(T-t)}E_Q[\max(S_T-K,0)]=S_t\Phi(d_1)-Ke^{-r(T-t)}\Phi(d_2)$$

where Q is the risk-neutral probability measure for S_t.

In the presence of dividends, we need to work out:

$$V_t = e^{-r(T-t)} E_{\tilde{Q}}[\max(S_T - K, 0)]$$

where $\tilde{Q}$ is the risk-neutral probability measure for $\tilde{S}_t$. Note that the payoff is still based on S_T, not $\tilde{S}_T$, because if the option is exercised, the option holder will just receive the basic share without the accumulated dividends.

If we use the relationship $\tilde{S}_T = S_T e^{qT}$ and we define $\tilde{K} = Ke^{qT}$, we can write:

$$V_t = e^{-r(T-t)} E_{\tilde{Q}}[\max(\tilde{S}_T e^{-qT} - \tilde{K}e^{-qT}, 0)]$$

$$= e^{-qT} \times e^{-r(T-t)} E_{\tilde{Q}}[\max(\tilde{S}_T - \tilde{K}, 0)]$$

The part of this expression after the multiplication sign looks exactly like the pricing formula for a call option, except that we have put squiggles on the S_T, K and Q. (Note also that $\tilde{Q}$ is the correct risk-neutral probability measure for $\tilde{S}_t$.)

This means that we can calculate this part using the Black-Scholes formula, provided that we replace all the S's and K's with $\tilde{S}$'s and $\tilde{K}$'s. This gives us:

$$V_t = e^{-qT} \times [\tilde{S}_t \Phi(d_1) - \tilde{K}e^{-r(T-t)} \Phi(d_2)]$$

where d_1 and d_2 are now calculated based on $\tilde{S}$ and $\tilde{K}$.

Suppose that:

$$\boldsymbol{X = \max\{S_T - K, 0\} = e^{-qT} \max\{\tilde{S}_T - \tilde{K}, 0\}}$$

where $\tilde{K} = Ke^{qT}$.

By analogy with the non-dividend-paying stock:

$$\boldsymbol{V_t = e^{-r(T-t)} E_{\tilde{Q}}[X \mid F_t]}$$

$$\boldsymbol{= e^{-qT} \left[\tilde{S}_t \Phi(d_1) - \tilde{K}e^{-r(T-t)} \Phi(d_2) \right]}$$

where:

$$\boldsymbol{d_1 = \frac{\log \frac{\tilde{S}_t}{\tilde{K}} + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}}$$

$$= \frac{\log \frac{e^{qt} S_t}{e^{qT} K} + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

$$= \frac{\log \frac{S_t}{K} - q(T-t) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

$$= \frac{\log \frac{S_t}{K} + (r - q + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

and $\quad d_2 = d_1 - \sigma\sqrt{T-t}$

Also:

$$e^{-qT}\tilde{S}_t = S_t e^{-q(T-t)}$$

and:

$$e^{-qT}\tilde{K} = K$$

$$\Rightarrow \qquad V_t = S_t e^{-q(T-t)}\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

So the difference when dividends are present is that we have to 'strip out' the dividends with an $e^{-q(T-t)}$ factor, and change r to $r-q$ in the calculation of d_1 and d_2. The formula for a put option works in the same way.

By looking at this formula we can see that it may be optimal to exercise early an American call option on a continuous-dividend-paying stock. This is because the value of the equivalent European call option can be less than the option's intrinsic value. In particular, for any $t < T$, as S_t gets large (relative to *K*), V_t is approximately equal to:

$$S_t e^{-q(T-t)} - Ke^{-r(T-t)} < S_t - K$$

for large enough S_t.

This is because d_1 and d_2 would be large, so that $\Phi(d_1)$ and $\Phi(d_2)$ are approximately equal to 1.

Question

According to this approximation, how large would S_t need to be?

Solution

If we rearrange this inequality, we have:

$$K\left\{1 - e^{-r(T-t)}\right\} < S_t\left\{1 - e^{-q(T-t)}\right\}$$

So we would need to have:

$$S_t > \frac{1 - e^{-r(T-t)}}{1 - e^{-q(T-t)}}K$$

Question

Can you spot any other situations where this 'reverse' situation could apply?

Solution

One example would be if r is close to zero, but q is high. However, for economic reasons this situation is less likely to occur with the shares of major companies.

We can equally derive the price of a European put option on a dividend-paying stock:

ie $$V_t = Ke^{-r(T-t)}\Phi(-d_2) - S_t e^{-q(T-t)}\Phi(-d_1)$$

where d_1 and d_2 are defined above.

The chapter summary starts on the next page so that you can keep all the chapter summaries together for revision purposes.

Chapter 17 Summary

The martingale (5-step) approach (with dividends)

The derivative pricing formula $V_t = e^{-r(T-t)}E_Q[X \mid F_t]$ can be derived using the martingale approach, which consists of five steps:

Step 1

Find the unique equivalent martingale measure Q under which $\tilde{D}_t = e^{-rt}\tilde{S}_t$ is a martingale.

Step 2

Let $V_t = e^{-r(T-t)}E_Q\left[X|F_t\right]$ where X is the derivative payoff at time T. This is proposed as the fair price of the derivative at time t.

Step 3

Let $E_t = e^{-rT}E_Q\left[X|F_t\right] = e^{-rt}V_t$. This is a martingale under Q.

Step 4

By the martingale representation theorem, there exists a previsible process $\tilde{\phi}_t$ such that $dE_t = \tilde{\phi}_t d\tilde{D}_t$.

Step 5

Let $\psi_t = E_t - \tilde{\phi}_t\tilde{D}_t$ and at time t hold the portfolio consisting of:

- $\tilde{\phi}_t$ units of the tradeable $\tilde{S}_t$
- ψ_t units of the cash account.

At time t the value of this portfolio is equal to V_t. Also $V_T = X$. Therefore, the hedging strategy $(\tilde{\phi}_t, \psi_t)$ is replicating and so, by no arbitrage, V_t is the fair price of the derivative at time t.

Garman-Kohlhagen formula for a European option on a dividend-paying share

Call option

$$f(S_t) = S_t \Phi(d_1) e^{-q(T-t)} - Ke^{-r(T-t)} \Phi(d_2)$$

Put option

$$f(S_t) = Ke^{-r(T-t)} \Phi(-d_2) - S_t \Phi(-d_1) e^{-q(T-t)}$$

where:

- $d_1 = \dfrac{\ln \dfrac{S_t}{K} + (r - q + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$
- $d_2 = d_1 - \sigma\sqrt{T-t}$

The Black-Scholes formulae for a non-dividend-paying share are the same but using $q = 0$. These formulae can be derived by direct evaluation of the expected value using the general option pricing formula found using the 5-step method.

Delta hedging

$$\phi_t = \frac{\partial V}{\partial s}(t, S_t) = \Delta$$

- We start at time 0 with V_0 invested in cash and shares.
- We follow a self-financing portfolio strategy.
- We continually rebalance the portfolio to hold exactly ϕ_t units of S_t with the rest in cash.

By following these steps, we precisely replicate the derivative payoff, without risk.

The martingale approach *vs* the PDE approach

- In the PDE approach we have to 'guess' the solution, whereas with the martingale approach we do not.
- The martingale approach provides an expectation that can be evaluated explicitly in some cases and in a straightforward numerical way in other cases.
- The martingale approach also gives the replicating strategy for the derivative.
- The martingale approach can be applied to any F_T -measurable derivative payment, whereas the PDE approach cannot always.
- However, the PDE approach is much quicker and easier to construct, and more easily understood.

State price deflator approach

Corollary to the Cameron-Martin-Girsanov Theorem

There exists a process η_t such that, for any F_T -measurable derivative payoff X_T at time T, we have:

$$E_Q[X_T \mid F_t] = E_P\left[\frac{\eta_T}{\eta_t} X_T \middle| F_t\right]$$

The *state price deflator* A_t is defined by:

$$A_t = e^{-rt}\eta_t$$

where $\eta_t = e^{-\gamma Z_t - ½\gamma^2 t}$, which is a martingale under *P*.

Derivative prices can be calculated using the state price deflator formula:

$$V_t = \frac{E_P[A_T X_T \mid F_t]}{A_t}$$

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 17 Practice Questions

17.1 You are given that the fair price to pay at time t for a derivative paying X at time T is

Exam style

$V_t = e^{-r(T-t)}E_Q\left[X \middle| F_t\right]$, where Q is the risk-neutral probability measure and F_t is the filtration with respect to the underlying process. The price movements of a non-dividend-paying share are governed by the stochastic differential equation $dS_t = S_t\left(rdt + \sigma dB_t\right)$, where B_t is standard Brownian motion under the risk-neutral probability measure.

(i) Solve the above stochastic differential equation. [4]

(ii) Determine the probability distribution of $S_T \mid S_t$. [2]

(iii) Hence show that the fair price to pay at time t for a forward on this share, with forward price K and time to expiry $T-t$, is:

$$V_t = S_t - Ke^{-r(T-t)}$$ [4]

[Total 10]

17.2 The process S_t is defined by $S_t = S_0 e^{(\alpha - ½\sigma^2)t + \sigma W_t}$, where W_t is standard Brownian motion under

Exam style

a probability measure P, and α and σ^2 are constants.

(i) (a) State the name given to the process S_t.

(b) Give two real-world quantities that are commonly modelled using such a process. [3]

(ii) (a) State what is meant by 'equivalent probability measures'.

(b) State how the Cameron-Martin-Girsanov Theorem could be applied here if we wished to work with a process of the form $S_t = S_0 e^{(r - ½\sigma^2)t + \sigma \tilde{W}_t}$, where $\tilde{W}_t$ is standard Brownian motion and r is the risk-free rate of interest. [3]

(iii) (a) Determine the stochastic differential equation for dS_t in terms of dt and $d\tilde{W}_t$.

(b) State the drift of the process in (iii)(a) and comment on your answer. [5]

[Total 11]

17.3 Exam style

The random variable S_T is defined by $S_T = S_0 e^{(r-½\sigma^2)T+\sigma\tilde{Z}_T}$, where $\tilde{Z}_T \sim N(0,T)$ under a probability measure Q. The random variable X is defined by $X = \max(S_T - K, 0)$, where K is a positive constant.

(i) Show that:

$$E_Q[X|F_0] = S_0 e^{rT}\Phi\left(d_2 + \sigma\sqrt{T}\right) - K\Phi\left(d_2\right)$$

where $d_2 = \dfrac{\log(S_0/K)+(r-½\sigma^2)T}{\sigma\sqrt{T}}$. [7]

Hint: You may wish to use the lognormal integral formulae given on page 18 of the Tables.

(ii) Explain carefully the relevance of this result in financial mathematics. [3]

[Total 10]

17.4 Exam style

(i) State the general risk-neutral pricing formula for the price of a derivative at time t in terms of the derivative payoff X_T at the maturity date T and the constant risk-free force of interest r. [1]

Assume that the price of a share, which pays a constant force of dividend yield q, follows geometric Brownian motion.

(ii) (a) Derive the formula for the price at time t of Derivative 1, which pays one at time T provided the share price at that time is less than K.

(b) Derive the formula for the price at time t of Derivative 2, which pays the share price at time T provided the share price at that time is less than K.

(c) Hence derive the formula for the price of a European put option with strike price K. [12]

Hint for (ii)(a) and (ii)(b): You may wish to use the lognormal integral formulae given on page 18 of the Tables.

(iii) An exotic forward contract provides a payoff equal to the value of the square of the share price at the maturity date T in return for a payment equal to the square of the forward price. Derive the formula for the value of this contract at time t. [3]

[Total 16]

17.5 An exotic forward provides a payoff equal to the square root of the share price at maturity time T less the square root of the delivery price, K.

Exam style

(i) (a) Assuming that the Black-Scholes assumptions apply, use risk-neutral pricing to derive a formula for the price at time t of the forward on a non-dividend-paying share.

(b) Derive the corresponding formula for the vega of the forward. [6]

(ii) (a) Explain why an investor might want to vega hedge their portfolio.

(b) Use the result that $S_t\phi(d_1) - Ke^{-r(T-t)}\phi(d_2) = 0$, where d_1 and d_2 are defined as on page 47 in the *Tables*, to show that the formula for the vega of a European call option is $\nu_{call} = S_t\phi(d_1)\sqrt{T-t}$. [6]

The current price of the share is \$1, which is also the delivery price of the forward. The risk-free force of interest is 5%, the volatility of the underlying share, which pays no dividends, is 20% and the forward has one year to delivery.

(iii) An investor has a long position in 1,000 exotic forwards. Find the vega-hedged portfolio for this position involving standard European call options on the underlying share and also the underlying share itself. [8]

[Total 20]

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 17 Solutions

17.1 (i) ***Solving the SDE***

The process is geometric Brownian motion. To solve it we consider the function $f(S_t) = \log S_t$.

Applying the Taylor's series formula to the above function, we get:

$$df(S_t) = d(\log S_t) = \frac{1}{S_t} dS_t + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)(dS_t)^2$$

$$= \frac{1}{S_t}(rS_t\,dt + \sigma S_t\,dB_t) - \frac{1}{2S_t^2}(rS_t\,dt + \sigma S_t\,dB_t)^2$$

$$= r\,dt + \sigma\,dB_t - \tfrac{1}{2}\sigma^2\,dt$$

$$= (r - \tfrac{1}{2}\sigma^2)dt + \sigma dB_t \qquad [2]$$

Changing the t's to s's and integrating this equation between limits of $s=0$ and $s=t$, we get:

$$\left[\log S_s\right]_{s=0}^{s=t} = (r - \tfrac{1}{2}\sigma^2)\int_0^t ds + \sigma \int_0^t dB_s$$

$$\Rightarrow \qquad \log S_t - \log S_0 = (r - \tfrac{1}{2}\sigma^2)t + \sigma B_t$$

$$\Rightarrow \qquad S_t = S_0\, e^{(r - ½\sigma^2)t + \sigma B_t} \qquad [2]$$

[Total 4]

(ii) ***Probability distribution***

From part (i), we know that, under the risk-neutral probability measure Q,

$$\log S_t - \log S_0 = (r - \tfrac{1}{2}\sigma^2)t + \sigma B_t$$

Since $B_t \sim N(0,t)$ then:

$$\log S_t - \log S_0 \sim N\left((r - ½\sigma^2)t,\ \sigma^2 t\right) \qquad [1]$$

Replacing 0 and t with t and $T-t$ we get:

$$\log S_T - \log S_t \sim N\left((r - ½\sigma^2)(T-t),\ \sigma^2(T-t)\right)$$

$$\Rightarrow \quad \log S_T \mid S_t \sim N\left(\log S_t + (r - ½\sigma^2)(T-t),\ \sigma^2(T-t)\right)$$

$$\Rightarrow \quad S_T \mid S_t \sim \log N\left(\log S_t + (r - ½\sigma^2)(T-t),\ \sigma^2(T-t)\right) \qquad [1]$$

[Total 2]

(iii) ***Fair price for a forward***

We are given that the fair price to pay at time t for a derivative paying X at time T is $V_t = e^{-r(T-t)} E_Q\left[X \middle| F_t\right]$, where Q is the risk-neutral probability measure.

The random variable payoff of a forward on a non-dividend-paying share is:

$$X = S_T - K$$

Substituting this into the fair price formula, we get:

$$\begin{aligned} V_t = e^{-r(T-t)} E_Q\left[X \middle| F_t\right] &= e^{-r(T-t)} E_Q\left[S_T - K \middle| F_t\right] \\ &= e^{-r(T-t)} \left(E_Q\left[S_T \middle| F_t\right] - K\right) \end{aligned} \qquad [2]$$

Now $E_Q\left[S_T \middle| F_t\right]$ is the conditional mean of the random variable S_T, and from part (ii), we know that $S_T \mid S_t \sim \log N\left(\log S_t + (r - ½\sigma^2)(T-t),\ \sigma^2(T-t)\right)$.

Using the formula for the expectation of the lognormal distribution on page 14 of the *Tables*:

$$\begin{aligned} E_Q\left[S_T \middle| F_t\right] &= e^{\left\{\log S_t + (r - \frac{1}{2}\sigma^2)(T-t) + \frac{1}{2}\sigma^2(T-t)\right\}} \\ &= S_t e^{r(T-t)} \end{aligned} \qquad [1]$$

So:

$$\begin{aligned} V_t = e^{-r(T-t)} E_Q\left[X \middle| F_t\right] &= e^{-r(T-t)} \left(E_Q\left[S_T \middle| F_t\right] - K\right) \\ &= S_t - K e^{-r(T-t)} \end{aligned} \qquad [1]$$

[Total 4]

17.2 (i)(a) ***Name of the process***

This process is *geometric Brownian* motion (also known as the continuous-time lognormal model). [1]

(i)(b) ***Two real-world quantities***

This process is commonly used to model share prices (with or without dividends) and currency exchange rates. [2]

[Total 3]

(ii)(a) ***State what is meant by 'equivalent probability measures'***

In words, two probability measures are equivalent if they are defined on the same sample space and have the same null sets (*ie* sets that have probability zero).

Mathematically, P and Q are equivalent if $P(A)>0 \Leftrightarrow Q(A)>0$, where $P(A)$ denotes the probability under measure P and $Q(A)$ denotes the probability under measure Q. [1]

This is sometimes expressed as 'Equivalent measures agree on what is possible'.

(ii)(b) ***State how the CMG Theorem could be applied***

The CMG Theorem tells us that, for any γ, there is a probability measure Q (equivalent to P) such that $\tilde{W}_t = W_t + \gamma t$ is standard Brownian motion under Q.

So we could change probability measures and work with Q.

Since $W_t = \tilde{W}_t - \gamma t$, under the measure Q, the formula for S_t in terms of $\tilde{W}_t$ becomes:

$$S_t = S_0 e^{(\alpha - ½\sigma^2)t + \sigma(\tilde{W}_t - \gamma t)} = S_0 e^{(\alpha - \gamma\sigma - ½\sigma^2)t + \sigma\tilde{W}_t}$$

If we want this to equal $S_0 e^{(r-½\sigma^2)t+\sigma\tilde{W}_t}$, we need to set $\alpha - \gamma\sigma = r$, *ie* $\gamma = \dfrac{\alpha - r}{\sigma}$. [2]

[Total 3]

(iii)(a) ***SDE for dS_t***

We can write the equation $S_t = S_0 e^{(r-½\sigma^2)t+\sigma\tilde{W}_t}$ in the form:

$$S_t = f(X_t)$$

where $X_t = (r - ½\sigma^2)t + \sigma\tilde{W}_t$ and $f(x) = S_0 e^x$

So: $dX_t = (r - ½\sigma^2)dt + \sigma d\tilde{W}_t$ and $f'(x) = f''(x) = S_0 e^x$

Using a Taylor Series expansion, we can write:

$$dS_t = df(X_t) = f'(X_t)dX_t + \tfrac{1}{2}f''(X_t)(dX_t)^2$$

$$= S_0 e^{X_t}\left\{dX_t + \tfrac{1}{2}(dX_t)^2\right\}$$

$$= S_t\left\{dX_t + \tfrac{1}{2}(dX_t)^2\right\}$$

Substituting the SDE for X_t gives:

$$dS_t = S_t\left\{[(r - ½\sigma^2)dt + \sigma d\tilde{W}_t] + \tfrac{1}{2}[(r - ½\sigma^2)dt + \sigma d\tilde{W}_t]^2\right\}$$

Simplifying using the 2×2 multiplication grid, we get:

$$dS_t = S_t\left\{[(r - ½\sigma^2)dt + \sigma d\tilde{W}_t] + \tfrac{1}{2}\sigma^2 dt\right\}$$

$$= S_t(rdt + \sigma d\tilde{W}_t) \qquad [3]$$

(iii)(b) ***SDE for dS$_t$***

The drift in this equation is rS_t, or just r, if we're thinking in units of S_t. [1]

Since $\tilde{W}_t$ is standard Brownian motion, the increment $d\tilde{W}_t$ has mean zero. This means that the expected value of dS_t under the measure Q is $rS_t dt$, *ie* the share price is drifting upwards at the risk-free rate. This means that Q is the risk-neutral probability measure for the process S_t. [1]

[Total 5]

17.3 (i) ***Formula for $E_Q[X]$***

If we take logs of the equation given, we get:

$$\log S_T = \log S_0 + (r - ½\sigma^2)T + \sigma\tilde{Z}_T$$

Since $\tilde{Z}_T \sim N(0,T)$, we see that the distribution of S_T given F_0 under the probability measure Q is:

$$S_T \,|\, F_0 \sim logN[\log S_0 + (r - ½\sigma^2)T, \sigma^2 T] \qquad [1]$$

The expectation we require is:

$$E_Q[X|F_0] = E_Q[\max(S_T - K, 0)|F_0]$$

$$= \int_K^\infty (x-K)f(x)dx$$

$$= \int_K^\infty x\,f(x)dx - K\int_K^\infty f(x)dx \qquad [1]$$

where $f(x)$ is the density function of the above lognormal distribution.

We can evaluate these integrals using the formula on page 18 of the *Tables*, with $L = K$ and $U = \infty$, and with μ and σ^2 replaced with $\log S_0 + (r - ½\sigma^2)T$ and $\sigma^2 T$.

If we put $k = 1$, we get:

$$\int_K^\infty xf(x)dx = e^{\log S_0 + (r-½\sigma^2)T + ½\sigma^2 T}\left[\Phi(\infty) - \Phi\left(\frac{\log K - [\log S_0 + (r-½\sigma^2)T]}{\sigma\sqrt{T}} - \sigma\sqrt{T}\right)\right]$$

$$= S_0 e^{rT}\left[1 - \Phi\left(\frac{\log K - [\log S_0 + (r-½\sigma^2)T]}{\sigma\sqrt{T}} - \sigma\sqrt{T}\right)\right]$$

We can simplify this using the identity $1 - \Phi(x) = \Phi(-x)$ to get:

$$\int_K^\infty xf(x)dx = S_0 e^{rT}\Phi\left(\frac{-\log K + \log S_0 + (r-½\sigma^2)T}{\sigma\sqrt{T}} + \sigma\sqrt{T}\right)$$

$$= S_0 e^{rT}\Phi\left(\frac{\log(S_0/K) + (r-½\sigma^2)T}{\sigma\sqrt{T}} + \sigma\sqrt{T}\right)$$

$$= S_0 e^{rT}\Phi\left(d_2 + \sigma\sqrt{T}\right) \qquad [3]$$

where d_2 is as defined in the question.

Similarly, if we put $k = 0$, we get:

$$\int_K^\infty f(x)dx = \Phi\left(d_2\right) \qquad [1]$$

So the expectation is:

$$E_Q[X|F_0] = S_0 e^{rT}\Phi\left(d_2 + \sigma\sqrt{T}\right) - K\Phi\left(d_2\right) \qquad [1]$$

[Total 7]

(ii) ***Relevance of result***

If S_T denotes the value of a share at time T, and Q denotes the risk-neutral measure for this share, this formula allows us to work out the fair price of a European call option on the share with strike price K and time to expiry T. [1]

This price would be calculated as the discounted value of the expectation in part (i), namely:

$$e^{-rT}E_Q[X|F_0] = e^{-rT}\left[S_0 e^{rT}\Phi\left(d_2+\sigma\sqrt{T}\right) - K\Phi\left(d_2\right)\right]$$

$$= S_0\Phi\left(d_1\right) - Ke^{-rT}\Phi\left(d_2\right) \qquad [1]$$

where $d_1 = d_2 + \sigma\sqrt{T}$.

This matches the Black-Scholes formula for valuing a call option on a non-dividend-paying share. [1]

[Total 3]

17.4 (i) ***State the general risk-neutral pricing formula***

$$V_t = e^{-r(T-t)}E_Q\left[X_T \middle| F_t\right] \qquad [1]$$

where:

r	=	constant risk-free force of interest
T	=	maturity date of the derivative
t	=	today's date
X_T	=	derivative payoff at maturity date T
Q	=	risk-neutral probability measure
F_t	=	filtration at time t

(ii)(a) ***Derive the formula for the price of Derivative 1***

The payoff function of this non-standard derivative is:

$$X_{1T} = \begin{cases} 1 & S_T < K \\ 0 & S_T \geq K \end{cases} \qquad [½]$$

To derive the pricing formula of the derivative, we substitute the payoff function into the general risk-neutral pricing formula in part (i), *ie*:

$$V_{1t} = e^{-r(T-t)} E_Q\left[X_{1T} \middle| F_t\right]$$

which, as the derivative pays 1 provided the share price is between 0 and the strike price K, is equal to:

$$V_{1t} = e^{-r(T-t)} \int_0^K 1\, f(S_T | S_t)\, dS_T \qquad (1) \qquad [1]$$

where $f(S_T | S_t)$ is the probability density function of the share price at the maturity date, S_T, given the current share price, S_t.

Note that as the share price is assumed to follow geometric Brownian motion, the independence of increments means we do not have to condition on the full past history of the share price – only the current share price, S_t.

The distribution of S_T given S_t is:

$$S_T | S_t \sim \log N\left[\ln S_t + (r - q - ½\sigma^2)(T-t), \sigma^2(T-t)\right] \qquad [½]$$

We can simply state this result or obtain it using the following reasoning:

Under the risk-neutral probability measure, Q, the tradeable asset is expected to grow at the risk-free force of interest, r. Here, the tradeable asset is the share plus the dividends earned. Assuming the dividends are immediately reinvested in the asset, they will give a rate of growth of the tradeable asset of q. So this means the share price alone must grow at a rate $r-q$.

Therefore, if S_t denotes the share price at time t, then it has SDE:

$$dS_t = S_t\left((r-q)dt + \sigma dZ_t\right)$$

where Z_t denotes standard Brownian motion. This has solution:

$$S_T = S_t \exp\left(\left(r - q - ½\sigma^2\right)(T-t) + \sigma(Z_T - Z_t)\right)$$

giving the distribution:

$$\ln S_T | S_t \sim N\left[\ln S_t + (r - q - ½\sigma^2)(T-t), \sigma^2(T-t)\right]$$

$$\Leftrightarrow \quad S_T | S_t \sim \log N\left[\ln S_t + (r - q - ½\sigma^2)(T-t), \sigma^2(T-t)\right]$$

To work out the integral (1) above, we need to use the formula for evaluating the truncated moments of a lognormal distribution, which appears on page 18 of the *Tables.*

We have $1=(S_T)^0$ and so $k=0$ in the formula on page 18 of the *Tables*.

Thus, (1) above is equal to:

$$V_{1t}=e^{-r(T-t)}\left[e^0\right]\left[\Phi(U_0)-\Phi(L_0)\right] \qquad (2) \qquad [½]$$

where:

$$U_0=\frac{\ln K-\left\{\ln S_t+(r-q-½\sigma^2)(T-t)\right\}}{\sigma\sqrt{T-t}}-0$$

$$=-\left[\frac{\ln(S_t/K)+(r-q-½\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right]$$

$$=-d_2 \qquad [1]$$

$$L_0=\frac{\ln 0-\left\{\ln S_t+(r-q-½\sigma^2)(T-t)\right\}}{\sigma\sqrt{T-t}}-0=-\infty \qquad [½]$$

So, (2) above becomes:

$$V_{1t}=e^{-r(T-t)}\left[\Phi(-d_2)-\Phi(-\infty)\right]$$

$$=e^{-r(T-t)}\left[\Phi(-d_2)-0\right]$$

$$=e^{-r(T-t)}\Phi(-d_2) \qquad [1]$$

(ii)(b) ***Derive the formula for the price of Derivative 2***

The payoff function of this non-standard derivative is:

$$X_{2T}=\begin{cases} S_T & S_T<K \\ 0 & S_T\geq K \end{cases} \qquad [½]$$

To derive the pricing formula of this derivative, we substitute the payoff function into the risk-neutral pricing formula in part (i), *ie*:

$$V_{2t}=e^{-r(T-t)}E_Q\left[X_{2T}\,|F_t\right]$$

Substituting in for X_{2T} gives:

$$V_{2t}=e^{-r(T-t)}\int_0^K S_T\, f(S_T\,|S_t)\,dS_T \qquad (3) \qquad [1]$$

To evaluate this, we again use the formula for evaluating the truncated moments of a lognormal distribution on page 18 of the *Tables*, where:

$$S_T \mid S_t \sim \log N\left[\ln S_t + (r - q - ½\sigma^2)(T-t), \sigma^2(T-t)\right] \qquad [½]$$

In this instance, $S_T = (S_T)^1$, so $k=1$ in the formula on page 18 of the *Tables*. Thus, (3) above is equal to:

$$V_{2t} = e^{-r(T-t)}\left[e^{\ln S_t + (r-q-½\sigma^2)(T-t) + ½\sigma^2(T-t)}\right]\left[\Phi(U_1) - \Phi(L_1)\right] \qquad (4) \qquad [½]$$

where:

$$U_1 = \frac{\ln K - \left\{\ln S_t + (r-q-½\sigma^2)(T-t)\right\}}{\sigma\sqrt{T-t}} - \sigma\sqrt{T-t}$$

$$= \frac{\ln K - \ln S_t - (r-q-½\sigma^2)(T-t) - \sigma^2(T-t)}{\sigma\sqrt{T-t}}$$

$$= \frac{\ln K - \ln S_t - (r-q+½\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

$$= -\left[\frac{\ln(S_t/K) + (r-q+½\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right]$$

$$= -d_1 \qquad [1]$$

$$L_1 = \frac{\ln 0 - \left\{\ln S_t + (r-q-½\sigma^2)(T-t)\right\}}{\sigma\sqrt{T-t}} - \sigma\sqrt{T-t} = -\infty \qquad [½]$$

So, after some cancelling of terms involving r and σ^2, (4) above becomes:

$$V_{2t} = S_t e^{-q(T-t)}\left[\Phi(-d_1) - \Phi(-\infty)\right]$$

$$= S_t e^{-q(T-t)}\left[\Phi(-d_1) - 0\right]$$

$$= S_t e^{-q(T-t)}\Phi(-d_1) \qquad [1]$$

(ii)(c) ***Derive the pricing formula for a European put option***

The payoff function for a European put option can be written as:

$$X_{PT} = \begin{cases} K - S_T & S_T < K \\ 0 & S_T \geq K \end{cases} \qquad [½]$$

This payoff function can be replicated using a combination of $+K$ of the derivatives in part (ii)(a) and -1 of the derivatives in part (ii)(b). [½]

Consequently, and assuming that markets are arbitrage-free, the price of a European put option must be given by:

$$p_t = K \times V_{1t} - V_{2t}$$

$$= Ke^{-r(T-t)}\Phi(-d_2) - S_t e^{-q(T-t)}\Phi(-d_1)$$ [1]

[Total 12]

(iii) ***Price of exotic forward contract***

Here the payoff function is equal to:

$$X_T = S_T^2 - K^2$$ [1]

So, once again substituting this into the general risk-neutral pricing formula in part (i) gives:

$$V_t = e^{-r(T-t)} E_Q\left[S_T^2 - K^2 \middle| F_t \right]$$

$$= e^{-r(T-t)}\left(E_Q\left[S_T^2 \middle| F_t \right] - K^2 \right)$$

where:

$$S_T \middle| S_t \sim \log N\left[\ln S_t + (r - q - ½\sigma^2)(T-t), \sigma^2(T-t) \right]$$ [1]

As this derivative contract provides a non-zero payoff regardless of the share price at maturity, we can evaluate $E_Q\left[S_T^2 \middle| F_t \right]$ using the formula for the moments of a non-truncated lognormal distribution on page 14 of the *Tables*, *ie*:

$$V_t = e^{-r(T-t)}\left[e^{2(\ln S_t + (r-q-½\sigma^2)(T-t)) + ½\times 4\sigma^2(T-t)} - K^2 \right]$$

$$= S_t^2 e^{(r-2q+\sigma^2)(T-t)} - K^2 e^{-r(T-t)}$$ [1]

[Total 3]

17.5 (i)(a) ***Derive formula for price of forward***

The general risk-neutral formula for pricing a derivative at time $t < T$ is:

$$V_t = e^{-r(T-t)} E_Q\left[X_T \middle| F_t \right]$$ [½]

In this instance, the payoff function of the derivative is:

$$X_T = S_T^{½} - K^{½}$$

So, the price will be given by:

$$V_t = e^{-r(T-t)} E_Q\left[S_T^{½} - K^{½} \middle| F_t \right]$$ [1]

Black-Scholes assumes that under the risk-neutral measure Q, the underlying share price follows geometric Brownian motion with drift r and volatility σ and so:

$$\log S_T \,|\, S_t \sim N\left[\log S_t + \left(r - ½\sigma^2\right)(T-t), \sigma^2(T-t)\right] \qquad [½]$$

Hence, using the formula for the moments of a lognormal distribution from page 14 in the *Tables*, with $r = ½$, the price will be given by:

$$V_t = e^{-r(T-t)}\left[e^{½\left(\log S_t + (r-½\sigma^2)(T-t)\right) + ½\times½^2\sigma^2(T-t)} - K^{½}\right] \qquad [1]$$

Note that the formula in the Tables also works for non-integer values of r.

This simplifies to:

$$V_t = S_t^{½} e^{-½\left(r + ¼\sigma^2\right)(T-t)} - K^{½} e^{-r(T-t)} \qquad [1]$$

[Total 4]

(i)(b) ***Formula for the vega of the forward***

The vega of a derivative with price *f* based on an underlying share with volatility σ is defined as:

$$\nu = \frac{\partial f}{\partial \sigma} \qquad [½]$$

So, here:

$$\nu = -¼\sigma(T-t)\, S_t^{½} e^{-½\left(r + ¼\sigma^2\right)(T-t)} \qquad [1½]$$

[Total 2]

(ii)(a) ***Why vega hedge a portfolio?***

A vega-hedged portfolio is one whose overall vega, which is equal to the sum of the vegas of the constituent securities, is close to zero. [½]

Consequently, the value of such a portfolio will be relatively insensitive to changes in the volatility of the underlying share. [½]

An investor might therefore wish to vega hedge in order to:

- protect the value of a portfolio against (small) changes in the volatility of the underlying share. [½]
- compensate for the fact that the volatility of the share is unknown, as it cannot be observed directly. It is less important to have an accurate estimate of the volatility if vega is low and hence has little effect on the portfolio's value. [½]

[Total 2]

(ii)(b) ***Show that vega of European call option is*** $\boldsymbol{\nu_{call} = S_t \phi(d_1)\sqrt{T-t}}$

Differentiating the Black-Scholes formula for a European call option with respect to σ using the product and chain rules for differentiation gives:

$$\nu_{call} = S_t \phi(d_1)\frac{\partial d_1}{\partial \sigma} - Ke^{-r(T-t)}\phi(d_2)\frac{\partial d_2}{\partial \sigma}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)$ is the probability density function of the standard normal distribution (from page 11 of the *Tables*). [1]

Now:

$$d_2 = d_1 - \sigma\sqrt{T-t}$$

Thus:

$$\frac{\partial d_2}{\partial \sigma} = \frac{\partial d_1}{\partial \sigma} - \sqrt{T-t} \qquad [½]$$

Hence:

$$\nu_{call} = S_t \phi(d_1)\frac{\partial d_1}{\partial \sigma} - Ke^{-r(T-t)}\phi(d_2) \times \left[\frac{\partial d_1}{\partial \sigma} - \sqrt{T-t}\right]$$

$$= \frac{\partial d_1}{\partial \sigma}\left[S_t\phi(d_1) - Ke^{-r(T-t)}\phi(d_2)\right] + Ke^{-r(T-t)}\phi(d_2)\sqrt{T-t} \qquad [1]$$

So, using the result that:

$$S_t\phi(d_1) - Ke^{-r(T-t)}\phi(d_2) = 0$$

we have:

$$\nu_{call} = \frac{\partial d_1}{\partial \sigma} \times 0 + Ke^{-r(T-t)}\phi(d_2)\sqrt{T-t}$$

$$= Ke^{-r(T-t)}\phi(d_2)\sqrt{T-t} \qquad [1]$$

Or equally, given that:

$$S_t\phi(d_1) = Ke^{-r(T-t)}\phi(d_2)$$

this can be written as:

$$\nu_{call} = S_t\phi(d_1)\sqrt{T-t} \qquad [½]$$

[Total 4]

(iii) ***Find vega-hedged portfolio***

Using the information given in the question, together with the formulae found earlier in the question, the price and vega of the forward are:

$$V_t = S_t^{½} e^{-½\left(r+¼\sigma^2\right)(T-t)} - K^{½} e^{-r(T-t)}$$

$$= 1 \times e^{-½\left(0.05+¼\times 0.2^2\right)\times 1} - 1 \times e^{-0.05\times 1}$$

$$= 0.019216 \qquad [1]$$

$$\nu = -¼\sigma(T-t)\, S_t^{½} e^{-½\left(r+¼\sigma^2\right)(T-t)}$$

$$= -¼ \times 0.2 \times 1 \times 1 \times e^{-½\left(0.05+¼\times 0.2^2\right)\times 1}$$

$$= -0.048522 \qquad [1]$$

Using the Black-Scholes formula on page 47 in the *Tables*:

$$d_1 = \frac{\log(1) + \left(0.05 + ½ \times 0.2^2\right)\times 1}{0.2 \times 1} = 0.35 \qquad [½]$$

$$d_2 = 0.35 - 0.2 \times 1 = 0.15 \qquad [½]$$

So, the price of the European call option is:

$$c_t = 1 \times \Phi(0.35) - 1 \times e^{-0.05\times 1}\Phi(0.15)$$

$$= 0.63683 - 0.951229 \times 0.55962$$

$$= 0.104503 \qquad [1]$$

and its vega is equal to:

$$\nu_{call} = S_t \phi(d_1)\sqrt{T-t}$$

$$= 1 \times \frac{1}{\sqrt{2\pi}} e^{-½\times 0.35^2} \times 1$$

$$= 0.375240 \qquad [1]$$

Finally, the vega of a share is equal to zero. [½]

So, to vega hedge the long position in 1,000 forwards, we need to find the number of European calls x and shares y such that:

$$1{,}000 \times 0.019216 = 0.104503x + 1 \times y \qquad [½]$$

$$-1{,}000 \times 0.048522 = 0.375240x + 0 \times y \qquad [½]$$

Solving these equations gives:

$$x = -129.31$$ [½]

$$y = 32.729$$ [½]

ie we need to short sell 129.31 calls and buy 32.729 shares. [½]

This portfolio will respond in a similar way to the exotic forward to (small) changes in the volatility of the underlying share. Strictly speaking, in order to vega hedge the exotic forward we need to take the opposite positions to those found above, *ie* buy 129.31 calls and short sell 32.729 shares. [1]

[Maximum 8]

End of Part 3

What next?

1. Briefly **review** the key areas of Part 3 and/or re-read the **summaries** at the end of Chapters 13 to 17.
2. Ensure you have attempted some of the **Practice Questions** at the end of each chapter in Part 3. If you don't have time to do them all, you could save the remainder for use as part of your revision.
3. Attempt **Assignment X3**.

18

The term structure of interest rates

Syllabus objectives

4.5 Models of the term structures of interest rates

1. Explain the principal concepts and terms underlying the theory of a term structure of interest rates.
2. Describe the desirable characteristics of models for the term structure of interest rates.
3. Apply the term structure of interest rates to modelling various cash flows, including calculating the sensitivity of the value to changes in the term structure.
4. Describe, as a computational tool, the risk-neutral approach to the pricing of zero-coupon bonds and interest rate derivatives for a general one-factor diffusion model for the risk-free rate of interest.
5. Describe, as a computational tool, the approach using state price deflators to the pricing of zero-coupon bonds and interest rate derivatives for a general one-factor diffusion model for the risk-free rate of interest.
6. Demonstrate an awareness of the Vasicek, Cox-Ingersoll-Ross and Hull-White models for the term structure of interest rates.
7. Discuss the limitations of these one-factor models and show an awareness of how these issues can be addressed.

0 Introduction

In this chapter we will look at stochastic models for the term structure of interest rates. In particular, we will focus on models that are framed in continuous time and which are arbitrage-free.

Continuous-time models are generally used for their greater tractability compared to discrete-time ones. An arbitrage-free model of the term structure is one that generates arbitrage-free bond prices.

There are two main types of models used to describe interest rates mathematically:

1. The Heath-Jarrow-Morton approach uses an Ito process to model the forward rate for an investment with a fixed maturity. We will not consider this approach here.

2. *Short-rate models* use an Ito process to model the short rate. We will look at three specific models of this type: the Vasicek model, the Cox-Ingersoll-Ross model and the Hull-White model.

Ito processes are a key feature of these models, so it might be helpful to review the topics of Brownian motion, Ito's Lemma and stochastic differential equations from earlier in the course.

Interest rate modelling is the most important topic in derivative pricing. Interest rate derivatives account for around 80% of the value of derivative contracts outstanding, mainly swaps and credit derivatives used to support the securitisation of debt portfolios. More philosophically, is the fact that at any one time there will be a multitude of contracts (bonds) written on the same underlying (an interest rate), and derivative pricing is principally concerned with pricing derivatives coherently.

The Core Reading in this chapter is adapted from the course notes written by Timothy Johnson.

1 Notation and preliminaries

1.1 Zero-coupon bonds

The multitude of traded instruments leads to the first challenge in interest rate modelling: the multitude of definitions of interest rates.

Modelling interest rates is more complicated than modelling share prices because interest rates depend not only on the current time (which we will denote by t), but also on the term of the investment. For example, an investor with a 10-year bond will normally earn a different rate of interest than an investor with a 5-year bond.

The basic debt instrument is the discount bond (or, equivalently, the zero-coupon bond). This is an asset that will pay one unit of currency at time T and is traded at time $t<T$. If the interest rate, R, is constant between t and T then we can say that the price of the discount bond purchased at t and maturing at T is given by $P(t,T)$ where:

$$P(t,T)=\frac{1}{(1+R(t,T))^{(T-t)}}$$

The spot rate $R(t,T)$ is the *effective rate* of interest applicable over the period from time t to time T that is implied by the market prices at time t.

Observe that $P(T,T)=1$ and for all $t<T$, $P(t,T)<P(T,T)=1$. We define $\tau=T-t$ in what follows.

The discrete bond yield calculated from discount bond prices is:

$$R(t,t+\tau)=\frac{1}{P(t,t+\tau)^{1/\tau}}-1$$

Alternatively we could write this as:

$$R(t,T)=\frac{1}{P(t,T)^{1/(T-t)}}-1$$

If a 'spot' rate is paid m times a year, then:

$$\frac{1}{P(0,n)}=\left(1+\frac{R}{m}\right)^{nm}$$

The limit as $m\to\infty$ is a continuously compounded rate, $r(t,T)$ ('force of interest'), such that:

$$e^{-r(t,T)\tau}=P(t,T)=\frac{1}{(1+R(t,T))^{\tau}}$$

The continuously compounded bond yield is calculated as:

$$r(t,T)=-\frac{\ln P(t,T)}{\tau}$$

The spot rate $r(t,T)$ is the *continuously* compounded rate of interest applicable over the period from time t to time T that is implied by the market prices at time t.

1.2 Yield curves

Fixing $t = 0$ and plotting yield, $R(0,T)$ or $r(0,T)$, against maturity, T, gives the yield curve which gives information on the *term structure*, how interest rates for different maturities are related. Typically, the yield curve increases with maturity, reflecting uncertainty about far-future rates. However, if current rates are unusually high, the yield curve can be downward sloping, and is *inverted*.

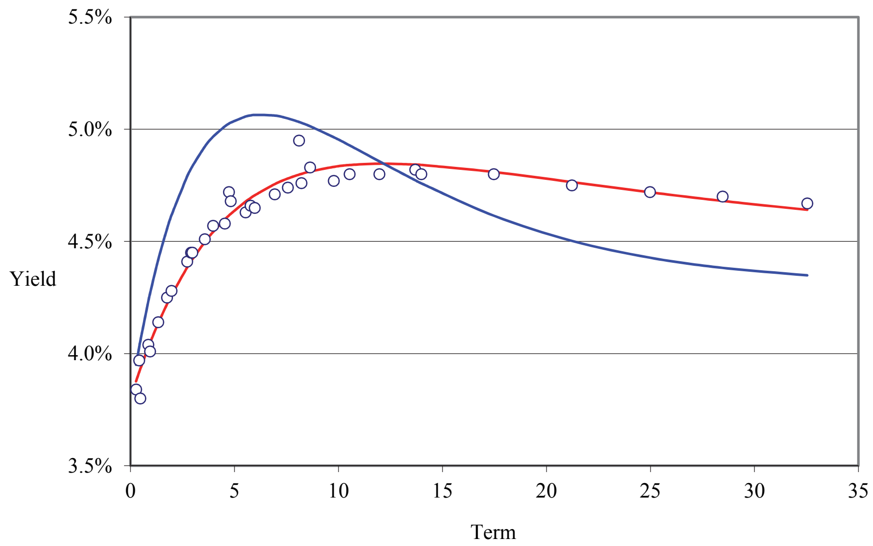
There are various theories explaining the shape of the yield curve. The *expectations theory* argues that the long-term rate is determined purely by current and future expected short-term rates, so that the expected final value of investing in a sequence of short-term bonds equals the final value of wealth from investing in long-term bonds.

The *market segmentation* theory argues that different agents in the market have different objectives: pension funds determine longer-term rates, market makers determine short-term rates, and businesses determine medium-term rates, which are all determined by the supply and demand of debt for these different market segments.

The *liquidity preference theory* argues that lenders want to lend short term while borrowers wish to borrow long term, and so forward rates are higher than expected future zero rates (and yield curves are upward sloping).

Zero rate is another name for the spot rate.

The figure below shows the yield curve for the UK Government bond market on 31 December 2003. The 'term' plotted on the x-axis corresponds to $T - t$.



UK yield curves 31 December 2003

The circles show the remaining term and the gross redemption yield for each of the available bonds. These approximate to the spot rates $r(t,T)$ for zero-coupon bonds with corresponding terms. A mathematical curve has been fitted to these points. At this particular time the yield curve had a humped shape.

The other curve shown is the forward rate curve $f(0,t,T)$, which is defined in the next section.

1.3 Short rate and forward rates

The *short* or *instantaneous* rate, $r(t)$, is the interest rate charged today for a very short period (*ie* overnight). This is defined (equivalently) as:

$$r(t) = r(t,t+\delta) \approx R(t,t+\delta)$$

where δ is a small positive quantity. So the short rate $r(t)$ is the force of interest that applies in the market at time t for an infinitesimally small period of time δ. Using the relationship developed in the opening section we have:

$$r(t) = -\frac{\partial}{\partial \delta} \ln P(t,t+\delta)$$

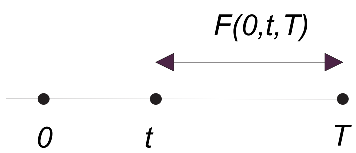
The short rate is often the basis of some interest rate models; however, it will not generate, on its own, discount bond prices.

The forward rate, $F(0,t,T)$ if discretely compounded and $f(0,t,T)$ if continuously compounded, relates to a loan starting at time *t*, for the fixed forward rate, the forward rate, repaid at maturity, *T*. It involves three times, the time at which the forward rate agreement is entered into (typically 0), the start time of the forward rate, *t* and the maturity of the forward rate agreement, *T*.

The law of one-price/the no-arbitrage principle, implies:

$$F(0,t,T) = \left(\frac{P(0,t)}{P(0,T)}\right)^{\frac{1}{T-t}} - 1$$

At time t we can consider the market prices of two investments, one maturing at time t and one at a later time T. These two prices will imply a certain rate of interest applicable between time t and time T. This is $F(0,t,T)$. In an arbitrage-free market, it represents the rate of interest at which we can agree at time 0 to borrow or lend over the period from t to T.

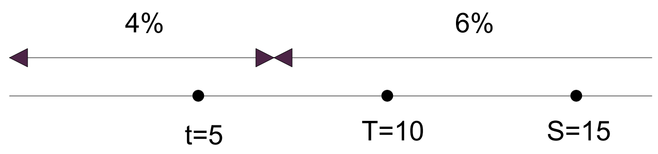


Question

Suppose that the current time corresponds to $t = 5$ and that the force of interest has been a constant 4% *pa* over the last 5 years. Suppose also that the force of interest implied by current market prices is a constant 4% *pa* for the next 2 years and a constant 6% *pa* thereafter.

If $T = 10$ and $S = 15$, write down or calculate each of the four quantities $P(t,T)$, $r(t)$, $f(t,T,S)$ and $r(t,T)$ using the notation above.

Solution



- $P(5,10) = e^{-(2\times 0.04 + 3\times 0.06)} = e^{-0.26} = 0.771$
- $r(5) = 0.04$
- $f(5,10,15) = 0.06$
- $r(5,10) = \dfrac{1}{5}(2\times 0.04 + 3\times 0.06) = 0.052$

For continuously compounded forward rates:

$$\boldsymbol{f(0,t,T) = \frac{r(0,T)T - r(0,t)t}{T-t}}$$

$$\boldsymbol{= r(0,t) + \frac{(r(0,T) - r(0,t))T}{T-t}}$$

The forward rate is related to the zero-coupon bond price as follows:

$$f(0,t,T) = \frac{1}{T-t}\log\frac{P(0,t)}{P(0,T)} \quad \text{for } t < T$$

Question

Derive this relationship.

Solution

One way to see this is to say that the price for a bond purchased at time 0 and maturing at time T should equal the price for a similar bond maturing at the earlier time t but discounted for the extra period $T-t$ at the forward rate applicable over that period. In other words:

$$P(0,T)=P(0,t)\exp\left[-f(0,t,T)(T-t)\right]$$

Rearranging this equation gives the required result:

$$f(0,t,T)=\frac{1}{T-t}\log\frac{P(0,t)}{P(0,T)}$$

Alternatively we could use the Core Reading results from above:

$$\begin{aligned}
f(0,t,T)&=\frac{r(0,T)T-r(0,t)t}{T-t}\\
&=\frac{\frac{-\log P(0,T)\times T}{T}-\frac{-\log P(0,t)\times t}{t}}{T-t}\\
&=\frac{\log P(0,t)-\log P(0,T)}{T-t}\\
&=\frac{1}{T-t}\log\frac{P(0,t)}{P(0,T)}
\end{aligned}$$

The instantaneous rate $f(0,T)$ is defined as $\lim_{t\to T} f(0,t,T)$.

From the Core Reading above we have that:

$$f(0,t,T)=r(0,t)+\frac{(r(0,T)-r(0,t))T}{T-t}$$

So, in the limit $t\to T$, we get the instantaneous forward rate:

$$\begin{aligned}
\boldsymbol{f(0,T)}&=\boldsymbol{r(0,T)+T\frac{\partial r(0,T)}{\partial T}}\\
&=\boldsymbol{-\frac{\partial}{\partial T}\ln(P(0,T))}
\end{aligned}$$

Question

Show that $\dfrac{\partial}{\partial T}r(0,T)=\dfrac{1}{T}\left(-\dfrac{\partial}{\partial T}\log P(0,T)-r(0,T)\right)$.

Solution

By applying the product rule to the equation $r(0,T) = -\frac{1}{T}\log P(0,T)$, we have:

$$\begin{aligned}
\frac{\partial}{\partial T} r(0,T) &= \frac{\partial}{\partial T}\left(-\frac{1}{T}\log P(0,T)\right) \\
&= -\frac{1}{T}\frac{\partial}{\partial T}(\log P(0,T)) + \log P(0,T)\frac{\partial}{\partial T}\left(-\frac{1}{T}\right) \\
&= -\frac{1}{T}\frac{\partial}{\partial T}(\log P(0,T)) + \log P(0,T)\left(\frac{1}{T^2}\right) \\
&= \frac{1}{T}\left(-\frac{\partial}{\partial T}\log P(0,T) - r(0,T)\right)
\end{aligned}$$

So $f(0,T)$, the instantaneous forward rate, is the force of interest at future time T implied by the current market prices at time 0. Then the short rate $r(T)$ is given by $r(T) = f(T,T)$.

We can generalise $\lim_{t \to T} f(0,t,T) = f(0,T) = -\frac{\partial}{\partial T}\ln(P(0,T))$ to give:

$$f(t,T,T) = f(t,T) = -\frac{\partial}{\partial T}\ln(P(t,T))$$

Hence, the fundamental theorem of calculus tells us that:

$$\begin{aligned}
\boldsymbol{P(t,T) = \exp\left\{-\int_t^T f(t,u,u)du\right\}} \\
= \exp\left\{-\int_t^T f(t,u)du\right\}
\end{aligned}$$

We can also deduce the following relationship, which shows that the spot rate is an average of the forward rates:

$$r(t,T) = -\frac{1}{T-t}\log P(t,T) = \frac{1}{T-t}\int_t^T f(t,u,u)du$$

Question

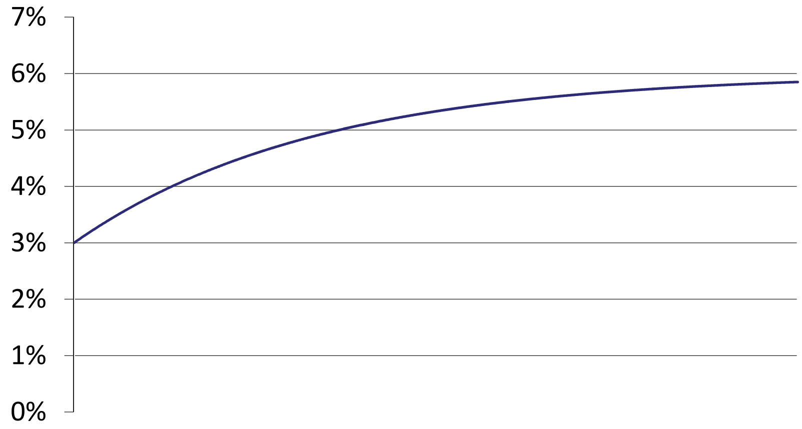
Under one particular term structure model:

$$f(t,T) = 0.03e^{-0.1(T-t)} + 0.06(1-e^{-0.1(T-t)}).$$

Sketch a graph of $f(t,T)$ as a function of T , and derive expressions for $P(t,T)$ and $r(t,T)$.

Solution

The factor $e^{-0.1(T-t)}$ equals 1 when $T = t$, but then decreases exponentially to zero as $T \to \infty$. So $f(t,T)$ is a weighted average of 0.03 and 0.06, and its graph will increase from 0.03 (*ie* a force of interest of 3%) to 0.06.



We can find $P(t,T)$ from the relationship $P(t,T) = \exp\left[-\int_t^T f(t,u)du\right]$:

$$
\begin{aligned}
P(t,T) &= \exp\left[-\int_t^T f(t,u)du\right] \\
&= \exp\left[-\int_t^T \left\{0.03e^{-0.1(u-t)} + 0.06(1-e^{-0.1(u-t)})\right\}du\right] \\
&= \exp\left[-\int_t^T \left\{0.06 - 0.03e^{-0.1(u-t)}\right\}du\right] \\
&= \exp\left(-\left[0.06u + 0.3e^{-0.1(u-t)}\right]_t^T\right) \\
&= \exp\left(-0.06(T-t) - 0.3e^{-0.1(T-t)} + 0.3\right)
\end{aligned}
$$

We can then find $r(t,T)$ from the relationship $r(t,T) = \dfrac{-1}{T-t}\log P(t,T)$:

$$
\begin{aligned}
r(t,T) &= \frac{-1}{T-t}\log P(t,T) \\
&= \frac{-1}{T-t}\left(-0.06(T-t) - 0.3e^{-0.1(T-t)} + 0.3\right) \\
&= 0.06 - 0.3\left[\frac{1-e^{-0.1(T-t)}}{T-t}\right]
\end{aligned}
$$

2 Desirable characteristics of a term structure model

Question

What do you think term structure models are used for?

Solution

The main uses of term structure (interest rate) models are:

- by bond traders looking to identify and exploit price inconsistencies
- for calculating the price of interest rate derivatives
- by investors with a portfolio involving bonds or loans who want to set up a hedged position
- for asset-liability modelling.

Equilibrium models **start with a theory about the economy, such that interest rates revert to some long-run average, are positive or their volatility is constant or geometric. Based on the model for (typically) the short rate, the implications for the pricing of assets is worked out. Examples of equilibrium models are Rendleman and Bartter, Vasicek and Cox-Ingersoll-Ross.**

Being based on 'economic fundamentals', equilibrium models rarely reproduce observed term structures. This is unsatisfactory.

No-arbitrage models **use the term structure as an input and are formulated to adhere to the no-arbitrage principle. An example of a no-arbitrage model is the Hull-White (one- and two-factor).**

The implication here is that the Vasicek and the Cox-Ingersoll-Ross models permit arbitrage opportunities. This is not actually true; both of these models result in no-arbitrage bond price formulae, even though their underlying construction principles are based in economic theory.

We will now discuss characteristics of a term structure model that are regarded as desirable features.

- The model should be arbitrage free.

  In very limited circumstances this is not essential, but in the majority of modern actuarial applications, it is essential. Most obviously, anything involving dynamic hedging would immediately identify and exploit any arbitrage opportunities.

  The markets for government bonds and interest rate derivatives are generally assumed to be pretty much arbitrage-free in practice.

- Interest rates should ideally be positive. Banks have to offer investors a positive return to prevent them from withdrawing paper cash and putting it 'under the bed'. This might be impractical for a large life office or pension fund but, nevertheless, it typically holds in practice.

  Some term structure models do allow interest rates to go negative.

  One such example is the Vasicek model we will see later in this chapter.

  Whether or not this is a problem depends on the probability of negative interest rates within the timescale of the problem in hand and their likely magnitude if they can go negative. It also depends on the economy being modelled, as negative interest rates have been seen in some countries.

- $r(t)$ and other interest rates should exhibit some form of mean-reverting behaviour.

  Again this is because the empirical evidence suggests that interest rates do tend to mean revert in practice.

  This might not be particularly strong mean reversion but it is essential for many actuarial applications where the time horizon of a problem might be very long.

- How easy is it to calculate the prices of bonds and certain derivative contracts?

  This is a computational issue. It is no good in a modelling exercise to have a wonderful model if it is impossible to perform pricing or hedging calculations within a reasonable amount of time.

  This is because we need to act quickly to identify any potential arbitrage opportunities or to rebalance a hedged position.

  Thus we aim for models that either give rise to simple formulae for bond and option prices, or that make it straightforward to compute prices using numerical techniques.

- Does the model produce realistic dynamics?

  For example, can it reproduce features that are similar to what we have seen in the past with reasonable probability? Does it give rise to a full range of plausible yield curves, *ie* upward-sloping, downward-sloping and humped?

- Does the model, with appropriate parameter estimates, fit historical interest rate data adequately?

- Can the model be calibrated easily to current market data?

  If so, is this calibration perfect or just a good approximation? This is an important point when we are attempting to establish the fair value of liabilities. If the model cannot fit observed yield curves accurately then it has no chance of providing us with a reliable fair value for a set of liabilities.

- Is the model flexible enough to cope properly with a range of derivative contracts?

3 Models for the term structure of interest rates

3.1 The risk-neutral approach to pricing

We will assume that the short rate is driven by an Ito diffusion:

$$dr_t = \mu(t,r_t)dt + \sigma(t,r_t)d\tilde{W}_t$$

where:

- $\mu(t,r_t)$ is the drift parameter
- $\sigma(t,r_t)$ is the volatility parameter
- $\tilde{W}_t$ **is a Wiener process under the martingale measure.**

In actuarial work we are used to assuming a *fixed* rate of interest in calculations. However, in this chapter we are considering *stochastic* models where future interest rates behave randomly. This means that we need to specify which probability measure we are using.

If we are to have a model that is arbitrage-free then we need to consider the prices of tradeable assets, with the most natural of these being the zero-coupon bond prices $P(t,T)$.

Modelling the short rate $r(t)$ does not tell us directly about the prices of the assets traded in the market. To see whether arbitrage opportunities exist or not, we need to examine these prices.

We can use an argument similar to the derivation of the Black-Scholes model using the martingale approach to demonstrate that:

$$P(t,T) = E_Q\left[\exp\left(-\int_t^T r_u du\right) \middle| r_t\right]$$

where Q is called the risk-neutral measure.

Since we are considering Markov models, any information about the value of the short rate before time t is irrelevant, and so we can replace r_t with its filtration F_t. We can then write this equation in the equivalent form:

$$P(t,T) = E_Q\left[\exp\left(-\int_t^T r_u du\right) \times 1 \middle| F_t\right]$$

Since the payoff from the bond at maturity is $P(T,T)=1$, we can see that this equation is analogous to the valuation formula $V_t = E_Q\left[e^{-r(T-t)}X \middle| F_t\right]$, which we met in an earlier chapter for derivatives based on shares.

If we had a derivative whose payoff X was dependent on the future value of the bond, then its value at time t would be:

$$V_t = E_Q\left[\exp\left(-\int_t^T r_u du\right)X \,\middle|\, F_t\right]$$

This formula is interesting because it tells us that, if we assume that the market is arbitrage-free, we can work out the value of the bond at the current time t based on the current short rate r_t and the possible future values of r_u for $t < u \le T$. We don't need to know the entire term structure, *ie* the values of $f(t,u)$ for $t < u \le T$, at the current time.

We define the bank, or money-market, account process as:

$$dB_t = r_t B_t dt\,,\ B_0 = 1$$

and:

$$B_t = \exp\left\{\int_0^t r_u du\right\}$$

According to the standard theory, all discounted assets must be martingales under the martingale measure, or:

$$\frac{P(t,T)}{B_t} = E_Q\left[\frac{1}{B_T}\middle| F_t\right]$$

Since B_t is known at time *t*, then we can re-write this as:

$$P(t,T) = E_Q\left[\frac{B_t}{B_T}\middle| F_t\right]$$

The martingale measure (*ie* the risk-neutral probability measure) is chosen so that this relationship holds *by definition*. It is the set of probabilities such that the expected future value of the payout (which is guaranteed to be $P(T,T)=1$) discounted from time *T* to time *t* (*ie* B_t / B_T) is equal to the value of the contract at time *t*, $P(t,T)$.

Then we have:

$$P(t,T) = E_Q\left[\exp\left\{-\int_t^T r_u du\right\}\middle| F_t\right]$$

Compare with:

$$P(t,T) = \exp\left\{-\int_t^T f(t,u)du\right\}$$

from earlier.

The difference between $P(t,T)$ and B_t is captured by noting:

$$P(t,T) = \exp\left\{-\int_t^T f(t,u)du\right\} \qquad \text{while} \qquad B_t = \exp\left\{\int_0^t r_u du\right\}$$

So the value of B_t, the cash account at time *t,* can be calculated using the known past short rates. In contrast, the price of the bond is the expectation of unknown future values of the short rate. However, given the known bond prices at time *t*, we can derive a set of consistent forward rates.

By considering the price of the bond at time $t=0$ we can see that:

$$P(0,t) = \exp\left\{-\int_0^t f(0,u)du\right\} \qquad \text{while} \qquad B_t = \exp\left\{\int_0^t r_u du\right\} \qquad \text{and} \qquad f(u,u) = r_u$$

We now assume that the discount bond price, $P(t,T)$, is some deterministic function of the short-rate process, r_t:

$$P(t,T) = g(t,r_t)$$

By using the stochastic product rule from the stochastic calculus chapter we have:

$$\begin{aligned}
d\left(\frac{P(t,T)}{B_t}\right) &= d\left(g(t,r_t)B_t^{-1}\right) \\
&= d\left(g(t,r_t)\right)B_t^{-1} + g(t,r_t)d\left(B_t^{-1}\right) \\
&= d\left(g(t,r_t)\right)B_t^{-1} + g(t,r_t)d\left(\exp\left\{-\int_0^t r_u du\right\}\right) \\
&= d\left(g(t,r_t)\right)B_t^{-1} - g(t,r_t)r_t\left(\exp\left\{-\int_0^t r_u du\right\}\right)dt \\
&= B_t^{-1}\left[d\left(g(t,r_t)\right) - r_t g(t,r_t)dt\right]
\end{aligned}$$

By applying Ito's Lemma to the function $g(t,r_t)$ we have:

$$\begin{aligned}
d\left(\frac{P(t,T)}{B_t}\right) = {} & B_t^{-1}\left(\frac{\partial g(t,r_t)}{\partial t} + \frac{\partial g(t,r_t)}{\partial r_t}\mu(t,r_t) + \frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}\sigma^2(t,r_t) - r_t g(t,r_t)\right)dt \\
& + B_t^{-1}\frac{\partial g(t,r_t)}{\partial r_t}\sigma(t,r_t)d\tilde{W}_t
\end{aligned}$$

Question

Verify the above result by using Taylor's theorem.

Solution

By Taylor's theorem in two variables we have:

$$d\left(\frac{P(t,T)}{B_t}\right)=B_t^{-1}\left(\frac{\partial g(t,r_t)}{\partial t}dt+\frac{\partial g(t,r_t)}{\partial r_t}dr_t+\frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}(dr_t)^2\right)-B_t^{-1}r_t g(t,r_t)dt$$

But $dr_t=\mu(t,r_t)dt+\sigma(t,r_t)d\tilde{W}_t$ so $(dr_t)^2=\sigma^2(t,r_t)dt$ and:

$$d\left(\frac{P(t,T)}{B_t}\right)=B_t^{-1}\left(\frac{\partial g(t,r_t)}{\partial t}dt+\frac{\partial g(t,r_t)}{\partial r_t}\left(\mu(t,r_t)dt+\sigma(t,r_t)d\tilde{W}_t\right)+\frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}\sigma^2(t,r_t)dt\right)$$

$$-B_t^{-1}r_t g(t,r_t)dt$$

$$=B_t^{-1}\left(\frac{\partial g(t,r_t)}{\partial t}+\frac{\partial g(t,r_t)}{\partial r_t}\mu(t,r_t)+\frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}\sigma^2(t,r_t)-r_t g(t,r_t)\right)dt$$

$$+B_t^{-1}\frac{\partial g(t,r_t)}{\partial r_t}\sigma(t,r_t)d\tilde{W}_t$$

In order for this to be a martingale, we require that:

$$\frac{\partial g(t,r_t)}{\partial t}+\frac{\partial g(t,r_t)}{\partial r_t}\mu(t,r_t)+\frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}\sigma^2(t,r_t)-r_t g(t,r_t)=0$$

with the boundary condition that $P(T,T)=g(T,r_T)=1$ for all r_T.

Question

Why does the partial differential equation above need to equal zero in order for $\dfrac{P(t,T)}{B_t}$ to be a martingale?

Solution

Martingale processes have zero drift. In order for $\dfrac{P(t,T)}{B_t}$ to have zero drift the coefficient of the dt term in the stochastic differential equation must be zero.

3.2 The Vasicek model (1977)

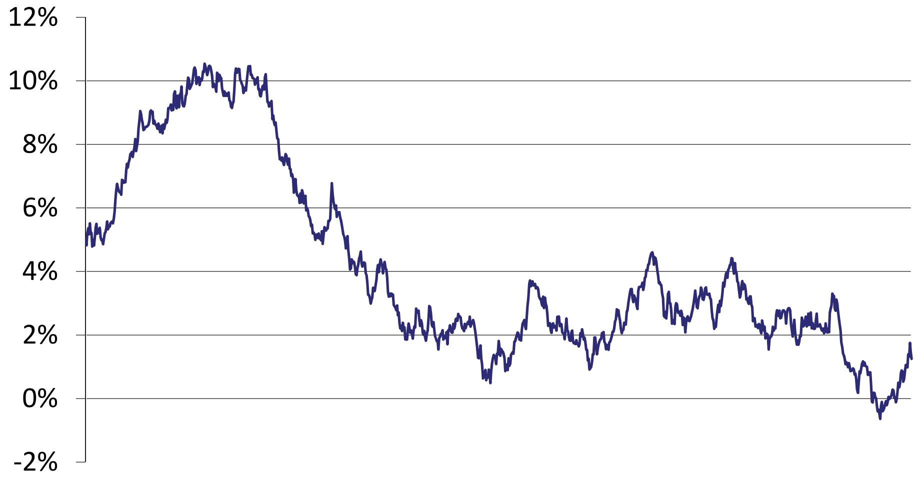
Vasicek assumes that:

$$dr_t = \alpha(\mu - r_t)dt + \sigma d\tilde{W}_t$$

for constants $\alpha > 0$, μ and, σ.

Here μ represents the 'mean' level of the short rate. If the short rate grows (driven by the stochastic term) the drift becomes negative, pulling the rate back to μ. The speed of the 'reversion' is determined by α. If α is high, the reversion will be very quick.

The graph below show a simulation of this process based on the parameter values $\alpha = 0.1$, $\mu = 0.06$ and $\sigma = 0.02$.



Example simulation of short rate from the Vasicek model

This yields the partial differential equation:

$$\frac{\partial g(t,r_t)}{\partial t} + \frac{\partial g(t,r_t)}{\partial r_t}\alpha(\mu - r_t) + \frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}\sigma^2 - r_t g(t,r_t) = 0$$

by making the substitution $\mu(t,r_t) = \alpha(\mu - r_t)$ and letting $\sigma(t,r_t)$ be constant.

To establish the form of $g(t,r_t) = P(t,T)$, recall that for the Ornstein-Uhlenbeck process:

$$r_t = \mu + (r_0 - \mu)e^{-\alpha t} + \sigma\int_0^t e^{-\alpha(t-u)}d\tilde{W}_u$$

and so, integrating again:

$$\int_0^T r_s ds = \mu T + \frac{1}{\alpha}(r_0 - \mu)\left(1 - e^{-\alpha T}\right) + \frac{\sigma}{\alpha}\int_0^T \left(1 - e^{-\alpha(T-u)}\right)d\tilde{W}_u$$

This implies that $\int_0^T r_s ds$ is normally distributed with conditional mean:

$$E\left[\int_0^T r_s ds\right] = \mu T + \frac{1}{\alpha}(r_0 - \mu)\left(1 - e^{-\alpha T}\right)$$

Using Ito isometry:

$$Var_Q\left[\int_0^T \nu(X_t,t)dW_t\right] = \int_0^T E\left[\nu^2(X_t,t)\right]dt$$

we can see that:

$$Var_Q\left[\int_0^T r_s ds\right] = \frac{\sigma^2}{\alpha^2}\left(T - \frac{2}{\alpha}(1 - e^{-\alpha T}) + \frac{1}{2\alpha}(1 - e^{-2\alpha T})\right)$$

Since $\int_0^T r_s ds$ is normally distributed, using the moment generating function of a normal, $N\left(\eta,\nu^2\right)$, random variable:

$$E\left[e^{tX}\right] = e^{\eta t + \frac{1}{2}\nu^2 t^2}$$

by putting $t = -1$, we can say that:

$$P(0,T) = E_Q\left[\exp\left(-\int_0^T r_s ds\right)\right]$$

$$= \exp\left(-\left(\mu T + \frac{1}{\alpha}(r_0 - \mu)\left(1 - e^{-\alpha T}\right)\right) + \frac{1}{2}\left(\frac{\sigma^2}{\alpha^2}\left(T - \frac{2}{\alpha}(1 - e^{-\alpha T}) + \frac{1}{2\alpha}(1 - e^{-2\alpha T})\right)\right)\right)$$

In general, by setting:

$$b(t,T) = \frac{1}{\alpha}\left(1 - e^{-\alpha(T-t)}\right)$$

and:

$$a(t,T) = \left(\mu - \frac{\sigma^2}{2\alpha^2}\right)\left(b(t,T) - T + t\right) - \frac{\sigma^2}{4\alpha}b^2(t,T)$$

we have:

$$P(t,T) = e^{a(t,T) - b(t,T)r_t}$$

By defining $\tau = T - t$, then equivalently we have:

$$P(t,T) = e^{a(\tau) - b(\tau) r_t}$$

where:

$$b(\tau) = \frac{1}{\alpha}\left(1 - e^{-\alpha\tau}\right)$$

and:

$$a(\tau) = \left(\mu - \frac{\sigma^2}{2\alpha^2}\right)\left(b(\tau) - \tau\right) - \frac{\sigma^2}{4\alpha} b^2(\tau)$$

Question

Show that the instantaneous forward rate for the Vasicek model can be expressed as:

$$f(t,T) = r(t)e^{-\alpha\tau} + \left(\mu - \frac{\sigma^2}{2\alpha^2}\right)(1 - e^{-\alpha\tau}) + \frac{\sigma^2}{2\alpha^2}(1 - e^{-\alpha\tau})e^{-\alpha\tau}$$

where $\tau = T - t$.

Solution

We have a formula for $P(t,T)$ under this model. So the instantaneous forward rate can be derived from this using the relationship:

$$f(t,T) = -\frac{\partial}{\partial T} \log P(t,T)$$

By use of the chain rule, and noting that $\frac{\partial \tau}{\partial T} = 1$, this gives:

$$\begin{aligned} f(t,T) &= -\frac{\partial}{\partial T}[a(\tau) - b(\tau)r(t)] \\ &= -\frac{\partial \tau}{\partial \tau}\frac{\partial}{\partial T}[a(\tau) - b(\tau)r(t)] \\ &= -\frac{\partial \tau}{\partial T}\frac{\partial}{\partial \tau}[a(\tau) - b(\tau)r(t)] = -a'(\tau) + b'(\tau)r(t) \end{aligned}$$

From the definitions of $b(\tau)$ and $a(\tau)$, we find that:

$$b'(\tau) = \frac{d}{d\tau}\left(\frac{1-e^{-\alpha\tau}}{\alpha}\right) = e^{-\alpha\tau}$$

and:

$$a'(\tau) = \frac{d}{d\tau}\left[(b(\tau)-\tau)\left(\mu - \frac{\sigma^2}{2\alpha^2}\right) - \frac{\sigma^2}{4\alpha}b(\tau)^2\right]$$

$$= (b'(\tau)-1)\left(\mu - \frac{\sigma^2}{2\alpha^2}\right) - \frac{\sigma^2}{4\alpha} \times 2b(\tau)b'(\tau)$$

$$= (e^{-\alpha\tau}-1)\left(\mu - \frac{\sigma^2}{2\alpha^2}\right) - \frac{\sigma^2}{2\alpha}\left(\frac{1-e^{-\alpha\tau}}{\alpha}\right)e^{-\alpha\tau}$$

$$= -(1-e^{-\alpha\tau})\left(\mu - \frac{\sigma^2}{2\alpha^2}\right) - \frac{\sigma^2}{2\alpha^2}(1-e^{-\alpha\tau})e^{-\alpha\tau}$$

Substituting these expressions into the general formula for $f(t,T)$ gives the required answer.

Question

Write down an expression in terms of the model parameters for the long rate, *ie* the instantaneous forward rate corresponding to $T-t=\infty$, according to the Vasicek model.

Solution

Letting $T\to\infty$ (and hence $\tau\to\infty$) in the equation for $f(t,T)$ gives:

$$f(t,\infty) = r(t)\times 0 + \left(\mu - \frac{\sigma^2}{2\alpha^2}\right)(1-0) + \frac{\sigma^2}{2\alpha^2}(1-0)\times 0 = \mu - \frac{\sigma^2}{2\alpha^2}$$

The curves shown on the graph of gilt yields in Section 1 were fitted using a Vasicek model with parameter values $\alpha = 0.131$, $\mu = 0.083$ and $\sigma = 0.037$.

Question

'The particular model used for the graph implies that interest rates are mean-reverting to the value $\mu = 0.083$.'

True or false?

Solution

The dynamics of $r(t)$ for this particular Vasicek model are:

$$dr(t) = -0.131\,[r(t) - 0.083]dt + 0.037d\tilde{W}(t)$$

under the *risk-neutral* probability measure Q. Under this measure $\tilde{W}(t)$ is standard Brownian motion and therefore has zero drift and the process mean-reverts to the value 0.083.

However, under the *real-world* probability measure P, $\tilde{W}(t)$ would have non-zero drift and the process will mean-revert to a different value. In fact, although we will not prove it here, the long-term rate in the real world can be found from the formula derived in the previous question, namely:

$$\mu - \frac{\sigma^2}{2\alpha^2} = 0.0431 \quad ie\ \ 4.31\%$$

3.3 The Cox-Ingersoll-Ross (CIR) model (1985)

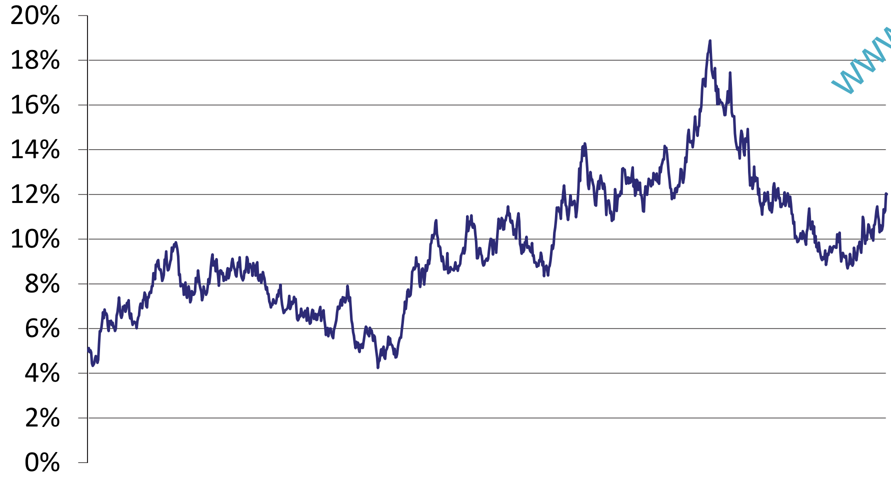
In Vasicek's model (and Hull-White, below) interest rates are not strictly positive. This assumption is not ideal for a short-rate model. CIR use the Feller, or square root mean reverting process which is positive (it can instantaneously touch 0 but immediately re-bounds):

$$\boldsymbol{dr_t = \alpha(\mu - r_t)dt + \sigma\sqrt{r_t}\,d\tilde{W}_t}$$

for constants $\alpha > 0$, $\mu > 0$ and, σ.

We can see that the form of the drift of r_t is the same as for the Vasicek model. The critical difference between the two models occurs in the volatility, which is increasing in line with the square root of r_t. Since this diminishes to zero as r_t approaches zero, and provided σ^2 is not too large (r_t will never hit zero provided $\sigma^2 \le 2\alpha\mu$), we can guarantee that r_t will not hit zero. Consequently all other interest rates will also remain strictly positive.

The graph below shows a simulation of this process based on the parameter values $\alpha = 0.1$, $\mu = 0.06$ and $\sigma = 0.1$.



Simulation from Cox-Ingersoll-Ross model

The associated PDE is:

$$\frac{\partial g(t,r_t)}{\partial t}+\frac{\partial g(t,r_t)}{\partial r_t}\alpha(\mu-r_t)+\frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}\sigma^2 r_t - r_t g(t,r_t)=0$$

The only difference between this and the Vasicek PDE is the inclusion of an r_t in the second derivative term.

Again, $P(t,T)=e^{a(t,T)-b(t,T)r_t}$ **, with:**

$$b(\tau)=\frac{2(e^{\gamma\tau}-1)}{(\gamma+\alpha)(e^{\gamma\tau}-1)+2\gamma}$$

and:

$$a(\tau)=\frac{2\alpha\mu}{\sigma^2}\ln\left(\frac{2\gamma e^{\frac{1}{2}(\gamma+\alpha)\tau}}{(\gamma+\alpha)(e^{\gamma\tau}-1)+2\gamma}\right)$$

$$=\frac{2\alpha\mu}{\sigma^2}\ln\left(b(\tau)\frac{\gamma e^{\frac{1}{2}(\gamma+\alpha)\tau}}{(e^{\gamma\tau}-1)}\right)$$

where:

$$\gamma=\sqrt{\alpha^2+2\sigma^2}$$

It turns out that these values for *a* and *b* are not that different from those in Vasicek's model.

The distribution of r_t is given by a 'non-central chi-squared' distribution. This is a 'fat-tailed' distribution.

If $X_1, X_2, \ldots, X_n$ are independent random variables, each with a $N(0,1)$ distribution, then $Y = X_1^2 + X_2^2 + \cdots + X_n^2$ has a chi-square distribution with n degrees of freedom.

If $X_1, X_2, \ldots, X_n$ are independent random variables and $X_i \sim N(d_i, 1)$, then $Y_d = X_1^2 + X_2^2 + \cdots + X_n^2$ is said to have a *non-central* chi-squared distribution with n degrees of freedom and non-centrality parameter $d = \sum_{i=1}^{n} d_i^2$.

So the non-central chi-squared distribution can be thought of as a lopsided version of the ordinary chi-square distribution.

Question

What is the mean of the non-central chi-squared distribution with n degrees of freedom and non-centrality parameter d ?

Solution

Since $X_i \sim N(d_i, 1)$, we find that:

$$E[X_i^2] = Var(X_i) + [E(X_i)]^2 = 1 + d_i^2$$

It follows that:

$$E[Y_d] = E[X_1^2 + X_2^2 + \cdots + X_n^2] = \sum_{i=1}^{n} (1 + d_i^2) = n + d$$

3.4 Vasicek and CIR yield curves

Recall:

$$\boldsymbol{r(t,T) = -\frac{\ln P(t,T)}{T-t}}$$

and:

$$P(t,T) = e^{a(t,T) - b(t,T) r_t}$$

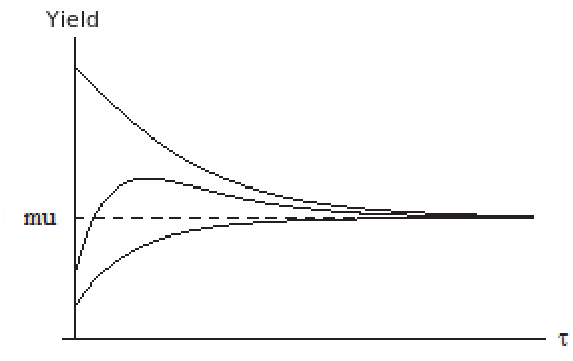
so we have:

$$\boldsymbol{r(t,T) = -\frac{a(t,T) - b(t,T) r_t}{T-t}}$$

Alternatively:

$$r(\tau) = -\frac{a(\tau) - b(\tau) r_t}{\tau}$$

Then the yield curves coming out of the Vasicek model are of three (related) types:



These yield curves are generally described as being:

- normal, with short-term yields being lower than long-term yields
- inverted, with short-term yields being higher than long-term yields
- humped, *ie* a yield curve with a turning point.

Recall that the duration, *D*, of an asset, whose interest rate dependent price is given by *B*, is defined by:

$$\frac{\Delta B}{B} = -D\delta y$$

where *y* is the instrument's yield. In the context of interest rate models, this is equivalent to:

$$\frac{\partial P(t,T)}{\partial r_t} = -DP(t,T) = -b(t,T)P(t,T)$$

and there is a connection between this duration D, and the function *b*.

The relationship above comes from the fact that:

$$\frac{\Delta B}{B} = -D\delta y \quad \Leftrightarrow \quad \frac{\Delta B}{\delta y} = -BD$$

and $\frac{\Delta B}{\delta y}$ is the rate at which the bond price changes with respect to a change in its yield, *ie* $\frac{\partial P(t,T)}{\partial r_t}$.

As $P(t,T) = e^{a(t,T) - b(t,T)r_t}$, then:

$$\frac{\partial P(t,T)}{\partial r_t} = \frac{\partial}{\partial r_t}\left(e^{a(t,T)-b(t,T)r_t}\right) = -b(t,T)e^{a(t,T)-b(t,T)r_t} = -b(t,T)P(t,T)$$

3.5 The Hull-White model (1990)

The Hull-White model is an extension of Vasicek where the mean-reversion level, μ, is a deterministic function of time:

$$dr_t = \alpha(\mu(t) - r_t)dt + \sigma d\tilde{W}_t$$

for constants $\alpha > 0$ and σ.

In some representations, the parameter α is also allowed to be a function of time. This is known as the *extended Vasicek model*.

This yields a PDE similar to Vasicek, and so we can start by 'guessing' that $P(t,T) = e^{a(t,T)-b(t,T)r_t} = g(t,r_t)$ and so:

$$\frac{\partial g(t,r_t)}{\partial t} + \frac{\partial g(t,r_t)}{\partial r_t}\alpha(\mu(t) - r_t) + \frac{1}{2}\frac{\partial^2 g(t,r_t)}{\partial r_t^2}\sigma^2 - r_t g(t,r_t) = 0$$

By noting that:

$$\frac{\partial g(t,r_t)}{\partial t} = g(t,r_t)\left(\frac{\partial a(t,T)}{\partial t} - \frac{\partial b(t,T)}{\partial t}r_t\right)$$

$$\frac{\partial g(t,r_t)}{\partial r_t} = -g(t,r_t)b(t,T)$$

$$\frac{\partial^2 g(t,r_t)}{\partial r_t^2} = g(t,r_t)b^2(t,T)$$

Then the Hull-White PDE becomes:

$$g(t,r_t)\left(\frac{\partial a(t,T)}{\partial t} - \frac{\partial b(t,T)}{\partial t}r_t\right) - \alpha g(t,r_t)b(t,T)(\mu(t) - r_t) + \frac{1}{2}g(t,r_t)b^2(t,T)\sigma^2 - r_t g(t,r_t) = 0$$

$$\Rightarrow \quad g(t,r_t)\left(\frac{\partial a(t,T)}{\partial t} - \frac{\partial b(t,T)}{\partial t}r_t - \alpha b(t,T)(\mu(t) - r_t) + \frac{1}{2}b^2(t,T)\sigma^2 - r_t\right) = 0$$

$$\Rightarrow \quad g(t,r_t)\left(r_t\left(\alpha b(t,T) - \frac{\partial b(t,T)}{\partial t} - 1\right) + \frac{\partial a(t,T)}{\partial t} - \alpha b(t,T)\mu(t) + \frac{1}{2}b^2(t,T)\sigma^2\right) = 0$$

By letting $a'(t,T) = \dfrac{\partial a(t,T)}{\partial t}$ and $b'(t,T) = \dfrac{\partial b(t,T)}{\partial t}$ then we have:

$$\boldsymbol{g(t,r_t)\left(r_t\left(\alpha b(t,T) - b'(t,T) - 1\right) + a'(t,T) - \alpha b(t,T)\mu(t) + \frac{1}{2}b^2(t,T)\sigma^2\right) = 0}$$

This is essentially Vasicek but we have:

$$\boldsymbol{b(t,T)=\int_t^T \exp\left(-\int_t^s \alpha(u)du\right)ds}$$

As α has been taken as a constant in the model above we have:

$$\begin{aligned}b(t,T)&=\int_t^T \exp\left(-\int_t^s \alpha du\right)ds\\&=\int_t^T \exp\left(-\alpha(s-t)\right)ds\\&=\frac{1}{\alpha}\left(1-e^{-\alpha(T-t)}\right)\end{aligned}$$

So $b(t,T)$ is the same as for the Vasicek model.

$$\boldsymbol{a(t,T)=\int_t^T \left(-\alpha\mu(s)b(s,T)+\tfrac{1}{2}\sigma^2 b^2(s,T)\right)ds}$$

Question

Show that $a(t,T)$ and $b(t,T)$ satisfy the equation

$$g(t,r_t)\left(r_t\left(\alpha b(t,T)-b'(t,T)-1\right)+a'(t,T)-\alpha b(t,T)\mu(t)+\frac{1}{2}b^2(t,T)\sigma^2\right)=0$$

Solution

By differentiating $a(t,T)$ and $b(t,T)$ we have:

$$a'(t,T)=\frac{\partial a(t,T)}{\partial t}=\alpha\mu(t)b(t,T)-\tfrac{1}{2}\sigma^2 b^2(t,T)$$

and

$$b'(t,T)=\frac{\partial b(t,T)}{\partial t}=-e^{-\alpha(T-t)}$$

Substituting these into the given equation gives:

$$g(t,r_t)\left(r_t\left(\alpha b(t,T)-b'(t,T)-1\right)+a'(t,T)-\alpha b(t,T)\mu(t)+\tfrac{1}{2}b^2(t,T)\sigma^2\right)$$

$$=g(t,r_t)\begin{pmatrix} r_t\left(\alpha b(t,T)+e^{-\alpha(T-t)}-1\right)+\alpha\mu(t)b(t,T) \\ -\tfrac{1}{2}\sigma^2 b^2(t,T)-\alpha b(t,T)\mu(t)+\tfrac{1}{2}b^2(t,T)\sigma^2 \end{pmatrix}$$

$$=g(t,r_t)\left(r_t\times 0+0\right)$$

$$=0$$

as required.

The advantage of this model over Vasicek is that $\mu(t)$ can be chosen to reproduce (as closely as possible) the exact yield curve, rather than the restricted forms of the Vasicek model.

4 Summary of short-rate modelling

4.1 Summary of models

The properties of these models are summarised below:

Model	**Dynamics**	$r_t > 0$ **for all** t	**Distribution of** r_t
Vasicek	$dr_t = \alpha(\mu - r_t)dt + \sigma d\tilde{W}_t$	**No**	**Normal**
CIR	$dr_t = \alpha(\mu - r_t)dt + \sigma\sqrt{r_t}d\tilde{W}_t$	**Yes**	**Non-central chi-squared**
Hull-White – Vasicek	$dr_t = \alpha(\mu(t) - r_t)dt + \sigma d\tilde{W}_t$	**No**	**Normal**
Hull-White – CIR	$dr_t = \alpha(\mu(t) - r_t)dt + \sigma\sqrt{r_t}d\tilde{W}_t$	**Yes**	**Non-central chi-squared**

The last row of the table introduces an alternative form of the Hull-White model that extends the CIR model.

There are analytic solutions for $P(t,T)$ and option prices for each of these four models.

4.2 Limitations

Question

What is a one-factor term structure model?

Solution

A one-factor model is a model in which interest rates are assumed to be influenced by a single source of randomness.

Bearing in mind that the purpose of interest rate models is to price interest rate derivatives, there are some short-comings of short-rate models:

- **Single factor short-rate models mean that all maturities behave in the same way - there is no independence. This essentially means they are useless for pricing swaptions (but OK for caps/floors).**
- **There is little consistency in valuation between the models.**
- **They are difficult to calibrate.**

One-factor models, such as Vasicek and CIR, have certain limitations with which it is important to be familiar.

First, if we look at historical interest rate data we can see that changes in the prices of bonds with different terms to maturity are not perfectly correlated as one would expect to see if a one-factor model was correct. Sometimes we even see, for example, that short-dated bonds fall in price while long-dated bonds go up. Recent research has suggested that around three factors, rather than one, are required to capture most of the randomness in bonds of different durations.

Second, if we look at the long run of historical data we find that there have been sustained periods of both high and low interest rates with periods of both high and low volatility. Again, these are features which are difficult to capture without introducing more random factors into a model.

This issue is especially important for two types of problem in insurance: the pricing and hedging of long-dated insurance contracts with interest rate guarantees; and asset-liability modelling and long-term risk management.

One-factor models do, nevertheless have their place as tools for the valuation of simple liabilities with no option characteristics; or short-term, straightforward derivatives contracts.

For other problems it is appropriate to make use of models which have more than one source of randomness: so-called multi-factor models. Hull-White is not really a multi-factor model, the μ and α parameters are deterministic and aid fitting. A multi-factor version of Vasicek would involve a multidimensional driving Wiener process and possibly stochastic μ.

Question

Summarise the characteristics of the Vasicek, Cox-Ingersoll-Ross and Hull-White models.

Solution

The following table summarises the characteristics of the Vasicek, Cox-Ingersoll-Ross (CIR) and Hull-White models.

	Vasicek	CIR	Hull-White
Arbitrage-free	Yes	Yes	Yes
Positive interest rates	No	Yes	No
Mean-reverting interest rates	Yes	Yes	Yes
Easy to price bonds and derivatives	Yes	Yes(1)	Yes
Realistic dynamics	No(2)	No(2)	No(2)
Adequate fit to historical data	No	No	Yes
Easy to calibrate to current data	No	No	Yes(3)
Can price a wide range of derivatives	No(4)	No(4)	No(4)

Notes:

(1) *Although the CIR model is harder to use than the other two models, it is more tractable than models with two or more factors.*

(2) *All three models produce perfectly correlated changes in bond prices, which is inconsistent with the empirical evidence, and fail to model periods of high and low interest rates and high and low volatility.*

(3) *Whilst one-factor models are generally difficult to calibrate, the Hull-White model is easier than the other two because its time-varying mean-reversion function aids fitting.*

(4) *All three models can be used to price short-term, straightforward derivatives, but not complex derivatives.*

5 State-price deflator approach to pricing

5.1 Short-rate models

Other modellers prefer to take a different approach to model-building using *state price deflators*.

With this approach we first specify a strictly positive diffusion process $A(t)$ with SDE under P :

$$dA(t) = A(t)\left(\mu_A(t)dt + \sigma_A(t)dW(t)\right)$$

where $\mu_A(t)$ and $\sigma_A(t)$ will be stochastic.

Note that, with this approach, the dynamics of the state price deflator $A(t)$ are defined in terms of the real-world probability measure.

We define zero-coupon bond prices according to the formula

$$P(t,T) = \frac{E_P\left[A(T) \mid F_t\right]}{A(t)}$$

It can be shown, although it isn't done here, that this method gives rise to an arbitrage-free model for bond prices.

Question

Check that this formula works when $t = T$.

Solution

When $t = T$ we have:

$$P(T,T) = \frac{E_P\left[A(T) \mid F_T\right]}{A(T)}$$

The expectation here is conditional on the history of the process $A(t)$ up to time T . Consequently, there is no additional randomness in $A(T)$, which can be considered to be a known value. So we have:

$$P(T,T) = \frac{A(T)}{A(T)} = 1$$

This is exactly what we would expect, since the bond matures at time T and has a value of 1 at that time.

Question

Suppose that $A(t)$ is a deterministic process with $\mu_A(t) = -\delta$ and $\sigma_A(t) = 0$, where δ is a constant. Show that $A(t) = ke^{-\delta t}$ and evaluate $P(t,T)$.

Solution

Here the SDE for $A(t)$ becomes $dA(t) = -A(t)\delta\, dt$, which is an ordinary differential equation. This can be written in the form:

$$\frac{1}{A(t)}\frac{dA(t)}{dt} = -\delta \quad \text{or} \quad \frac{d}{dt}\log A(t) = -\delta$$

Integrating gives:

$$\log A(t) = -\delta t + \text{constant}$$

So: $A(t) = ke^{-\delta t}$, where k is a constant.

$P(t,T)$ can now be calculated from the formula $P(t,T) = \dfrac{E_P\left[A(T) \mid F_t\right]}{A(t)}$, *ie*:

$$P(t,T) = \frac{E_P\left[ke^{-\delta T} \mid F_t\right]}{ke^{-\delta t}}$$

Since $A(t)$ is not random here, we can simplify this to get:

$$P(t,T) = \frac{ke^{-\delta T}}{ke^{-\delta t}} = e^{-\delta(T-t)}$$

This question illustrates that the process $A(t)$ acts like a discount factor that applies to payments at time t. It is a generalisation of the familiar v^t factor.

The formula for $P(t,T)$ is a very simple looking formula, but any potential difficulty comes in working out $E_P\left[A(T) \mid F_t\right]$ as we need 'interesting' models for $A(t)$ in order to get interesting and useful models for $r(t)$ and $P(t,T)$.

Question

Suppose that $\mu_A(t) = -\delta$ and $\sigma_A(t) = \sigma$, where δ and σ are constants.

Show that $A(t) = A(0)e^{(-\delta - ½\sigma^2)t + \sigma W_t}$ and $P(t,T) = e^{-\delta(T-t)}$.

Solution

The SDE for $A(t)$ now becomes:

$$dA(t) = A(t)\left[-\delta dt + \sigma dW_t\right]$$

This has the same form as the SDE for the lognormal model for share prices. So its solution is:

$$A(t) = A(0)e^{(-\delta - ½\sigma^2)t + \sigma W_t}$$

$P(t,T)$ can now be calculated from the formula $P(t,T) = \dfrac{E_P\left[A(T) \mid F_t\right]}{A(t)}$:

$$P(t,T) = \frac{E_P\left[A(0)e^{(-\delta-½\sigma^2)T+\sigma W_T} \mid F_t\right]}{A(0)e^{(-\delta-½\sigma^2)t+\sigma W_t}}$$

$$= \frac{e^{(-\delta-½\sigma^2)(T-t)}}{e^{\sigma W_t}} E_P\left[e^{\sigma W_T} \mid F_t\right]$$

$$= e^{(-\delta-½\sigma^2)(T-t)} E_P\left[e^{\sigma(W_T - W_t)} \mid F_t\right]$$

This expectation is conditional on the history up to time t. But $W_T - W_t$ relates to the 'future' time interval (t,T), and so is statistically independent. So we can write:

$$P(t,T) = e^{(-\delta-½\sigma^2)(T-t)} E_P\left[e^{\sigma(W_T - W_t)}\right]$$

Since $W_T - W_t \sim N(0, T-t)$, the expectation is now just the MGF of a normal distribution, so that:

$$P(t,T) = e^{(-\delta-½\sigma^2)(T-t)} e^{½\sigma^2(T-t)}$$

$$= e^{-\delta(T-t)}$$

Using the state price deflator approach the risk-free rate of interest can be shown to be simply:

$$r(t) = -\mu_A(t)$$

Again, this result is not derived here.

It follows that our model will have positive interest rates if and only if $\mu_A(t)$ is negative for all t with probability 1. This means that $A(t)$ must be a strictly positive supermartingale (that is, $E_P\left[A(T) \mid F_t\right] < A(t)$ for all t, T).

Recall that martingales have the property that $E\left[X_t \mid F_s\right] = X_s$ whenever $t > s$, *ie* the process has zero drift. A *supermartingale* has the property that $E\left[X_t \mid F_s\right] \leq X_s$, *ie* the drift is negative or zero.

Question

How do we know that $A(t)$ is a supermartingale?

Solution

$A(t)$ takes positive values and $\mu_A(t)$ is negative. Therefore the drift coefficient $A(t)\mu_A(t)$ in the SDE for $A(t)$ is negative. So the expected value of $dA(t)$ is negative, *ie* the process has negative drift, which means that it is a supermartingale.

The state price deflator approach looks like it is quite different from the risk-neutral approach to pricing. However, the two approaches are, in fact, exactly equivalent: any model developed in one framework has an equivalent form under the other framework.

5.2 Market models

Today, 'market models' have superseded short-rate models for situations where the correlation between different maturity rates is critical, such as the pricing of swaptions. These treat each maturity instrument (such as forward rate) as a distinct object, correlated to other similar assets using a multi-dimensional Wiener process in a no-arbitrage set-up based on the 'state price deflator' approach.

Define a traded asset, based on the (traded) discount bonds $P(s,t)$ and $P(s,T)$:

$$X(s) = \frac{1}{\tau}\left(P(s,t) - P(s,T)\right)$$

The price of a traded asset divided by another traded asset must be a martingale under the measure associated with the numeraire.

A numeraire is a quantity whose values are used as the units for expressing the price of a security. Commonly used numeraires are:

- the accumulated value of a risk-free cash account
- the price of a zero-coupon bond
- the price of a foreign currency.

So, labelling Q^T as the measure associated with using $P(s,T)$ as the numeraire, we have:

$$F(s,t,T) = \frac{X(s)}{P(s,T)} = E_{Q^T}\left[\frac{X(t)}{P(t,T)}\right] = E_{Q^T}\left[F(t,t,T)\right]$$

Under Q^T, the *forward rates*, $F(s,t,T)$ are martingales, and:

$$dF(s,t,T) = \zeta(s,t)F(s,t,T)dW_s^{Q^T}$$

Note that we use t as the parameter for the volatility, rather than T, because the forward matures at t and does not exist in $(t,T]$.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 18 Summary

Notation

- t current time
- $P(t,T)$ zero-coupon bond price
- $r(t)$ instantaneous risk-free rate
- $f(t,T,S)$ continuously compounded forward rate for the period (T,S)
- $f(t,T)$ instantaneous forward rate
- $r(t,T)$ continuously compounded spot rate

Relationships

- $r(t,T) = -\dfrac{1}{T-t}\log P(t,T) = \dfrac{1}{T-t}\int_t^T f(t,u)du$
- $P(t,T) = \exp[-r(t,T)(T-t)]$
- $f(t,T,S) = \dfrac{1}{S-T}\log\dfrac{P(t,T)}{P(t,S)}$
- $f(t,T) = \lim\limits_{T\to S} F(t,T,S) = -\dfrac{\partial}{\partial T}\log P(t,T)$
- $P(t,T) = \exp\left[-\int_t^T f(t,u)du\right]$

Desirable characteristics of a term structure model

- The model should be arbitrage-free.
- Interest rates should ideally be positive.
- Interest rates should be mean-reverting.
- Bonds and derivative contracts should be easy to price.
- It should produce realistic interest rate dynamics.
- It should fit historical interest rate data adequately.
- It should be easy to calibrate to current market data.
- It should be flexible enough to cope with a range of derivatives.

Examples of one-factor models

Vasicek model

$$dr(t) = \alpha\left[\mu - r(t)\right]dt + \sigma d\tilde{W}(t) \quad \text{under } Q$$

Cox-Ingersoll-Ross model

$$dr(t) = \alpha\left[\mu - r(t)\right]dt + \sigma\sqrt{r(t)}\,d\tilde{W}(t) \quad \text{under } Q$$

Hull-White model

$$dr(t) = \alpha\left[\mu(t) - r(t)\right]dt + \sigma d\tilde{W}(t) \quad \text{under } Q$$

where $\mu(t)$ is a deterministic function.

Chapter 18 Practice Questions

18.1 Explain the similarities and differences in the following three interest rate models:

- the Hull-White model
- the Cox-Ingersoll-Ross model
- the Vasicek model

18.2 Explain the following formulae as they are used in interest rate modelling:

(a) $P(t,T) = E_Q\left[\exp\left(-\int_t^T r(u)du\right) \middle| r(t)\right]$

(b) $P(t,T) = \dfrac{1}{A(t)} E_P[A(T) \mid F_t]$

18.3 The stochastic differential equations defining the short-rate process assumed in three commonly used models for the term structure of interest rates are shown below:

$$\textit{Model 1}: \quad dr(t) = \alpha[\mu - r(t)]dt + \sigma d\tilde{W}(t)$$

$$\textit{Model 2}: \quad dr(t) = \alpha[\mu - r(t)]dt + \sigma\sqrt{r(t)}d\tilde{W}(t)$$

$$\textit{Model 3}: \quad dr(t) = \alpha[\mu(t) - r(t)]dt + \sigma d\tilde{W}(t)$$

In each case, $\tilde{W}(t)$ denotes a standard Brownian motion under the risk-neutral probability measure.

(i) Identify these three models.

(ii) Outline the key statistical properties of the short-rate processes for each of these models.

The dynamics of a fourth model are defined by:

$$\textit{Model 4}: \quad dr(t) = \theta dt + \sigma d\tilde{W}(t)$$

where θ and σ are constants.

(iii) State the limitations of this model.

18.4 (i) List the desirable characteristics of interest rate models. [4]

Exam style

(ii) Discuss the limitations of one-factor models of interest rates. [5]

(iii) The short rate of interest is governed by the stochastic differential equation (SDE):

$$dr_t = 0.6\left(0.04 - r_t\right)dt + 0.006dB_t$$

where B_t is a standard Brownian motion.

By considering the function $f\left(r_t, t\right) = r_t e^{0.6t}$, or otherwise, solve this SDE. [5]

[Total 14]

18.5 A bond trader assumes that $f(t,T)$, the instantaneous forward rate of interest at time T implied by the market prices of bonds at the current time t, can be modelled by:

Exam style

$$f(t,T) = 0.04e^{-0.2\tau} + 0.06(1 - e^{-0.2\tau}) + 0.1(1 - e^{-0.2\tau})e^{-0.2\tau}$$

where $\tau = T - t$.

(i) Sketch a graph of $f(t,T)$ as a function of τ. [3]

(ii) Calculate the following quantities using this model:

(a) the instantaneous forward rate of interest in two years' time

(b) the current price of a 10-year zero-coupon bond

(c) the current 10-year spot rate.

You should express your answers to (a) and (c) as annualised continuously compounded rates. [6]

[Total 9]

18.6 If $A(t)$ is a strictly positive supermartingale, then zero-coupon bond prices can be modelled using the formula $B(t,T) = \dfrac{E_P[A(T) \mid F_t]}{A(t)}$, where P is a suitably-chosen probability measure.

Exam style

(i) (a) Express mathematically the fact that $A(t)$ is a strictly positive supermartingale.

(b) Verify that the function $A(t) = e^{-0.05t + 0.02W(t)}$, where $W(t)$ denotes standard Brownian motion, satisfies the properties in (i)(a).

(c) State why the supermartingale property is required.

(d) Write down the name given to this type of process. [7]

(ii) By writing $A(t)$ in the form $A(t) = e^{X(t)}$, or otherwise, show that $A(t)$ satisfies a stochastic differential equation of the form:

$$dA(t) = A(t)[\mu_A(t)dt + \sigma_A(t)dW(t)]$$

State the forms of the functions $\mu_A(t)$ and $\sigma_A(t)$. [4]

(iii) (a) Write down or derive a formula for $P(t,T)$ based on the process $A(t)$ specified in (i)(b).

(b) Write down expressions for the instantaneous forward rate $f(t,T)$ and the spot rate $r(t,T)$ based on this model.

(c) State one problem that this model of interest rates has. [4]

(iv) Calculate the prices at time 5 according to the model in (ii) of the following risk-free bonds:

(a) a 10-year zero-coupon bond

(b) a 10-year bond that pays a coupon of 5% at the end of each year. [4]

[Total 19]

18.7 (i) Explain what is meant by a short rate model of interest rates and how such models can be used to price interest rate derivatives.

(ii) Explain what is meant by a one-factor short rate model of interest rates.

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 18 Solutions

18.1 These three models are all one-factor models used for modelling the short rate of interest $r(t)$.

All three models assume that $r(t)$ has the dynamics of an Ito process under the risk-neutral probability measure Q.

The equations defining the three models are:

Vasicek: $dr(t) = \alpha[\mu - r(t)]dt + \sigma d\tilde{W}(t)$

Cox-Ingersoll-Ross: $dr(t) = \alpha[\mu - r(t)]dt + \sigma\sqrt{r(t)}d\tilde{W}(t)$

Hull-White: $dr(t) = \alpha[\mu(t) - r(t)]dt + \sigma d\tilde{W}(t)$

All three models are mean reverting to μ, where…

… μ is time-dependent for the Hull-White model, but constant for the other two.

The Hull-White model is easier to fit to past and current data than the Vasicek or Cox-Ingersoll-Ross models due to the fact that there is more choice of parameters (since μ is time-dependent).

All three models generate arbitrage-free zero-coupon bond prices.

All three models can be used to price simple options on zero-coupon bonds.

The Cox-Ingersoll-Ross model includes the factor $\sqrt{r(t)}$ in the volatility coefficient. This prevents $r(t)$ taking negative values.

The Vasicek model is more tractable mathematically than the other two.

Over long periods the distribution of $r(t)$ under the Cox-Ingersoll-Ross model involves the non-central chi-square distribution, whereas the distribution under the other two models is normal.

Since these are all one-factor models:

- they cannot be used to price derivatives whose payoffs depend on more than one interest rate.
- there will be positive correlation between bond prices of all durations, which is unrealistic.

18.2 (a) $P(t,T) = E_Q\left[\exp\left(-\int_t^T r(u)du\right)\middle| r(t)\right]$

This formula is used to find the price at time t of a zero-coupon bond maturing at time T when a short rate model is used.

Q denotes the risk-neutral probability measure for this process.

(b) $$P(t,T) = \frac{1}{A(t)} E_P[A(T) \mid F_t]$$

This formula is used to find the price at time t of a zero-coupon bond maturing at time T when a state-price deflator model is used.

$A(t)$ is the state-price deflator, which is a stochastic process.

$A(t)$ must be a strictly positive supermartingale.

P denotes a suitably chosen probability measure.

The stochastic differential equation for $A(t)$ has the following form under P :

$$dA(t) = A(t)[\mu_A(t)dt + \sigma_A(t)dW(t)]$$

where $\mu_A(t)$ and $\sigma_A(t)$ are appropriately chosen stochastic processes.

18.3 (i) ***Identify the models***

These are the Vasicek, Cox-Ingersoll-Ross and Hull-White models, respectively.

(ii) ***Key statistical properties***

In each case the short rate of interest $r(t)$ is modelled as an Ito process.

The process therefore operates in continuous time and has normally-distributed increments over short time intervals.

The volatility parameter σ controls the size of the random movements.

The coefficient of dt represents the expected annual drift under the risk-neutral probability measure (since the increments $d\tilde{W}(t)$ have zero mean under this measure). But under the real-world probability measure the drift will have a different value.

All three models exhibit mean reversion.

In Model 1 and Model 2 the long-term "target" rate μ is constant, while in Model 3 it is a function of the time t .

With Model 1 and Model 3 it is possible to get a negative value for $r(t)$, which is unrealistic.

The inclusion of the $\sqrt{r(t)}$ factor in Model 2 prevents $r(t)$ taking negative values.

Over longer periods the distribution of $r(t)$ under Model 2 involves the non-central chi-square distribution, whereas the distribution under the other two models is normal.

Model 3 is easier to fit to past and current data than the other models due to the fact that there is more choice of parameters (since μ is time-dependent).

(iii) ***Limitations of Model 4***

Model 4 does not exhibit mean reversion.

If θ is non-zero, the trend value of $r(t)$ will increase or decrease steadily, which is unrealistic.

The model allows negative interest rates, which is not necessarily realistic.

The model involves only two parameters, θ and σ, and so may be difficult to calibrate to past and current data.

Since this is a one-factor model:

- it cannot be used to price derivatives whose payoffs depend on more than one interest rate.
- there will be positive correlation between bond prices of all durations, which is unrealistic.

18.4 (i) ***Desirable characteristics of interest rate models***

The model should be arbitrage-free.

Interest rates should ideally be positive.

Interest rates should be mean-reverting over the long term.

Bonds and derivative contracts should be easy to price.

The model should produce realistic interest rate dynamics.

It should fit historical interest rate data adequately.

It should be easy to calibrate to current market data.

It should be flexible enough to cope with a range of derivatives.

[½ each, Total 4]

(ii) ***Limitations of one-factor models***

A result of one-factor models is that yields on bonds of different durations are perfectly correlated. This is not realistic. [1]

In fact, they need not even be positively correlated. Sometimes we see, for example, that short-dated bonds fall in price while long-dated bonds increase in price. [½]

If we look at the long run of historical data we find that there have been sustained periods of both high and low interest rates with periods of both high and low volatility. These are features that are difficult to capture in one-factor models. [1]

This issue is especially important for two types of problem in insurance:

- the pricing and hedging of long-dated insurance contracts with interest rate guarantees [½]
- asset-liability modelling and long-term risk management. [½]

One-factor models struggle to cope with valuing derivative contracts. We need more complex models to deal effectively with these. For example, any contract that makes reference to more than one interest rate should allow these rates to be less than perfectly correlated. [1½]

[Total 5]

(iii) ***Short rate of interest***

Using a Taylor Series expansion for the given function we get:

$$d\left(r_t e^{0.6t}\right) = df\left(r_t, t\right)$$

$$= \frac{\partial f}{\partial r_t} dr_t + \frac{1}{2}\frac{\partial^2 f}{\partial r_t^2}\left(dr_t\right)^2 + \frac{\partial f}{\partial t} dt$$

$$= e^{0.6t} \times \left[0.6\left(0.04 - r_t\right)dt + 0.006 dB_t\right] + \frac{1}{2} \times 0 \times \left(dr_t\right)^2 + 0.6 r_t e^{0.6t} dt$$

$$= 0.024 e^{0.6t} dt + 0.006 e^{0.6t} dB_t - 0.6 r_t e^{0.6t} dt + 0.6 r_t e^{0.6t} dt$$

$$= 0.024 e^{0.6t} dt + 0.006 e^{0.6t} dB_t \qquad [2]$$

Alternatively, you can use the general form of Ito's Lemma, the product rule or an integrating factor to derive this equation.

Substituting s for t and integrating between 0 and t we get:

$$\int_0^t d\left(r_s e^{0.6s}\right) = \int_0^t 0.024 e^{0.6s} ds + \int_0^t 0.006 e^{0.6s} dB_s$$

$$\Leftrightarrow \quad \left[r_s e^{0.6s}\right]_0^t = 0.04\left[e^{0.6s}\right]_0^t + 0.006\int_0^t e^{0.6s} dB_s \qquad [2]$$

$$\Leftrightarrow \quad r_t e^{0.6t} - r_0 = 0.04\left(e^{0.6t} - 1\right) + 0.006\int_0^t e^{0.6s} dB_s$$

We can then rearrange to get the required solution:

$$\Leftrightarrow \quad r_t = r_0 e^{-0.6t} + 0.04\left(1 - e^{-0.6t}\right) + 0.006\int_0^t e^{-0.6(t-s)} dB_s \qquad [1]$$

[Total 5]

Note that this is an example of the Vasicek model, which is also an Ornstein-Uhlenbeck process.

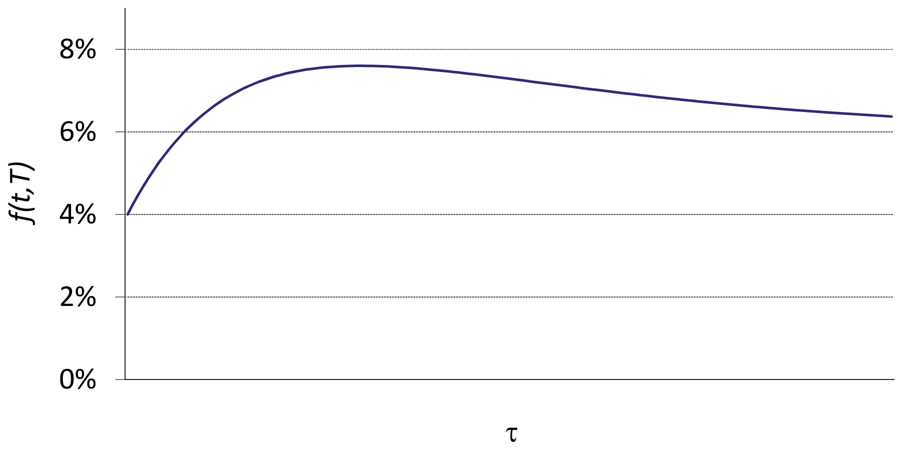
18.5 (i) ***Graph***

The factor $e^{-0.2\tau}$ reduces from 1 to zero as the term τ increases from zero to infinity. So the first two terms represent a weighted average of 0.04 (the short rate) and 0.06 (the long rate). [½]

The final term is zero when $\tau = 0$ or $\tau = \infty$, but positive in between. So this adds a 'hump' to the graph. [½]

In fact, if you differentiate the function given, you will find that the maximum occurs when $e^{-0.2\tau} = 0.4$*, which corresponds to the point where* $\tau = 4.58$ *and* $f = 7.6\%$*.*

The graph looks like this:



[2]

[Total 3]

(ii)(a) ***Calculate the instantaneous forward rate in 2 years' time***

This is:

$$f(t,t+2) = 0.04e^{-0.2\times2} + 0.06(1-e^{-0.2\times2}) + 0.1(1-e^{-0.2\times2})e^{-0.2\times2} = 0.0687$$

ie 6.87% [1]

(ii)(b) ***Calculate the price of a 10-year zero-coupon bond***

This is:

$$P(t,t+10) = \exp\left[-\int_t^{t+10} f(t,u)\,du\right]$$

The integral, using the substitution $\tau = u - t$, is:

$$\int_t^{t+10} f(t,u)du = \int_0^{10} [0.04e^{-0.2\tau} + 0.06(1-e^{-0.2\tau}) + 0.1(e^{-0.2\tau} - e^{-0.4\tau})]d\tau$$

$$= \int_0^{10} [0.06 + 0.08e^{-0.2\tau} - 0.1e^{-0.4\tau}]d\tau$$

$$= \left[0.06\tau - 0.4e^{-0.2\tau} + 0.25e^{-0.4\tau}\right]_0^{10}$$

$$= 0.55044 - (-0.15) = 0.70044$$

So: $P(t,t+10) = e^{-0.70044} = 0.4964$

ie £49.64 per £100 nominal [3]

(ii)(c) ***Calculate the 10-year spot rate***

This can be calculated as the average of the forward rates:

$$r(t,t+10) = \frac{1}{10}\int_t^{t+10} f(t,u)du = \frac{1}{10} \times 0.70044 = 0.070044$$

ie 7.00% [2]

Alternatively, we could solve:

$$P(t,t+10) = e^{-0.70044} = 1 \times e^{-10R(t,t+10)}$$

[Total 6]

18.6 (i)(a) ***Express these properties mathematically***

'Strictly positive' simply means that:

$A(t) > 0$ for all times t [1]

The 'supermartingale' property means that, whenever $t < T$:

$E_P[A(T) | F_t] \le A(t)$ [1]

(i)(b) ***Verify that this function has these properties***

The presence of the exponential function ensures that this function is strictly positive. [½]

Since $A(t)$ is strictly positive, the supermartingale property is equivalent to:

$$\frac{E_P[A(T) | F_t]}{A(t)} < 1$$ [½]

With the definition given for $A(t)$, the left-hand side is:

$$LHS = \frac{E_P[A(T) \mid F_t]}{A(t)}$$

$$= e^{0.05t - 0.02W(t)} E_P\left[e^{-0.05T + 0.02W(T)} \middle| F_t \right]$$

$$= e^{-0.05(T-t)} E_P\left[e^{0.02[W(T) - W(t)]} \middle| F_t \right] \qquad [1]$$

Because of the independent increments property of Brownian motion, we can drop the F_t. We can then use the fact that $W(T) - W(t) \sim N(0, T-t)$ under P to evaluate the expectation on the right-hand side, which corresponds to an MGF based on a normal distribution. Using the formula given on page 11 of the *Tables*, we get:

$$LHS = e^{-0.05(T-t)} e^{½(0.02)^2 (T-t)} = e^{-0.0498(T-t)} \qquad [½]$$

When $t < T$ (which we have assumed throughout), this is indeed less than 1. So the supermartingale property is satisfied. [½]

(i)(c) ***State why the supermartingale property is required***

The supermartingale property is equivalent to:

$$\frac{E_P[A(T) \mid F_t]}{A(t)} < 1$$

The left-hand side matches the formula for the bond price $B(t,T)$. So this property ensures that the price of a zero-coupon bond is always less than 1. This is equivalent to prohibiting negative interest rates. [1]

(i)(d) ***Name of the process***

The process $A(t)$ is a state price deflator. [1]

[Total 7]

(ii) ***Stochastic differential equation for A(t)***

We can write:

$$A(t) = e^{X(t)}$$

where $X(t) = -0.05t + 0.02W(t)$, so that $dX(t) = -0.05dt + 0.02dW(t)$. [1]

So, using a Taylor Series expansion, we can write:

$$dA(t) = d\left[e^{X(t)} \right] = e^{X(t)} dX(t) + \tfrac{1}{2} e^{X(t)} [dX(t)]^2$$

$$= A(t) \left\{ dX(t) + \tfrac{1}{2} [dX(t)]^2 \right\} \qquad [1]$$

Substituting the SDE for $X(t)$ gives:

$$dA(t) = A(t)\left\{-0.05dt + 0.02dW(t) + \tfrac{1}{2}[-0.05dt + 0.02dW(t)]^2\right\}$$

Simplifying using the 2×2 multiplication grid, we get:

$$dA(t) = A(t)\left\{-0.05dt + 0.02dW(t) + \tfrac{1}{2}(0.02)^2 dt\right\}$$

$$= A(t)\{-0.0498dt + 0.02dW(t)\} \qquad [1]$$

So, in this case, the drift and volatility coefficients are:

$\mu_A(t) = -0.0498$ and $\sigma_A(t) = 0.02$ [1]

[Total 4]

(iii)(a) ***Formula for $P(t,T)$***

We have already evaluated the formula for $P(t,T)$ in part (i)(b), which gave:

$$P(t,T) = \frac{E_P[A(T)|F_t]}{A(t)} = e^{-0.0498(T-t)} \qquad [1]$$

(iii)(b) ***Expressions for $f(t,T)$ and $r(t,T)$***

The instantaneous forward rate is:

$$f(t,T) = -\frac{\partial}{\partial T}\log P(t,T) = -\frac{\partial}{\partial T}[-0.0498(T-t)] = 0.0498$$

ie a constant rate of 4.98%. [1]

The spot rate therefore also takes a constant value of 4.98%. [1]

(iii)(c) ***One problem with this model***

A model with constant interest rates over all terms is not arbitrage-free. This would be a serious problem if the model was used in practical applications. [1]

[Total 4]

(iv)(a) ***Price of a zero-coupon bond***

According to this model, the price at time 5 (or indeed, at *any* time) of a 10-year zero-coupon bond is:

$$P(5,15) = e^{-0.0498(15-5)} = e^{-0.498} = 0.6077$$

ie 60.77 per 100 nominal. [2]

(iv)(b) ***Price of a 5% annual coupon bond***

The price of a 5% annual coupon bond per 100 nominal is:

$$P = 5[v + v^2 + \cdots + v^{10}] + 100v^{10}$$

where $v = e^{-0.0498} = 0.95142$.

Evaluating the sum as a geometric progression, we get:

$$P = 5v\left(\frac{1-v^{10}}{1-v}\right) + 60.77 = 5(7.682) + 60.77 = 99.18$$ [2]

[Total 4]

18.7 (i) ***Short rate model of interest rates***

Short rate models are used to model the term structure of interest rates as a function of the short rate $r(t)$.

In theory, this is the interest rate that applies over the next instant of time, of length dt.

In practice, the short rate is an overnight rate, *ie* the force of interest earned when money is lent today and received back with interest the following day.

$r(t)$ itself is usually assumed to be an Ito process, with a stochastic differential equation of the form:

$$dr(t) = a(t, r(t))\, dt + b(t, r(t))\, d\tilde{W}_t$$

where $\tilde{W}_t$ is a standard Brownian motion under the risk-neutral probability measure, Q.

The price at time t of an interest rate derivative that pays X_T at maturity date $T > t$ can then be found from:

$$V_t = E_Q\left[\exp\left(-\int_t^T r(u)du\right)X_T \,\middle|\, F_t\right]$$

(ii) ***Explain what is meant by a one-factor short rate model***

A *one-factor short rate model* is one in which the short rate, and hence the term structure as a whole, is assumed to be influenced by a single source of randomness.

The prices of all bonds (of all maturities) and interest-rate derivatives must therefore move together.

The single source of randomness is typically assumed to be a standard Brownian motion process.

The short rate is therefore usually modelled as an Ito process.

19

Credit risk

Syllabus objectives

4.6 Simple models for credit risk

4.6.1 Define the terms credit event and recovery rate.

4.6.2 Describe the different approaches to modelling credit risk:

- structural models
- reduced-form models
- intensity-based models.

4.6.3 Demonstrate a knowledge and understanding of the Merton model.

4.6.4 Demonstrate a knowledge and understanding of a two-state model for credit ratings with a constant transition intensity.

4.6.5 Describe how the two-state model can be generalised to the Jarrow-Lando-Turnbull model for credit ratings.

4.6.6 Describe how the two-state model can be generalised to incorporate a stochastic transition intensity.

0 Introduction

This chapter addresses credit risk – the risk that a person or an organisation will fail to make a payment they have promised. This is quite a new area (*eg* the Jarrow-Lando-Turnbull (JLT) model was only published in 1997) that has grown in importance with the introduction of credit derivatives (which are covered in some later subjects, not in this course).

We start with some definitions to set the scene, then go on to look at the Merton model, which is a simple model that relates the price of a bond subject to default to the price of an option. We then look at two models that are applications of the theory of continuous-time jump processes. The first model is a two-state model. The second is the JLT model, which is a more general multi-state model.

The Core Reading in this chapter is adapted from course notes written by Timothy Johnson.

1 Credit events and recovery rates

In earlier chapters it has been assumed that bonds are default-free.

Question

Explain what it means for a bond to be default-free.

Solution

A bond is default-free if the stream of payments due from the bond will *definitely* be paid *in full* and *on time*.

Credit risk exists when a party may default on its obligations. Credit risk is usually ignored with respect to payments by a sovereign government in its own currency, but needs to be accommodated for if an obligation is met in a currency issued by a third-party (such as corporate obligations, obligations by a government in a currency it does not control).

Corporate entities issuing bonds consist mainly of large companies and banks. For example, in May 2014, Barclays had a 5¾% bond redeemable in September 2026 and Tesco had a 5½% bond redeemable in December 2019. These companies entered into a contract to make interest payments on set dates to the bondholders and to repay the face value of the bond on the redemption date. Failure to do this would result in the bonds being in default.

A credit loss only exists if the counter-party defaults *and* the contract has value. In a forward or swap contract, both long/receiving and short/paying parties are exposed to a credit risk, since either party could default if the market moves against them. For options and bonds, the purchaser of the option/bond is exposed to default by the writer/issuer, but they do not have an obligation to the writer/issuer.

Credit risk is calculated as an expected loss:

$$\textbf{Expected Loss} = \textbf{Exposure} \times \textbf{Probability of Default} \times \textbf{Loss Given Default}$$

All the parameters have an implicit time dependence. The Loss Given Default (LGD) is the percentage of the exposure that will be lost on a default, the recovery rate is the reciprocal of the LGD (Recovery Rate = 100%-LGD). Usually some value can be recovered when a counter-party defaults.

Credit risk changes with the market and good practice is to assess both *current* and *potential* exposure. The current exposure is the current market value of the asset, the future exposure should be based on a wide range of future scenarios, with different default probabilities.

For the remainder of the chapter we'll only be considering credit risk with respect to bonds.

The outcome of a default may be that the contracted payment stream:

- is rescheduled
- is cancelled by the payment of an amount which is less than the default-free value of the original contract
- continued but at a reduced rate
- is totally wiped out.

An example of cancellation would be where the bondholders agree to accept a reduced one-off cash payment of 75% of the face value instead of the contractual payments.

***Credit events*, which might result in a failure to meet an obligation (defined for the purposes of credit derivatives), include:**

- **actions that are associated with bankruptcy or insolvency laws**

  ie the bond issuer becomes insolvent.

- **downgrade by 'Nationally Recognised Statistical Rating Organisations', (NRSROs such as Moody's, S&P and Fitch)**

  This is of particular concern when a bond is issued with a guaranteed minimum credit rating.

- **failure to pay**

  ie either a coupon or the capital amount is not paid in full and on time.

- **repudiation / moratorium**

  ie the validity of the contract is disputed or a temporary suspension of activity is imposed on the issuer.

- **restructuring – when the terms of the obligation are altered so as to make the new terms less attractive to the debt holder, such as a reduction in the interest rate, re-scheduling, change in principal, change in the level of seniority.**

2 Approaches to modelling credit risk

2.1 Structural models

Structural models are explicit models for a corporate entity issuing both equity and debt. They aim to link default events explicitly to the fortunes of the issuing corporate entity.

Structural, or firm-value, models are used to represent a firm's assets and liabilities and define a mechanism for default. Typically, default occurs when a stochastic variable (or process) hits a barrier representing default. The main example of a structural model is the Merton model.

These models deliver an explicit link between a firm's default and the economic conditions and provide a sound basis for estimating default correlations amongst different firms. The disadvantage is identifying the correct model and estimating its parameters.

These models are called 'structural' because they focus on the financial structure – the split between debt and equity – of the company issuing the bond. We will discuss the Merton model in the next section.

2.2 Reduced-form models

Reduced-form models do not attempt to deliver a representation of a firm, like structural models do. Rather they are statistical models that use observed data, both macro and micro, and so can usually be 'fitted' to data.

The market statistics most commonly used are the credit ratings issued by NRSROs. The credit rating agencies will have used detailed data specific to the issuing corporate entity when setting their rating. They will also regularly review the data to ensure that the rating remains appropriate and will re-rate the bond either up or down as necessary.

Default is no longer tied to the firm value falling below a threshold-level, as in structural models. Rather, default occurs according to some exogenous hazard rate process.

These models are called 'reduced-form' because they do not attempt to model the inner financial workings of the particular company issuing the bond. Instead, they model the different levels of creditworthiness and how companies move from one status to another.

2.3 Intensity-based models

An intensity-based model is a particular type of continuous-time reduced-form model. It typically models the jumps between different states (usually credit ratings) using transition intensities. The disadvantage of reduced-form models is that they sometimes lack the clarity of structural models.

This approach uses a continuous-time multiple state model (jump process) where the states correspond to the creditworthiness of the company. The 'intensities' (denoted by λ) are the rates driving the company to switch from one state to another as time passes. Intensity-based models are an example of a reduced-form model.

3 The Merton model

The Merton model is a simple example of a structural model.

Classical finance defines the value of a firm $F(t)$ as the sum of its debt, $B(t)$ and equity $E(t)$, so:

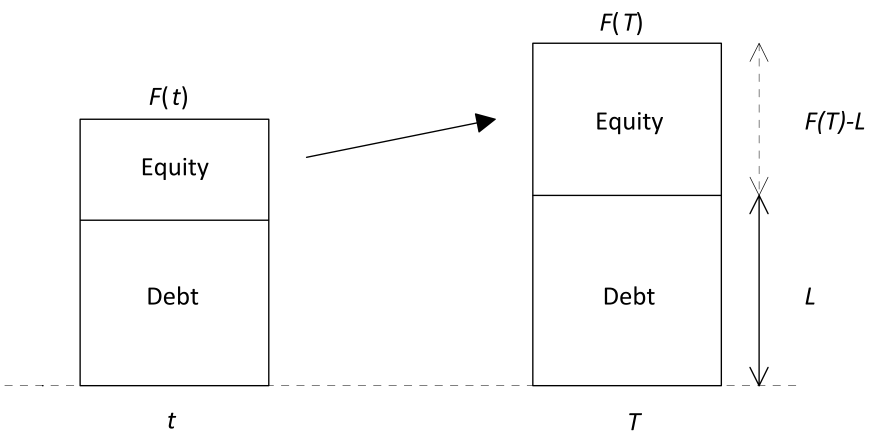
$$F(t) = B(t) + E(t)$$

Merton's model assumes that a corporate entity has issued both equity and debt such that its total value at time t is $F(t)$. This value varies over time as a result of actions by the corporate entity, which does not pay dividends on its equity or coupons on its bonds.

For example, the value of a company will change along with investors' perceptions of the future prospects of that company.

Assume a firm has issued a single zero-coupon bond with face value of *L* which matures at time *T*.

Debt holders rank above shareholders in the wind-up of a company. So, provided the company has sufficient funds to pay the debt, the shareholders will receive $F(T) - L$.



The corporate entity will default if the total value of its assets, $F(T)$, is less than the promised debt repayment at time T:

ie $\quad F(T) < L$

In this situation, the bondholders will receive $F(T)$ instead of L and the shareholders will receive nothing.

Combining these two cases, we see that the shareholders will receive a payoff of $\max\{F(T) - L, 0\}$ at time T.

This can be regarded as treating the shareholders of the corporate entity as having a European call option on the assets of the company with maturity T and a strike price equal to the value of the debt. The Merton model can be used to estimate either the risk-neutral probability that the company will default or the credit spread on the debt.

Since debt is senior to equity, the value of equity at maturity:

$$\boldsymbol{E(T) = \max\{F(T) - L, 0\}}$$

and so the value of a firm's equity is a call option on the value of the firm with a strike of the debt.

Credit spread is a measure of the excess of the yield on a risky security over a risk-free yield. It largely relates to the expected cost of default. However, in practice it will also typically reflect other factors, such as a risk premium relating to the risk of default and a liquidity premium.

Because of the risk of default, a bond issued by a company will have a lower market price than a similar bond issued by a government. So the yield on the company bond will be higher than the yield on the government bond. The credit spread refers to the difference between these two rates.

Question

Suggest how this model could be used to calculate the value of the risky corporate bonds at time t.

Solution

Since we are viewing the equity as a call option on the total value of the company, we could use an option pricing method, such as the Black-Scholes option pricing formula, to calculate how much the equity at time T is worth now (*ie* at time t).

The value of the bonds could then be found by subtracting this from the current value of the company $F(t)$, *ie*:

$$B(t) = F(t) - E(t) .$$

where $B(t)$ and $E(t)$ are the current value of the company's risky corporate bonds and the equity respectively.

Consequently,

$$E(t) = F(t)\Phi(d_1) - Le^{-r(T-t)}\Phi(d_2) \qquad (1)$$

with:

$$d_1 = \frac{\log\left(F(t)/L\right) + (r + \frac{1}{2}\sigma_F^2)(T-t)}{\sigma_F\sqrt{T-t}}, \quad d_2 = d_1 - \sigma_F\sqrt{T-t}$$

where σ_F is the volatility of the firm value.

The value of the debt today is $F(t) - E(t)$.

Unfortunately, $F(t)$ (and σ_F) are unobservable, since they depend on the market's assessment of $B(t)$. However, if we assume that the value of the firm and the equity both follow geometric Brownian motion, and $E(t) = f(F(t))$ then by Ito:

$$dE(t) = \left[\mu_F F(t)\frac{\partial E(t)}{\partial F(t)} + \frac{1}{2}\sigma_F^2 F^2(t)\frac{\partial^2 E(t)}{\partial F^2(t)}\right]dt + \sigma_F F(t)\frac{\partial E(t)}{\partial F(t)}dW_t$$

where the values μ_F and σ_F come from the stochastic differential equation for $F(t)$:

$$dF(t) = F(t)\left(\mu_F dt + \sigma_F dW_t\right)$$

However, $E(t)$ also has its own stochastic differential equation:

$$dE(t) = E(t)\left(\mu_E dt + \sigma_E dW_t\right)$$

Comparing terms leaves:

$$\sigma_E E(t) = \sigma_F F(t)\frac{\partial E(t)}{\partial F(t)} = \sigma_F F(t)\Phi\left(d_1\right) \qquad (2)$$

Now $E(t)$ and σ_E *can* be observed from market data.

Using equations (1) and (2) we have:

$$\begin{aligned}\sigma_E E(t) &= \sigma_F F(t)\Phi\left(d_1\right)\\ &= \sigma_F\left(E(t) + Le^{-r(T-t)}\Phi(d_2)\right)\end{aligned}$$

$$\Rightarrow \qquad \sigma_F = \frac{\sigma_E E(t)}{E(t) + Le^{-r(T-t)}\Phi(d_2)}$$

This is not trivial since σ_F is used to calculate d_2.

Recall that, in the Black-Scholes world, $\Phi(d_2)$ is the risk-neutral probability that a call option will be exercised, that is it expires in the money. In this context, this means that:

$$1-\Phi(d_2)=\Phi(-d_2)$$

is the risk-neutral probability that the firm is in default at time T (but not that it has defaulted in $[0,T)$ and then recovered). This method provides a rough estimate of the probability of default.

In general it may be possible for default bonds to 'recover', *ie* to re-start coupon payments. However, the Merton model is only concerned with the state of the bond at time T. In the next section it is assumed that once a bond defaults then it always remains in such a state.

One limitation is that the default probability is given in the abstract risk-neutral world. The real-world probability can be derived using:

$$d_1=\frac{\log\left(F(t)/L\right)+(\mu_F+\frac{1}{2}\sigma_F^2)(T-t)}{\sigma_F\sqrt{T-t}}, \quad d_2=d_1-\sigma_F\sqrt{T-t}$$

where the real-world drift, μ_F, replaces the risk-less drift. However, μ_F is not observable.

The pricing equation (1) uses the risk-neutral probability measure, under which $F(t)$ is expected to grow at the risk-free rate. This is the reason why the unobservable quantity μ_F is not required, and the above calculations have been to find only σ_F (which is used in (1)).

4 Two-state models for credit risk

4.1 Interest rates as hazard rates

In Europe, before the Reformation of the Catholic Church in the sixteenth century, the charging of interest was only permissible as a compensation for the risk that the lender took on; this is captured in the opening observation of the Black-Scholes paper:

It should not be possible to make a risk-less profit.

We will now consider a two-state intensity-based model, which is the simplest continuous-time reduced-form model.

A model can be set up, in continuous time, with two states:

1. N = not previously defaulted
2. D = defaulted.

If the transition intensity, under the risk-neutral measure Q, from N to D at time t is denoted by $\lambda(t)$, this model can be represented as:

$$\boxed{\textit{No default, N}} \xrightarrow{\lambda(t)} \boxed{\textit{Default, D}}$$

and D is an absorbing state.

This is a two-state continuous-time Markov model. It has the same structure as the two-state mortality model, with 'No default' corresponding to 'Alive', 'Default' corresponding to 'Dead' and $\lambda(t)$ corresponding to the force of mortality.

Question

Write down a formula for the probability (under Q) that a company that is in state N at time t will remain in state N until time T.

Hint: *Use the survival probability formula* ${}_tp_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)$ *for the corresponding two-state mortality model from page 32 of the Tables.*

Solution

The probability we are interested in can be written as $Q\left(X(T) = "N" \mid X(t) = "N"\right)$.

The formula given in the hint tells us the probability that a person who is alive at age x will still be alive t years later.

So the corresponding formula here is:

$$Q\left(X(T)="N" \mid X(t)="N"\right)=\exp\left(-\int_t^T \lambda(s)ds\right)$$

Consider a lender lending a sum of money, *L*. The lender is concerned that the borrower will not default, a hopefully rare event, and will eventually pay back the loan. Poisson worked out that if the rate of a rare event occurring was λ then the chance of there being *k* rare events in *n* time periods was given by:

$$P(k \text{ wins in } n \text{ rounds})=\frac{(\lambda n)^k e^{-\lambda n}}{k!}$$

Say the lender assesses that the borrower will default at a rate of λ defaults a day – known as the *hazard rate* – and the loan will last *T* days. The lender might also assume that they will get all their money back, providing the borrower makes no defaults in the *T* days, and nothing if the borrower makes one or more defaults.

In this model the borrower can make at most one default.

On this basis the lender's mathematical expectation of the value of the loan is:

$$E\left[\text{value of loan}\right]=P(\text{No default})\times L + P(\text{Default})\times 0$$

Using the *Law of Rare Events*, the probability of no defaults is given when $k=0$, so:

$$E\left[\text{value of loan}\right]=\frac{(\lambda T)^0 e^{-\lambda T}}{0!}\times L + P(\text{Default})\times 0$$

We can ignore the second expression, since it is zero, then:

$$E\left[\text{value of loan}\right]=Le^{-\lambda T}$$

So, the lender is handing over *L* with the expectation of only getting $Le^{-\lambda T}<L$ back. To make the loan equal the expected repayment, the banker needs to inflate the expected repayment by $e^{\lambda T}$:

$$\left(Le^{\lambda T}\right)e^{-\lambda T}=L$$

Another way to view this is to keep the payout *L* the same, and to reduce the amount the lender is prepared to give for it, *ie*: $Le^{-\lambda T}$. This discounting can be viewed as the interest rate obtained on the debt.

4.2 Incorporating recovery rates

Consider a simple situation whereby an asset is due to pay-out at some time *T*, but is subject to credit-risk and may default at a time τ. In the event of default, the investor recovers a fraction δ at time *T*.

So, for a zero-coupon bond that is due to pay 1 at time T, the actual payment at time T will be:

$$\text{payment} = \begin{cases} 1 & \text{if } T < \tau \\ \delta & \text{if } \tau \le T \end{cases}$$

For example, if $\delta = 0.9$, we are assuming that, once the company has gone into default, all future interest payments and the redemption payment will be reduced by 10% (since $1-\delta = 0.1$). So bondholders will receive 90% (the recovery rate) of the full amounts.

Question

Write down a single formula (*ie* one without if's) for the payment at time T.

Solution

We can use the indicator variables $1_{\{T<\tau\}}$ and $1_{\{\tau \le T\}}$ to do this. The correct formula for the payment is:

$$1_{\{T<\tau\}} + \delta 1_{\{\tau \le T\}}.$$

If default has not occurred by time T, (*ie* $T < \tau$) then $1_{\{T<\tau\}} = 1$ and $1_{\{\tau \le T\}} = 0$, and the formula gives 1, *ie* the full payment will be made.

If default *has* occurred (*ie* $\tau \le T$), then $1_{\{T<\tau\}} = 0$ and $1_{\{\tau \le T\}} = 1$ and the formula gives δ, *ie* the reduced or 'recovered' payment of δ will be made.

So a bond with credit risk can be viewed as a derivative, whose future payout is uncertain. This allows the use of the general risk-neutral pricing formula:

$$\text{value of the loan at time } t = e^{-r(T-t)} E_Q\left[\text{payout} \mid F_t\right]$$

assuming a constant risk-free rate of interest r, and a risk-neutral probability measure Q.

More generally, let C_t represent a money market cash account which is a unit of currency invested at time 0 and accumulated at the (possibly varying) risk-free rate. In the case of a constant risk-free rate we have $C_t = e^{rt}$. Then we have:

$$\text{value of the loan at time } t = C_t E_Q\left[\frac{\text{payout} \mid F_t}{C_T}\right]$$

The value, at time *t*, of this asset to its holder is:

$$v(t,T) = e^{-r(T-t)} E_Q\left[1_{\{T<\tau\}} + \delta 1_{\{\tau \le T\}} \mid F_t\right]$$

$$= C_t E_Q\left[\frac{1_{\{T<\tau\}} + \delta 1_{\{\tau \le T\}} \mid F_t}{C_T}\right]$$

This implies that:

$$\frac{v(t,T)}{C_t} = E_Q\left[\frac{1_{\{T<\tau\}} + \delta 1_{\{\tau \le T\}}}{C_T} \mid F_t\right]$$

Assume that the probability of default is independent of the money market account.

This means that the C_T doesn't need to be inside the expectation operator. Let $P(t,T)$ be the price at time *t* of a risk-free zero-coupon bond that pays out 1 at future time *T*.

Question

Write $P(t,T)$ in terms of the money market cash account.

Solution

$$P(t,T) = \frac{C_t}{C_T}$$

Then the value of the credit-risky asset can be written as:

$$v(t,T) = \frac{C_t}{C_T} E_Q\left[1_{\{T<\tau\}} + \delta 1_{\{\tau \le T\}} \mid F_t\right]$$

$$= \boldsymbol{P(t,T)\left[1 - q(t,T) + \delta q(t,T)\right]}$$

where *q* is the probability, in the risk-neutral measure, that default occurs before *T*, *ie*: $Q(\tau \le T)$.

This is algebraically equivalent to:

$$v(t,T) = P(t,T)\left[1-(1-\delta)q(t,T)\right]$$

where $q(t,T)$ is the risk-neutral probability of default. We saw earlier that the risk-neutral probability of not defaulting (*ie* surviving) is $\exp\left(-\int_t^T \lambda(s)ds\right)$. So we have:

$$v(t,T) = P(t,T)\left[1-(1-\delta)\left(1-\exp\left(-\int_t^T \lambda(s)ds\right)\right)\right]$$

This expression implies that the probability of default, *q*, is given by:

$$q(t,T) = \frac{1}{1-\delta}\left(1 - \frac{v(t,T)}{P(t,T)}\right)$$

Employing this model is not as straightforward as it appears because it is practically difficult to identify the risk-less bond, $P(t,T)$, pertaining to the credit-risky asset, $v(t,T)$, and the recovery rate, δ, at some specific maturity, *T*.

To get around this, credit ratings are employed, and it is assumed that the same yield curve is applied to firms in the same credit rating. This allows bonds issued by different firms in the same risky class to be considered as if there were a single issuer.

Dividing companies into just two states – 'No default' (N) and 'Default' (D) – is a rather crude, black and white approach. In reality, the 'N' state is a heterogeneous category and we can improve this model by introducing some shades of grey. This involves subdividing the 'N' state.

In the next section we will extend the two-state model to cover multiple credit ratings.

5 The Jarrow-Lando-Turnbull (JLT) model

There are several established credit rating agencies, such as Standard & Poor's and Moody's, who publish credit ratings for all the major companies. On Standard & Poor's scale, companies that have already defaulted are Grade D. Companies that have not defaulted are given one of the seven grades: AAA, AA, A, BBB, BB, B, CCC (which can be fine-tuned further with +'s and –'s). For example, in September 2014, the Barclays bond mentioned earlier had a credit rating of BBB– with Standard & Poor's.

Using credit ratings develops the two-state (default/no default) model into an *n*-state model. The state dynamics are represented by a time-homogenous Markov chain. The probability of going from one state to another depends only on the two states themselves (the *Markov property*) and this transition probability is assumed to be independent of time (*time homogeneity*).

The first state is the best credit quality and the n th state represents default, which is an absorbing state – there is no chance of recovering from default – and a payment of δ is made at maturity. The $n \times n$ transition matrix, $Q(t,T)$, of the Markov chain, can be obtained from credit-ratings agencies along with information on recovery rates.

So, for example, we could model the Standard & Poor's rating system if we used $n = 8$. This would give $n - 1 = 7$ credit ratings, from AAA (= State 1) down to CCC (= State 7), for companies that are not already in default. The nth state (= State 8) would be for companies that are already in default (which we assume they stay in for ever).

The following Core Reading table gives an example of a transition matrix, showing the probabilities (%) of jumping from one credit rating to another (and ultimately into 'default').

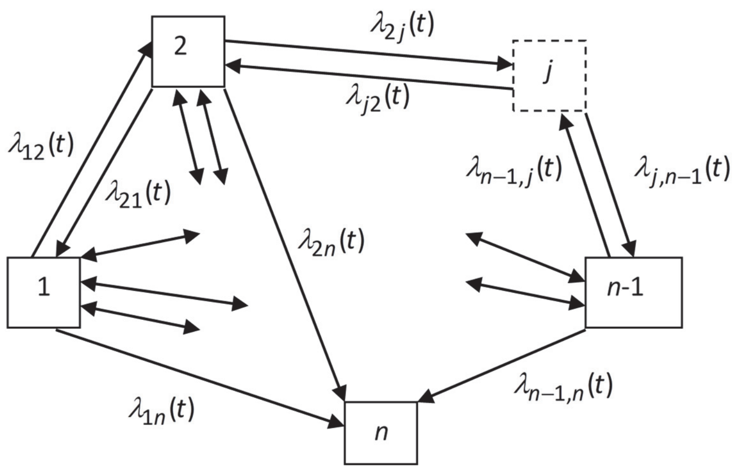
	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90.81	8.33	0.6	0.06	0.12	0.08	0.00	0.00
AA	0.70	90.65	7.79	0.64	0.06	0.14	0.02	0.00
A	0.09	2.27	91.05	5.52	0.74	0.26	0.01	0.06
BBB	0.02	0.33	5.95	86.93	5.30	1.17	0.12	0.18
BB	0.03	0.14	0.67	7.73	80.53	8.84	1.00	1.06
B	0.00	0.12	0.24	0.43	6.48	83.46	4.07	5.20
CCC	0.22	0.00	0.22	1.30	2.38	11.24	64.86	19.78

This means that the value of a credit-risky asset issued by a firm with rating *i* is given by:

$$v^i(t,T) = P(t,T)\left[1 - q_{in}(t,T) + \delta q_{in}(t,T)\right]$$

where q_{in} is the risk-neutral probability of a bond in State *i* at time *t* being in State *n* at time *T* (*ie* having defaulted). So bonds with the same credit rating and term are equally priced.

We define the transition intensity, under the risk-neutral measure Q, from State i to State j at time t to be $\lambda_{ij}(t)$. If the transition intensities $\lambda_{ij}(t)$ are assumed to be deterministic, then this model for default risk can be represented by the following diagram:



In this n-state model, transfer is possible between all states except for the default State n, which is absorbing. The λ's here correspond to the entries in the *generator matrix*. As shorthand we write:

$$\lambda_{ii}(t) = -\sum_{j \neq i} \lambda_{ij}(t)$$

The value of λ when the two subscripts are equal (*ie* λ_{ii}) does not correspond to an actual transition – if the process starts and ends in State i, it hasn't actually made a jump. However, defining λ_{ii} to equal minus the sum of all the actual transition rates out of State i allows us to write the set of equations for this process in a compact form.

Question

For a particular 4-state version of the JLT model:

$$\lambda_{12} = \lambda_{21} = 0.1\,,\ \lambda_{23} = 0.2\,,\ \lambda_{24} = 0.1\,,\ \lambda_{32} = 0.1\,,\ \lambda_{34} = 0.4\,,\ \lambda_{13} = \lambda_{14} = \lambda_{31} = 0$$

Construct the complete generator matrix.

Solution

Entering the values we've been given, we get:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} ? & 0.1 & 0 & 0 \\ 0.1 & ? & 0.2 & 0.1 \\ 0 & 0.1 & ? & 0.4 \\ ? & ? & ? & ? \end{bmatrix} \end{array}$$

We know that State 4 (the default state) is absorbing. So the first three entries along the bottom row will be zero.

The entries down the diagonal are equal to minus the sum of the other entries for that row. For example, $\lambda_{22} = -(0.1+0.2+0.1) = -0.4$.

So the final generator matrix looks like this:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} -0.1 & 0.1 & 0 & 0 \\ 0.1 & -0.4 & 0.2 & 0.1 \\ 0 & 0.1 & -0.5 & 0.4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

The generator matrix can be used to write down a set of differential equations (the Kolmogorov differential equations), which can then be solved to find the probabilities of making specified transitions over specified periods.

In fact it is possible to calculate the transition probabilities directly from the generator matrix, using a method involving matrix calculations, which we will describe now.

We introduce the following notation:

- $\Lambda(t)$ is an $n \times n$ generator matrix, $\Lambda(t) = \left(\lambda_{ij}(t)\right)_{i,j=1}^{n}$
- $q_{ij}(s,t) = P\left[X(t) = j \middle| X(s) = i\right]$ for $t > s$
- $Q(s,t) = \left(q_{ij}(s,t)\right)_{i,j=1}^{n}$ is the matrix of transition probabilities

Question

State in words what $\Lambda(t)$, $q_{ij}(s,t)$ and $Q(s,t)$ represent.

Solution

$\Lambda(t)$ is the generator matrix at time t. The positive entries in this matrix represent the transition rates (intensities) from one state to another at that time. The negative entries on the main diagonal are notional values equal to minus the sum of the other entries in the same row (which means that the entries in each row add up to zero).

$q_{ij}(s,t)$ is the probability that a company that is in State i at a particular time s will be in State j at a specified future time t. If s is the current time and t is the time a bond payment is due, this probability tells us how likely it is that the company (currently in State i) will be in each state – and hence how likely it is to default. So these are what we want to use the model to work out.

$Q(s,t)$ is a matrix showing the complete set of $n \times n$ transition probabilities for the period from time s to time t. It gives us the same information as working out all the individual probabilities.

It can be shown that:

$$Q(s,t) = \exp\left[\int_s^t \Lambda(u)du\right]$$

where, for an $n \times n$ matrix M, we define:

$$e^M \equiv I + \sum_{k=1}^{\infty} \frac{1}{k!} M^k$$

The proof of this is not required for this course.

The e^x function on your calculator can be generalised to an exponential function for matrices, written as $\exp(M)$ or e^M. You can see that its definition corresponds to the familiar series $e^x = 1 + x + \frac{1}{2!}x^2 + \cdots$.

In the matrix version, $I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$ is the identity matrix, which has 1's down the main diagonal and zeros everywhere else. The matrix exponential function shares most of the familiar properties of the scalar version, such as $\exp(\mathbf{0}) = I$, where $\mathbf{0}$ is a matrix consisting entirely of zeros, and $\exp(A)\exp(A) = \exp(2A)$.

Question

In a particular two-state intensity-based model with constant transition intensities, the integrated generator matrix $M = \int_s^t \Lambda(u)du$ has the form $M = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$, where $a = 0.2(t-s)$ and $b = 0.1(t-s)$.

(i) Show that, in this case, $M^2 = -(a+b)M$.

(ii) Hence deduce an explicit formula for e^M in terms of a and b.

(iii) Hence deduce a set of formulae for the transition probabilities $q_{ij}(s,t)$.

Solution

(i) If $M = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$, we find that:

$$M^2 = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}\begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$$

$$= \begin{pmatrix} (-a)(-a)+ab & (-a)a+a(-b) \\ b(-a)+(-b)b & ba+(-b)(-b) \end{pmatrix}$$

$$= \begin{pmatrix} a^2+ab & -a^2-ab \\ -ba-b^2 & ba+b^2 \end{pmatrix}$$

$$= \begin{pmatrix} a(a+b) & -a(a+b) \\ -b(a+b) & b(a+b) \end{pmatrix}$$

$$= -(a+b)\begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$$

$$= -(a+b)M$$

(ii) Extending this to higher powers, we see that:

$$M^3 = M^2M = -(a+b)MM = -(a+b)M^2 = (a+b)^2 M$$

and $\quad M^k = [-(a+b)]^{k-1}M$

So, in this case, the exponential function is:

$$\exp(M) = I + \sum_{k=1}^{\infty} \frac{1}{k!} M^k$$

$$= I + \sum_{k=1}^{\infty} \frac{1}{k!}[-(a+b)]^{k-1}M$$

$$= I - \frac{1}{a+b}\left(\sum_{k=1}^{\infty} \frac{1}{k!}[-(a+b)]^k\right)M$$

$$= I - \left(\frac{e^{-(a+b)}-1}{a+b}\right)M = I + \left(\frac{1-e^{-(a+b)}}{a+b}\right)M$$

(iii) We can then substitute the values of a and b to find the matrix of transition probabilities:

$$Q(s,t) = I + \left(\frac{1-e^{-(a+b)}}{a+b}\right)M$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \left(\frac{1-e^{-(a+b)}}{a+b}\right)\begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \left(\frac{1-e^{-0.3(t-s)}}{0.3}\right)\begin{pmatrix} -0.2 & 0.2 \\ 0.1 & -0.1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{3}+\frac{2}{3}e^{-0.3(t-s)} & \frac{2}{3}-\frac{2}{3}e^{-0.3(t-s)} \\ \frac{1}{3}-\frac{1}{3}e^{-0.3(t-s)} & \frac{2}{3}+\frac{1}{3}e^{-0.3(t-s)} \end{pmatrix}$$

So the transition probabilities are:

$$q_{11}(s,t) = \tfrac{1}{3}+\tfrac{2}{3}e^{-0.3(t-s)}\,, \quad q_{12}(s,t) = \tfrac{2}{3}-\tfrac{2}{3}e^{-0.3(t-s)}$$

$$q_{21}(s,t) = \tfrac{1}{3}-\tfrac{1}{3}e^{-0.3(t-s)}\,, \quad q_{22}(s,t) = \tfrac{2}{3}+\tfrac{1}{3}e^{-0.3(t-s)}$$

This seems rather trivial, however the approach can be enhanced if it is assumed, as in Jarrow-Lando-Turnbull, that the transition matrix, $Q(t,T)$, obeys the following relationship:

$$Q(t,T) = e^{\Lambda(T-t)}$$

In addition, assume that Λ, a hazard rate, is the generator matrix and is 'diagonalisable', meaning it can be written:

$$\Lambda = \Sigma D \Sigma^{-1}$$

where D is a diagonal matrix of the eigenvalues of Λ and Σ is a matrix whose columns are the eigenvectors of Λ. This means:

$$Q(t,T) = \Sigma e^{D(T-t)}\Sigma^{-1}$$

The elements, λ_{ij} of Λ have the following properties:

- $\lambda_{ij} \geq 0$
- $\sum_{j=1}^{n} \lambda_{ij} = 0$
- $\lambda_{nj} = 0$
- **state $i+1$ is always more risky that state i**

  So $\lambda_{in} < \lambda_{jn}$ for all $i < j$.

As a result:

$$q_{in}(t,T)=\sum_{j=1}^{n-1}\sigma_{ij}\hat{\sigma}_{jn}\left(e^{d_j(T-t)}-1\right),\ 1\le i\le n-1$$

where:

- σ_{ij} **is an element of** Σ
- $\hat{\sigma}_{jn}$ **is an element of** Σ^{-1}
- d_j **is an eigenvalue of** Λ**.**

6 Stochastic transition probabilities

The Jarrow-Lando-Turnbull approach assumes that the transition intensities between states are deterministic.

Although we have allowed the values of the transition intensities $\lambda(t)$ between any two states to vary over time, we have assumed that the functions involved are known with certainty at the outset. In real life, however, economic conditions can change unpredictably. If, for example, a recession struck, we would expect the $\lambda(t)$'s corresponding to jumps to a higher-numbered state (*ie* a state closer to the default State n) to increase significantly, as companies struggled to remain profitable.

An alternative approach would be to assume that the transition intensity between states, $\lambda(t)$, is stochastic and dependent on a separate state variable process $U(t)$.

By using a stochastic approach, $\lambda(t)$, can be allowed to vary with company fortunes and other economic factors. For example, a rise in interest rates may make default more likely, so $U(t)$ could include appropriate allowance for changes in interest rates. This approach can be used to develop models for credit risk that combine the structural modelling and intensity-based approaches.

In the previous section we had a generator matrix of the form:

$$\Lambda = \Sigma D \Sigma^{-1}$$

The approach now is to allow this matrix to vary in a stochastic way.

This additional level of complexity means that Λ can be made stochastic:

$$\Lambda = \Sigma D U(t) \Sigma^{-1}$$

where $U(t)$ is a stochastic process. This means that the transition probabilities can be conditioned on $U(t)$:

$$q_{in}(t,T) \mid U(t) = \sum_{j=1}^{n-1} \sigma_{ij} \hat{\sigma}_{jn} \left(E\left[e^{d_j \int_t^T U(s)ds} \right] - 1 \right), \quad 1 \le i \le n-1$$

The process $U(t)$ can be made as complex as required, generating multi-factor credit-risk models.

Question

In the trivial case where $U(t) = 1$ for all t, show that the above result recovers the deterministic version of $q_{in}(t,T)$ found in the previous section.

Solution

$$q_{in}(t,T)\,|\,(U(t)=1)=\sum_{j=1}^{n-1}\sigma_{ij}\hat{\sigma}_{jn}\left(E\left[e^{d_j\int_t^T 1ds}\right]-1\right)$$

$$=\sum_{j=1}^{n-1}\sigma_{ij}\hat{\sigma}_{jn}\left(E\left[e^{d_j(T-t)}\right]-1\right)$$

$$=\sum_{j=1}^{n-1}\sigma_{ij}\hat{\sigma}_{jn}\left(e^{d_j(T-t)}-1\right)$$

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 19 Summary

Credit events and recovery rates

A *credit event* is an event that will trigger the default of a bond. In the event of a default, the fraction δ of the defaulted amount that can be recovered through bankruptcy proceedings or some other form of settlement is known as the *recovery rate*.

Modelling credit risk

Structural models are models for a company issuing both shares and bonds, which aim to link default events explicitly to the fortunes of the issuing company.

Reduced-form models are statistical models that use observed market statistics such as credit ratings.

Intensity-based models are a particular type of continuous-time reduced-form models. They typically model the 'jumps' between different states (usually credit ratings) using transition intensities.

The Merton model

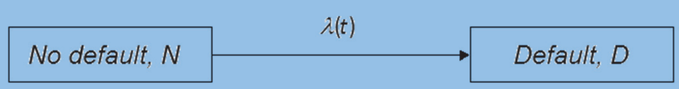
Merton's model is a structural model. It assumes that a company has issued both equity and debt such that its total value at time t is $F(t)$. The total value of the bonds issued and the shareholders' interests equals $F(t)$.

The shareholders of the company can be regarded as having a European call option on the assets of the company with maturity T and a strike price equal to the face value (L) of the debt.

The Merton model is tractable and gives us some insight into the nature of default and the interaction between bondholders and shareholders. It can be used to estimate either the risk-neutral probability that the company will default or the credit spread on the debt.

The two-state model for credit ratings

- State N = not previously defaulted
- State D = previously defaulted



Let $v(t,T)$ be the price at time t of a risky zero-coupon bond that matures at time T. Then:

$$v(t,T) = e^{-r(T-t)} E_Q\left[\text{payoff at time } T \middle| F_t\right]$$

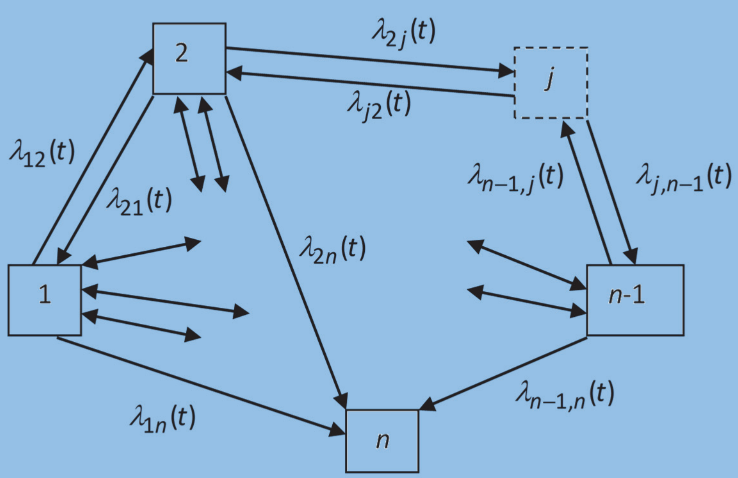
$$= e^{-r(T-t)} E_Q\left[1_{\{T<\tau\}} + \delta 1_{\{\tau \le T\}} \middle| F_t\right]$$

$$= e^{-r(T-t)} \left[1 - (1-\delta)\left\{1 - \exp\left(-\int_t^T \lambda(s)ds\right)\right\}\right]$$

where:

- δ is the recovery rate
- $1_{\{\}}$ is the indicator function
- $\lambda(s)$ is the risk-neutral transition rate or intensity.

The Jarrow-Lando-Turnbull (JLT) model



In this n-state model, transfer is possible between all states except for State n (default), which is absorbing. If $X(t)$ is the state or credit rating at time t, then, for $i = 1, 2, \ldots, n-1$, the transition probabilities over the time interval (s,t) are:

$$q_{ij}(s,t) = Q\left(X(t) = j \,\middle|\, X(s) = i\right) \text{ for } t > s$$

The matrix of transition probabilities is:

$$Q(s,t) = \left(q_{ij}(s,t)\right)_{i,j=1}^{n}$$

then:

$$Q(s,t) = \exp\left[\int_s^t \Lambda(u)\,du\right]$$

where:

$$\Lambda(t) = \left(\lambda_{ij}(t)\right)_{i,j=1}^{n} \text{ is the matrix of transition intensities.}$$

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 19 Practice Questions

19.1 To fund an expansion in its operations, a company has just issued 5-year zero-coupon bonds with a total face value of £10 million, taking its total asset value up to £15 million.

Exam style

(i) Explain how the value of the bonds can be expressed in terms of a European put option. [3]

(ii) Hence calculate the fair price of a holding of the company bonds with a face value of £100 using the Black-Scholes model, given that the price of a 5-year zero-coupon government bond is £77.88. Assume that the annualised volatility of the company's assets over the 5-year period is 25%. [4]

(iii) Explain what is meant by a credit spread and calculate its value for the company bonds. [3]

[Total 10]

19.2 Company X has just issued some 5-year zero-coupon bonds. A continuous-time two-state model is to be used to model the status of the company and to calculate the fair price of the bonds. It is believed that the risk-neutral transition rate for failure of the company is $\lambda(t) = 0.002t$, where t is the time in years since the issue of the bonds. The 5-year risk-free spot yield is 5.25% expressed as an annual effective rate.

Exam style

(i) Calculate the risk-neutral probability that the company will have failed by the end of 5 years. [2]

(ii) In the event of failure of the company, the bonds will make a reduced payment at the maturity date. The recovery rate for a payment due at time t is:

$$\delta(t) = 1 - 0.05t$$

Calculate the fair price to pay for £100 nominal of a Company X bond, taking into account the possibility of company failure. [3]

(iii) An analyst is concerned that the estimate of $\lambda(t)$ may be too simplistic. Explain the possible reasons for his concern and how the model could be developed to deal with this. [3]

[Total 8]

19.3 (i) Explain what is meant by a default-free bond.

(ii) State the possible outcomes of a default.

(iii) List five types of credit event.

(iv) Explain what is meant by the recovery rate for a bond.

19.4 Company X has the following financial structure at time 0:

Debt	£3m (current book value)
Equity	£6m (issued share capital)

The debt is a zero-coupon bond with face value £5m that is repayable at par at time 10.

There are 400,000 shares in circulation.

(i) Explain how the Merton model could be used to value shares in Company *X*.

(ii) Assuming that the debt is repaid directly from the company's funds at that time, state the share price at time 10 if the total value of Company *X* at that time is:

(a) £15m

(b) £4m

19.5 A two-state model is to be used to model the probability that a bond defaults:

$$\boxed{\textit{No default, N}} \xrightarrow{\lambda(t)} \boxed{\textit{Default, D}}$$

where $\lambda(t) = \dfrac{5+20t-t^2}{500}$, $0 \le t \le 20$.

(i) Calculate the probability that the bond does not default between times 5 and 10.

(ii) Explain how the model may be modified to allow the default intensity $\lambda(t)$ to depend on future unforeseen events such as a sudden downturn in the economy.

19.6 A company has just issued 4-year zero-coupon bonds with a nominal value of £4 million. The total value of the company now stands at £7.5 million. A constant risk-free rate of return of 2% *pa* continuously-compounded is available in the market.

Exam style

(i) Use the Merton model to calculate the theoretical price of £100 nominal of the company's bonds, assuming that the annual volatility of the value of the company's assets is 30%. [4]

(ii) Estimate the risk-neutral probability of default on the company's bonds. [3]
[Total 7]

19.7 An analyst is using a two-state continuous-time model to study the credit risk of zero-coupon bonds issued by different companies.

Exam style

The risk-neutral transition intensity function is:

- $\lambda_A(s) = 0.0148$ for Company A, and
- $\lambda_B(s) = 0.01s^2$ for Company B

where s measures time in years from now.

The analyst observes that the credit spread on a 3-year zero-coupon bond just issued by Company B is twice that on a 3-year zero-coupon bond just issued by Company A.

(i) Given that the risk-free force of interest is 5% *pa,* and that the average recovery rate in the event of default, δ, where $0 < \delta < 1$, is the same for both companies, calculate δ. [7]

(ii) Explain how the two-state model for credit risk can be generalised to give the Jarrow-Lando-Turnbull model. [2]

[Total 9]

19.8 The credit-worthiness of debt issued by companies is assessed at the end of each year by a credit rating agency. The ratings are A (the most credit-worthy), B and D (debt defaulted). Historical evidence supports the view that the credit rating of a debt can be modelled as a Markov chain with the following matrix of one-year transition probabilities:

Exam style

$$\mathbf{X} = \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}$$

(i) Determine the probability that a company rated A will never be rated B in the future. [2]

(ii) (a) Calculate the two-year transition probabilities of the Markov chain.

(b) Hence calculate the expected number of defaults within the next two years from a group of 100 companies, all initially rated A. [2]

The manager of a portfolio investing in company debt follows a "downgrade trigger" strategy. Under this strategy, any debt in a company whose rating has fallen to B at the end of a year is sold and replaced with debt in an A-rated company.

(iii) Calculate the expected number of defaults for this investment manager over the next two years, given that the portfolio initially consists of 100 A-rated bonds. [2]

(iv) Comment on the suggestion that the downgrade trigger strategy will improve the return on the portfolio. [2]

[Total 8]

19.9 A bond analyst is going to use a two-state intensity-based model to derive risk-neutral transition intensities.

(i) State the general risk-neutral pricing formula for a ZCB subject to default risk in a two-state intensity-based model with a deterministic transition intensity. Define all notation used.

(ii) Starting from the formula in (i), derive the corresponding formula for the risk-neutral transition intensity function in terms of the ZCB price.

Chapter 19 Solutions

19.1 (i) ***Expressing the bond value as an option***

This is the Merton model. After the bond issue the total current value of the company will be $F(0)=15$ (working in £million). In 5 years' time the company will have an unknown value of $F(5)$. [1]

The bondholders have first call on the company's assets at that time. So they will receive 10 if $F(5)\geq 10$ and $F(5)$ if $F(5)<10$. [1]

So the redemption payment from the bonds is $\min[F(5),10]$, which can be written as $10-\max[10-F(5),0]$. The function $\max[10-F(5),0]$ is the payoff for a European put option on $F(t)$ maturing at time 5 with strike price 10. [1]

[Total 3]

(ii) ***Calculate the fair price of the bonds***

The parameters for valuing the put option with payoff $\max[10-F(5),0]$ using the Black-Scholes formula are:

$F(0)=15\ (=S_0)$, $K=10$, $T-t=5$, $\sigma=0.25$ (and $q=0$) [1]

The risk-free interest rate is found from the equation:

$100e^{-5r}=77.88 \quad \Rightarrow r=0.050$ [½]

Using page 47 of the *Tables*, we find that:

$d_1=1.45204$, $d_2=0.89302$ [½]

$$p=10e^{-0.25}\Phi(-0.89302)-15\Phi(-1.45204)$$

$$=7.788\times 0.18592-15\times 0.07325=0.349$$ [1]

So the total value of the bonds now (remembering that the 10 in the payoff function is a future payment and needs to be discounted) is:

$Ke^{-r(T-t)}-p=7.788-0.349=7.439$ *ie* £7.439 million [½]

So the fair price is £74.39 per £100 face value. [½]

[Total 4]

(iii) ***Credit spread***

The price of the company bonds (£74.39) is less than the price of the government bonds (£77.88) because of the risk of default, *ie* the company may not make the redemption payment in full on the due date. [1]

As a result, the yield r_B (assuming full payment) will be slightly higher. It can be found from the equation:

$$100e^{-5r_B} = 74.39 \quad \Rightarrow r = 0.0592 \qquad [1]$$

The difference between the bond yield of 5.92% and the default-free rate of 5% is called the *credit spread*. So here the credit spread is 0.92% (per annum, continuously compounded). [1]

[Total 3]

19.2 (i) ***Probability of company failure***

The risk-neutral probability of company failure, by time n, can be expressed in terms of the transition rate $\lambda(t)$:

$$p(n) = 1 - \exp\left(-\int_0^n \lambda(t)dt\right) \qquad [½]$$

With the transition rate given, we have:

$$p(n) = 1 - \exp\left(-\int_0^n 0.002t\,dt\right)$$

$$= 1 - \exp\left[-0.001t^2\right]_0^n$$

$$= 1 - e^{-0.001n^2} \qquad [1]$$

For $n = 5$ this is:

$$p(5) = 1 - e^{-0.001 \times 5^2} = 0.02469 \qquad [½]$$

[Total 2]

(ii) ***Fair price of a Company X bond***

The recovery rate at time 5 is:

$$\delta(5) = 1 - 0.05(5) = 0.75 \qquad [1]$$

The risk-neutral expected payment at maturity can then be found using the probability calculated in part (i):

$$0.02469 \times 0.75 + (1 - 0.02469) \times 1 = 0.99383 \qquad [1]$$

We can discount this using the 5-year spot rate to get the fair price of the bond:

$$0.99383 \times (1.0525)^{-5} = 0.7695$$

ie £76.95 per £100 nominal. [1]

[Total 3]

(iii) ***Stochastic transition rates***

The analyst may be concerned because the fair price of the bond is critically dependent on an accurate assessment of the default transition intensity. If this assessment is incorrect then the bond could be mispriced. [1]

An alternative approach would be to assume that the transition intensity $\lambda(t)$ is stochastic and dependent on a separate state variable process, $X(t)$ say. [½]

By using a stochastic approach, $\lambda(t)$ can be allowed to vary with company fortunes and other economic factors. [½]

For example, a rise in interest rates may make default more likely and so $X(t)$ could include appropriate allowance for changes in interest rates. This approach can be used to develop models for credit risk that combine the structural modelling and intensity-based approaches. [1]
[Total 3]

19.3 (i) ***Meaning of default-free***

A bond is default-free if the stream of payments due from the bond will definitely be paid in full and on time.

(ii) ***Outcomes of a default***

The outcome of a default may be that the contracted payment stream is:

- rescheduled
- cancelled by the payment of an amount which is less than the default-free value of the original contract
- continued but at a reduced rate
- totally wiped out.

(iii) ***Types of credit events***

A credit event is an event that will trigger the default of a bond and includes the following:

- actions that are associated with bankruptcy or insolvency laws
- rating downgrade of the bond by a rating agency such as Standard & Poor's or Moody's
- failure to pay
- repudiation/moratorium
- restructuring – where the terms are changed to become less favourable to the bond holder.

(iv) ***Meaning of recovery rate***

In the event of a default, the fraction of the defaulted amount that can be recovered through bankruptcy proceedings or some other form of settlement is known as the recovery rate.

19.4 (i) ***Expressing the equity as a call option***

The Merton model values shares as call options on the company's assets with a strike price equal to the face value of Company X's debt.

The equity value at time 10 will be:

$$\max\{F(10)-5,0\}$$

where $F(10)=$ total value of Company X at time 10.

Company X will default on payment of the debt if the total value of its assets at time 10 is less than the promised debt repayment at that time.

There are 400,000 shares, so the (theoretical) share price at time 10 in £m will be:

$$\frac{\max\{F(10)-5,0\}}{400{,}000}$$

An appropriate option pricing formula can then be used to value this 'call option' at time 0.

(ii) ***Share price at time 10***

The share price at time 10 will be:

(a) $$\frac{\max\{15-5,0\}}{400{,}000}=\frac{\text{£}10\text{m}}{400{,}000}=\text{£}25 \text{ per share}$$

(b) Here the share price at time 10 will be 0 because the value of the outstanding debt exceeds the total value of the company.

19.5 (i) ***Probability the bond does not default between times 5 and 10***

The probability that the bond does not default between times 5 and 10 is:

$$\exp\left\{-\int_5^{10}\lambda(t)dt\right\}=\exp\left\{-\frac{1}{500}\int_5^{10}\left(5+20t-t^2\right)dt\right\}$$

$$=\exp\left\{-\frac{1}{500}\left[5t+10t^2-\frac{t^3}{3}\right]_5^{10}\right\}$$

$$=\exp\left\{-\frac{1}{500}\left[716\tfrac{2}{3}-233\tfrac{1}{3}\right]\right\}$$

$$=e^{-0.9667}=0.3803$$

(ii) ***Incorporating unforeseen events***

Unforeseen events can be considered as random and so a stochastic approach would be needed.

By using a stochastic approach, $\lambda(t)$ can be allowed to vary with company fortunes and other economic factors.

This is done by introducing an additional stochastic process $X(t)$, which models a suitable economic indicator, such as interest rates.

For example, a downturn in the economy may make default more likely and so $\lambda(t)$ could include appropriate allowance for this possibility.

19.6 (i) ***Theoretical price of the bonds***

The approach we will take is to value the shareholders' funds and then subtract this from the total value of the company to determine the bondholders' funds.

Under the Merton model, the shareholders in the company receive a payoff after 4 years equivalent to that from a call option with strike price equal to the amount to be repaid to the bondholders. [1]

The current value of the shareholding can be assessed using the Black-Scholes formula for the value of a call option, with parameters:

$$S_0 = 7.5, K = 4, \sigma = 30\%, r = 2\%, T - t = 4, q = 0$$ [½]

Letting $E(0)$ represent the value of the shareholding at time 0:

$$E(0) = 7.5\Phi(d_1) - 4e^{-0.02\times4}\Phi(d_2)$$

where:

$$d_1 = \frac{\ln\left(7.5/4\right) + \left(0.02 + \frac{1}{2}(0.3)^2\right)\times 4}{0.3\times 2} = 1.4810 \;\Rightarrow\; \Phi(d_1) = 0.93069$$

and: $d_2 = d_1 - 0.3\times 2 = 0.8810 \;\Rightarrow\; \Phi(d_2) = 0.81084$

So:

$$E(0) = 7.5\times 0.93069 - 4e^{-0.02\times4}\times 0.81084 = £3.986 \text{ million}$$ [1½]

Therefore the value of £4 million nominal of bonds at time 0, $B(0)$, is:

$$B(0) = 7.5 - 3.986 = £3.514 \text{ million}$$ [½]

The theoretical price of £100 nominal of these bonds is:

$$\frac{3.514}{4}\times 100 = £87.85$$ [½]

[Total 4]

Another way of doing this question is to note that the payoff on the bonds in 4 years' time is:

$$\min(4, F(4))$$

where $F(t)$ *is the value of the company's assets at time* t *.*

This can be re-expressed as:

$$4 - \max(4 - F(4), 0)$$

where $\max(4 - F(4), 0)$ *is the payoff on a 4-year put option on the company's assets with strike price of 4.*

The value of this payoff at time 0 is:

$$B(0) = 4e^{-0.02\times4} - p_0$$

where:

$$p_0 = 4e^{-0.02\times4}\Phi(-d_2) - 7.5\Phi(-d_1)$$

As before, $d_1 = 1.4810$ *and* $d_2 = 0.8810$ *, so:*

$$\Phi(-d_1) = \Phi(-1.4810) = 1 - \Phi(1.4810) = 0.06931$$

and:

$$\Phi(-d_2) = \Phi(-0.8810) = 1 - \Phi(0.8810) = 0.18916$$

This gives:

$$p_0 = 4e^{-0.02\times4} \times 0.18916 - 7.5 \times 0.06931 = 0.17866$$

so:

$$B(0) = 4e^{-0.02\times4} - 0.17866 = 3.514$$

as before.

(ii) ***Risk-neutral probability of default***

In the standard Black-Scholes formula for the price of a call option, $\Phi(d_2)$ represents the risk-neutral probability that the option $\Phi(d_2)$ will be exercised, or, equivalently, the risk-neutral probability that the share price at expiry exceeds the strike price. Under the Merton model approach, the call option replicates the shareholders' position and is exercised if the shareholders repay the bondholders in full. [1]

So, is equal to the probability that the bondholders are repaid in full, or, equivalently, the probability that the company does not default. This means that the probability of default is:

$$1 - \Phi(d_2)$$ [1]

In this case, the probability of default is:

$$1-\Phi(0.8810)=1-0.81084=0.18916$$ [1]

[Total 3]

Alternatively, this can be derived from first principles.

Under the assumptions of the Black-Scholes formula, the value of the company's assets at time t, $F(t)$, given $F(0)$, follows a lognormal distribution:

$$F(t)|F(0) \sim \log N\left(\ln F(0)+(r-\tfrac{1}{2}\sigma^2)t, \sigma^2 t\right)$$

Here:

$$F(4)|F(0) \sim \log N\left(\ln 7.5+(0.02-\tfrac{1}{2}0.3^2)\times 4, 0.3^2\times 4\right)$$

$$\Rightarrow \quad F(4)|F(0) \sim \log N\left(\ln 7.5-0.1, 0.36\right)$$

The company will default on the bonds if the value of the company at time 4, $F(4)$, is less than the amount to be repaid of £4 million.

So, the probability of default is:

$$\begin{aligned} P(F(4)<4) &= P(\ln F(4)<\ln 4) \\ &= P\left(Z<\frac{\ln 4-(\ln 7.5-0.1)}{\sqrt{0.36}}\right) \\ &= P(Z<-0.8810) \\ &= 1-\Phi(0.8810) \\ &= 0.18916 \end{aligned}$$

19.7 (i) ***Value of δ***

The general formula for the price of a zero-coupon bond under the two-state model for credit risk using a risk-neutral probability measure is:

$$v(t,T)=e^{-r(T-t)}\left[1-(1-\delta)\left(1-\exp\left(-\int_t^T \tilde{\lambda}(s)ds\right)\right)\right]$$ [1]

For the bond issued by Company A:

$$B_A=e^{-0.05\times 3}\left[1-(1-\delta)\left(1-\exp\left(-\int_0^3 0.0148\,ds\right)\right)\right]$$

$$=e^{-0.15}\left[1-(1-\delta)\left(1-e^{-0.0444}\right)\right]$$ [1]

For the bond issued by Company B:

$$B_B = e^{-0.05\times 3}\left[1-(1-\delta)\left(1-\exp\left(-\int_0^3 0.01s^2\,ds\right)\right)\right]$$

$$= e^{-0.15}\left[1-(1-\delta)\left(1-\exp\left(-\left[\frac{0.01}{3}s^3\right]_0^3\right)\right)\right]$$

$$= e^{-0.15}\left[1-(1-\delta)\left(1-e^{-0.09}\right)\right] \qquad [1]$$

The credit spread on a zero-coupon bond is the difference between the yield on the bond and the yield on a similar bond issued by the government, which we take here to be the risk-free force of interest of 5% *pa*.

If C_i is the credit spread on the zero-coupon bond issued by Company i and r_i is the continuously-compounded yield on the zero-coupon bond issued by Company i, then:

$$C_A = r_A - 0.05 \;\Rightarrow\; r_A = C_A + 0.05$$

$$C_B = r_B - 0.05 = 2C_A \;\Rightarrow\; r_B = 2C_A + 0.05 \qquad [½]$$

We can express the price of each zero-coupon bond in terms of the continuously-compounded yield on the bond, so:

$$B_A = e^{-3r_A} = e^{-3(C_A+0.05)}$$

$$B_B = e^{-3r_B} = e^{-3(2C_A+0.05)} \qquad [½]$$

This gives the simultaneous equations:

$$e^{-3(C_A+0.05)} = e^{-0.15}\left[1-(1-\delta)\left(1-e^{-0.0444}\right)\right]$$

$$e^{-3(2C_A+0.05)} = e^{-0.15}\left[1-(1-\delta)\left(1-e^{-0.09}\right)\right]$$

Cancelling the $e^{-0.15}$ terms gives:

$$e^{-3C_A} = 1-(1-\delta)\left(1-e^{-0.0444}\right) \qquad (1)$$

$$e^{-6C_A} = 1-(1-\delta)\left(1-e^{-0.09}\right) \qquad (2) \qquad [1]$$

Squaring Equation (1) and substituting it into Equation (2):

$$\left(1-(1-\delta)\left(1-e^{-0.0444}\right)\right)^2 = 1-(1-\delta)\left(1-e^{-0.09}\right)$$

Expanding the left-hand side:

$$1-2(1-\delta)\left(1-e^{-0.0444}\right)+(1-\delta)^2\left(1-e^{-0.0444}\right)^2=1-(1-\delta)\left(1-e^{-0.09}\right)$$

$$\Rightarrow -2(1-\delta)\left(1-e^{-0.0444}\right)+(1-\delta)^2\left(1-e^{-0.0444}\right)^2=-(1-\delta)\left(1-e^{-0.09}\right)$$

Cancelling $(1-\delta)$ on both sides, since $0<\delta<1$, and solving for δ:

$$-2\left(1-e^{-0.0444}\right)+(1-\delta)\left(1-e^{-0.0444}\right)^2=-\left(1-e^{-0.09}\right)$$

$$\Rightarrow \quad 1-\delta=\frac{2\left(1-e^{-0.0444}\right)-\left(1-e^{-0.09}\right)}{\left(1-e^{-0.0444}\right)^2}=\frac{0.0007887}{0.0018861}=0.418164$$

$$\Rightarrow \quad \delta=0.5818$$ [2]

[Total 7]

(ii) ***Jarrow-Lando-Turnbull model***

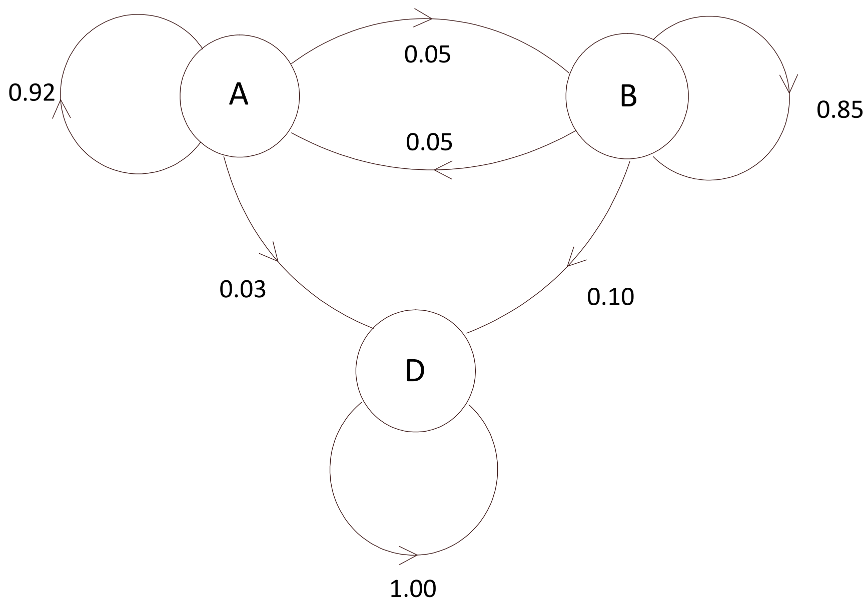
Instead of the simple default / no default two-state model, a more general model has been developed by Jarrow, Lando and Turnbull, in which there are n states. The n states relate to $n-1$ possible credit ratings for a non-defaulted company, and one default state. [1½]

Transitions are possible between all states, except for the default state, which is absorbing (*ie* once a company has entered the default state, it cannot leave it). [½]

[Total 2]

19.8 (i) ***Probability A never rated B in the future***

We have the following diagram of the one-year transition probabilities:



A company that is never rated B in the future will:

(a) remain in State A for some period of time, and

(b) will then move to State D and remain there.

So we can sum over all the times at which the single transition from State A to State D can take place. This gives us the following expression:

$$0.03+0.92\times 0.03+(0.92)^2\times 0.03+(0.92)^3\times 0.03+\cdots \qquad [1]$$

This is an infinite geometric progression, whose sum is:

$$\frac{0.03}{1-0.92}=0.375$$

So the probability that a company is never rated B in the future is 0.375. [1]

[Total 2]

(ii)(a) ***Two-year transition probabilities***

The two-year transition probabilities are given by:

$$\mathbf{X}^2=\begin{pmatrix}0.92 & 0.05 & 0.03\\ 0.05 & 0.85 & 0.1\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}0.92 & 0.05 & 0.03\\ 0.05 & 0.85 & 0.1\\ 0 & 0 & 1\end{pmatrix}=\begin{pmatrix}0.8489 & 0.0885 & 0.0626\\ 0.0885 & 0.7250 & 0.1865\\ 0 & 0 & 1\end{pmatrix} \qquad [1]$$

(ii)(b) ***Expected number of defaults***

The probability that a company rated A at time zero is in State D at time 2 is 0.0626. So the expected number of companies in this state out of 100 is 6.26. [1]

[Total 2]

(iii) ***Expected number of defaults***

For this manager we use the original matrix $\mathbf{X}$. After one year, the expected number of companies in each state will be:

$$\begin{pmatrix}100 & 0 & 0\end{pmatrix}\begin{pmatrix}0.92 & 0.05 & 0.03\\ 0.05 & 0.85 & 0.1\\ 0 & 0 & 1\end{pmatrix}=\begin{pmatrix}92 & 5 & 3\end{pmatrix} \qquad [1]$$

If the five state B's are replaced with State A's and the process repeated, we have:

$$\begin{pmatrix}97 & 0 & 3\end{pmatrix}\begin{pmatrix}0.92 & 0.05 & 0.03\\ 0.05 & 0.85 & 0.1\\ 0 & 0 & 1\end{pmatrix}=\begin{pmatrix}89.24 & 4.85 & 5.91\end{pmatrix}$$

So the expected number of defaults by the end of the second year under this arrangement is 5.91. [1]

[Total 2]

Alternatively we could have used the revised transition matrix:

$$\mathbf{X} = \begin{pmatrix} 0.97 & 0 & 0.03 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and calculated $\mathbf{X}^2$ *as before.*

Note that the question does not say that the companies in default at the end of the first year are replaced by grade A companies in the second year. If it had done so, then there would have been three expected defaults each year, and the answer would be 6.

(iv) ***Comment***

The downgrade trigger strategy will reduce the expected number of defaults, as we have seen. However, the return on the portfolio will also be a function of the yields on the debt. Companies rated B are likely to have bonds with a higher yield (because of the higher risk), so excluding these may in fact reduce the yield on the portfolio. [1½]

Also, the actual number of defaults may not match the expected number. The return depends on the actual progress of the portfolio, rather than the expected outcome. [½]

[Total 2]

19.9 (i) ***General risk-neutral pricing formula for a ZCB***

$$v(t,T) = e^{-r(T-t)}\left[1-(1-\delta)\left\{1-\exp\left(-\int_t^T \lambda(s)ds\right)\right\}\right]$$

where:

- t and T are the current time and the maturity date of the ZCB
- r is the constant risk-free force of interest
- δ is the assumed (constant) recovery rate
- $\lambda(s)$ is the risk-neutral transition intensity.

(ii) ***Derive formula for risk-neutral transition intensity function***

Multiplying both sides of the ZCB pricing formula by $e^{r(T-t)}$ and then multiplying out the right-hand side, gives:

$$e^{r(T-t)}v(t,T)=1-\left\{1-\exp\left(-\int_t^T \lambda(s)ds\right)\right\}+\delta\left\{1-\exp\left(-\int_t^T \lambda(s)ds\right)\right\}$$

$$=\exp\left(-\int_t^T \lambda(s)ds\right)+\delta-\delta\exp\left(-\int_t^T \lambda(s)ds\right)$$

$$=(1-\delta)\exp\left(-\int_t^T \lambda(s)ds\right)+\delta$$

Moving the constant δ to the left-hand side, then taking logs, we get:

$$e^{r(T-t)}v(t,T)-\delta=(1-\delta)\exp\left(-\int_t^T \lambda(s)ds\right)$$

$$\log\left[e^{r(T-t)}v(t,T)-\delta\right]=\log(1-\delta)-\int_t^T \lambda(s)ds$$

Differentiating both sides with respect to T then gives:

$$\frac{\partial}{\partial T}\log\left[e^{r(T-t)}v(t,T)-\delta\right]=-\lambda(T)$$

Remember that differentiating an integral (with respect to the upper limit) takes you back to the original function.

Finally, renaming the variable T as s, swapping the two sides and flipping the signs gives:

$$\lambda(s)=-\frac{\partial}{\partial s}\log\left[e^{r(s-t)}v(t,s)-\delta\right]$$

20

Ruin theory

Syllabus objectives

5.1 Ruin theory

5.1.1 Explain what is meant by the aggregate claim process and the cash-flow process for a risk.

5.1.2 Use the Poisson process and the distribution of inter-event times to calculate probabilities of the number of events in a given time interval and waiting times.

5.1.3 Define a compound Poisson process and calculate probabilities using simulation.

5.1.4 Define the probability of ruin in infinite/finite and continuous/discrete time and state and explain relationships between the different probabilities of ruin.

5.1.5 Describe the effect on the probability of ruin, in both finite and infinite time, of changing parameter values by reasoning or simulation.

5.1.6 Calculate probabilities of ruin by simulation.

0 Introduction

In this chapter we consider the aggregate claims $S(t)$ arising up to time t. If $N(t)$ is the number of claims arising by time t, and X_i is the amount of the i-th claim, then $S(t) = X_1 + X_2 + \cdots + X_{N(t)}$. $N(t)$ is called a Poisson process and $S(t)$ is called a compound Poisson process. We can use $S(t)$ to model claims received by an insurance company and hence consider the probability that this insurance company is *ruined*.

We start with the notation and the other basic concepts before giving formal definitions of both the Poisson process and the compound Poisson process. We will also introduce the concept of a premium security loading. Briefly, this is an additional amount charged on an insurance premium to reduce the likelihood of an insurance company becoming ruined.

Later we will introduce the *adjustment coefficient*, a parameter associated with risk, and *Lundberg's inequality*.

We will consider the effect of changing parameter values on the probability of ruin for an insurance company before finally considering the impact of introducing reinsurance.

1 Basic concepts

1.1 Notation

One technical point needed later in this chapter is that a function $f(x)$ is described as being $o(x)$ as x goes to zero, if:

$$\lim_{x \to 0} \frac{f(x)}{x} = 0$$

You can use this notation to simplify your working. For example, the function $g(x) = 3x + 0.5x^2 + 0.004x^3$ can be rewritten as $g(x) = 3x + o(x)$, since $\dfrac{0.5x^2 + 0.004x^3}{x} \to 0$ as $x \to 0$. Note that $o(x)$ does not represent an actual number so that $c \times o(x)$ (c is a constant), $-o(x)$ and $o(x)$ are all equivalent.

Question

Which of the following functions are $o(x)$ as $x \to 0$?

(i) x^2 (ii) e^x (iii) $e^{-x} - 1 + x$

Solution

(i) Yes

(ii) No

(iii) Yes, because if we expand e^{-x} as a power series and simplify, we get:

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots$$

This gives $\dfrac{e^{-x} - 1 + x}{x} = \dfrac{1}{2}x +$ terms in x^2 and higher powers.

In the actuarial literature, the word 'risk' is often used instead of the phrase 'portfolio of policies'. In this chapter both terms will be used, so that by a 'risk' will be meant either a single policy or a collection of policies. This chapter will focus on claims generated by a portfolio over successive time periods. Some notation is needed.

- **$N(t)$ the number of claims generated by the portfolio in the time interval [0, t], for all $t \geq 0$**
- **X_i the amount of the i-th claim, i = 1, 2, 3, ...**
- **$S(t)$ the aggregate claims in the time interval [0, t], for all $t \geq 0$.**

$\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables. $\{N(t)\}_{t\geq 0}$ and $\{S(t)\}_{t\geq 0}$ are both families of random variables, one for each time $t \geq 0$; in other words $\{N(t)\}_{t\geq 0}$ and $\{S(t)\}_{t\geq 0}$ are stochastic processes.

You can think of a stochastic process as being a whole family of different random variables. Consider a time line. On the line there are an infinite number of different time intervals. For each interval of time, there is a random variable that corresponds to the aggregate claim amount arising in that time interval. This is what we mean here by a stochastic process.

It can be seen that:

$$S(t) = \sum_{i=1}^{N(t)} X_i$$

with the understanding that $S(t)$ is zero if $N(t)$ is zero.

The stochastic process $\{S(t)\}_{t\geq 0}$ as defined above is known as the aggregate claims process for the risk. The random variables $N(1)$ and $S(1)$ represent the number of claims and the aggregate claims respectively from the portfolio in the first unit of time.

The insurer of this portfolio will receive premiums from the policyholders. It is convenient at this stage to assume, as will be assumed throughout this chapter, that the premium income is received continuously and at a constant rate. Here is some more notation:

c = the rate of premium income per unit time

so that the total premium income received in the time interval $[0, t]$ is ct. It will also be assumed that c is strictly positive.

1.2 The surplus process

Suppose that at time 0 the insurer has an amount of money set aside for this portfolio. This amount of money is called the initial surplus and is denoted by U. It will always be assumed that $U \geq 0$. The insurer needs this initial surplus because the future premium income on its own may not be sufficient to cover the future claims. Here we are ignoring expenses. The insurer's surplus at any future time $t\,(>0)$ is a random variable since its value depends on the claims experience up to time t. The insurer's surplus at time t is denoted by $U(t)$.

The following formula for $U(t)$ can be written:

$$U(t) = U + ct - S(t)$$

In words this formula says that the insurer's surplus at time *t* is the initial surplus plus the premium income up to time *t* minus the aggregate claims up to time *t*. Notice that the initial surplus and the premium income are not random variables since they are determined before the risk process starts. The above formula is valid for $t \geq 0$ with the understanding that $U(0)$ is equal to *U*. For a given value of *t*, $U(t)$ is a random variable because $S(t)$ is a random variable. Hence $\{U(t)\}_{t \geq 0}$ is a stochastic process, which is known as the cash flow process or surplus process.

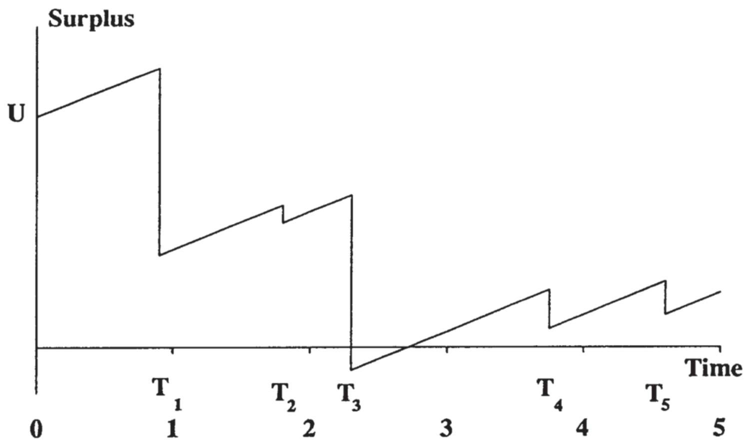


Figure 1

Figure 1 shows one possible outcome of the surplus process. Claims occur at times T_1, T_2, T_3, T_4 and T_5 and at these times the surplus immediately falls by the amount of the claim. Between claims the surplus increases at constant rate *c* per unit time. The model being used for the insurer's surplus incorporates many simplifications, as will any model of a complex real-life operation. Some important simplifications are that it is assumed that claims are settled as soon as they occur and that no interest is earned on the insurer's surplus. Despite its simplicity this model can give an interesting insight into the mathematics of an insurance operation.

We are also assuming that there are no expenses associated with the process (or, equivalently, that $S(t)$ makes allowance for expense amounts as well as claim amounts), and that the insurer cannot vary the premium rate c.

We are also ignoring the possibility of reinsurance. Simple forms of reinsurance will be incorporated into the model later in this chapter.

1.3 The probability of ruin in continuous time

It can be seen from Figure 1 that the insurer's surplus falls below zero as a result of the claim at time T_3. Speaking loosely for the moment, when the surplus falls below zero the insurer has run out of money and it is said that *ruin* has occurred. In this simplified model, the insurer will want to keep the probability of this event, that is, the probability of ruin, as small as possible, or at least below a predetermined bound. Still speaking loosely, ruin can be thought of as meaning insolvency, although determining whether or not an insurance company is insolvent is, in practice, a very complex problem. Another way of looking at the probability of ruin is to think of it as the probability that, at some future time, the insurance company will need to provide more capital to finance this particular portfolio.

Now to be more precise. The following two probabilities are defined:

$$\psi(U) = P[U(t) < 0, \text{ for some } t, \ 0 < t < \infty]$$

$$\psi(U, t) = P[U(\tau) < 0, \text{ for some } \tau, \ 0 < \tau \leq t]$$

$\psi(U)$ is the probability of ultimate ruin (given initial surplus U) and $\psi(U,t)$ is the probability of ruin within time t (given initial surplus U). These probabilities are sometimes referred to as the probability of ruin in infinite time and the probability of ruin in finite time. Here are some important logical relationships between these probabilities for $0 < t_1 \leq t_2 < \infty$ and for $0 \leq U_1 \leq U_2$:

$$\psi(U_2, t) \leq \psi(U_1, t) \qquad (1.1)$$

$$\psi(U_2) \leq \psi(U_1) \qquad (1.2)$$

$$\psi(U,t_1) \leq \psi(U,t_2) \leq \psi(U) \qquad (1.3)$$

$$\lim_{t\to\infty} \psi(U,t) = \psi(U) \qquad (1.4)$$

The intuitive explanations for these relationships are as follows:

The larger the initial surplus, the less likely it is that ruin will occur either in a finite time period, hence (1.1), or an unlimited time period, hence (1.2).

For a given initial surplus U, the longer the period considered when checking for ruin, the more likely it is that ruin will occur, hence (1.3).

Finally, the probability of ultimate ruin can be approximated by the probability of ruin within finite time t provided t is sufficiently large, hence (1.4).

Question

What is $\lim_{u\to\infty} \psi(u,t)$?

Solution

As the amount of initial surplus increases, ruin will become less and less likely. So the limit is zero.

You may be wondering whether it is possible to find numerical values for these ruin probabilities. In some very simple cases it is. However, for most practical situations, finding an exact value for the probability of ruin is impossible. In some cases there are useful approximations to $\psi(u)$, even if calculation of an exact value is not possible.

1.4 The probability of ruin in discrete time

The two probabilities of ruin considered so far have been continuous time probabilities of ruin, so-called because they check for ruin in continuous time. In practice it may be possible (or even desirable) to check for ruin only at discrete intervals of time.

For a given interval of time, denoted h, the following two discrete time probabilities of ruin are defined:

$$\psi_h(U) = P\left[U(t) < 0, \text{ for some } t,\ t = h, 2h, 3h, \ldots\right]$$

$$\psi_h(U,t) = P\left[U(\tau) < 0, \text{ for some } \tau,\ \tau = h, 2h, \ldots, t-h, t\right]$$

Note that it is assumed for convenience in the definition of $\psi_h(U, t)$ that t is an integer multiple of h. Figure 2 shows the same realisation of the surplus process as given in Figure 1 but assuming now that the process is checked only at discrete time intervals. The black markers show the values of the surplus process at integer time intervals (*ie* $h = 1$); the black markers together with the white ones show the values of the surplus process at time intervals of length ½.

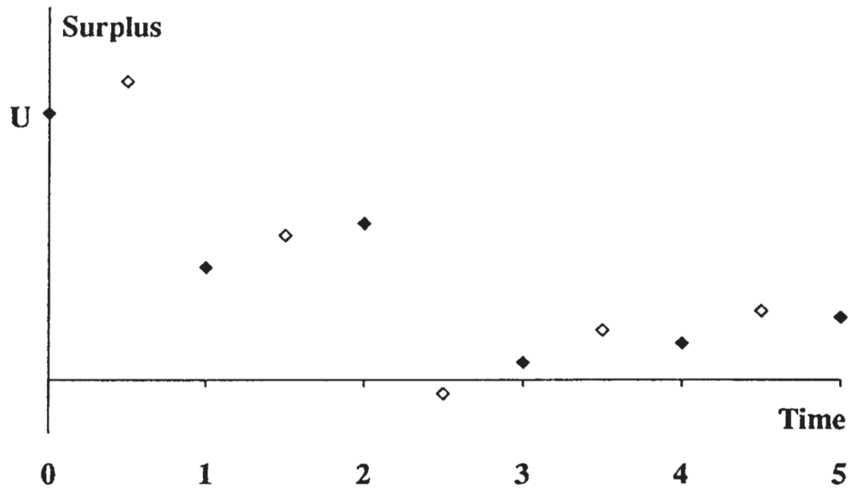


Figure 2

It can be seen from Figure 2 that in discrete time with $h = 1$, ruin does not occur for this realisation of the surplus process before time 5, but ruin does occur (at time 2½) in discrete time with $h = ½$.

Listed below are five relationships between different discrete time probabilities of ruin for $0 \le U_1 \le U_2$ and for $0 \le t_1 \le t_2 \le \infty$. Formulae (1.5), (1.6), (1.7) and (1.8) are the discrete time versions of formulae (1.1), (1.2), (1.3) and (1.4) above and their intuitive explanations are similar. The intuitive explanation of (1.9) comes from Figure 2.

$$\psi_h(U_2,t) \le \psi_h(U_1,t) \quad (1.5)$$

$$\psi_h(U_2) \le \psi_h(U_1) \quad (1.6)$$

$$\psi_h(U,t_1) \le \psi_h(U,t_2) \le \psi_h(U) \quad (1.7)$$

$$\lim_{t\to\infty} \psi_h(U,t) = \psi_h(U) \quad (1.8)$$

$$\psi_h(U,t) \le \psi(U,t) \quad (1.9)$$

Question

Explain why Equation 1.9 is true.

Solution

$\psi(U,t)$ involves checking for ruin at all possible times. Since the more often we check for ruin, the more likely we are to find it, we would expect that $\psi(U,t)$ would be greater than $\psi_h(U,t)$.

Intuitively, it is expected that the following two relationships are true since the probability of ruin in continuous time could be approximated by the probability of ruin in discrete time, with the same initial surplus, *U*, and time horizon, *t*, provided ruin is checked for sufficiently often, *ie* provided *h* is sufficiently small.

$$\lim_{h\to 0+} \psi_h(U,t) = \psi(U,t) \quad (1.10)$$

$$\lim_{h\to 0+} \psi_h(U) = \psi(U) \quad (1.11)$$

Formulae (1.10) and (1.11) are true but the proofs are rather messy and will not be given here.

2 The Poisson and compound Poisson processes

2.1 Introduction

In this section some assumptions will be made about the claim number process, $\{N(t)\}_{t\geq 0}$, and the claim amounts, $\{X_i\}_{i=1}^{\infty}$. The claim number process will be assumed to be a Poisson process, leading to a compound Poisson process $\{S(t)\}_{t\geq 0}$ for aggregate claims. The assumptions made in this section will hold for the remainder of this chapter.

2.2 The Poisson process

We use the term "Poisson process" to describe the number of claims arising from a time period of length t.

If the number of claims N arising from a single time period has a Poisson distribution with parameter λ then the number of claims $N(t)$ which arise over a time period of length t is a Poisson process, *ie* $N(t)$ has a Poisson distribution with parameter λt .

The Poisson process is an example of a counting process. Here the number of claims arising from a risk is of interest. Since the number of claims is being counted over time, the claim number process $\{N(t)\}_{t\geq 0}$ must satisfy the following conditions:

(i) $N(0)=0$, ***ie* there are no claims at time 0**

(ii) **for any $t>0$, $N(t)$ must be integer valued**

(iii) **when $s<t$, $N(s)\leq N(t)$, *ie* the number of claims over time is non-decreasing**

(iv) **when $s<t$, $N(t)-N(s)$ represents the number of claims occurring in the time interval $(s,t]$.**

The claim number process $\{N(t)\}_{t\geq 0}$ is defined to be a Poisson process with parameter λ if the following conditions are satisfied:

(i) $N(0)=0$, **and** $N(s)\leq N(t)$ **when** $s<t$

(ii) $P\left(N(t+h)=r \mid N(t)=r\right)=1-\lambda h+o(h)$

$$P\left(N(t+h)=r+1 \mid N(t)=r\right)=\lambda h+o(h) \tag{2.1}$$

$$P\left(N(t+h)>r+1 \mid N(t)=r\right)=o(h)$$

(iii) **when $s<t$, the number of claims in the time interval $(s,t]$ is independent of the number of claims up to time s.** **(2.2)**

Condition (ii) implies that there can be a maximum of one claim in a very short time interval h. It also implies that the number of claims in a time interval of length h does not depend on when that time interval starts.

Question

Explain how motor insurance claims could be represented by a Poisson process.

Solution

The events in this case are occurrences of claim events (*ie* accidents, fires, thefts *etc*) or claims reported to the insurer. The parameter λ represents the average rate of occurrence of claims (*eg* 50 per day), which we are assuming remains constant throughout the year and at different times of day. The assumption that, in a sufficiently short time interval, there can be at most one claim is satisfied if we assume that claim events cannot lead to multiple claims (*ie* no motorway pile-ups *etc*).

When studying a Poisson process the distribution of the time to the first claim and the times between claims is often of particular interest.

Time to the first claim

This section will show that the time to the first claim has an exponential distribution with parameter λ.

Let the random variable T_1 denote the time of the first claim. Then, for a fixed value of t, if no claims have occurred by time t, $T_1 > t$. Hence:

$$P(T_1 > t) = P(N(t) = 0) = \exp\{-\lambda t\}$$

This last step follows from the formula for the probability function of a $Poisson(\lambda t)$ distribution with $x = 0$.

And:

$$P(T_1 \leq t) = 1 - \exp\{-\lambda t\}$$

so that T_1 has an exponential distribution with parameter λ.

This is because the RHS matches the formula for the distribution function of an exponential distribution.

The time to the first claim in a Poisson process has an exponential distribution with parameter λ.

Time between claims

This section will show that the time between claims has an exponential distribution with parameter λ.

For $i = 2,3,\ldots$, let the random variable T_i denote the time between the $(i-1)$th and the ith claims. Then:

$$P\left(T_{n+1} > t \,\middle|\, \sum_{i=1}^{n} T_i = r\right) = P\left(\sum_{i=1}^{n+1} T_i > t + r \,\middle|\, \sum_{i=1}^{n} T_i = r\right)$$
$$= P\big(N(t+r) = n \,\big|\, N(r) = n\big)$$
$$= P\big(N(t+r) - N(r) = 0 \,\big|\, N(r) = n\big)$$

Then we use the independence of claim numbers in different time periods to remove the conditioning.

By condition (2.2):

$$P(N(t+r) - N(r) = 0 \,|\, N(r) = n) = P(N(t+r) - N(r) = 0)$$

Finally:

$$P(N(t+r) - N(r) = 0) = P(N(t) = 0) = \exp\{-\lambda t\}$$

since the number of claims in a time interval of length r does not depend on when that time interval starts (condition (2.1)). Thus inter-event times also have an exponential distribution with parameter λ.

The time between claims in a Poisson process has an exponential distribution with parameter λ.

Note that the inter-event time is independent of the absolute time. In other words the time until the next event has the same distribution, irrespective of the time since the last event or the number of events that have already occurred. This is referred to as the *memoryless property* of the exponential distribution.

Question

If reported claims follow a Poisson process with rate 5 per day (and the insurer has a 24 hour hotline), calculate:

(i) the probability that there will be fewer than 2 claims reported on a given day

(ii) the probability that another claim will be reported during the next hour.

Solution

(i) The expected number of claims reported on a given day is 5. So the number of claims reported on a given day has a $Poisson(5)$ distribution and the probability that there will be fewer than 2 claims is:

$$P(N<2) = P(N=0) + P(N=1) = e^{-5} + 5e^{-5} = 0.040$$

Here we have used the formula for the Poisson probability, but alternatively, using page 176 of the *Tables*, $P(N<2) = P(N \le 1) = 0.04043$.

(ii) The waiting time until the next event has an $Exp(5)$ distribution. We need to find a probability using the exponential distribution. To do this, we can use the cumulative distribution function:

$$P(T<t)=1-e^{-\lambda t}$$

So the probability that there will be a claim (or several claims) during the next hour ($\frac{1}{24}$ of a day) is:

$$P(T<\tfrac{1}{24})=1-e^{-\frac{5}{24}}=0.1881$$

Note that it is unlikely that the rate would be constant over time in reality.

2.3 The compound Poisson process

In this section the Poisson process for the number of claims will be combined with a claim amount distribution to give a compound Poisson process for the aggregate claims.

The following three important assumptions are made:

- **the random variables $\{X_i\}_{i=1}^{\infty}$ are independent and identically distributed**
- **the random variables $\{X_i\}_{i=1}^{\infty}$ are independent of $N(t)$ for all $t \geq 0$**
- **the stochastic process $\{N(t)\}_{t\geq 0}$ is a Poisson process whose parameter is denoted λ.**

This last assumption means that for any $t \geq 0$, the random variable $N(t)$ has a Poisson distribution with parameter λt, so that:

$$P[N(t)=k]=\exp\{-\lambda t\}\frac{(\lambda t)^k}{k!} \quad \text{for } k=0,1,2,\ldots$$

With these assumptions the aggregate claims process, $\{S(t)\}_{t\geq 0}$, is called a compound Poisson process with Poisson parameter λ. By comparing the assumptions above with the assumptions in Section 0, it can be seen that the connection between the two is that if $\{S(t)\}_{t\geq 0}$ is a compound Poisson process with Poisson parameter λ, then, for a fixed value of t (≥ 0), $S(t)$ has a compound Poisson distribution with Poisson parameter λt.

Note the slight change in terminology here: 'Poisson parameter λ' becomes 'Poisson parameter λt' when a change is made from the process to the distribution.

The common distribution function of the X_i s will be denoted $F(x)$ and it will be assumed for the remainder of this chapter that $F(0)=0$ so that all claims are for positive amounts.

$F(x)$ is defined to be $P(X \le x)$. In the continuous case we would find $F(x)$ by integrating the probability density function (PDF):

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

The probability density function of the X_is, if it exists, will be denoted $f(x)$ and the kth moment about zero of the X_is, if it exists, will be denoted m_k, so that:

$$m_k = E\left[X_i^k\right] \text{ for } k = 1, 2, 3, \ldots$$

Whenever the common moment generating function of the X_is exists, its value at the point r will be denoted by $M_X(r)$.

The definition of a moment generating function is:

$$M_X(r) = E[e^{rX}]$$

Note that we are using r for the dummy variable to avoid confusion with time t.

Since, for a fixed value of t, $S(t)$ has a compound Poisson distribution, it follows from an earlier subject that the process $\{S(t)\}_{t \ge 0}$ has mean $\lambda t m_1$, variance $\lambda t m_2$, and moment generating function $M_S(r)$, where:

$$M_S(r) = \exp\{\lambda t(M_X(r) - 1)\}$$

Subject CS2 shows that if:

$$S = X_1 + X_2 + \cdots + X_N$$

where $N \sim Poisson(\lambda)$, then:

$$E[S] = \lambda E[X] = \lambda m_1 \qquad Var(S) = \lambda E\left[X^2\right] = \lambda m_2$$

$$M_S(r) = M_N\left[\ln M_X(r)\right]$$

These formulae can be found on page 16 of the *Tables*.

Here $N(t) \sim Poisson(\lambda t)$, therefore $E\left[S(t)\right] = \lambda t m_1$ and $Var\left(S(t)\right) = \lambda t m_2$. Also $M_{N(t)}(r) = \exp\left(\lambda t\left(e^r - 1\right)\right)$, so:

$$M_{S(t)}(r) = \exp\left(\lambda t(e^{\ln M_x(r)} - 1)\right) = \exp\left(\lambda t\left(M_X(r) - 1\right)\right)$$

For the remainder of this chapter the following (intuitively reasonable) assumption will be made concerning the rate of premium income:

$$c > \lambda m_1 \tag{2.3}$$

so that the insurer's premium income (per unit time) is greater than the expected claims outgo (per unit time).

Question

Why is this intuitively reasonable?

Solution

Otherwise the insurer would be charging premiums that were less than the amount it expected to pay out in claims.

In the real world this assumption may not always be true, especially during periods of competitive pressure when premium rates are soft.

2.4 Probability of ruin in the short term

If we know the distribution of the aggregate claims $S(t)$, we can often determine the probability of ruin for the discrete model over a finite time horizon directly (without reference to the models), by looking at the cashflows involved.

Question

The claims arising during each year from a particular type of annual insurance policy are assumed to follow a normal distribution with mean $0.7P$ and standard deviation $2.0P$, where P is the annual premium. Claims are assumed to arise independently. Insurers assess their solvency position at the end of each year.

A small insurer with an initial surplus of £0.1m expects to sell 100 policies at the beginning of the coming year in respect of identical risks for an annual premium of £5,000. The insurer incurs expenses of $0.2P$ at the time of writing each policy. Calculate the probability that the insurer will prove to be insolvent at the end of the coming year. Ignore interest.

Solution

Using the information given, the insurer's surplus at the end of the coming year will be:

$$U(1) = \text{initial surplus} + \text{premiums} - \text{expenses} - \text{claims}$$

$$= 0.1\text{m} + 100 \times 5{,}000 - 100 \times 0.2 \times 5{,}000 - S(1)$$

$$= 0.5\text{m} - S(1)$$

The distribution of $S(1)$ is:

$$S(1) \sim N\left[100 \times 0.7 \times 5{,}000,\ 100 \times \left(2.0 \times 5{,}000\right)^2\right] = N\left[0.35\text{m}, \left(0.1\text{m}\right)^2\right]$$

So the probability that the surplus will be negative is:

$$P\left[U(1) < 0\right] = P\left[S(1) > 0.5\text{m}\right]$$

$$= P\left[N\left[0.35\text{m}, \left(0.1\text{m}\right)^2\right] > 0.5\text{m}\right]$$

$$= 1 - \Phi\left(\frac{0.5\text{m} - 0.35\text{m}}{0.1\text{m}}\right) = 1 - 0.93319 = 0.067$$

Question

If the insurer expects to sell 200 policies during the second year for the same premium and expects to incur expenses at the same rate, calculate the probability that the insurer will prove to be insolvent at the end of the second year.

Solution

The insurer's surplus at the end of the second year will be:

$$U(2) = \text{initial surplus + premiums} - \text{expenses} - \text{claims}$$

$$= 0.1\text{m} + (100 + 200) \times 5{,}000 - (100 + 200) \times 0.2 \times 5{,}000 - S(2)$$

$$= 1.3\text{m} - S(2)$$

The distribution of $S(2)$ is:

$$S(2) \sim N(300 \times 0.7 \times 5{,}000,\ 300 \times (2.0 \times 5{,}000)^2) = N[1.05\text{m}, (0.173\text{m})^2]$$

The probability that the surplus will be negative at the end of the second year is:

$$P[U(2) < 0] = P[S(2) > 1.3\text{m}]$$

$$= P(N[1.05\text{m}, (0.173\text{m})^2] > 1.3\text{m})$$

$$= 1 - \Phi\left(\frac{1.3\text{m} - 1.05\text{m}}{0.173\text{m}}\right)$$

$$= 1 - \Phi(1.443) = 0.074$$

The normal distribution is probably not a very realistic distribution to use for the claim amount distribution in most portfolios, as it is symmetrical, whereas many claim amount portfolios will have skewed underlying distributions. However, it is commonly used in exam questions.

Question

The number of claims from a portfolio of policies has a Poisson distribution with parameter 30 per year. The individual claim amount distribution is lognormal with parameters $\mu = 3$ and $\sigma^2 = 1.1$. The rate of premium income from the portfolio is 1,200 per year.

If the insurer has an initial surplus of 1,000, estimate the probability that the insurer's surplus at time 2 will be negative, by assuming that the aggregate claims distribution is approximately normal.

Solution

First we need the mean and variance of the aggregate claims in a two-year period. The expected number of claims will be 60. So the mean and variance are (using the formulae for the first two moments of a lognormal distribution):

$$E\left[S(2)\right] = 60e^{3+0.55} = 2{,}088.80$$

and $$Var\left(S(2)\right) = 60e^{6+2.2} = 218{,}457$$

Ruin will occur if $S(2)$ is greater than the initial surplus plus premiums received. So we want:

$$P\left[S(2) > 1{,}000 + 2\times 1{,}200\right] \approx P\left[N(0,1) > \frac{3{,}400 - 2{,}088.80}{\sqrt{218{,}457}}\right] = 1 - \Phi(2.8053) = 0.0025$$

The probability of ruin is approximately 0.25%.

2.5 Premium security loadings

So far, we have used c to denote the rate of premium income per unit time, independent of the claims outgo. In some circumstances it is more useful to think of the rate of premium income as being related to the rate of claims outgo.

For the insurer to survive, the rate at which premium income comes in needs to be greater than the rate at which claims are paid out. If this is not true, the insurer is certain to be ruined at some point.

Sometimes c will be written as:

$$c = (1+\theta)\lambda m_1$$

where $\theta\,(>0)$ is the premium loading factor.

The security loading is the percentage by which the rate of premium income exceeds the rate of claims outgo. So, for the Poisson process outlined above, we have:

$$c = (1+\theta)E(S) = (1+\theta)\lambda m_1$$

where θ is the security loading. θ is also sometimes called the 'relative security loading'. It might typically be a figure such as 0.2, *ie* 20%.

The insurer will need to adopt a positive security loading when pricing policies, in order to cover expenses, profit, contingency margins and so on.

Note that this does not mean that ruin is impossible. It is quite possible for the actual claims outgo to exceed substantially its expected value. So even in this situation the insurer's probability of ruin is non-zero.

Mean, variance and MGF of the total claim amount

For a compound Poisson process $S(t)$, the mean and variance of the total claim amount are given by:

$$E\left[S(t)\right] = \lambda t E\left[X\right] \qquad Var\left(S(t)\right) = \lambda t E\left[X^2\right]$$

The moment generating function of the process is given by:

$$M_{S(t)}(r) = \exp\left(\lambda t(M_X(r) - 1\right)$$

2.6 A technicality

In the next section a technical result will be needed concerning $M_X(r)$ (the moment generating function of the individual claim amount distribution), which, for convenience, will be presented here. It will be assumed throughout the remainder of this chapter that there is some number γ $(0 < \gamma \le \infty)$ such that $M_X(r)$ is finite for all $r < \gamma$ and:

$$\lim_{r \to \gamma^-} M_X(r) = \infty \qquad \textbf{(2.4)}$$

(For example, if the X_i s have a range bounded above by some finite number, then γ will be ∞; if the X_i s have an exponential distribution with parameter α, then γ will be equal to α.)

Suppose for example that claim amounts have a continuous uniform distribution on the interval $(0,10)$, so that they are bounded above by 10. Then the moment generating function of the claim distribution is (from the *Tables*):

$$M_X(r) = \frac{e^{10r} - 1}{10r}$$

This is defined for all positive values of r, and so in this case $\gamma = \infty$. We can see that as $r \to \infty$, the limit of the MGF is infinite. If the claim distribution is $Exp(\alpha)$, the MGF (as stated in the *Tables*) is:

$$M_X(r) = (1 - r/\alpha)^{-1} = \frac{\alpha}{\alpha - r}$$

This tends to infinity as r tends to α from below.

In the next section the following result will be needed:

$$\lim_{r \to \gamma^-} \left(\lambda M_X(r) - cr \right) = \infty \qquad \textbf{(2.5)}$$

If γ is finite, (2.5) follows immediately from (2.4).

This is because λ, c and r would all have finite values in the limit.

Now it will be shown that (2.5) holds when γ is infinite. This requires a little more care. First note that there is a positive number, ε say, such that:

$$P\left[X_i > \varepsilon\right] > 0$$

The reason for this is that all claim amounts are positive.

So, if we pick a small enough number ($\varepsilon = 0.01$ maybe), we're bound to get a proportion of claims whose amount exceeds this.

This probability will be denoted by π. Then:

$$M_X(r) \geq e^{r\varepsilon}\pi$$

This follows by considering the claims below and above ε:

$$\begin{aligned} M_X(r) &= E\left[e^{rX}\right] = E\left[e^{rX} \mid X \leq \varepsilon\right] P(X \leq \varepsilon) + E\left[e^{rX} \mid X > \varepsilon\right] P(X > \varepsilon) \\ &\geq 0 + e^{r\varepsilon}\pi \end{aligned}$$

Hence:

$$\lim_{r \to \infty} (\lambda M_X(r) - cr) \geq \lim_{r \to \infty} (\lambda e^{r\varepsilon}\pi - cr) = \infty$$

Here the $e^{r\varepsilon}$ term is tending to $+\infty$, while the $-cr$ term is tending to $-\infty$. Remember that in such cases the exponential term always 'wins'. So the limit is $+\infty$.

3 The adjustment coefficient and Lundberg's inequality

This section will look at the probability of ruin and introduce the *adjustment coefficient*, a parameter associated with risk. The letters R and r will be used interchangeably for the adjustment coefficient.

3.1 Lundberg's inequality

Lundberg's inequality states that:

$$\psi(U) \leq \exp\{-RU\}$$

where U is the insurer's initial surplus and $\psi(U)$ is the probability of ultimate ruin. R is a parameter associated with a surplus process known as the adjustment coefficient. Its value depends upon the distribution of aggregate claims and on the rate of premium income. Before defining R the importance of the result and some features of the adjustment coefficient will be illustrated.

Don't worry at this stage about what R actually represents. It will be defined shortly. Until then just think of it as a parameter associated with the surplus process.

Note that if we can find a value for R, then Lundberg's inequality tells us that we can find an upper bound for the probability of ruin. This is a very useful result.

Figure 3 shows a graph of both $\exp\{-RU\}$ and $\psi(U)$ against U when claim amounts are exponentially distributed with mean 1, and when the premium loading factor is 10%. (The solution for R will be found in Section 3.2. The formula for $\psi(U)$ is given in Section 4.2.)

In fact it can be shown that the value of R in this case is $\dfrac{10}{11}$.

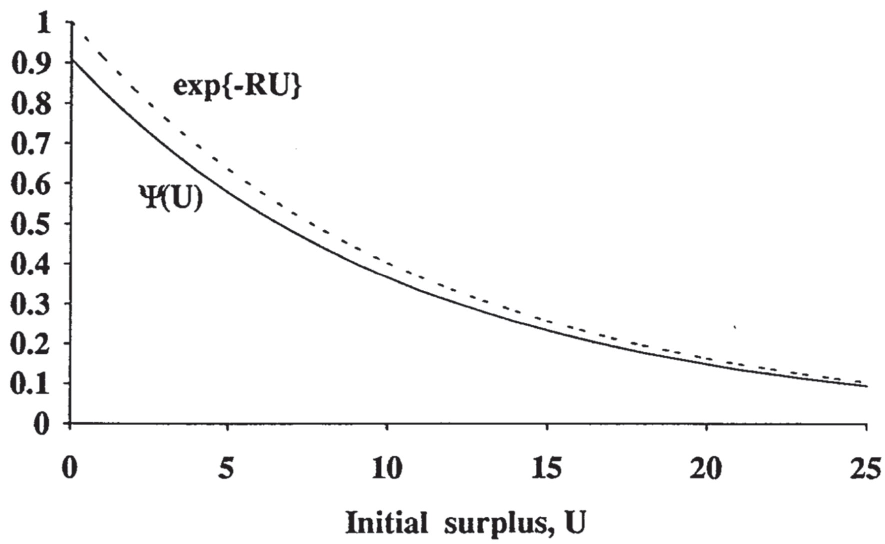


Figure 3 – Exponentially distributed claims with mean 1 and premium loading factor 10%

It can be seen that, for large values of U, $\psi(U)$ is very close to the upper bound, so that $\psi(U) \approx \exp\{-RU\}$.

In the actuarial literature, $\exp\{-RU\}$ is often used as an approximation to $\psi(U)$.

R can be interpreted as measuring risk. The larger the value of R, the smaller the upper bound for $\psi(U)$ will be. Hence, $\psi(U)$ would be expected to decrease as R increases. R is a function of the parameters that affect the probability of ruin, and R's behaviour as a function of these parameters can be observed.

Note that R is an *inverse* measure of risk. Larger values of R imply smaller ruin probabilities, and *vice versa*.

Figure 4 shows a graph of R as a function of the loading factor, θ, when:

(i) the claim amount distribution is exponential with mean 10, and

(ii) all claims are of amount 10.

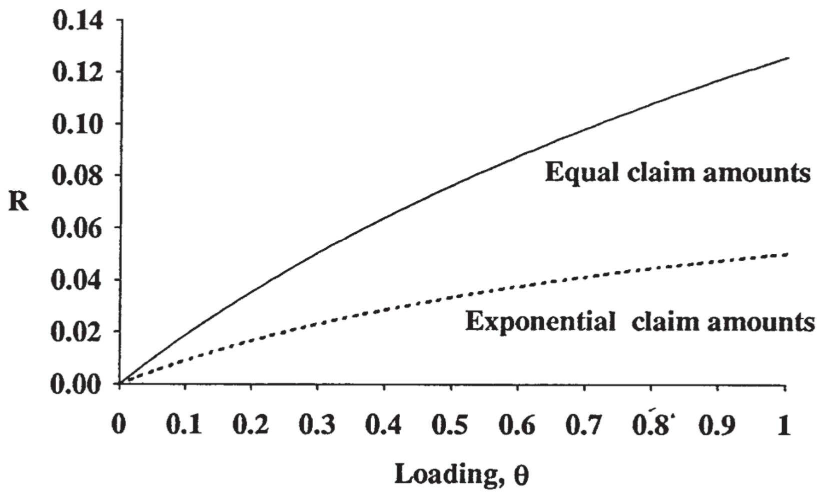


Figure 4 – R by loading

Note that in both cases, R is an increasing function of θ. This is not surprising as $\psi(U)$ would be expected to be a decreasing function of θ, and since $\psi(U) \approx \exp\{-RU\}$, any factor causing a decrease in $\psi(U)$ would cause R to increase.

Question

Why would $\psi(U)$ be expected to be a decreasing function of θ?

Solution

θ is the security loading, and $\psi(U)$ is the probability of ruin for a fixed level of surplus U. If θ increases the premiums we charge will increase, and we should become more secure, *ie* the probability of ruin should fall.

Note also that the value of R when claim amounts are exponentially distributed is less than the value of R when all claim amounts are 10. Again, this result is not surprising. Both claim amount distributions have the same mean, but the exponential distribution has greater variability. Greater variability is associated with greater risk, and hence a larger value of $\psi(U)$ would be expected for the exponential distribution, and a lower value of R.

This example illustrates that R is affected by the premium loading factor and by the characteristics of the individual claim amount distribution. R is now defined and shown, in general, to encapsulate all the factors affecting a surplus process.

3.2 The adjustment coefficient – compound Poisson processes

The surplus process depends on the initial surplus, on the aggregate claims process and on the rate of premium income. The adjustment coefficient is a parameter associated with a surplus process which takes account of two of these factors: aggregate claims and premium income. The adjustment coefficient gives a measure of risk for a surplus process. When aggregate claims are a compound Poisson process, the adjustment coefficient is defined in terms of the Poisson parameter, the moment generating function of individual claim amounts and the premium income per unit time.

The adjustment coefficient, denoted R, is defined to be the unique positive root of:

$$\lambda M_X(r) - \lambda - cr = 0 \qquad (3.1)$$

So, R is given by:

$$\lambda M_X(R) = \lambda + cR \qquad (3.2)$$

Note that, although R relates to the aggregate claims, the MGF used in the definition is for the *individual* claim amount.

It is probably not at all obvious to you at this stage *why* R is defined in this way. The reason is bound up with the proof of Lundberg's inequality, which you are not required to know. Please accept the definition, so that you can find the value of R in simple cases.

Note that equation (3.1) implies that the value of the adjustment coefficient depends on the Poisson parameter, the individual claim amount distribution and the rate of premium income. However, writing $c = (1+\theta)\lambda m_1$ gives:

$$M_X(r) = 1 + (1+\theta)m_1 r$$

so that R is independent of the Poisson parameter and simply depends on the loading factor, θ, and the individual claim amount distribution.

You can see from this equation that all the λ's have cancelled.

Question

An insurer knows from past experience that the number of claims received per month has a Poisson distribution with mean 15, and that claim amounts have an exponential distribution with mean 500. The insurer uses a security loading of 30%. Calculate the insurer's adjustment coefficient and give an upper bound for the insurer's probability of ruin, if the insurer sets aside an initial surplus of 1,000.

Solution

The equation for the adjustment coefficient is:

$$M_X(r) = 1 + (1+\theta)m_1 r$$

We have $X \sim Exp\left(\dfrac{1}{500}\right)$, so that $M_X(r) = (1-500r)^{-1}$ (this comes from the *Tables*), $\theta = 0.3$, and $m_1 = E[X] = 500$. Substituting these into the equation:

$$(1-500r)^{-1} = 1 + 1.3 \times 500r = 1 + 650r$$

Rearranging:

$$1 = (1-500r)(1+650r)$$

$$1 = 1 - 500r + 650r - 325{,}000r^2$$

$$0 = 150 - 325{,}000r$$

$$\Rightarrow \quad r = 0.000462$$

From Lundberg's inequality, $\psi(U) \le e^{-rU}$, so here:

$$\psi(U) \le e^{-0.000462 \times 1{,}000} = 0.630$$

Note that we didn't use the Poisson parameter in our solution.

We will see another example of finding the adjustment coefficient shortly.

Question

In the previous question we ignored the fact that r could be 0. Why?

Solution

The first reason is that Core Reading defines R to be the unique *positive* root, so R cannot be zero.

The second reason is that $R = 0$ should *always* be a solution to the equation. Why? Consider the LHS. $M_X(0) = E[e^0] = 1$. Consider the RHS. $1 + (1+\theta)m_1 \times 0 = 1$. So $R = 0$ is *always* a solution, but why do we ignore it?

Consider Lundberg's inequality, $\psi(U) \le e^{-RU}$. If $R = 0$, we have an upper bound of 1 for the probability of ruin. That should have been obvious!

In practice, ignore $R = 0$, just as we have done here.

It may not be obvious to you why R does not depend on the Poisson parameter. The basic reason is that increasing the Poisson parameter speeds up the whole process, so that claims arise more quickly. This means that ruin, if it is going to happen, will happen sooner, rather than later. However it does not affect the probability that ruin does actually occur, when we are considering ruin at any time in the future.

It can be shown that there is indeed only one positive root of (3.1) as follows.

Define $g(r) = \lambda M_X(r) - \lambda - cr$ and consider the graph of $g(r)$ over the interval $[0, \gamma]$. Note first that $g(0) = 0$.

This is what we discovered in the previous example.

Further, $g(r)$ is a decreasing function at $r = 0$ since:

$$\frac{d}{dr}g(r) = \lambda \frac{d}{dr}M_X(r) - c$$

so that the derivative of $g(r)$ at $r = 0$ is $\lambda m_1 - c$ which is less than zero by assumption (2.3).

$M_X'(0)$ gives the mean of X, which is $E[X]$ or m_1.

It can also be shown that if the function $g(r)$ has a turning point, it must be at the minimum of the function. The second derivative is:

$$\frac{d^2}{dr^2}g(r) = \lambda \frac{d^2}{dr^2}M_X(r)$$

which is always strictly positive.

The second derivative can be written:

$$M_X''(r) = E\left[X^2 e^{rX}\right]$$

The function in this expectation is made up of two positive factors, and hence the expectation must have a positive value.

Hence, there can only be one turning point, since any turning point is a minimum. To show that there is a turning point, note from (2.5) that $\lim\limits_{r \to \gamma -} g(r) = \infty$. Since $g(r)$ is a decreasing function at $r = 0$, it must have a minimum turning point and so the graph of $g(r)$ is as shown in Figure 5.

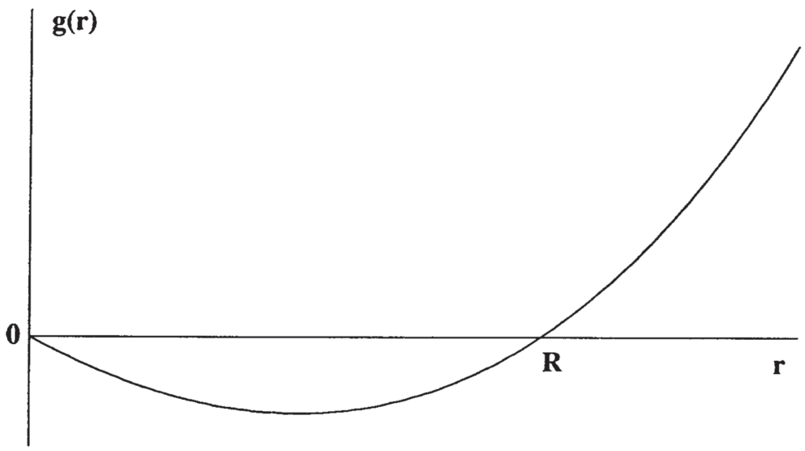


Figure 5

Thus there is a unique positive number R satisfying equation (3.1).

Equation (3.2) is an implicit equation for R. For some forms of $F(x)$ it is possible to solve explicitly for R; otherwise the equation has to be solved numerically.

Consider the exponential distribution where $F(x) = 1 - e^{-\alpha x}$.

This is in the *Tables*, using α as the parameter for the exponential distribution, which avoids confusion with the Poisson parameter.

For this distribution, $M_X(r) = \dfrac{\alpha}{\alpha - r}$, so:

$$\lambda + cR = \frac{\lambda\alpha}{\alpha - R}$$

$$\Rightarrow \quad \lambda\alpha - \lambda R + cR\alpha - cR^2 = \lambda\alpha$$

$$\Rightarrow \quad R^2 - \left(\alpha - \frac{\lambda}{c}\right)R = 0 \Rightarrow \quad R = \alpha - \frac{\lambda}{c} \qquad (3.3)$$

since R is the positive root of (3.1).

If $c = \dfrac{(1+\theta)\lambda}{\alpha}$, then $R = \dfrac{\alpha\theta}{(1+\theta)}$.

Question

Write down the equation for the adjustment coefficient for personal accident claims if 90% of claims are for £10,000 and 10% of claims are for £25,000, assuming a proportional security loading of 20%.

Show that this equation has a solution in the range $0.00002599 < R < 0.00002601$.

Solution

The adjustment coefficient satisfies:

$$1+(1+\theta)m_1R=M_X(R)$$

The distribution of the individual claim sizes X is:

$$X=\begin{cases}10{,}000 \text{ with probability } 0.9\\ 25{,}000 \text{ with probability } 0.1\end{cases}$$

So $\quad E[X]=\sum xP(X=x)=0.9\times 10{,}000+0.1\times 25{,}000=11{,}500$

and $\quad M_X(R)=E\left[e^{RX}\right]=\sum e^{Rx}P(X=x)=0.9\,e^{10{,}000R}+0.1\,e^{25{,}000R}$

The security loading is $\theta=0.2$.

So the equation for the adjustment coefficient is:

$$1+1.2\times 11{,}500R=0.9e^{10{,}000R}+0.1e^{25{,}000R}$$

ie $\quad 1+13{,}800R=0.9e^{10{,}000R}+0.1e^{25{,}000R}$

We can show that there is a solution in the range stated by looking at the values of $LHS-RHS$:

At $R=0.00002599$: $\quad 1+13{,}800R-(0.9e^{10{,}000R}+0.1e^{25{,}000R})=\ 0.000035$

At $R=0.00002601$: $\quad 1+13{,}800R-(0.9e^{10{,}000R}+0.1e^{25{,}000R})=-0.000018$

Since there is a reversal of signs (and we are dealing with a continuous function), the difference must be zero at some point between these two values, *ie* there is a solution of the equation in the range $0.00002599<R<0.00002601$.

If the equation for R has to be solved numerically, it is useful to have a rough idea of R's value. Equation (3.2) can be used to find a simple upper bound for R as follows:

$$\begin{aligned}\lambda+cR&=\lambda M_X(R)\\ &=\lambda\int_0^\infty e^{Rx}f(x)dx\\ &>\lambda\int_0^\infty (1+Rx+\tfrac{1}{2}R^2x^2)f(x)dx\\ &=\lambda(1+Rm_1+\tfrac{1}{2}R^2m_2)\end{aligned}$$

The inequality is true because all the terms in the series for e^{Rx} are positive. So e^{Rx} must always be greater than the total of the first few terms.

Alternatively, we could present this proof as:

$$\lambda + cR = \lambda E[e^{RX}]$$
$$= \lambda E\left[1 + RX + \frac{1}{2}R^2X^2 + \cdots\right]$$
$$= \lambda\left[E[1] + E[RX] + E\left[\frac{1}{2}R^2X^2\right] + \cdots\right]$$
$$> \lambda\left[1 + RE[X] + \frac{1}{2}R^2E[X^2]\right]$$

So that $(c - \lambda m_1)R > \frac{1}{2}\lambda R^2 m_2$, giving:

$$R < 2(c - \lambda m_1) / \lambda m_2 \qquad \textbf{(3.4)}$$

so that $R < 2\theta m_1 / m_2$ when $c = (1+\theta)\lambda m_1$. Notice that if the value of R is small, then it should be very close to this upper bound since the approximation to e^{Rx} should be good.

Question

Find an upper limit for the adjustment coefficient in the previous question, and comment on your answer.

Solution

Here:

$$E[X] = 11{,}500$$

Also:

$$E[X^2] = \sum x^2 P(X = x) = 0.9 \times 10{,}000^2 + 0.1 \times 25{,}000^2 = 152{,}500{,}000$$

So:

$$R < \frac{2\theta m_1}{m_2} = \frac{2 \times 0.2 \times 11{,}500}{152{,}500{,}000} = 0.0000302$$

So this is a reasonable initial estimate, compared with the correct value of approximately 0.000026.

$$R > \frac{1}{M}\log(c / \lambda m_1) = \frac{1}{M}\log(1+\theta) = \frac{1}{25{,}000}\log 1.2 = 0.00000729$$

The steps used in the proof are equally valid for a discrete claims distribution. However, note that the lower bound obtained here is not very close to the accurate value for R.

3.3 A lower bound for R

A lower bound for R can be derived when there is an upper limit, say M, to the amount of an individual claim. For example, if individual claim amounts are uniformly distributed on $(0,100)$, then $M=100$. The result is proved in a similar fashion to Result (3.4). The lower bound is found by applying the inequality:

$$\exp(Rx) \le \frac{x}{M}\exp(RM) + 1 - \frac{x}{M} \quad \text{for} \quad 0 \le x \le M \tag{3.5}$$

The inequality is proved through the series expansion of $\exp(RM)$:

$$\begin{aligned}
\frac{x}{M}\exp(RM) + 1 - \frac{x}{M} &= \frac{x}{M}\sum_{j=0}^{\infty}\frac{(RM)^j}{j!} + 1 - \frac{x}{M} \\
&= 1 + \sum_{j=1}^{\infty}\frac{R^j M^{j-1} x}{j!} \\
&\ge 1 + \sum_{j=1}^{\infty}\frac{(Rx)^j}{j!} \qquad \text{for } 0 \le x \le M \\
&= \exp(Rx)
\end{aligned}$$

since $x^j \le M^{j-1}x$ if $0 \le x \le M$.

Inequality (3.5) can be used to show that:

$$R > \frac{1}{M}\log(c/\lambda m_1)$$

when individual claim amounts have a continuous distribution on $(0,M)$.

This is the lower bound for R that we are trying to find.

The starting point is the equation defining R,

ie Equation 3.2:

$$\begin{aligned}
\lambda + cR &= \lambda\int_0^M \exp(Rx) f(x)dx \\
&\le \lambda\int_0^M \left(\frac{x}{M}\exp(RM) + 1 - \frac{x}{M}\right) f(x)dx \\
&= \frac{\lambda}{M}\exp(RM)m_1 + \lambda - \frac{\lambda}{M}m_1
\end{aligned}$$

Hence, rearranging:

$$\frac{c}{\lambda m_1} \le \frac{1}{RM}(\exp(RM) - 1) = 1 + \frac{RM}{2} + \frac{(RM)^2}{3!} + \cdots$$

$$< 1 + \frac{RM}{1!} + \frac{(RM)^2}{2!} + \cdots$$

$$= \exp(RM)$$

This gives $R > \frac{1}{M} \log\left(\frac{c}{\lambda m_1}\right)$ as required.

If $c = (1+\theta)\lambda m_1$, this is just $R > \frac{1}{M}\log(1+\theta)$.

Other approximations for R can be found, particularly when R is small, by truncating the series expansion of $\exp(Rx)$.

3.4 The adjustment coefficient – general aggregate claims processes

In Section 3.2 the existence of the adjustment coefficient, R, was proved for a compound Poisson aggregate claims process. In this section the existence of the adjustment coefficient for a general aggregate claims process is proved.

Let $\{S_i\}_{i=1}^{\infty}$ be a sequence of independent identically distributed random variables:

$S_i \equiv$ aggregate claims from a risk in time period i.

c is the constant premium charged to insure this risk

The following assumptions are made:

$$c > E[S_i] \qquad (3.6)$$

There is some number $\gamma > 0$ such that:

$$\lim_{r \to \gamma-} E\left[e^{r(S_i - c)}\right] = \infty \qquad (3.7)$$

S_i has density function $h(x)$, $-\infty < x < \infty$ (3.8)

In the general situation the adjustment coefficient is the positive number R that can be shown to satisfy the following:

$$E[e^{R(S_i - c)}] = 1$$

The proof that there is one, and only one, positive number R to satisfy this is as follows.

Let $f(r) = E[e^{r(S_i - c)}]$ for $-\infty < r < \gamma$.

Then $f(0) = 1$, $f'(0) = E[S_i - c] < 0$, $f''(x) > 0$, and:

$$\lim_{r \to \gamma-} f(r) = +\infty$$

We then use the same argument as for the compound Poisson case, based on the graph below. This graph forms part of the Core Reading.

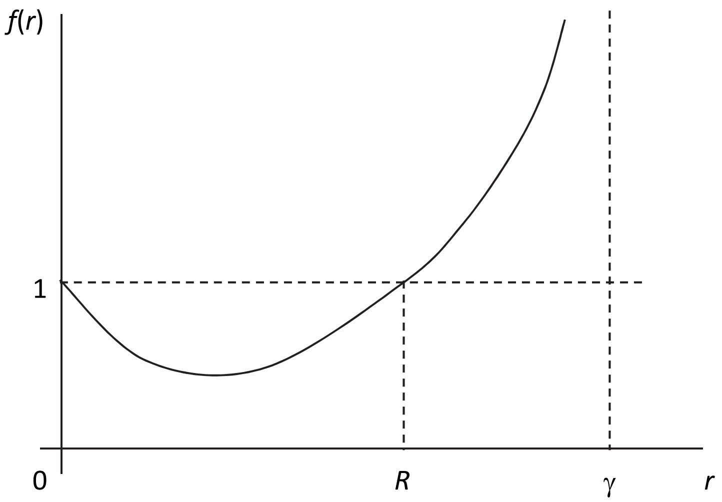


Figure 6 – *f*(*r*) by *r*

Suppose S_i has a compound Poisson distribution with Poisson parameter λ and claim size random variable X.

Then:

$$E\left[e^{R(S_i-c)}\right]=1$$

$$\Rightarrow \quad E\left[e^{RS_i}\right]=e^{Rc}$$

$$\Rightarrow \quad M_S(R)=e^{Rc}$$

$$\Rightarrow \quad e^{\lambda(M_X(R)-1)}=e^{Rc}$$

$$\Rightarrow \quad \lambda M_X(R)=Rc+\lambda$$

which is the same as (3.2).

4 The effect of changing parameter values on ruin probabilities

4.1 Introduction

Recall that $\psi(U)$ was defined to be $P\left(U(\tau)<0, \tau>0\right)$, and $\psi(U,t)$ was defined to be $\psi(U,t)=P\left(U(\tau)<0, 0<\tau<t\right)$.

In this section the effect of changing parameter values on $\psi(U,t)$ and $\psi(U)$ will be discussed.

No new theory will be introduced and the method for obtaining numerical values for $\psi(U,t)$ will not be discussed. Features of $\psi(U,t)$, and in some cases of $\psi(U)$, will be illustrated by a series of numerical examples. In these examples the same basic assumptions will be made as in previous sections. In particular, it will be assumed that the aggregate claims process is a compound Poisson process. In addition it will be assumed throughout Section 4.3, Section 4.4 and Section 4.5 that:

- **the Poisson parameter for the number of claims is 1 (4.1)**
- **the expected value of an individual claim is 1 (4.2)**
- **individual claims have an exponential distribution. (4.3)**

In Section 4.6, Assumptions (4.2) and (4.3) will be made, but the Poisson parameter will be allowed to vary.

The implication of Assumption (4.1) is that the unit of time has been chosen to be such that the expected number of claims in a unit of time is 1. Hence $\psi(U,500)$ is the probability of ruin (given initial surplus U) over a time period in which 500 claims are expected. The actual number of claims over this time period has a Poisson distribution (with parameter 500) and could take any non-negative integer value.

The implication of Assumption (4.2) is that the monetary unit has been chosen to be equal to the expected amount of a single claim. Hence $\psi(20,500)$ is the probability of ruin (over a time period in which 500 claims are expected) given an initial surplus equal to 20 times the expected amount of a single claim.

The advantage of using an exponential distribution for individual claims (Assumption (4.3)) is that both $\exp(-RU)$ and $\psi(U)$ can be calculated for these examples. See Section 3.2 and Section 4.2.

4.2 A formula for $\psi(U)$ when $F(x)$ is the exponential distribution

The formula for $\psi(U)$ when individual claims amounts are exponentially distributed with mean 1, and when the premium loading factor is θ, is given by the following result.

When $F(x)=1-\exp(-x)$:

$$\psi(U)=\frac{1}{1+\theta}\exp\left(-\frac{\theta U}{1+\theta}\right) \tag{4.4}$$

The syllabus does not require this result to be derived or memorised.

This result has been stated in order to illustrate how, for this particular distribution, the ultimate ruin probability is affected by changes in parameter values.

4.3 $\psi(U,t)$ as a function of t

Question

Is $\psi(U,t)$ an increasing or decreasing function of t ?

Solution

$\psi(U,t)$ is the probability of ruin at some point between times 0 and t. This should increase with time since the longer the time period, the more chance there is of ruin. It should be intuitively obvious that $\psi(U,t_1)<\psi(U,t_2)$ for $t_1<t_2$, since if a scenario produces ruin before time t_1, then ruin has also occurred before time t_2.

Figure 7 shows a graph of $\psi(15,t)$ for $0 \le t \le 500$. The premium loading factor, θ, is 0.1 so that the premium income per unit time is 1.1. Also shown in Figure 7 are $\psi(15)$ (dotted line) and $\theta = 0.1$ (solid line) for this portfolio. These last two values are shown as lines parallel to the time axis since their values are independent of time.

Here:

- $\psi(15,t)$ has been worked out using a numerical method not described here
- $\psi(15)$ has been calculated using equation 4.4
- $\exp\{-15R\}$ is the upper bound given by Lundberg's inequality.

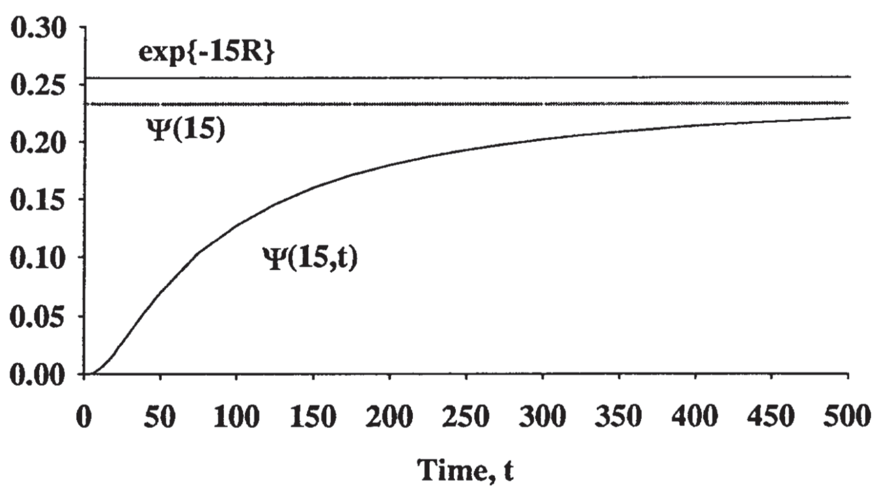


Figure 7 – Values over time

Question

Calculate the value of $\psi(15)$.

Solution

$$\psi(15) = \frac{1}{1.1} e^{-0.1\times\frac{15}{1.1}} = 0.23248$$

Question

Calculate the value of e^{-15R} .

Solution

R is worked out from $R = \dfrac{\alpha\theta}{1+\theta} = \dfrac{1\times 0.1}{1.1}$.

So $e^{-15R} = e^{-15\times\frac{0.1}{1.1}} = 0.2557$

The features of note in Figure 7 are:

(i) $\psi(15,t)$ **is an increasing function of** ***t***

(ii) **for small values of** ***t***, $\psi(15,t)$ **increases very quickly (its value doubles as** ***t*** **increases from 25 to 50 and doubles again as** ***t*** **increases from 50 to 100)**

(iii) **for larger values of** ***t***, $\psi(15,t)$ **increases less quickly and approaches asymptotically the value of** $\psi(15)$ **.**

General reasoning should help you to understand (ii) and (iii). You would expect a much higher probability of ruin before time 50 than before time 25 since the overall performance of the fund could easily change in such a short time period. However, if premium rates are expected to be profitable in the long term, then at time 400, say, a significant surplus will have built up in most cases and so the probability of ruin at time 425 won't be that much higher than at time 400. We are assuming here that accumulated surpluses stay in the fund and are not, for example, distributed to shareholders.

4.4 Ruin probability as a function of initial surplus

Question

Is $\psi(U,t)$ an increasing or decreasing function of U ?

Solution

Decreasing. The bigger the initial surplus, the less chance there should be of ruin.

Figure 8 shows values of $\psi(U,t)$ for α and for three values of the initial surplus, $U = 15$, 20, and 25. The premium loading factor is 0.1 as in Figure 7. For $U = 15$ the graph of α is as in Figure 7.

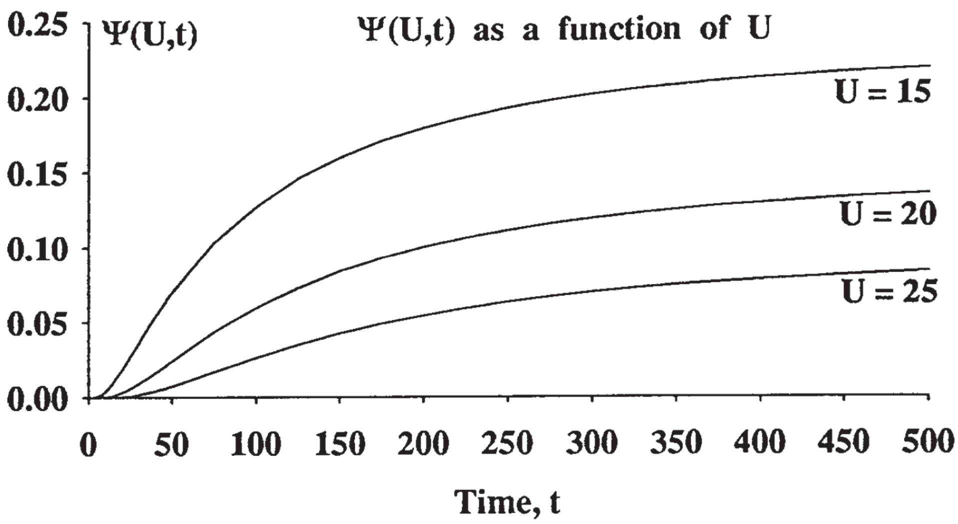


Figure 8 – α for different values of U

Question

Calculate the value of $\psi(20)$.

Solution

$$\psi(20) = \frac{1}{1.1} e^{-0.1 \times \frac{20}{1.1}} = 0.14756$$

This is the limit to which the middle of the three lines is tending to as t tends to ∞.

The features of note in Figure 8 are:

(i) the graphs all have the same general shape

(ii) increasing the value of U decreases the value of $\psi(U,t)$ for any value of t

(iii) each of the three graphs approaches an asymptotic limit as t increases (as has already been noted for U equal to 15 in the discussion of Figure 4). Note that $\psi(20) = 0.1476$ and $\psi(25) = 0.0937$.

$\psi(U)$ is a non-increasing function of U.

In the case of exponentially distributed individual claim amounts, the derivative with respect to U of $\psi(U)$ is:

$$\frac{d}{dU}\psi(U) = \frac{-\theta}{1+\theta}\psi(U)$$

which is negative since $\theta > 0$. Hence $\psi(U)$ is a decreasing function of U.

It is intuitively clear that $\psi(U,t)$ (of which $\psi(U)$ is a special case) should be a decreasing function of U. An increase in U represents an increase in the insurer's surplus without any corresponding increase in claim amounts. Thus, an increase in U represents an increase in the insurer's security and hence will reduce the probability of ruin.

4.5 Ruin probability as a function of premium loading

Question

Is $\psi(U,t)$ an increasing or decreasing function of θ?

Solution

Decreasing. If everything else remains unchanged, then increasing the premium income will reduce the probability of ruin.

Figure 9 shows values of $\psi(15,t)$ for $0 \le t \le 500$ and for three values of the premium loading factor, $\theta = 0.1$, 0.2 and 0.3. The graph of $\psi(15,t)$ for $\theta = 0.1$ is the same as the graph in Figure 7 and the same as one of the graphs in Figure 8. Figure 9 is, in many respects, similar to Figure 8. The features of note in Figure 9 are:

(i) the graphs of $\psi(15,t)$ all have the same general shape,

(ii) increasing the value of θ decreases the value of $\psi(15,t)$ for any given value of t; this is in fact true for any value of U, and is an obvious result since an increase in θ is equivalent to an increase in the rate of premium income with no change in the aggregate claims process,

(iii) it can be seen that when $\theta = 0.1$, 0.2 and 0.3, $\psi(15,t)$ is more or less constant for t greater than about 150. For $t_1 \le t_2$, the difference $\psi(15,t_2) - \psi(15,t_1)$ represents the probability that ruin occurs between times t_1 and t_2. Hence for these values of θ, 0.2 and 0.3, (and for this value of the initial surplus, 15, and for this aggregate claims process) ruin, if it occurs at all, is far more likely to occur before time 150, *ie* within the time period for 150 claims to be expected, than after time 150. This point will be discussed further in Section 4.6.

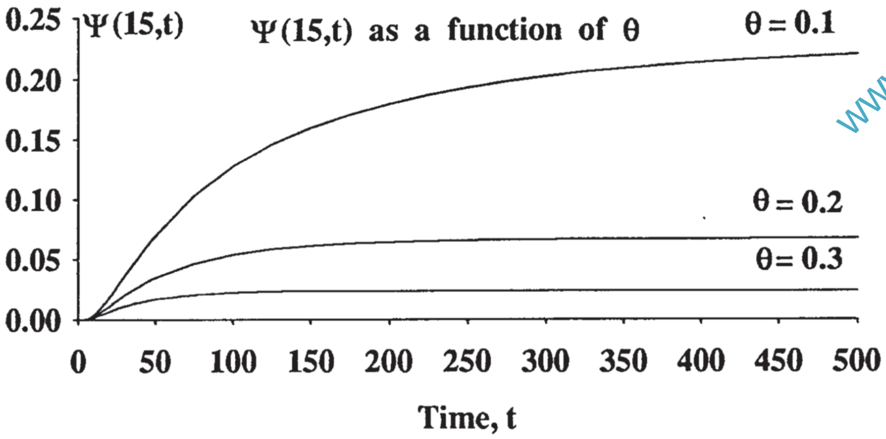


Figure 9 – $\psi(15,t)$ for different values of θ

It is clear by general reasoning that $\psi(U)$ must be a non-increasing function of θ. In the case of exponential individual claim amounts, $\psi(U)$ is a decreasing function of θ.

$$\frac{d}{d\theta}\psi(U) = -(1+\theta)^{-1}\psi(U) - U(1+\theta)^{-2}\psi(U)$$

Question

Verify this expression for $\frac{d}{d\theta}\psi(U)$.

Solution

This is probably easiest if we start by writing:

$$\psi(U) = \frac{1}{1+\theta}\exp\left\{U(1+\theta)^{-1} - U\right\}$$

We need to differentiate this using the product rule:

$$\frac{d}{d\theta}(uv) = u\frac{dv}{d\theta} + v\frac{du}{d\theta}$$

Differentiating $\psi(U)$ with respect to θ:

$$\frac{d}{d\theta}\psi(U) = (1+\theta)^{-1}\left[-U(1+\theta)^{-2}\exp\left\{U(1+\theta)^{-1} - U\right\}\right] - (1+\theta)^{-2}\exp\left\{U(1+\theta)^{-1} - U\right\}$$

Now substituting back in each term for $\psi(U)$:

$$\frac{d}{d\theta}\psi(U) = \left[-U(1+\theta)^{-2}\right]\psi(U) - (1+\theta)^{-1}\psi(U)$$

This is clearly negative since θ, U and $\psi(U)$ are all positive quantities. Since the derivative is less than zero for all values of θ, $\psi(U)$ is a decreasing function of θ.

Figure 10 shows $\psi(10)$ as a function of θ.

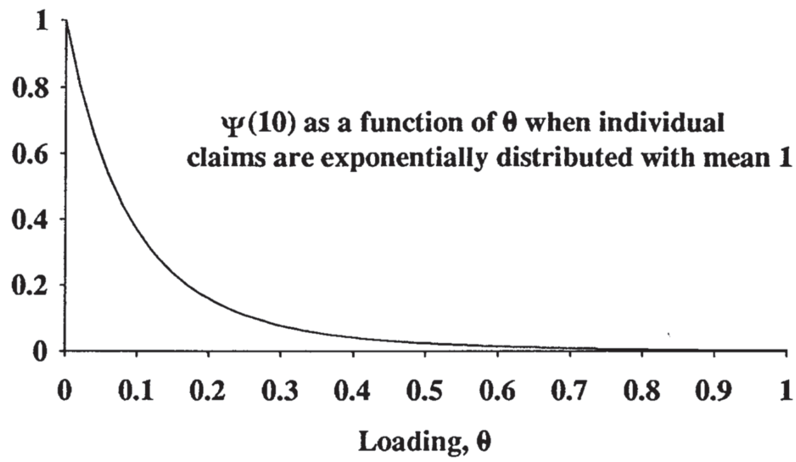


Figure 10

4.6 Ruin probability as a function of the Poisson parameter

Figure 11 shows $\psi(15,10)$ as a function of λ for three values of the premium loading factor, $\theta = 0.1$, 0.2 and 0.3. This graph is identical to Figure 9 apart from the labelling of the *x*-axis. This can be explained by considering the following two risks.

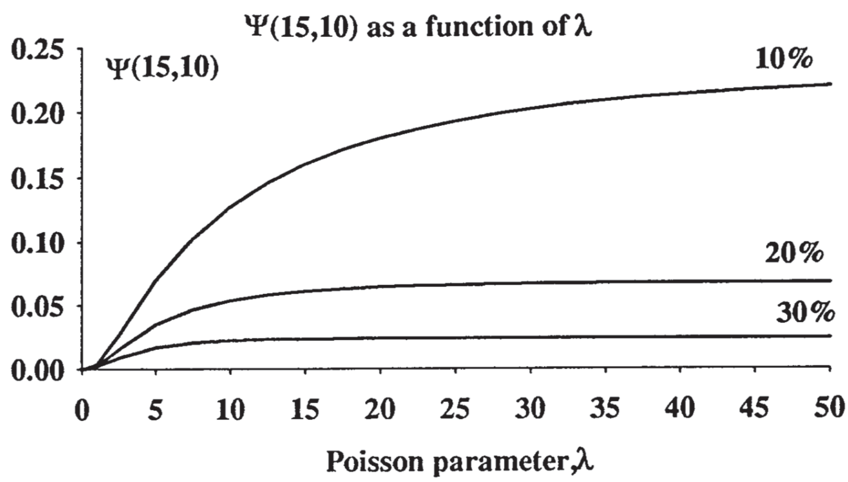


Figure 11

Risk 1: aggregate claims are a compound Poisson process with Poisson parameter 1 and $F(x) = 1 - e^{-x}$. The premium income per unit time to cover this risk is $(1+\theta)$.

Risk 2: aggregate claims are a compound Poisson process with Poisson parameter 0.5 and $F(x) = 1 - e^{-x}$. The premium income per unit time to cover this risk is $0.5(1+\theta)$.

The unit of time is taken to be one year. The only difference between these risks is that twice as many claims are expected each year under Risk 1. This is reflected in the two premiums.

Consider Risk 2 over a new time unit equivalent to two years. Then the distribution of aggregate claims and the premium income per unit time are now identical to the corresponding quantities for Risk 1. Hence, the probability of ruin over an infinite time span is the same for both risks.

The solid line in Figure 12 shows an outcome of the surplus process for Risk 1 when $\theta = 0.1$. The dotted line shows the same surplus process when the unit of time is two years. This illustrates that any outcome of the surplus process that causes ultimate ruin for Risk 1 will also cause ultimate ruin for Risk 2. There is thus no difference in the probability of ultimate ruin for these two risks. It is only the time (in years) until ruin that will differ. Measuring times in years, the probability of ruin by time 1 for Risk 1 is the same as the probability of ruin by time 2 for Risk 2. This explains why Figures 9 and 11 show the same functions. For example, the value of $\psi(15,10)$ when $\lambda = 50$ (Figure 11) is the same as the value of $\psi(15,500)$ when $\lambda = 1$ (Figure 9).

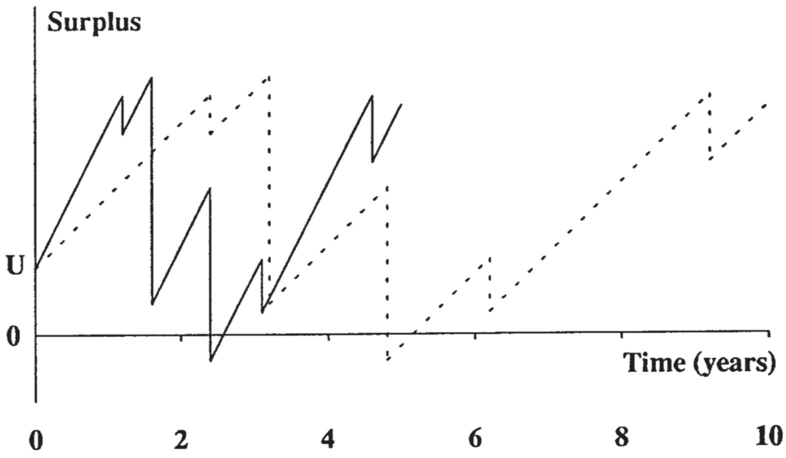


Figure 12

Point (iii) in Section 4.5 will now be investigated, where it was noted that values of $\psi(15,t)$ were more or less constant for values of t greater than 150 when $\theta = 0.2$ and 0.3. In particular, the situation will be considered when the premium loading factor is 0.2.

Consider a second aggregate claims process, which is the same as the process considered throughout this section except that its Poisson parameter is 150 and not 1. (This second process is really identical to the original one; all that has happened is that the time unit has been changed.) Use ψ^* to denote ruin probabilities for the second process and ψ to denote, as before, ruin probabilities for the original process. The change of time unit means that for any $t \geq 0$:

$$\psi^*(U,t) = \psi(U,150t)$$

but it has no effect on the probability of ultimate ruin (put $t = \infty$ in the relationship above) so that:

$$\psi^*(U) = \psi(U)$$

The point made in (iii) above was that:

$$\psi(15,150) \approx \psi(15)$$

From this and the previous two relations it can be seen that:

$$\psi^*(15,1) \approx \psi^*(15)$$

In words this relation says that for the second process, starting with initial surplus 15, the probability of ruin within one time period is almost equal to (actually a little less than) the probability of ultimate ruin. This conclusion depends crucially on the fact that $\psi^*(15,1)$ is a continuous time probability of ruin. To see this, consider $\psi^*(15,1)$, which is just the probability that for the second process the surplus at the end of one time period is negative. $\psi^*(15,1)$ can be calculated approximately by assuming that the aggregate claims in one time period, which will be denoted *S(1), have a normal distribution. Recall that individual claims have an exponential distribution with mean 1 and that the number of claims in one time period has a Poisson distribution with mean 150. From this:**

$$E\left[S^*(1)\right] = 150 \qquad \text{and} \qquad Var\left(S^*(1)\right) = 300$$

These are calculated as $\lambda m_1 = 150 \times 1 = 150$ and $\lambda m_2 = 150 \times 2 = 300$, where $E\left[X^2\right] = 2$ for an $Exp(1)$ distribution.

Now, using tables of the normal distribution:

$$\begin{aligned}\psi^*(15,1) &= P\left[S^*(1) > 15 + 1.2 \times 150\right] \\ &= P\left[(S^*(1) - 150)/17.32 > 45/17.32\right] \\ &\approx 0.005\end{aligned}$$

Recall that if $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma}$ has a standard normal distribution, *ie* $Z \sim N(0,1)$. Probabilities for this distribution can be looked up in the *Tables*.

From Figure 9 it can be seen that the value of $\psi(15,150)$, and hence of $\psi^*(15,1)$, is about 0.07 which is very different from the (approximate) value of the discrete time probability of ruin $\psi^*(15,1)$ calculated above.

4.7 Concluding remarks

When individual claim amounts are exponentially distributed with mean 1, first note that if $\theta = 0$, then $\psi(U) = 1$ irrespective of the value of U.

We're thinking here of substituting $\theta = 0$ into equation 4.4.

This result is in fact true for any form of $F(x)$ (it trivially follows that if $\theta < 0$, then $\psi(U) = 1$). In other words a positive premium loading is essential if ultimate ruin is not to be certain.

Also note that throughout this section it has been assumed that individual claim amounts are exponentially distributed with mean 1. This mean could be measured in units of £100, £1,000 or perhaps £1,000,000. The parameter of the exponential distribution can be set to 1 without loss of generality provided that the monetary unit is correctly specified. In simple terms, the probability of ruin when U is £1 is the same as the probability of ruin when U is 100 pence. It can be said that:

$$\psi(U) \text{ when } F(x) = 1 - e^{-\alpha x}$$

is the same as:

$$\psi(\alpha U) \text{ when } F(x) = 1 - e^{-x}$$

In other words, if the expected claims per unit time increase by a factor α, so too must the initial surplus if the probability of ultimate ruin is to be unchanged.

5 Reinsurance and ruin

5.1 Introduction

One of the options open to an insurer who wishes to reduce the variability of aggregate claims from a risk is to effect reinsurance. This is a form of insurance in which an insurance company obtains insurance cover from other insurance companies (reinsurers) against the risk of losses.

A reduction in variability would be expected to increase an insurer's security, and hence reduce the probability of ruin. A reinsurance arrangement could be considered optimal if it minimises the probability of ruin. As it is difficult to find explicit solutions for the probability of ruin, the effect of reinsurance on the adjustment coefficient will be considered instead.

If a reinsurance arrangement can be found that maximises the value of the adjustment coefficient, the upper bound for the probability of ultimate ruin will be minimised. As the adjustment coefficient is a measure of risk, it seems a reasonable objective to maximise its value. In the following, the effect on the adjustment coefficient of proportional and of excess of loss reinsurance arrangements will be considered.

Throughout this section we will use the notation $X =$ individual claim amount, $Y =$ amount paid by the direct insurer and $Z =$ amount paid by the reinsurer.

5.2 Proportional reinsurance

Under proportional reinsurance, the reinsurer covers an agreed proportion of each risk and the reinsurance premium is in proportion to this risk ceded.

For example, the reinsurer might agree to pay 20% of each claim. The insurer would then pay 80% of the claim amount.

If we have a retained proportion α then:

$$Y = \alpha X \qquad Z = (1-\alpha)X$$

$$\Rightarrow \qquad E[Y] = E[\alpha X] = \alpha E[X] \qquad E[Z] = E[(1-\alpha)X] = (1-\alpha)E[X]$$

Question

Write down expressions for $Var(Y)$ and $Var(Z)$.

Solution

$$Var(Y) = Var(\alpha X) = \alpha^2 Var(X)$$

$$Var(Z) = Var((1-\alpha)X) = (1-\alpha)^2 Var(X)$$

Let us consider the idea of a proportional reinsurance approach by way of a question:

Question

Consider the insurer in the third question of Section 2.4. The number of claims has a Poisson distribution with parameter 30 per year. The individual claim amount distribution is lognormal with parameters $\mu = 3$ and $\sigma^2 = 1.1$. The rate of premium income from the portfolio is 1,200 per year. The insurer has an initial surplus of 1,000

This insurer is investigating the possibility of using proportional reinsurance. It has approached a reinsurer, who uses a security loading of 50% to calculate its reinsurance premiums. If the insurer decides to reinsure 20% of each risk in the portfolio, estimate the effect the reinsurance will have on its probability of ruin at Time 2. Again you can assume that the aggregate claim distribution is approximately normal.

Solution

We first need to calculate the reinsurance premium. Since the reinsurer takes responsibility for 20% of each risk, and uses a loading factor of 50%, the reinsurance premium (per annum) will be:

$$RP = (1+\theta_R)\lambda\alpha m_1 = 1.5 \times 30 \times 0.2 \times e^{3.55} = 313.32$$

So over a two year period, the insurer will pay 626.64 for the reinsurance.

We now use $S_{net}(2)$ for the insurers aggregate payments (net of reinsurance). We need the mean and variance of $S_{net}(2)$, which are, using the formulae for a compound Poisson distribution as before:

$$E[S_{net}(2)] = 60 \times 0.8 \times e^{3.55} = 1{,}671.04$$

and: $$Var\left(S_{net}(2)\right) = 60 \times 0.8^2 \times e^{8.2} = 139{,}812.49$$

So ruin will occur if:

$$S_{net}(2) > 1{,}000 + 2{,}400 - 626.64 = 2{,}773.36$$

Using a normal distribution approach, we have:

$$P\left[S_{net}(2) > 2{,}773.36\right] = P\left[N(0,1) > \frac{2{,}773.36 - 1{,}671.04}{\sqrt{139{,}812.49}}\right]$$

$$= 1 - \Phi(2.9481) = 0.0016$$

Question

Comment on the usefulness of reinsurance in this context.

Solution

The reinsurance has reduced the probability of ruin to some extent, *ie* from about 0.25% (which was calculated in Section 2.4) to about 0.16%. However, this result is probably quite sensitive to the assumptions made (we are near the tail of the normal distribution), and slightly different assumptions might give us very different results.

We will also want to look at the effect of reinsurance on profitability. As we are paying a reinsurance premium, it is likely that the overall effect on profitability is negative (although the effect on security is positive, as we have seen). There is likely to be a trade-off between security and profitability here.

5.3 Excess of loss reinsurance

Under excess of loss reinsurance, the cost to an insurer of a large claim, a cluster of claims arising from a single event (such as an explosion) or claims over a given period (such as a catastrophe) is capped with the liability above a certain level being passed to a reinsurer. In most cases the reinsurer's maximum liability is limited, but we will assume there is no such limit within this chapter.

We refer to the capped amount as the retention of the insurer.

For example, the insurer might retain the first ten million of claims arising during the year, with the reinsurer paying any excess above this amount. If the claim amount turns out to be less than ten million, then the insurer will pay this amount in full and the reinsurer will pay nothing. However, if the claim amount is twelve million, then the insurer will pay ten million and the reinsurer will pay the excess of two million.

Subject SP7 discusses the different forms of reinsurance contracts and their relative merits in greater detail.

We can apply the same type of logic as used in the previous section if the insurer decides to buy excess of loss reinsurance. You might like to think about the effect on the probability of ruin if the insurer in the previous example purchases excess of loss reinsurance with an individual retention of 2,000, say, and a security loading of 50% as before.

If we have a retention limit M, and no upper limit, then:

$$Y=\begin{cases} X & X<M \\ M & X\geq M \end{cases} \qquad\qquad Z=\begin{cases} 0 & X<M \\ X-M & X\geq M \end{cases}$$

Y can be also be written as $\min(X,M)$, and Z can also be written as $\max(0,X-M)$.

Question

Calculate $E[Y]$ if X has an exponential distribution with parameter 0.01, and the insurer has an excess of loss reinsurance arrangement with retention limit M.

Solution

The formula for an expectation is $\int_x xf(x)\,dx$. We have to calculate $E[Y]$ by carrying out two integrals, to allow for the two different ranges of X:

$$E[Y]=\int_0^M xf(x)\,dx+\int_M^\infty Mf(x)\,dx$$

$$=\int_0^M 0.01xe^{-0.01x}\,dx+\int_M^\infty 0.01Me^{-0.01x}\,dx$$

Using integration by parts:

$$=\left[-\frac{0.01x}{0.01}e^{-0.01x}\right]_0^M-\int_0^M-\frac{0.01}{0.01}e^{-0.01x}dx+\left[-M\frac{0.01}{0.01}e^{-0.01x}\right]_M^\infty$$

$$=\left[-xe^{-0.01x}\right]_0^M-\left[\frac{1}{0.01}e^{-0.01x}\right]_0^M+\left[-Me^{-0.01x}\right]_M^\infty$$

$$=-Me^{-0.01M}-\frac{1}{0.01}e^{-0.01M}+\frac{1}{0.01}+Me^{-0.01M}$$

$$=\frac{1}{0.01}(1-e^{-0.01M})=100\left(1-e^{-0.01M}\right)$$

Question

Calculate $Var(Z)$ (in terms of M) if $X\sim U(0,100)$, where the insurer has an excess of loss reinsurance arrangement with retention limit M, $0<M<100$.

Solution

To find $Var(Z)$, we need to find $E[Z^2]$, since $Var(Z)=E[Z^2]-E^2[Z]$.

$$E[Z^2]=\int_0^M 0^2 f(x)\,dx+\int_M^{100}(x-M)^2 f(x)\,dx$$

The PDF of the $U(0,100)$ distribution is $\frac{1}{100}$ (see page 13 of the *Tables*), so:

$$E[Z^2]=\int_M^{100}\frac{(x-M)^2}{100}dx=\left[\frac{(x-M)^3}{300}\right]_M^{100}=\frac{(100-M)^3}{300}$$

We now need $E[Z]$:

$$E[Z]=\int_M^{100}\frac{(x-M)}{100}dx=\left[\frac{(x-M)^2}{200}\right]_M^{100}=\frac{(100-M)^2}{200}$$

So $$Var(Z)=\frac{(100-M)^3}{300}-\left(\frac{(100-M)^2}{200}\right)^2$$

$$=\frac{(100-M)^3}{300}-\frac{(100-M)^4}{40,000}$$

5.4 Maximising the adjustment coefficient under proportional reinsurance

First consider the effect of proportional reinsurance with retention α on the insurer's adjustment coefficient. Throughout Section 5.4 the insurer's premium income per unit time, before payment of the reinsurance premium, will be written as $(1+\theta)\lambda m_1$, which represents the expected aggregate claims per unit time for the compound Poisson process with a loading factor θ. It will also be assumed that the reinsurance premium is calculated as $(1+\xi)(1-\alpha)\lambda m_1$. Since the reinsurer pays proportion $1-\alpha$ of each claim, $(1-\alpha)\lambda m_1$ represents the reinsurer's expected claims per unit time.

Thus, ξ is the premium loading factor used by the reinsurer. Hence, the insurer's premium income, net of reinsurance, is:

$$\left[(1+\theta)-(1+\xi)(1-\alpha)\right]\lambda m_1 \qquad \textbf{(5.1)}$$

Question

Explain this formula.

Solution

The insurer will charge a premium of:

$$(1+\theta)\lambda E[X]=(1+\theta)\lambda m_1$$

The reinsurer will charge a premium of:

$$(1+\xi)\lambda E[Z]$$

But $E[Z]=(1-\alpha)E[X]=(1-\alpha)m_1$, so the reinsurer's premium is:

$$(1+\xi)\lambda(1-\alpha)m_1$$

So the net premium received by the insurer is the difference:

$$(1+\theta)\lambda m_1 - (1+\xi)\lambda(1-\alpha)m_1 = \{(1+\theta)-(1+\xi)(1-\alpha)\}\lambda m_1$$

It will also be assumed that $\xi \geq \theta$. If this were not true, it would be possible for the insurer to pass the entire risk on to the reinsurer and to make a certain profit.

This of course ignores commission, expenses and other adjustments to the theoretical risk premium.

For the insurer's premium income, net of reinsurance, to be positive:

$$1+\theta > (1+\xi)(1-\alpha)$$

ie $\quad \alpha > (\xi-\theta)/(1+\xi)$

Question

What range of values is possible for α if $\theta = 0.2$ and $\xi = 0.4$?

Solution

$$\alpha > \frac{0.2}{1.4} = 0.1429$$

But since α cannot exceed 1, the possible range of values is $0.1429 < \alpha \leq 1$.

There is, however, a more important constraint on the insurer. The insurer's net of reinsurance premium income per unit time must exceed the expected aggregate claims per unit time. Otherwise ultimate ruin is certain (as noted in Section 0). Net of reinsurance, the insurer's expected aggregate claims per unit time are $\alpha\lambda m_1$.

Thus:

$$(1+\theta)-(1+\xi)(1-\alpha) > \alpha$$

or $\quad \alpha > 1 - \dfrac{\theta}{\xi} \qquad\qquad$ **(5.2)**

Question

So what is the range of possible values of α now, given the figures in the previous question?

Solution

$\alpha > 1 - \dfrac{0.2}{0.4} = 0.5$. So $0.5 < \alpha \leq 1$.

Equation (5.2) specifies the insurer's minimum retention level since:

$$1-\theta/\xi \geq (\xi-\theta)/(1+\xi) \text{ when } \theta \leq \xi$$

the only case of interest. If the premium loading factors are equal, then inequality (5.2) becomes $\alpha > 0$. In this case there exists a risk sharing arrangement and any retention level is possible. If, however, $\xi > \theta$ then the insurer has to retain part of the risk.

Same loadings

First consider the case where both the insurer and the reinsurer use θ as the premium loading factor. The adjustment coefficient will be found as a function of the retention level α, when $F(x) = 1-e^{-0.1x}$.

The distribution of the insurer's individual claims net of reinsurance is exponential with parameter $0.1/\alpha$. This can be seen by noting that if $Y = \alpha X$, then:

$$P[Y \leq y] = P[X \leq y/\alpha] = 1-\exp\{-0.1y/\alpha\}$$

Note that the assumptions for the claims process and the adjustment coefficient equation apply equally well in the presence of reinsurance, provided that we use the net premium and the net claim amounts in the adjustment coefficient equation.

Question

What will be the general equation for R, the direct insurer's adjustment coefficient, when there is reinsurance?

Solution

From the previous work, we know that the equation for R is:

$$\lambda + cr = \lambda M_X(r)$$

With reinsurance this will become:

$$\lambda + c_{net} r = \lambda M_Y(r)$$

for the direct insurer. But we know that $c_{net} = (1+\theta)\lambda E[X] - (1+\xi)\lambda E[Z]$, so the equation for R becomes:

$$\lambda + \big((1+\theta)\lambda E[X] - (1+\xi)\lambda E[Z]\big) r = \lambda M_Y(r)$$

or $$1 + \big((1+\theta)E[X] - (1+\xi)E[Z]\big) r = M_Y(r)$$

In the case of proportional reinsurance this is:

$$1 + \big((1+\theta)E[X] - (1+\xi)(1-\alpha)E[X]\big) r = M_Y(r)$$

Hence, the equation defining R (see formula (3.2)) is:

$$\lambda + (1+\theta)\lambda 10\alpha R = \lambda \int_0^\infty e^{Rx}(0.1/\alpha)e^{-0.1x/\alpha}\,dx$$

$$\Rightarrow 1+(1+\theta)10\alpha R = \frac{1}{1-10\alpha R} \tag{5.3}$$

$$\Rightarrow R = \frac{\theta}{(1+\theta)10\alpha} \text{ for } 0 < \alpha \le 1 \tag{5.4}$$

It can be seen that R is a decreasing function of α. This is sensible as the larger the retention α, the larger the risk for the insurer and so $\psi(U)$ would be expected to increase, and R to decrease, with α.

Different loadings

Now consider what happens when $\xi > \theta$. For the rest of Section 5.4 assume that:

- **the individual claim amount distribution is $F(x) = 1 - e^{-0.1x}$**
- **the insurer's premium loading factor is $\theta = 0.1$.**

Case A: $\xi = 0.2$

Suppose first that the reinsurer's premium loading factor is $\xi = 0.2$, so that the insurer's (net) premium income per unit time is $(12\alpha - 1)\lambda$.

This comes from Equation 5.1.

Equation (5.2) shows that the insurer must retain at least 50% of each claim. Hence, a value of α will be sought in the interval $[0.5,1]$ that maximises the value of R. The equation defining R is:

$$\lambda + (12\alpha - 1)\lambda R = \frac{\lambda}{1-10\alpha R}$$

Question

Derive this formula for R.

Solution

Using the equation from the previous question where $\theta = 0.1$, $E[X] = 1/0.1 = 10$ and $\xi = 0.2$, we get:

$$1+\left(1.1\times 10 - 1.2(1-\alpha)\times 10\right)r = M_Y(r)$$

$$\Rightarrow \quad 1+\left(11-12(1-\alpha)\right)r = M_Y(r)$$

$$\Rightarrow \quad 1+(12\alpha - 1)r = M_Y(r)$$

But what about $M_Y(r)$?

$$M_Y(r)=E[e^{rY}]=E[e^{r\alpha X}]=M_X(\alpha r)=\left(1-\frac{r\alpha}{0.1}\right)^{-1}=\frac{1}{1-10r\alpha}$$

Therefore the equation is:

$$1+(12\alpha-1)r=\frac{1}{1-10r\alpha}$$

This follows from (5.3) as only the premium is different – which leads to:

$$\boldsymbol{R=\frac{2\alpha-1}{10(12\alpha^2-\alpha)} \quad \text{for } 0.5<\alpha\leq 1} \qquad \textbf{(5.5)}$$

The right hand side is based on the MGF of the net claim amounts which have an $\xi > \theta$ distribution.

Question

Derive this formula for R .

Solution

Multiplying the equation in the previous question through by $1-10\alpha R$ gives:

$$(1+12\alpha R-R)(1-10\alpha R)=1$$

Multiplying out the left hand side and subtracting 1 from both sides gives:

$$(10\alpha-120\alpha^2)R^2+(2\alpha-1)R=0$$

Dividing through by R and rearranging gives the required expression.

As when the loading factors were equal, the adjustment coefficient depends on the retention level.

The value of α that maximises R in (5.5) is sought.

Differentiate R with respect to α (using the quotient rule for differentiation) to give:

$$\boldsymbol{\frac{dR}{d\alpha}=\frac{20(12\alpha^2-\alpha)-(2\alpha-1)10(24\alpha-1)}{100(12\alpha^2-\alpha)^2}}$$

The quotient rule is $\dfrac{d}{d\alpha}\left(\dfrac{u}{v}\right)=\dfrac{v\dfrac{du}{d\alpha}-u\dfrac{dv}{d\alpha}}{v^2}$.

Alternatively the algebra is a little easier if you write $R = \frac{1}{10\alpha} - \frac{1}{12\alpha - 1}$.

Now the denominator is always positive for values of α in $[0.5,1]$, so there will be a turning point of the function when:

$$20(12\alpha^2 - \alpha) = (2\alpha - 1)10(24\alpha - 1)$$

***ie* when $24\alpha^2 - 24\alpha + 1 = 0$**

The roots of this quadratic are 0.9564 and 0.0436, and so the turning point which is of interest is 0.9564.

Remember that α must lie in the range (0.5,1).

Consider the following values:

α	R
0.5	0
0.9564	0.00911
1.0	0.00909

This shows that *R* has a maximum in $[0.5,1]$ at 0.9564.

Alternatively you can show that the second derivative is negative.

Figure 13 shows R as a function of α (as given by (5.5)) for values of α greater than 0.85. This range of α values has been chosen to highlight the important features of the graph. The dotted line shows the value of R when $\alpha = 1$ (*ie* no reinsurance).

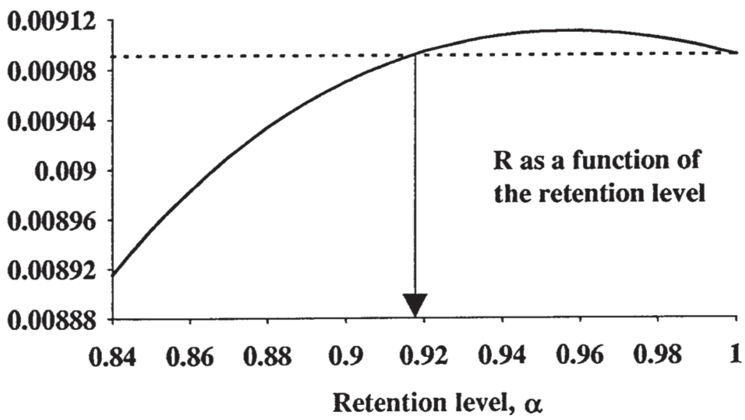


Figure 13 – R as a function of α

It can be seen from Figure 13 that there is a range of values for α, $\beta < \alpha < 1$, such that if the retention level is in this range, the value of the adjustment coefficient exceeds the value when $\alpha = 1$. The value of β can be calculated from (5.5) by setting the value of R at β equal to the value of R at 1, giving $\beta = 0.9167$. The arrow in Figure 13 indicates the value of β.

You can check for yourself that $\beta = 11/12$.

In terms of maximising the adjustment coefficient, the optimal retention level is $\alpha = 0.9564$. It should be noted, however, that optimality in one sense does not imply optimality in another. For example, if the insurer does not effect reinsurance, then the expected profit per unit time is $\theta\lambda m_1$ (*ie* λ, since $\theta = 0.1$ and $m_1 = 10$).

$\theta\lambda m_1$ is the 'loading' bit.

If the insurer effects reinsurance with retention level 0.9564, then the expected profit per unit time is 0.9128λ (*ie* premium income, from (5.1), less expected claims).

The expected profit per unit time is now found in terms of α and λ.

It has already been calculated from (5.1), that with $\theta = 0.1$, $\xi = 0.2$ and $m_1 = 10$, the insurer's net premium income is $(12\alpha - 1)\lambda$. The insurer's expected claims per unit time are $10\alpha\lambda$. Hence, the expected profit per unit time is $(2\alpha - 1)\lambda$.

If we put $\alpha = 0.9564$, this gives 0.9128λ, as stated above.

This shows that expected profit per unit time is an increasing function of α, and if the insurer were to choose α to maximise the expected profit per unit time, the choice would be $\alpha = 1$. This example illustrates a general point – the level of reinsurance is a *trade-off* between security and profit.

Case B: $\xi = 0.3$

The value of α is now found that maximises R when the reinsurer's premium loading factor is 0.3.

The calculations are very similar to the previous case.

From (5.1), the insurer's net premium income is $(13\alpha - 2)\lambda$, so that the equation defining R is:

$$\lambda + (13\alpha - 2)\lambda R = \frac{\lambda}{1 - 10\alpha R}$$

which leads to:

$$R = \frac{3\alpha - 2}{10(13\alpha^2 - 2\alpha)} \qquad \text{for } 0.67 < \alpha \le 1$$

Or $R = \dfrac{1}{10\alpha} - \dfrac{1}{13\alpha - 2}$, adopting the same approach as before.

By (5.2), the insurer must retain at least 2/3 of each claim so the value of α in the range $[2/3,1]$ that maximises R is sought. Differentiation gives:

$$\frac{dR}{d\alpha}=\frac{30(13\alpha^2-2\alpha)-(3\alpha-2)10(26\alpha-2)}{100(13\alpha^2-2\alpha)^2}$$

and the function has a turning point when:

$$30(13\alpha^2-2\alpha)=(3\alpha-2)10(26\alpha-2)$$

***ie* when:**

$$39\alpha^2-52\alpha+4=0$$

The roots of this quadratic are 0.0820 and 1.2514, so there are no turning points in the interval [2/3,1] and R as a function of α in this interval increases from 0 at $\alpha=2/3$ to 0.00909 at $\alpha=1$. Thus, the value of α which maximises the adjustment coefficient is 1.

It is not always possible to increase the value of the adjustment coefficient by effecting reinsurance. Note that when an insurer effects reinsurance, this reduces the variability of the insurer's aggregate claims. A reduction in variability is associated with an increase in the value of the adjustment coefficient. However, when $\xi>\theta$, the insurer's premium loading factor, net of reinsurance, decreases, and the value of the adjustment coefficient is expected to decrease with the loading factor. When the reinsurer's premium loading factor was 0.3, the reduction in the insurer's security caused by the reduction in the loading factor has a greater effect on the adjustment coefficient than the increase resulting from reinsurance for all values of α.

Loading factor (net of reinsurance)

The insurer's premium loading factor, net of reinsurance, implied by (5.1) is now found, and shown to be an increasing function of α.

The loading factor is found by dividing the expected profit per unit time by the expected claims per unit time. The expected profit per unit time is:

$$[(1+\theta)-(1+\xi)(1-\alpha)]\lambda m_1-\alpha\lambda m_1$$

This is just an algebraic expression for net premiums less expected net claims.

So the loading factor is:

$$\begin{aligned}\theta'&=[(1+\theta)-(1+\xi)(1-\alpha)-\alpha]/\alpha\\&=\xi-(\xi-\theta)/\alpha\end{aligned}$$

Now $\dfrac{d\theta'}{d\alpha}=(\xi-\theta)/\alpha^2$ which is positive since $\xi>\theta$, so that θ' is an increasing function of α. Thus, the net loading factor increases as the retention level increases.

5.5 Maximising the adjustment coefficient under excess of loss reinsurance

In this section the effect of excess of loss reinsurance on the adjustment coefficient will be considered. The following assumptions will be made for Section 0:

- **the insurer's premium income (before reinsurance) per unit time is $(1+\theta)\lambda m_1$**
- **the reinsurance premium per unit time is $(1+\xi)\lambda E(Z)$, where $\xi\ (\geq \theta)$ is the reinsurer's premium loading factor, and $Z = \max(0, X - M)$.**

The insurer's individual net claim payments are distributed as $Y = \min(X, M)$, and the insurer's premium income, net of reinsurance, is:

$$c^* = (1+\theta)\lambda m_1 - (1+\xi)\lambda E(Z)$$

which gives the equation defining R as:

$$\lambda + c^* R = \lambda\left[\int_0^M e^{Rx} f(x)dx + e^{RM}[1 - F(M)]\right]$$

You may see c^* referred to as c_{net}.

Question

Explain where the right hand side of this equation comes from.

Solution

We need $\lambda M_Y(r)$, where $Y = \begin{cases} X & X < M \\ M & X \geq M \end{cases}$ and X has PDF $f(x)$. By definition, $M_Y(r) = E[e^{rY}]$, but we need to express this as two separate integrals to take into account the different ranges of X:

$$E[e^{rY}] = \int_0^M e^{rx} f(x)\,dx + \int_M^\infty e^{rM} f(x)\,dx$$

$$= \int_0^M e^{rx} f(x)\,dx + e^{rM} \int_M^\infty f(x)\,dx$$

But the second integral is just integrating the PDF from M to ∞. This is the same as $P(X > M)$, which can be written as $1 - F(M)$. So the right hand side of the equation is:

$$\lambda\left[\int_0^M e^{rx} f(x)\,dx + e^{rM}(1 - F(M))\right]$$

This is formula (3.2) with a truncated claim amount distribution as a result of the excess of loss reinsurance.

To illustrate ideas, look at the situation when $X \sim U(0,20)$, so that $f(x) = 0.05$ for $0 < x < 20$. Then for $0 < M \le 20$:

$$E[Z] = \int_M^{20} (x - M)\, 0.05\, dx = 10 - M + 0.025M^2$$

and $$M_Y(R) = \int_0^M e^{Rx}\, 0.05 dx + e^{RM}(1 - 0.05M)$$

$$= \frac{0.05}{R}(e^{RM} - 1) + e^{RM}(1 - 0.05M)$$

The equation for R must be solved numerically for given values of θ and ξ. Figure 14 shows R as a function of M when $\theta = \xi = 0.1$. As in Section 5.4, any retention level is possible when the premium loading factors are equal. R is a decreasing function of M.

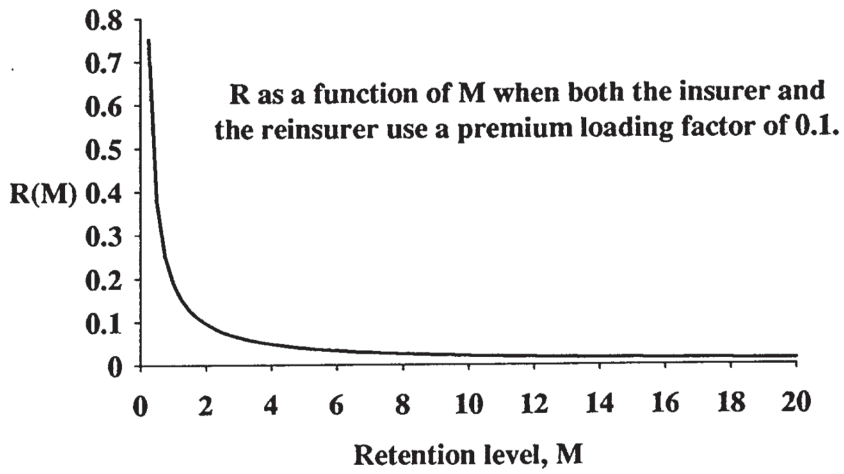


Figure 14 – R as a function of M

In Figure 14, R goes to ∞ as M goes to 0.

When $\theta < \xi$, there is a minimum retention level for the same reason as in the previous section.

Recall that the lower limit for α given by Equation 5.2 applied when we were considering proportional reinsurance.

For example, when $\theta = 0.1$ and $\xi = 0.2$ the insurer's net premium income, c^*, is $11\lambda - 1.2\lambda(10 - M + 0.025M^2)$ and this must exceed the insurer's expected claims, net of reinsurance. The insurer's expected net claims equal $\lambda E[X] - \lambda E[Z]$, which gives $\lambda(M - 0.025M^2)$.

Thus:

$$-1+1.2M-0.03M^2 > M-0.025M^2$$

$$\Rightarrow \quad M^2-40M+200<0$$

$$\Rightarrow \quad 5.8579 < M < 34.1421$$

Hence, the minimum retention level is 5.8579. Similarly, when $\xi = 0.4$, the minimum retention level is 10.

Figure 15 shows R as a function of M for the following combinations of θ and ξ:

(a) $\theta = 0.1$ and $\xi = 0.2$ (solid line)

(b) $\theta = 0.1$ and $\xi = 0.4$ (dotted line).

Without reinsurance, *ie* for $M = 20$, the insurer's adjustment coefficient is 0.014 (irrespective of the reinsurer's loading factor).

From Figure 15, it can be seen that, for $\xi = 0.2$:

$$R(M) > R(20) \text{ for } 9.6 < M < 20$$

$$R(M) < R(20) \text{ for } M < 9.6$$

and for $\xi = 0.4$:

$$R(M) < R(20) \text{ for } R < 20$$

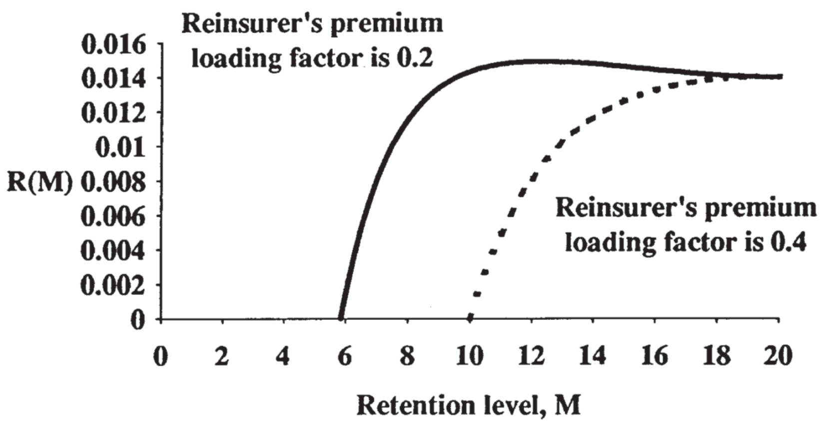


Figure 15 – *R* as a function of *M*

Hence, for $\xi = 0.2$ it is possible for the insurer to increase the value of the adjustment coefficient by effecting reinsurance, provided that the retention level is above 9.6. However, when $\xi = 0.4$, the insurer should retain the entire risk in order to maximise the value of the adjustment coefficient. As in the case of proportional reinsurance, the insurer's expected profit per unit time is reduced if reinsurance is effected.

Question

Claims occur as a Poisson process with rate λ and individual claim sizes X follow an $Exp(\beta)$ distribution. The office premium includes a security loading θ_1. An individual excess of loss arrangement operates under which the reinsurer pays the excess of individual claims above an amount M in return for a premium equal to the reinsurer's risk premium increased by a proportionate security loading θ_2. Derive and simplify as far as possible an equation satisfied by the adjustment coefficient for the direct insurer.

Solution

The adjustment coefficient equation is $\lambda + cR = \lambda M_X(R)$. The net rate of premium income for the direct insurer equals the rate of premiums charged to the policyholder minus the rate of premiums paid to the reinsurer:

$$c = (1+\theta_1)\lambda\frac{1}{\beta} - (1+\theta_2)\lambda\int_M^\infty (x-M)\beta e^{-\beta x}dx$$

The second term can be integrated using the substitution $y = x - M$, and identifying the integral as the mean of an $Exp(\beta)$ distribution. This gives:

$$c = \lambda\frac{1}{\beta}[(1+\theta_1) - (1+\theta_2)e^{-\beta M}]$$

The individual net claims are the claims paid to policyholders minus the recoveries from the reinsurer. So the MGF (which is valid for all values of $R < \beta$) is:

$$M_X(R) = \int_0^M e^{Rx}\beta e^{-\beta x}dx + \int_M^\infty e^{RM}\beta e^{-\beta x}dx = \frac{1}{\beta - R}[\beta - R\,e^{-(\beta-R)M}]$$

So the equation for the adjustment coefficient is:

$$\lambda + \lambda\frac{1}{\beta}[(1+\theta_1) - (1+\theta_2)e^{-\beta M}]R = \lambda\frac{1}{\beta - R}[\beta - R\,e^{-(\beta-R)M}]$$

Cancelling λ's and multiplying through by $\beta(\beta - R)$ gives:

$$\beta(\beta - R) + (\beta - R)\left[(1+\theta_1) - (1+\theta_2)e^{-\beta M}\right]R = \beta\left[\beta - Re^{-(\beta-R)M}\right]$$

Cancelling the β^2's from both sides gives:

$$-\beta R + (\beta - R)[(1+\theta_1) - (1+\theta_2)e^{-\beta M}]R = -\beta R\,e^{-(\beta-R)M}$$

Cancelling R's to exclude the trivial solution gives:

$$-\beta + (\beta - R)[(1+\theta_1) - (1+\theta_2)e^{-\beta M}] = -\beta\,e^{-(\beta-R)M}$$

The adjustment coefficient R is the smallest positive solution of this equation.

Question

Use the approximation $e^x \approx 1+x+x^2/2$ to find an approximate numerical value for the adjustment coefficient for the previous example in the case where $\beta=0.05$, $\theta_1=0.3$, $\theta_2=0.4$ and $M=10$.

Solution

Using the values given, the adjustment coefficient equation becomes:

$$-0.05+(0.05-R)(1.3-1.4e^{-0.5})=-0.05e^{-0.5+10R}$$

Multiplying by $-20e^{0.5}$ to clear some of the fractions gives:

$$e^{0.5}-(1-20R)(1.3e^{0.5}-1.4)=e^{10R}$$

Expanding the LHS and applying the approximation to the RHS:

$$0.90538+14.8668R=1+10R+50R^2$$

ie $\quad -0.09462+4.8668R-50R^2=0$

Solving this using the quadratic formula (taking the smallest positive root) gives:

$$R=\frac{-4.8668+\sqrt{4.8668^2-4(-50)(-0.09462)}}{2(-50)}=0.0268$$

Chapter 20 Summary

Poisson process – claim numbers

Claim numbers $\{N(t)\}_{t\geq 0}$ can be modelled using a Poisson process with parameter λ so that $N(t) \sim Poisson(\lambda t)$. A Poisson process is an example of a counting process.

Time until first claim and inter claim time

If $\{N(t)\}_{t\geq 0}$ is a Poisson process with parameter λ, then the time until the first claim and the inter claim time follow exponential distributions with parameter λ:

$$f_T(t) = \lambda e^{-\lambda t} \quad (t > 0)$$

Compound Poisson process - aggregate claim amounts

$$S(t) = X_1 + X_2 + \cdots + X_{N(t)}$$

$$E\left[S(t)\right] = \lambda t E(X) \qquad Var\left(S(t)\right) = \lambda t E\left[X^2\right] \qquad M_{S(t)}(u) = e^{\lambda t[M_X(u)-1]}$$

Surplus process

$$U(t) = U + ct - S(t), \qquad t \geq 0$$

U is the initial surplus and c is the premium income per unit time

$c = (1+\theta)E[S]$ where θ is the insurer's premium loading

Ruin probabilities

$$\psi(u) = P[U(t) < 0 \text{ for some } t]$$

$$\psi(u, t_0) = P[U(t) < 0 \text{ for some } t \leq t_0]$$

$$\psi_h(u) = P[U(t) < 0 \text{ for some } t = h, 2h, 3h, \ldots]$$

$$\psi_h(u, t_0) = P[U(t) < 0 \text{ for some } t = h, 2h, 3h, \ldots \text{ and } t \leq t_0]$$

Lundberg's inequality and the adjustment coefficient

For the continuous time model with an infinite time horizon, Lundberg's inequality, which uses a parameter *R* called the adjustment coefficient, provides an upper bound for the probability of ultimate ruin. The adjustment coefficient *R* is an inverse measure of risk, *ie* the higher the value of *R*, the lower the upper bound on the probability of ultimate ruin.

For a compound Poisson process with parameter λ, the adjustment coefficient *R* is the unique positive root of the equation:

$$\lambda + cr = \lambda M_X(r)$$

where λ is the Poisson parameter, c is the premium rate per unit of time and $M_X(r)$ is the MGF of the individual claim amounts at point r.

It is possible to derive upper and lower bounds for *R*.

The adjustment coefficient in the presence of reinsurance

In the presence of reinsurance, for a compound Poisson process with parameter λ, the adjustment coefficient *R* is the unique positive root *r* of the equation:

$$\lambda + c_{net}r = \lambda M_Y(r)$$

where λ is the Poisson parameter, c_{net} is the premium rate per unit of time net of the rate paid to the reinsurer and $M_Y(r)$ is the MGF of the individual claim amounts paid by the insurer (net of reinsurance) at point r.

In order to maximise security, the insurer will want to find a reinsurance arrangement that maximises the adjustment coefficient *R*. However, this will not necessarily be the arrangement that maximises expected profits. There is a trade-off between security and profit.

Effect of changes in parameter values on the probability of ruin

The probability of ultimate ruin decreases if the insurer's premium loading θ is increased or if the insurer's initial surplus U is increased. This is because the insurer has more of a buffer against claims.

An increase in the value of the Poisson parameter λ will not affect the probability of ultimate ruin since the expected aggregate claims $E[S] = \lambda E[X]$, the variance of aggregate claims $Var(S) = \lambda E\left[X^2\right]$ and the premium rate $(1+\theta)\lambda E[X]$ all increase proportionately in line with λ. However, it will reduce the time it takes for ruin to occur.

An increase in the variance $Var(X)$ of the individual claim amounts will increase the probability of ruin as it will increase the uncertainty associated with the aggregate claims process without any corresponding increase in premium.

An increase in the expected individual claim amount $E[X]$ will increase the probability of ruin. The expected aggregate claims and the premium rate both increase proportionately in line with $E[X]$, however the variance of the aggregate claims amount increases disproportionately since $Var(S) = \lambda E\left[X^2\right] = \lambda\left\{Var\left(X\right) + \left(E[X]\right)^2\right\}$. The variance of the aggregate claim amount increases in line with $\left(E[X]\right)^2$.

The practice questions start on the next page so that you can keep the chapter summaries together for revision purposes.

Chapter 20 Practice Questions

20.1 A general insurance company is planning to set up a new class of travel insurance. It plans to start the business with £2 million and expects claims to occur according to a Poisson process with parameter 50. Individual claims are thought to have a gamma distribution with parameters $\alpha = 150$ and $\lambda = ¼$. A premium loading factor of 30% is applied.

Explain how each the following changes to the company's model will affect the probability of *ultimate* ruin:

(i) A 28% premium loading factor is applied instead.

(ii) Individual claims are found to have a gamma distribution with parameters $\alpha = 150$ and $\lambda = ½$.

(iii) The Poisson parameter is now believed to be 60.

20.2 Claims occur according to a compound Poisson process at a rate of 0.2 claims per year. Individual claim amounts, X, have probability function:

$$P(X = 50) = 0.7$$

$$P(X = 100) = 0.3$$

The insurer's surplus at time 0 is 75 and the insurer charges a premium of 120% of the expected annual aggregate claim amount at the beginning of each year. The insurer's surplus at time t is denoted $U(t)$. Find:

$$P[U(2) < 0]$$

20.3 $\{S(t)\}_{t \geq 0}$ and $\{S^*(t)\}_{t \geq 0}$ are compound Poisson processes representing the aggregate claims up to time t from two risks. Individual claim amounts have the same distribution for the two risks and premiums are calculated using the same premium loading factor for the two risks. The Poisson parameters for the two risks are λ and λ^*, respectively. The probability of ruin in finite time and in infinite time for these two risks, given initial surplus U, are $\psi(U,t)$ and $\psi^*(U,t)$ and $\psi(U)$ and $\psi^*(U)$, respectively. You are given that $\lambda / \lambda^* = 2$.

Which of the following equations is/are true:

I $\psi(U) = 2\psi^*(U)$

II $\psi(U) = \psi^*(U)$

III $\psi(U,t) = \psi^*(U,2t)$

IV $\psi(U,t) = \psi^*(U,t)$

20.4 For a Poisson process with intensity λ, determine the probability that exactly one event will occur during a finite time interval of length t.

20.5 The aggregate claims process for a risk is a compound Poisson process. The expected number of claims each year is 0.5 and individual claim amounts have the following distribution:

$$P[\text{claim amount} = 1] = 0.5$$

$$P[\text{claim amount} = 2] = 0.25$$

$$P[\text{claim amount} = 3] = 0.25$$

Let $U(t)$ denote the insurer's surplus at time t. The insurer's surplus at time 0 is 0.5 and the insurer charges a premium of 1 each year to insure this risk.

Calculate the probability:

$$P[U(t) < 0 \quad \text{for } t = 1 \text{ or } 2]$$

20.6 An insurance company has a portfolio of policies, for which claims occur as a Poisson process at a rate of 25 claims per year. The claim amounts in pounds follow a generalised (three parameter) Pareto distribution with parameters $k = 3$, $\lambda = 500$ and $\alpha = 4$. The insurer includes a premium loading of 15% in its premiums for this portfolio. You may assume that the aggregate claim amount for a year is approximately normally distributed.

Exam style

(i) Find u, the initial capital required in order to ensure that the probability of ruin at the end of the first year is 2%.

(ii) If the insurer takes out proportional reinsurance, reinsuring 30% of the loss with a reinsurer which loads its premiums by 45%, find the new level of initial capital required, and compare your answer with that in part (i).

[Total 9]

20.7 Aggregate annual claims from a portfolio of general insurance policies have a compound Poisson distribution with Poisson parameter λ. Individual claim amounts have an exponential distribution with mean 1. The premium loading factor used to calculate the premium for these policies is 0.30. Given an initial surplus of 2, calculate the probability of ruin at the first claim. [6]

Exam style

20.8 The aggregate claims produced by a risk have a compound Poisson distribution with Poisson parameter 100 and individual claim size density, $f(x)$, where:

Exam style

$$f(x) = 0.2e^{-0.2(x-5)} \qquad x > 5$$

The premium charged by the insurer to insure the risk is calculated using a premium loading factor of 0.15. The insurer is considering excess of loss reinsurance for this risk. The reinsurer's premium would be calculated using a premium loading factor of 0.30. The table below shows, for various values of the retention limit M, the insurer's expected profit in one year net of reinsurance, with some missing values indicated by asterisks.

Retention limit M	Expected annual profit	Adjustment coefficient
7.5	59	0.0227
10	*	0.0252
15	*	0.0240
*	147	0.0220
∞	*	0.0213

(i) Calculate the missing values of M and of the insurer's expected profit in one year and set out the complete table. [12]

(ii) Using the values in the completed table comment on the effect on the insurer of the choice of different values for M. [3]

[Total 15]

20.9 Exam style (i) The random variable W has a compound negative binomial distribution, so that W can be written in the form:

$$W = \sum_{i=1}^{N} Y_i$$

where $\{Y_i\}$ for $i = 1, 2, \ldots$, is a sequence of independent and identically distributed random variables, each with mean m and variance s^2, and N is independent of that sequence and has the following probability function:

$$P[N = x] = \binom{k+x-1}{x} p^k (1-p)^x \qquad x = 0, 1, \ldots$$

for some parameters k $(k > 0)$ and p $(0 < p < 1)$.

Show that:

(a) $E(W) = k(1-p)m / p$

(b) $\text{var}(W) = k(1-p)(m^2 + ps^2) / p^2$ [6]

(ii) An insurer plans to issue 5,000 one-year policies at the start of a year. For each policy the annual aggregate claims have a compound negative binomial distribution; the negative binomial parameters are $k = 0.5$ and $p = 0.5$, and individual claim amounts, in pounds, have a lognormal distribution with parameters $\mu = 5.04$ and $\sigma = 1.15$.

The premium for each policy is £160 and is payable at the start of the year. Claims are assumed to be paid at the mid-point of the year. Calculate the minimum annual rate of interest the insurer must earn throughout the year if the accumulation to the end of the year of premiums minus claims is to exceed £52,500 with probability 90%. You may assume that the distribution of total aggregate claims in the year may be approximated by a normal distribution. [15]

[Total 21]

20.10 Individual claim amounts in a Poisson claims process with a frequency of 50 claims per year for the whole portfolio have mean £5,000 and standard deviation £2,500. If the annual premium rate is £300,000, calculate the tightest upper bound for the adjustment coefficient.

20.11 An insurer calculates the annual premiums for fire insurance of flats by increasing the risk premium by 30% and adding a £30 loading. The claim frequency is 3% and individual claim amounts can be assumed to be:

- £2,000 with probability 0.9
- £15,000 with probability 0.1.

Calculate the insurer's adjustment coefficient for these policies, to 2 significant figures.

20.12 A Poisson claims process has security loading $\theta = 2/5$ and claim size density function:

Exam style

$$f(x) = \tfrac{3}{2}e^{-3x} + \tfrac{7}{2}e^{-7x}, \quad x > 0$$

(i) Derive the moment generating function (MGF) for the claim size distribution, and state the values of t for which it is valid. [3]

(ii) Calculate the value of the adjustment coefficient. [4]

[Total 7]

20.13 (i) Show that the adjustment coefficient for a compound Poisson claims process satisfies the inequality:

Exam style

$$r < \frac{2[c/\lambda - E(X)]}{E(X^2)}$$

and define what each of the symbols represents. [9]

(ii) An insurer considers that claims of a certain type occur in accordance with a compound Poisson process. The claim frequency for the whole portfolio is 100 per annum and individual claims have an exponential distribution with a mean of £8,000.

(a) Calculate the adjustment coefficient if the total premium rate for the portfolio is £1,000,000 per annum.

(b) Verify that the value calculated in (ii)(a) satisfies the inequality in (i).

(c) The insurer decides to take out excess of loss reinsurance for this portfolio. The reinsurer has agreed to pay the excess of any individual claim above £20,000 in return for an annual premium of £80,000. Calculate the adjustment coefficient for the direct insurer when the reinsurance is in operation.

(d) Estimate the direct insurer's probability of ultimate ruin with and without the reinsurance arrangement, assuming that the initial surplus is £20,000 and that future premiums remain at the same level.

(e) Comment briefly on the effect of the reinsurance on the probability of ruin. [14]
[Total 23]

20.14 Claims occur on a portfolio of insurance policies according to a Poisson process with Poisson parameter λ. Claim amounts, $X_1, X_2, \ldots,$ are assumed to be identically distributed with moment generating function $M_X(t)$. The insurer calculates premiums using a loading factor $\theta (>0)$. The insurer's adjustment coefficient, R, is defined to be the smallest positive root of the equation:

Exam style

$$\lambda + cr = \lambda M_X(r)$$

where c is the insurer's premium income rate.

(i) Using the above equation for R, or otherwise, show that, provided R is small, an approximation to R is $\hat{R}$, where:

$$\hat{R} = \frac{2(c/\lambda - \mu)}{\sigma^2 + \mu^2}$$

where $\mu = E[X_i]$ and $\sigma^2 = \text{var}[X_i]$. [4]

(ii) Describe how the adjustment coefficient can be used to assess reinsurance arrangements on the basis of security. [3]

(iii) The Poisson parameter, λ, for this portfolio is 20 and all individual claims are for a fixed amount of £5,000. The insurer's premium loading factor, θ, is 0.15 and proportional reinsurance can be purchased from a reinsurer who calculates premiums using a loading factor of 0.25.

Calculate the maximum proportion of each claim that could be reinsured so that the insurer's security, measured by $\hat{R}$, is greater than the insurer's security without reinsurance. [9]
[Total 16]

The solutions start on the next page so that you can separate the questions and solutions.

Chapter 20 Solutions

20.1 (i) ***Reduction in premium loading factor***

Since there is a smaller loading factor, the premiums will be reduced even though claims remain the same. Hence, the probability of ultimate ruin will increase.

(ii) ***Change in claims distribution***

The mean of the distribution has decreased from 600 to 300, and the variance has decreased from 2,400 to 600. Therefore the claims are smaller on average and less uncertain. Both of these factors will decrease the probability of ultimate ruin.

(iii) ***Change in the Poisson parameter***

The Poisson parameter has increased so claims occur more often (but their size is unchanged). However, the premium received will also increase proportionally (as $c=(1+\theta)\lambda m_1$). Hence, the timing at which ruin may occur will be earlier, but not the probability of it occurring in the first place. Therefore, the probability of ultimate ruin will be unchanged.

20.2 The annual premium is 120% of $E(S)$, where S is the aggregate claim amount in a single year. Since S has a compound Poisson distribution with Poisson parameter 0.2, we have $E(S)=0.2E(X)$ where:

$$E(X)=50\times 0.7+100\times 0.3=65$$

Hence the annual premium is:

$$1.2E(S)=1.2\times 0.2E(X)=1.2\times 0.2\times 65=15.6$$

The initial surplus is 75, so the surplus at time 2 is:

$$U(2)=75+2\times 15.6-S(2)=106.2-S(2)$$

So the probability of ruin is:

$$P[U(2)<0]=P[106.2-S(2)<0]=P[S(2)>106.2]=1-P[S(2)\leq 106.2]$$

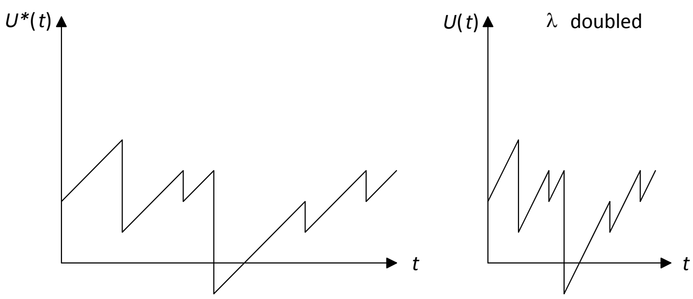
Considering $P[S(2) \leq 106.2]$, and remembering that $N(2) \sim Poi(2 \times 0.2)$ we have the following scenarios where the aggregate claim amount by time 2 is less than 106.2:

number of claims	amount of claim(s)	probability
0 claims	0	$e^{-0.4} = 0.67032$
1 claim	50	$0.4e^{-0.4} \times 0.7 = 0.18769$
	100	$0.4e^{-0.4} \times 0.3 = 0.08044$
2 claims	50, 50	$\frac{0.4^2}{2} e^{-0.4} \times 0.7^2 = 0.02628$

Hence $P[S(2) < 106.2] = 0.9647 \quad \Rightarrow \quad P[U(2) < 0] = 0.0353$

20.3 Equations II and III are true.

Doubling λ^* means claims occur in half the time, but premium income comes in at double the rate. So graph of $U(t)$ is squashed compared to $U^*(t)$:



No change to $\psi(u)$, but increases $\psi(u,t)$.

20.4 The number of events occurring has a $Poisson(\lambda t)$ distribution.

So the probability of exactly one event is:

$$\frac{(\lambda t)^1 e^{-\lambda t}}{1!} = \lambda t e^{-\lambda t}$$

20.5 The surplus is:

$$U(1) = 0.5 + 1 - S(1) = 1.5 - S(1)$$

$$U(2) = 0.5 + 2 \times 1 - S(2) = 2.5 - S(2)$$

Considering the probability of *non-ruin* we require:

$$S(1) < 1.5 \qquad \text{and} \qquad S(2) < 2.5$$

1st period		2nd period		probability
no. of claims	claim amount	no. of claims	claim amount	
0 claims	0	0 claims	0	$(e^{-0.5})(e^{-0.5}) = 0.36788$
		1 claim	1	$(e^{-0.5})(0.5e^{-0.5} \times ½) = 0.09197$
			2	$(e^{-0.5})(0.5e^{-0.5} \times ¼) = 0.04598$
		2 claims	1, 1	$(e^{-0.5})(\frac{0.5^2}{2}e^{-0.5} \times 0.5^2) = 0.01150$
1 claim	1	0 claims	0	$(0.5e^{-0.5} \times ½)(e^{-0.5}) = 0.09197$
		1 claim	1	$(0.5e^{-0.5} \times ½)(0.5e^{-0.5} \times ½) = 0.02299$

Hence $P[U(t) > 0 \text{ for } t = 1 \text{ or } 2] = 0.6323 \quad \Rightarrow \quad P[U(t) < 0 \text{ for } t = 1 \text{ or } 2] = 0.3677$

20.6 (i) ***Initial capital required***

We first need the moments of the generalised Pareto distribution. Using the formulae in the *Tables*, we have:

$$E(X) = \frac{k\lambda}{\alpha - 1} = \frac{3 \times 500}{3} = 500$$

and: $$E(X^2) = \frac{\Gamma(\alpha-2)\Gamma(k+2)}{\Gamma(\alpha)\Gamma(k)}\lambda^2 = \frac{\Gamma(2)\Gamma(5)}{\Gamma(4)\Gamma(3)}500^2 = 500{,}000$$ [1]

So:

$$E(S) = 25 \times 500 = 12{,}500 \qquad \text{and:} \qquad \text{var}(S) = 25 \times 500{,}000 = 12{,}500{,}000$$

where S is the aggregate claim amount in one year. [1]

For ruin, S must exceed the initial capital u plus the premiums received, $1.15E(S)$. So we want:

$$P\left(S > u + 1.15E(S)\right) = 0.02$$

Standardising this in the usual way, we have:

$$P\left(S > u + 1.15E(S)\right) = P\left(\frac{S - E(S)}{\sqrt{\text{var}(S)}} > \frac{u + 0.15E(S)}{\sqrt{\text{var}(S)}}\right)$$

$$\approx P\left(N(0,1) > \frac{u + 0.15E(S)}{\sqrt{\text{var}(S)}}\right) = 0.02$$ [1]

Using the tables of the standard normal distribution, this gives us the equation:

$$\frac{u+0.15E(S)}{\sqrt{\text{var}(S)}}=2.0537$$

Using the values of the mean and variance found earlier to solve this for u, we find that:

$$u=2.0537\sqrt{12,500,000}-(0.15\times 12,500)=7,260.93-1,875=5,385.93$$

So the initial capital required is £5,386. [1]

(ii) ***Initial capital with reinsurance***

We now need to allow for the reinsurance. The reinsurer takes on 30% of the risks. So the reinsurance premium is:

$$P'=1.45\times 0.3E(S)=5.437.5$$ [1]

The insurer retains 70% of the risks, so the amount paid out on claims by the insurer now has moments:

$$E(S')=E(0.7S)=0.7\times 12,500=8,750$$

$$\text{var}(S')=\text{var}(0.7S)=0.7^2\times 12,500,000=6,125,000$$ [1]

So the condition for ruin is now:

$$P\left(S'>u+1.15E(S)-5,437.5\right)=0.02$$

which becomes:

$$P\left(\frac{S'-E(S')}{\sqrt{\text{var}(S')}}>\frac{u+8,937.5-8,750}{\sqrt{6,125,000}}\right)=0.02$$

Again using the normal distribution tables, this gives us the equation;

$$\frac{u+8,937.5-8,750}{\sqrt{6,125,000}}=2.0537$$ [1]

Solving this equation, we find that $u=4,895.15$. So the initial capital required is now £4,895. [1]

The initial capital required has decreased. With the reinsurer taking on more of the risks the variance of the claim amounts paid by the insurer will have reduced and hence the amount of capital the insurer needs to write the business has decreased. This may make it worthwhile for the insurer to take out reinsurance (although the effect on profits and cashflows will also need to be considered). [1]

20.7 First, assume that the first claim occurs at time t. We want the probability that the first claim is big enough to cause ruin, *ie* it is bigger than the surplus accumulated at time t. Since the premium loading factor is 0.3 and the rate of premium income is 1.3λ, the probability that the first claim will cause ruin if it occurs at time t is:

$$\int_{2+1.3\lambda t}^{\infty} e^{-x}dx = \left[-e^{-x}\right]_{2+1.3\lambda t}^{\infty} = e^{-(2+1.3\lambda t)} \qquad [3]$$

We now consider all the times at which the first claim could occur. Since we have a Poisson process, the time to the first claim has an exponential distribution with parameter λ, and the unconditional probability of ruin at the first claim is:

$$\int_0^{\infty} \lambda e^{-\lambda t} e^{-(2+1.3\lambda t)}dt = \frac{e^{-2}}{2.3}\int_0^{\infty} 2.3\lambda e^{-2.3\lambda t}dt \qquad [2]$$

But since this integral is just the density function of another exponential distribution, it integrates to 1. So the unconditional probability of ruin at the first claim is:

$$\frac{e^{-2}}{2.3} = 0.05884 \qquad [1]$$

20.8 (i) ***Calculate the missing values***

The insurer's expected net annual profit will be:

$$c_{\text{net}} - E(S_I) = c_{\text{net}} - 100E(Y) \qquad [1]$$

where:

$$c_{\text{net}} = 1.15E(S) - 1.30E(S_R) = 115E(X) - 130E(Z) \qquad [1]$$

Hence, the expected annual profit is:

$$115E(X) - 130E(Z) - 100E(Y)$$

Since $E(X) = E(Y) + E(Z)$ we can substitute for $E(Y)$ to simplify to:

$$15E(X) - 30E(Z) \qquad [1]$$

Now, integrating by parts:

$$E(X) = \int_5^\infty x0.2e^{-0.2(x-5)}\,dx$$

$$= \left[-xe^{-0.2(x-5)}\right]_5^\infty + \int_5^\infty e^{-0.2(x-5)}\,dx$$

$$= 5 + \left[-5e^{-0.2(x-5)}\right]_5^\infty$$

$$= 5 + 5 = 10 \qquad [2]$$

And:

$$Z = \begin{cases} 0 & X < M \\ X - M & X > M \end{cases}$$

Hence, integrating parts:

$$E(Z) = \int_M^\infty (x-M)0.2e^{-0.2(x-5)}\,dx$$

$$= \left[-(x-M)e^{-0.2(x-5)}\right]_M^\infty + \int_M^\infty e^{-0.2(x-5)}\,dx$$

$$= \left[-5e^{-0.2(x-5)}\right]_M^\infty$$

$$= 5e^{-0.2(M-5)} \qquad [2]$$

So the expected annual profit is:

$$15 \times 10 - 30 \times 5e^{-0.2(M-5)} = 150\left(1 - e^{-0.2(M-5)}\right) \qquad [1]$$

The completed table is:

Retention limit M	Expected annual profit	Adjustment coefficient
7.5	59	0.0227
10	**94.8**	0.0252
15	**129.7**	0.0240
24.56	147	0.0220
∞	**150**	0.0213

[4]

(ii) ***Comment on the effect of different values of M***

The adjustment coefficient is a measure of security (*ie* a higher adjustment coefficient implies a reduced probability of ruin). We can see that decreasing the retention limit (*ie* passing on more of the claims to a reinsurer) at first increases the adjustment coefficient and then decreases it. So decreasing the retention limit will at first reduce the probability of ruin and then eventually increase it. [1]

A higher retention limit implies that the insurer is holding more of the claims (and more of the premium). However, by holding more of the claims they also shoulder more of the risk of claims not being as expected. [1]

A lower retention limit means more of the claim is passed onto the reinsurer which reduces the risk and should reduce the probability of ruin. However in the extreme case of a retention limit of 7.5 so much of the claim is passed onto the reinsurer that their larger premium (due to the larger premium loading factor) eats into the surplus of the insurer and results in a greater probability of ruin. [1]

20.9 (i)(a) ***E(W)***

Using the fact that $E(W)=E[E(W|N)]$, we have:

$$E(W|N=n)=E(Y_1+\cdots+Y_n)=E(Y_1)+\cdots+E(Y_n)=mn \qquad [1]$$

So, for the random variable N:

$$E(W|N)=mN$$

and:

$$E(W)=E(mN)=m\,E(N)=\frac{mk(1-p)}{p} \qquad [1]$$

using the formula for the mean of the negative binomial distribution.

(i)(b) ***var(W)***

Using the fact that $\text{var}(W)=\text{var}[E(W|N)]+E[\text{var}(W|N)]$, we have, since the Y_i are independent and identically distributed:

$$\text{var}(W|N=n)=\text{var}(Y_1+\cdots+Y_n)=\text{var}(Y_1)+\cdots+\text{var}(Y_n)=n\,s^2 \qquad [1]$$

So:

$$\text{var}(W|N)=Ns^2 \qquad [1]$$

So, again using the standard formulae for the negative binomial distribution, we find that:

$$\text{var}(W) = \text{var}(mN) + E(Ns^2) = m^2 \text{var}(N) + s^2 E(N)$$

$$= m^2 \frac{k(1-p)}{p^2} + s^2 \frac{k(1-p)}{p} = \frac{k(1-p)}{p^2}(m^2 + ps^2) \quad [2]$$

(ii) ***Minimum rate of interest required***

Consider a single policy. If W is the aggregate claim amount from one policy, then:

$$W = Y_1 + \cdots + Y_N$$

where $E(N) = \frac{k(1-p)}{p} = 0.5$ and $\text{var}(N) = \frac{k(1-p)}{p^2} = 1$.

Using standard formulae to calculate the moments of Y:

$$E(Y) = m = e^{\mu + \frac{1}{2}\sigma^2} = e^{5.70125} = 299.24122 \quad [1]$$

and $\text{var}(Y) = s^2 = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) = 246{,}499.57$ [1]

So the mean and variance of W are:

$$E(W) = 0.5 \times 299.24122 = 149.62061 \quad [1]$$

and $\text{var}(W) = 1 \times (299.24122^2 + 0.5 \times 246{,}499.57) = 212{,}795.09$ [1]

So the total aggregate claim amount from all policies is $S = W_1 + \cdots + W_{5000}$, and:

$$E(S) = 5000 \times 149.62061 = 748{,}103.05 \quad [1]$$

and $\text{var}(S) = 5000 \times 212{,}795.09 = 1.0639755 \times 10^9$ [1]

assuming that individual policies operate independently.

Suppose that the required annual rate of interest is i. We want:

$$P\left[5000 \times 160 \times (1+i) - S \times (1+i)^{\frac{1}{2}} > 52{,}500\right] = 0.90 \quad [2]$$

Rearranging this inequality gives:

$$P\left[S < 800{,}000(1+i)^{\frac{1}{2}} - 52{,}500(1+i)^{-\frac{1}{2}}\right] = 0.90$$

Standardising, we now have:

$$P\left[\frac{S - E(S)}{\sqrt{\text{var}(S)}} < \frac{800{,}000(1+i)^{\frac{1}{2}} - 52{,}500(1+i)^{-\frac{1}{2}} - 748{,}103.05}{\sqrt{1.0639755 \times 10^9}}\right] \quad [2]$$

But $\dfrac{S-E(S)}{\sqrt{\text{var}(S)}}$ is approximately standard normal. So we must have:

$$\frac{800,000(1+i)^{½}-52,500(1+i)^{-½}-748,103.05}{\sqrt{1.0639755\times 10^9}}=1.28155$$

$$800,000(1+i)^{½}-52,500(1+i)^{-½}-748,103.05=41,802.415$$

$$800,000(1+i)-789,905.46(1+i)^{½}-52,500 \quad =0$$ [2]

Solving this quadratic in $(1+i)^{½}$ gives:

$$(1+i)^{½}=\frac{789,905.46\pm\sqrt{789,905.46^2-4\times 800,000\times(-52,500)}}{2\times 800,000}$$

$$=1.0499 \quad (\text{or } -0.0625)$$

So: $i=1.0499^2-1=0.1023$ *ie* the required interest rate is 10.23% per annum. [3]

20.10 The adjustment coefficient satisfies the inequality:

$$r<\frac{2[c/\lambda-\text{E}(X)]}{\text{E}(X^2)}=\frac{2[300,000/50-5,000]}{5,000^2+2,500^2}=0.000064$$

Since we don't know the precise distribution of the individual claim amounts, this is the best we can do.

20.11 The adjustment coefficient equation is:

$$\lambda+cr=\lambda E(e^{rX})$$

The average claim size is:

$$E(X)=0.9\times 2,000+0.1\times 15,000=£3,300$$

So the annual risk premium is:

$$\lambda E(X)=0.03\times 3,300=£99$$

So the annual office premium is $c=1.3\times 99+30=£158.70$.

So the adjustment coefficient equation is:

$$0.03+158.70r=0.03(0.9e^{2,000r}+0.1e^{15,000r})$$

Simplifying by dividing by 0.03 and writing $R=1,000r$ gives:

$$1+5.29R=0.9e^{2R}+0.1e^{15R}$$

Expanding the RHS as a series to get a first approximation:

$$1+5.29R=0.9(1+2R+2R^2+\cdots)+0.1(1+15R+112.5R^2+\cdots)$$

$$=1+3.3R+13.05R^2+\cdots$$

So: $R\approx\dfrac{5.29-3.3}{13.05}=0.1525$

Evaluating $f(R)=0.9e^{2R}+0.1e^{15R}-1-5.29R$ for different values of R, we get:

$f(0.15)=0.37$	$f(0.1)=0.018$	$f(0.09)=-0.013$
$f(0.094)=-0.0015$	$f(0.095)=0.00156$	$f(0.0945)=0.000011$

So R lies between 0.094 and 0.0945. So the value of the adjustment coefficient correct to 2SF is 0.000094 (in units of £$^{-1}$).

20.12 (i) ***Moment generating function***

The moment generating function is given by:

$$E(e^{tX})=\int_0^\infty e^{tx}\left[\tfrac{3}{2}e^{-3x}+\tfrac{7}{2}e^{-7x}\right]dx=\tfrac{3}{2}\int_0^\infty e^{(t-3)x}dx+\tfrac{7}{2}\int_0^\infty e^{(t-7)x}dx$$

$$=\frac{3}{2}\left[\frac{e^{(t-3)x}}{t-3}\right]_0^\infty+\frac{7}{2}\left[\frac{e^{(t-7)x}}{t-7}\right]_0^\infty=\frac{3}{2(3-t)}+\frac{7}{2(7-t)} \qquad [2]$$

The first integral converges if $t<3$ and the second if $t<7$. So the MGF is valid for values of $t<3$. [1]

(ii) ***Adjustment coefficient***

For the adjustment coefficient, we require:

$$1+(1+\theta)E(X)r=M_X(r)$$

To find $E(X)$, we can differentiate the MGF:

$$M_X(t)=\tfrac{3}{2}(3-t)^{-1}+\tfrac{7}{2}(7-t)^{-1}$$

$$M'_X(t)=\tfrac{3}{2}(3-t)^{-2}+\tfrac{7}{2}(7-t)^{-2}$$

$$M'_X(0)=\frac{3}{18}+\frac{1}{14}=\frac{5}{21} \qquad [1]$$

Alternatively we can note that this distribution is a mixture of two exponential distributions. So the mean is given by a weighted average of two exponential means, ie $\frac{1}{2}\times\frac{1}{3}+\frac{1}{2}\times\frac{1}{7}=\frac{5}{21}$. *Or we could integrate from first principles.*

So $E(X)=\frac{5}{21}$ and the equation for r is:

$$1+\frac{7}{5}\times\frac{5r}{21}=\frac{3}{2(3-r)}+\frac{7}{2(7-r)}=\frac{3(7-r)+7(3-r)}{2(3-r)(7-r)}=\frac{21-5r}{(3-r)(7-r)} \qquad [1]$$

Multiplying through by 3:

$$3+r=\frac{63-15r}{(3-r)(7-r)}$$

Multiplying through by $(3-r)(7-r)$ and multiplying out brackets, we obtain:

$$63-9r-7r^2+r^3=63-15r$$

Gathering up the terms, we obtain:

$$6r-7r^2+r^3=0 \qquad [1]$$

Factorising:

$$r(6-r)(1-r)=0$$

which gives $r=1$ as the positive solution that satisfies the inequality $t<3$. So the value of the adjustment coefficient in this case is 1. [1]

20.13 (i) ***Adjustment coefficient***

The adjustment coefficient r is the smallest positive solution of the equation:

$$\lambda+cr=\lambda E(e^{rX}) \qquad [1]$$

The expected claim frequency λ is the expected number of claims occurring per unit of time. [1]

The premium rate c is the constant amount of premium actually received per unit of time. [1]

The claim size X is a random variable representing the amount of an individual claim. [1]

Expanding the RHS of the equation defining the adjustment coefficient gives:

$$\lambda+cr=\lambda E(e^{rX})=\lambda[1+rE(X)+\frac{r^2}{2}E(X^2)+\cdots] \qquad [1]$$

Since the individual claim sizes X take positive values, the terms on the RHS are all positive. So, ignoring terms in powers higher than X^2 gives:

$$\lambda+cr>\lambda[1+rE(X)+\frac{r^2}{2}E(X^2)] \qquad [1]$$

Subtracting a λ from both sides:

$$cr > \lambda[rE(X) + \frac{r^2}{2}E(X^2)]$$ [1]

Dividing by r (which must be a positive number):

$$c > \lambda[E(X) + \frac{r}{2}E(X^2)]$$ [1]

Rearranging to get an inequality for r gives:

$$r < \frac{2[c/\lambda - E(X)]}{E(X^2)}$$ [1]

(ii)(a) ***Adjustment coefficient***

The adjustment coefficient equation is:

$$\lambda + cr = \lambda E(e^{rX})$$

Since individual claims have an exponential distribution with a mean of 8,000:

$$E(e^{rX}) = M_X(r) = \frac{1}{1-8{,}000r} \qquad (t < 1/8000)$$

So the adjustment coefficient satisfies:

$$100 + 1{,}000{,}000r = \frac{100}{1-8{,}000r}$$ [1]

Dividing by 100:

$$1 + 10{,}000r = \frac{1}{1-8{,}000r}$$

Rearranging:

$$(1+10{,}000r)(1-8{,}000r) = 1$$

$$1 + 2{,}000r - 80{,}000{,}000r^2 = 1$$ [1]

Cancelling the ones and factorising:

$$2{,}000r(1-40{,}000r) = 0$$

The adjustment coefficient is the smallest positive solution, *ie*:

$$r = \frac{1}{40{,}000} = 0.000025$$ [1]

(ii)(b) ***Verify inequality***

Since X has an exponential distribution:

$$E(X)=8,000$$

and:

$$E(X^2)=\text{var}(X)+[E(X)]^2=8,000^2+8,000^2=2\times 8,000^2$$

So the inequality in (i) states that:

$$r<\frac{2[1,000,000/100-8,000]}{2\times 8,000^2}=0.00003125 \qquad [1]$$

The exact value of r (*ie* 0.000025) is indeed less than this.

(ii)(c) ***Adjustment coefficient with reinsurance***

The adjustment coefficient equation (with the reinsurance in effect) is:

$$\lambda+c_{net}r=\lambda E(e^{rX_{net}}) \qquad [1]$$

Since the direct insurer will still have to pay a part of every claim, the claim frequency λ is unchanged.

The *net* rate of premium income (after paying the reinsurance costs) for the direct insurer is:

$$c_{net}=1,000,000-80,000=920,000 \qquad [1]$$

The MGF for the *net* claim amount X_{net} paid by the direct insurer is:

$$E(e^{rX_{net}})=\int_0^{20,000} e^{rx}e^{-x/8000}/8,000\,dx+\int_{20,000}^{\infty} e^{20,000r}e^{-x/8000}/8,000\,dx$$

$$=\frac{1}{8,000}\int_0^{20,000} e^{-x(1/8000-r)}\,dx+\frac{e^{20,000r}}{8,000}\int_{20,000}^{\infty} e^{-x/8000}\,dx$$

$$=\frac{1-e^{-20,000(1/8000-r)}}{8,000(1/8,000-r)}+e^{20,000r}e^{-20,000/8,000}$$

$$=\frac{1-e^{20,000r-2.5}}{1-8,000r}+e^{20,000r-2.5} \qquad [2]$$

So the adjustment coefficient equation becomes:

$$100+920,000r=100\left[\frac{1-e^{20,000r-2.5}}{1-8,000r}+e^{20,000r-2.5}\right] \qquad [1]$$

Multiplying through by $1-8{,}000r$ and dividing through by 100 gives:

$$(1+9{,}200r)(1-8{,}000r)=1-e^{20{,}000r-2.5}+e^{20{,}000r-2.5}(1-8{,}000r)$$

Expanding and cancelling the 1's:

$$1{,}200r-73{,}600{,}000r^2=-8{,}000\, r\, e^{20{,}000r-2.5}$$

Dividing by $100r$:

$$12-736{,}000r=-80\,e^{20{,}000r-2.5}$$

Evaluating the function $f(r)=80e^{20{,}000r-2.5}-736{,}000r+12$ for trial values of r, we find: [1]

$$\begin{aligned} f(0.000025) &= 4.43 \\ f(0.00003) &= 1.89 \\ f(0.00004) &= -2.83 \\ f(0.000033) &= 0.42 \\ f(0.000034) &= -0.06 \\ f(0.000035) &= -0.54 \end{aligned}$$ [1]

So r is approximately 0.000034 (measured in units of £$^{-1}$). [1]

(ii)(d) ***Probability of ruin***

So the probability of ruin with and without the reinsurance are approximately:

$$\psi_{without}\approx e^{-0.000025\times 20{,}000}=0.61 \text{ and } \psi_{with}\approx e^{-0.000034\times 20{,}000}=0.51$$ [1]

(ii)(e) ***Comment***

So the reinsurance reduces the probability of ultimate ruin. [1]

20.14 (i) ***Approximation to R***

The MGF of X can be written:

$$M_X(t)=E(e^{tX})=E(1+tX+\frac{(tX)^2}{2!}+\cdots)=1+t\,E(X)+\frac{t^2}{2!}E(X^2)+\cdots$$

So: $$M_X(R)=1+R\,E(X)+\frac{R^2}{2!}E(X^2)+\cdots$$ [1]

Assuming that R is small enough for terms in R^3 and higher powers to be neglected, we can write:

$$M_X(R)\approx 1+R\mu+\frac{R^2}{2!}(\mu^2+\sigma^2)$$ [1]

Substituting this into the defining equation for R :

$$\lambda + cR = \lambda[1 + R\mu + \frac{R^2}{2!}(\mu^2 + \sigma^2)]$$

Subtracting λ from both sides and rearranging, we get:

$$(c - \lambda\mu)R = \frac{\lambda R^2}{2!}(\mu^2 + \sigma^2)$$ [1]

Dividing through by R and simplifying, we get:

$$R = \frac{2(c - \lambda\mu)}{\lambda(\mu^2 + \sigma^2)} = \frac{2(c/\lambda - \mu)}{\sigma^2 + \mu^2}$$ [1]

This is the required expression.

(ii) ***Assessment of reinsurance***

The adjustment coefficient can be used to assess the effectiveness of different reinsurance arrangements, using Lundberg's inequality to find an upper bound for the probability of ruin for the insurer under different reinsurance arrangements. An arrangement that produces a lower upper bound for the probability of ruin is in some sense more secure for the insurer than an arrangement that has a higher upper bound for the probability of ruin. [3]

Note however that the adjustment coefficient cannot tell us anything about the relative profitability of different reinsurance arrangements. This will need to be assessed using other means.

(iii) ***Maximum reinsurance***

First we consider the insurer's security without reinsurance. The equation for the adjustment coefficient is:

$$1 + (1+\theta)m_1 r = M_X(r)$$

Substituting in $\theta = 0.15$, $m_1 = 5000$ and $M_X(r) = e^{5000r}$, we get:

$$1 + 5750r = e^{5000r}$$ [1]

The rate of premium income is:

$$c = (1+\theta)\lambda m_1 = 1.15 \times 20 \times 5000 = 115{,}000$$

So, using the approximation derived for the adjustment coefficient in part (i), we have:

$$R \approx \frac{2(115000/20 - 5000)}{5000^2} = 6 \times 10^{-5}$$ [1]

Now assume that a proportion k of each risk is reinsured. The net premium income is now:

$$[1.15\times 5000-1.25\times k\times 5000]\lambda =(5750-6250k)\lambda \qquad [1]$$

The MGF of the insurer's net claim payments is now:

$$M_X(t)=e^{5000(1-k)R} \qquad [1]$$

So the equation for the insurer's adjustment coefficient is now:

$$1+(5750-6250k)R=e^{5000(1-k)t} \qquad [1]$$

Using the same approximation to R as before, we have:

$$R\approx \frac{2[(5750-6250k)-5000(1-k)]}{5000^2(1-k)^2}=\frac{3-5k}{50000(1-k)^2} \qquad [1]$$

If we want the insurer's security with reinsurance to be greater than without reinsurance, we want the adjustment coefficient with reinsurance to be larger, *ie*:

$$\frac{3-5k}{50000(1-k)^2}>6\times 10^{-5} \qquad [1]$$

Rearranging this inequality, we get:

$$3k^2-k<0 \qquad [1]$$

The solution of this is $0<k<1/3$. So the maximum proportion to be reinsured is $33\frac{1}{3}$%. [1]

21

Run-off triangles

Syllabus objectives

5.2 Run-off triangles

5.2.1 Define a development factor and show how a set of assumed development factors can be used to project the future development of a delay triangle.

5.2.2 Describe and apply the basic chain ladder method for completing the delay triangle using development factors.

5.2.3 Show how the basic chain ladder method can be adjusted to make explicit allowance for inflation.

5.2.4 Describe and apply the average cost per claim method for estimating outstanding claim amounts.

5.2.5 Describe and apply the Bornhuetter-Ferguson method for estimating outstanding claim amounts.

5.2.6 Describe how a statistical model can be used to underpin a run-off triangles approach.

5.2.7 Discuss the assumptions underlying the application of the methods in 5.2.1 to 5.2.6 above.

0 Introduction

0.1 The origins of run-off triangles

Run-off triangles (or delay triangles) are an important topic in the practical work of actuaries working in general insurance who make use of spreadsheets and other computer packages to forecast future claim numbers and amounts. In this section we will look at four standard methods for projecting run-off triangles – the basic chain ladder method, the inflation-adjusted chain ladder method, the average cost per claim method and the Bornhuetter-Ferguson method.

For these last two methods, there is no one way of applying the method that is universally agreed, *ie* you may find variations in the methods if you see these in everyday use. Be careful in the exam to apply them sensibly to the data you are given.

The techniques we will study in this chapter are directly relevant to general insurance. You are unlikely to be asked questions about the types of reserves given in the next section. However, this is useful background material which you will meet again in much more detail if you study the later general insurance subjects.

Run-off triangles (delay triangles) usually arise in types of insurance (particularly non-life insurance) where it may take some time after a loss until the full extent of the claims which have to be paid is known. It is important that the claims are attributed to the year in which the policy was written.

The claims are analysed in cohorts, and it is important that each claim is allocated to the correct cohort. You will see in the general insurance subjects that several different cohort definitions can be used. For example, claims can be grouped by the year in which the policy was originally written, or by the year in which the accident occurred, or in a number of other ways.

The insurance company needs to know how much it is liable to pay in claims so that it can calculate how much surplus it has made. However, it may be many years before it knows the exact claims totals. There are many causes for the delays in the claim totals being finalised. The delay may occur before notification of the claim and/or between notification and final settlement.

It is clear that although the insurance company does not know the exact figure for total claims each year, it must try to estimate that figure with as much confidence and accuracy as possible.

So the question that we shall attempt to answer in this chapter is this: how much needs to be set aside now (as a reserve) to meet future payments to be made on claims that have arisen during some recent past period?

0.2 Types of reserves

General insurers need to be able to estimate the *ultimate* cost of claims for several purposes. For example, they need to know the full cost of paying claims in order to set future premium rates. They also need to set up reserves in their accounts to make sure that they have sufficient assets to cover their liabilities.

The normal steps involved in settling a general insurance claim are shown in the diagram:

Claim event occurred	→→→	*Claim reported*	→→→	*Claim payment(s) made*	→→→	*Claim file closed*

After the claim event has occurred (*eg* a policyholder has been involved in a motor accident or has been burgled), the policyholder will report the incident to the insurer.

In due course the insurer will make any payments required (*eg* paying for repairs to a vehicle, compensation to an injured person or the cost of replacing stolen belongings). There may be several payments made under a single claim.

When the insurer considers that no further payments will be required for this claim, the claim file will be closed.

A general insurer will need to set up reserves to cover its liabilities for future payments in respect of accidents that have already occurred. These reserves will relate to claims at different stages in the settlement process. In particular, reserves will be required for *outstanding reported claims* and *IBNR* claims.

Types of reserves

An *IBNR* (pronounced 'I.B.N.R.') claims reserve is required in respect of claims that have been *incurred but not reported*, *ie* the claim event has occurred, but the claim has not yet been reported to the insurer.

An *outstanding reported* claims reserve is required in respect of claims that have been reported, but have not yet been closed.

Question

Identify where each of these reserves fits in on the diagram above.

Solution

The IBNR reserve corresponds to the claims at the stage indicated by the left hand arrow.

The outstanding reported claims reserve corresponds to the middle arrow.

(No reserve would be required for the right hand arrow, since all payments for these claims have already been made. However, in practice, insurers may hold a *reopened claims* reserve to cover the possibility that the claim file is closed 'too soon' and that further payments are required.)

0.3 Presentation of claims data

There are several ways of presenting claims data, which emphasise different aspects of the data. Here they will be presented as a triangle, which is the most commonly used method. The year in which the incident happened and the insurer was on risk is called the accident year. The number of years until a payment is made is called the delay, or development period. The claims data is divided up by the accident year and the development year. The following table is an example of claims data referenced by accident year and development year. In some types of insurance it might be relevant to look at development of claims by month or quarter, but the principles are unchanged.

Also, as we shall see, data may be presented cumulatively, or on an individual year basis.

Example

The figures in the table below are in units of £1,000, but for convenience we will not write them out in full in our calculations.

Cumulative claim payments

Accident Year	Development Year				
	0	1	2	3	4
2008	786	1,410	2,216	2,440	2,519
2009	904	1,575	2,515	2,796	
2010	995	1,814	2,880		
2011	1,220	2,142			
2012	1,182				

Figure 1: Sample claims data

Each row in the triangle represents an *origin* year which defines a *cohort* of claims. This example uses an *accident year* cohort. Because a lot of statistical theory in general insurance was developed in relation to motor insurance, the term 'accident year' is extended to situations where the claim event is clearly not an accident, *eg* car thefts, arson, burglaries. The 2008 row includes all claims relating to accidents that occurred during the 2008 calendar year.

In practice most general insurers use an accounting year starting on 1st January, so the rows do really represent calendar years.

The columns represent *development* years, which show how the cohort of claims relating to a particular origin year 'develop' over time. Column 0 represents the year in which the accident occurred. Column 1 represents the year after the accident occurred, *etc*.

Each entry in the table can be defined by its *accident year* (row) and its *development year* (column) *eg* the figure of 2,216 is for Accident Year 2008, Development Year 2, which we will write as 2008/2 or $C_{2008,2}$. Note that this figure includes payments made in 2008, 2009 and 2010, since it is a cumulative table.

The figures given are cumulative and represent total amounts paid by the end of each development year. They have been compiled after the end of the 2012 accident year. For the 2012 accident year, only payments with delay 0 have been reported. For the 2011 accident year, payments with delay 0 and delay 1 have been reported, and so on.

Question

Use the delay triangle above to determine:

(i) the total amount of claims paid in 2012 in respect of accidents that occurred in 2010, and

(ii) the total amount of claims paid during 2012.

Solution

Because the triangle shows cumulative amounts, we need to subtract neighbouring columns ('disaccumulate').

(i) $2,880 - 1,814 = 1,066$

(ii) $1,182 + (2,142 - 1,220) + (2,880 - 1,814) + (2,796 - 2,515) + (2,519 - 2,440)$

$= 3,530$

Note that particular calendar years are represented by the *diagonals* in the triangle. For example, the long diagonal (1182, 2142, 2880, 2796, 2519) includes all payments made during the most recent calendar year shown, *ie* 2012 (as well as past calendar years). Note also that the upper left corner represents 'known' past payments, while the lower right corner represents 'unknown' future payments. Our task in the remainder of this chapter is to look at methods of estimating these unknown figures to complete the lower right triangle. For each accident year, the difference between the figure in the extreme right hand column and the total amount paid so far will give us the estimate of the amount we need to hold currently to meet future liabilities arising.

The table in the example above showed the amounts of paid claims tabulated by accident year. Various alternative tabulations could have been used. For example:

1. The cohorts could be defined by *reporting year* ('all claims *reported* in year X') or by *written year* ('all claims from policies *written* in year X').

2. The origin years might be the company's *financial* years or might be origin *quarters* or origin *months*.

The entries in the table might show *numbers* of claims or estimated *ultimate* cost or claims-related *expenses*.

Question

If we projected the claims given in the run-off triangle, which of the reserves described in the previous section would be included if the table showed:

(i) the amounts of claims paid, tabulated by accident year, and

(ii) the amounts of claims paid, tabulated by reporting year?

Solution

(i) The table would include all claims in respect of *accidents that occurred* between 1 January 2008 and 31 December 2012. So this would include both outstanding reported claims and IBNR claims.

(ii) The table would include all claims that were *reported* between 1 January 2008 and 31 December 2012. So this would include outstanding reported claims, but not IBNR claims.

0.4 Estimating future claims

We now look at the methods we might use to try to estimate the missing figures in the table.

The task is to decide the amounts yet to be paid in respect of the given accident years. This can be done for 2012 by looking at previous accident years. If the cumulative payments increase in a similar way, it is possible to say that they are likely to be about 3,788 in 4 years' time. This figure is obtained by assuming that the 2012 accident year is similar to the 2008 accident year in the pattern of making payments, and estimating cumulative payments at the end of Development Year 4 by:

$$\mathbf{1{,}182 \times \frac{2{,}519}{786} = 3{,}788}$$

This is not necessarily the 'best' estimate, but it is possible to fill in the lower triangle in Figure 1 by comparing present figures with past experience. This process is the main object of this chapter.

Note that the figure of 3,788 actually includes what we have already paid out. The estimate of the amount yet to be paid, using this method, would be $3{,}788 - 1{,}182 = 2{,}606$.

Question

Using the same assumption about claims in 2011, what would be the cumulative claim payment at the end of Development Year 4 for 2011 accidents?

Solution

If we assume as before that payments for 2011 will have a similar pattern to those of 2008, we would get:

$$2,142 \times \frac{2,519}{1,410} = 3,827$$

Note that, since we have the more up-to-date total of 2,142 for this row, this calculation is likely to be more accurate than calculating $1,220 \times \dfrac{2,519}{786}$, based on the older figure of 1,220.

0.5 Other ways of recording data

The method used above is based on an accident-year basis where claims development is clustered by the year an accident has occurred. In that respect, such data would include incurred but not yet reported (IBNR) claims.

Another method of recording data is by underwriting year. This procedure would group claims by the time policies were written rather than when they occurred.

So this method would include IBNR claims, and claims that are yet to occur, if they relate to a policy written in the underwriting year.

A third method of grouping claims is by reporting year.

This method would not include IBNR claims as all claims would be allocated to the year in which they were reported.

Example

Consider a claim on an incident occurring on 1st December 2017 for an insurance policy written on 21st December 2016. Assume that this was reported to the insurer on 15th January 2018 and settled in July 2018.

Then, assuming calendar years are used, it would show:

- **in development year 1 for 2017 under an accident-year basis**
- **in development year 2 for 2016 under an underwriting-year basis**
- **in development year 0 for 2018 under a reporting-year basis.**

The relative merits of each basis are discussed further in Subject SP7.

1 Projections using development factors

1.1 Run-off patterns

The basic assumption made in estimating outstanding claims concerns the run-off pattern. The simplest assumption is that payments will emerge in a similar way in each accident year. The proportionate increases in the known cumulative payments from one development year to the next can then be used to calculate the expected cumulative payments for future development years.

However, as the example below illustrates, there are a number of choices as to which such ratio should be used to project future claims.

Note that the ratios that are used to project future claims are known as development factors or link ratios.

A development factor may describe the ratio between cumulative claim amounts in consecutive years or between years over a longer period. In the table below we calculate development factors for the ratios between consecutive years.

Example

Proportionate increases in cumulative payments

Accident Year	Development Year								
	0		1		2		3		4
2008	786	*1.794*	1,410	*1.572*	2,216	*1.101*	2,440	*1.032*	2,519
2009	904	*1.742*	1,575	*1.597*	2,515	*1.112*	2,796		
2010	995	*1.823*	1,814	*1.588*	2,880				
2011	1,220	*1.756*	2,142						
2012	1,182								

Figure 2: Increase in cumulative payments

For each accident year from 2008 to 2011 there is a different ratio for the increase in cumulative payments from Development Year 0 to Development Year 1. It is not clear which is the 'correct' one to use when projecting forward for Accident Year 2012. For a conservative estimate of cumulative payments, it might be best to take the largest ratio, *ie* 1.823.

Question

Project the known figure for Accident Year 2012 across to Development Year 4 using:

(i) the largest ratio for each development year

(ii) the smallest ratio for each year.

Comment on the difference between your two results.

Solution

(i) Using the largest ratio for each development year we get:

$$1{,}182 \times 1.823 \times 1.597 \times 1.112 \times 1.032 = 3{,}949$$

(ii) Using the smallest ratio in each case we get:

$$1{,}182 \times 1.742 \times 1.572 \times 1.101 \times 1.032 = 3{,}678$$

Note that the results are substantially different. The estimate for the outstanding claims reserve will be very sensitive to the method used to calculate the development factors.

However, some sort of average of the ratios would seem more appropriate. It is possible to use a simple arithmetic average:

$$\frac{1.794 + 1.742 + 1.823 + 1.756}{4} = 1.779$$

The disadvantage of this is that it does not take into account that the years in which more claims occur provide more information. Thus, the greater the amount of claims, the more confidence you can have in the ratio.

Note that we're assuming a large *number* of claims here, which would lead to a more predictable average, not a small number of very large claims, which would probably have the opposite effect.

This suggests using a weighted average and the usual choice of weights are the cumulative claims values.

Accident Year	Ratio	Weight
2008	1.794	786
2009	1.742	904
2010	1.823	995
2011	1.756	1,220

$$\frac{1.794 \times 786 + 1.742 \times 904 + 1.823 \times 995 + 1.756 \times 1{,}220}{786 + 904 + 995 + 1{,}220} = 1.777$$

This method of estimating the ratios which describe the run-off pattern is called the chain-ladder method. The most efficient mode of calculating the ratios is given in Section 1.3.

Question

Using this method, what estimate would you give for the ratio to be used for calculating figures for Development Year 2 from Development Year 1?

Solution

$$\frac{1.572 \times 1,410 + 1.597 \times 1,575 + 1.588 \times 1,814}{1,410 + 1,575 + 1,814} = 1.586$$

1.2 A statistical model for run-off triangles

The general form of a run-off triangle can be expressed as follows:

Accident Year	Development Year 0	1	…	j	....	n
0	$C_{0,0}$	$C_{0,1}$	…	$C_{0,j}$	....	$C_{0,n}$
1	$C_{1,0}$	$C_{1,1}$	…	$C_{1,j}$	.... $C_{1,n-1}$	
⋮	⋮	⋮				
i	$C_{i,0}$	$C_{i,1}$			$C_{i,n-i}$	
⋮	⋮					
⋮	⋮					
⋮	⋮	$C_{n-1,1}$				
n	$C_{n,0}$					

Each entry, C_{ij}, in the run-off triangle represents the incremental claims (as opposed to cumulative claims) **and can be expressed in general terms**

$$C_{ij} = r_j s_i x_{i+j} + e_{ij}$$

where:

- r_j **is the development factor for year j, representing the proportion of claim payments in Development Year j. Each r_j is independent of the Origin Year i.**
- s_i **is a parameter varying by Origin Year, i, representing the exposure, for example the number of claims** (or claim amount) **incurred in the Origin Year i.**
- x_{i+j} **is a parameter varying by calendar year, for example representing inflation.**
- e_{ij} **is an error term.**

Note that some of the terminology above is being used in a different context to previously. The development factors r_j in this general statistical model are defined differently to the development factors that we have met previously.

The development factors that we met previously were used to project forward *cumulative* claims data. The development factors in the general statistical model are being used to model *incremental* data. They are defined above as the proportion of claims from a particular accident year that are paid in the j^{th} development year. As such, they are a set of factors that add up to 1.

Question

An actuary using this model has estimated the parameters for a run-off triangle as follows:

$$s_{08} = £1.50m, \quad r_0 = 0.6, \quad x_{08} = 1.00$$
$$s_{09} = £1.75m, \quad r_1 = 0.3, \quad x_{09} = 1.10$$
$$s_{10} = £1.60m, \quad r_2 = 0.1, \quad x_{10} = 1.20$$
$$x_{11} = 1.25$$
$$x_{12} = 1.30$$

Use these estimates (ignoring error terms) to construct the complete table of incremental claim amounts for Accident Years 2008-2010, and hence estimate the amount of outstanding claims at the end of 2010.

Solution

In this example, there are only 3 origin years and development years. Multiplying the appropriate parameters gives the following incremental table:

Claim payments made during year (£000)		Development Year		
		0	1	2
Accident Year	2008	900	495	180
	2009	1,155	630	219
	2010	1,152	600	208

The outstanding claim amount at the end of 2010 is $219 + 600 + 208 = 1,027$ *ie* £1,027,000.

Question

Explain what each of the parameters s_i , r_j and x_{i+j} would represent if you were applying this model to a run-off triangle showing the incremental *number* of claims paid, tabulated by *reporting* year.

Solution

- s_i would represent the total number of claims reported in origin year i .
- r_j would represent the proportion of claims where a payment was completed in development year j .
- x_{i+j} would be 1, since we're looking at *numbers* of claims.

1.3 The chain ladder method

This section explains how you would carry out the calculations for completing the run-off triangle using the basic chain ladder method.

This method of calculating the development ratios is demonstrated in the following example.

Example

Recall that the ratio in Accident Year 2008 was calculated as follows:

$$1.794 = \frac{1,410}{786}$$

The ratios for the other accident years were calculated in a similar way. The numerator of the last Core Reading equation of Section 1.1 can therefore be written as:

$$\frac{1,410}{786} \times 786 + \frac{1,575}{904} \times 904 + \frac{1,814}{995} \times 995 + \frac{2,142}{1,220} \times 1,220$$

$$= 1,410 + 1,575 + 1,814 + 2,142$$

Thus, the development factor can be calculated using the cumulative claims in Development Years 0 and 1:

$$\frac{1,410 + 1,575 + 1,814 + 2,142}{786 + 904 + 995 + 1,220}$$

In other words the development factor is the sum of the figures in Column 1 divided by the sum of the *corresponding* figures from Column 0.

The name given to this method presumably arises from the ladder-like operations which are chained over the development years. The development factors for the chain ladder technique can be found for each development year by adding the appropriate number of terms. This is illustrated below.

Accident Year	Development Year				
	0	1	2	3	4
2008	786	1,410	2,216	2,440	2,519
2009	904	1,575	2,515	2,796	
2010	995	1,814	2,880		
2011	1,220	2,142			
2012	1,182				

$$\frac{6,941}{3,905} = 1.777 \qquad \frac{7,611}{4,799} = 1.586 \qquad \frac{5,236}{4,731} = 1.107 \qquad \frac{2,519}{2,440} = 1.032$$

Figure 3: Development factors

Note that the development factors will normally be 1-point-something.

Development factors have been calculated for each development year. It is now possible to project forward each accident year.

For Accident Year 2012, the projections of cumulative claims are:

$$1{,}182 \times 1.777 = 2{,}100$$

$$1{,}182 \times 1.777 \times 1.586 = 3{,}331$$

$$1{,}182 \times 1.777 \times 1.586 \times 1.107 = 3{,}688$$

$$1{,}182 \times 1.777 \times 1.586 \times 1.107 \times 1.032 = 3{,}806$$

Each calculation follows on from the previous one, so you don't have to multiply repeatedly by the same factors. Storing the development factors in your calculator's memories will also save time and maintain accuracy.

For Accident Year 2011, start from 2,142 in Development Year 1 and use only the last three link ratios (*ie* development factors)**.**

Accident Year	Development Year				
	0	1	2	3	4
2008					
2009					2,885
2010				3,188	3,290
2011			3,397	3,761	3,881
2012		2,100	3,331	3,688	3,806

Figure 4: Projections of cumulative payments

Note that no projection can be done for the first accident year because it is not possible to project beyond the highest development year. It is therefore assumed that all claims from this cohort have completely run off.

We would like to think that claims from Accident Year 2008 are now 'completely run off', *ie* that we can expect to make no further payments on these claims. So, in fact, we would not need to project the 2008 row. If we do not believe this to be the case, we should be using a triangle with a correspondingly greater number of columns.

In practice, where this assumption may not be appropriate, manual adjustments in the form of 'tail factors' may be used. These are beyond the scope of Subject CM2, and will be discussed in Subject SP7.

The reserve that needs to be held at the end of 2012 is the sum over all accident years for which a projection has been made of the difference between the cumulative payment at the end of Development Year 4 and the last known entry in the development triangle for that accident year.

So from Figures 1 and 4, the reserve at the end of 2012 is:

$$(2{,}885-2{,}796)+(3{,}290-2{,}880)+(3{,}881-2{,}142)+(3{,}806-1{,}182)=4{,}862$$

Note that no discount rate has been applied to the payments in different years.

This is the usual convention when using run-off triangles to calculate reserves. It is consistent with the way accounts are drawn up for other companies. However, for much insurance business of this type, the payments are usually fairly short-tail, *ie* we expect to have made the final payments on claims within a few years. If claims are paid quickly, there is less need to allow for any discounting. In addition, not discounting reserves errs on the side of prudence, rather than the reverse, which makes it a safe approach. There would be nothing to stop you including the appropriate v factors to calculate discounted reserves, if this was required.

The model can be applied in exactly the same way to the distribution of the number of claims, rather than to total claim amounts.

Question

The table below shows the numbers of household insurance claims reported in each development year for accident years from 2009 to 2012. Use the basic chain ladder method to estimate the total ultimate number of claims arising from accidents occurring between 1 January 2009 and 31 December 2012.

Number of claims reported		Development Year			
		0	1	2	3
Accident Year	2009	17,500	5,000	2,250	750
	2010	21,000	6,200	2,750	
	2011	18,800	5,500		
	2012	21,300			

Solution

The first thing we must do is to get the *cumulative* claim numbers.

If we write the development factors over the appropriate columns, the table of cumulative claim numbers looks like this:

$\times 1.2914$ $\times 1.1006$ $\times 1.0303$

Cumulative claim numbers		Development Year			
		0	1	2	3
Accident Year	2009	17,500	22,500	24,750	25,500
	2010	21,000	27,200	29,950	30,858
	2011	18,800	24,300	26,745	27,555
	2012	21,300	27,508	30,275	31,193

The projected ultimate number of claims is just the total of the last column:

$$25{,}500+30{,}858+27{,}555+31{,}193=115{,}106$$

1.4 Model checking

The chain ladder technique is used primarily to estimate the development of cumulative claim payments. However, it is useful to check whether it fits reasonably with the claims data which have already been received. To illustrate this, look at the data in Figure 2.

To check how well the chain ladder technique performs, the claims in Development Year 0 for Accident Years 2008-2011 will be considered in the example below.

Example

The actual claims in Development Year 0 are as follows:

2008	**786**
2009	**904**
2010	**995**
2011	**1,220**

The development factors calculated in Section 1.3 were 1.777, 1.586, 1.107 and 1.032. Using these, estimates of cumulative claim payments in each development year can be obtained. It is of particular interest to compare these with the actual values given in Figure 2.

Hence, the following table gives the 'fitted' values using the chain ladder technique.

Accident Year	**Development Year**				
	0	**1**	**2**	**3**	**4**
2008	**786**	**1,397**	**2,215**	**2,452**	**2,531**
2009	**904**	**1,606**	**2,548**	**2,820**	
2010	**995**	**1,768**	**2,804**		
2011	**1,220**	**2,168**			

Figure 5: Fitted cumulative claim payments

It is now possible to compare Figure 5 with Figure 1. However, it is preferable to look at the increases in cumulative payments when considering the fit of the model as it gives a more sensitive test.

The increases in cumulative payments with development year (both actual and fitted) are given in Figure 6.

The error is given by actual – fitted.

		Development Year				
		0	**1**	**2**	**3**	**4**
2008	**Actual**	786	624	806	224	79
	Fitted	786	611	818	237	79
	Error	–	*13*	*–12*	*–13*	*0*
2009	**Actual**	904	671	940	281	
	Fitted	904	702	942	272	
	Error	–	*–31*	*–2*	*9*	
2010	**Actual**	995	819	1,066		
	Fitted	995	773	1,036		
	Error	–	*46*	*30*		
2011	**Actual**	1,220	922			
	Fitted	1,220	948			
	Error	–	*–26*			

Figure 6: Errors

None of the errors is large enough to suggest that the model is inaccurate.

However, it is not clear exactly how large the differences should be before we start to doubt whether the model is accurate.

Question

Apply this technique to the data in the previous question and comment on any unusually large error figures.

Solution

Calculating the increase in the cumulative claim numbers as before gives the following figures.

		Development Year			
		0	1	2	3
2009	Actual	17,500	5,000	2,250	750
	Expected	17,500	5,100	2,274	754
	Error	0	−100	−24	−4
2010	Actual	21,000	6,200	2,750	
	Expected	21,000	6,119	2,728	
	Error	0	81	22	
2011	Actual	18,800	5,500		
	Expected	18,800	5,478		
	Error	0	22		
2012	Actual	21,300			
	Expected	21,300			
	Error	0			

Again, the fit looks pretty good.

Question

Is it possible to apply a chi square goodness of fit test here?

Solution

Possibly. But note that you are dealing with claim amounts here, and not frequencies. So some additional assumptions would be necessary.

If each claim amount is for a fixed amount m, and the number of claims N has a $Poisson(\lambda)$ distribution, then the total amount S for a given year would have mean λm and variance λm^2.

Therefore:

$$\frac{S-\lambda m}{\sqrt{\lambda m^2}}=\frac{Nm-\lambda m}{\sqrt{\lambda m^2}}=\frac{N-\lambda}{\sqrt{\lambda}}$$

This expression is of the form $\frac{A-E}{\sqrt{E}}$, where A is the actual claim number and E is the expected claim number.

This quantity would have an approximately $N(0,1)$ distribution, since N is approximately normal for large parameter values. So the square of this quantity would have an approximate chi square distribution, and the chi square test could be used.

However, despite this check it is quite possible that the estimate obtained may be a poor guide to the future.

Question

Why?

Solution

Because any of a number of factors may have changed. For example:

- Claims inflation may increase suddenly.
- Office expenses may also increase (if these are allowed for in the calculations).
- Weather patterns may change, altering the balance between short tail and long tail claims (*ie* poor weather conditions may increase the number of claims for damage, which are settled reasonably quickly, but have no effect on the number of claims for liability, which are generally settled much more slowly).
- The characteristics and behaviour of the insured lives in the portfolio may change over time.
- Catastrophe claims may occur.
- Underwriting procedures may change.

1.5 Other methods of deriving development factors

It is possible to adjust the calculated development factors in the light of other information. This method which uses prior knowledge can take a formal approach, but it is more often an ad hoc adjustment. There may be good reasons to change the development factors. For example, changes in accounting methods or claims administration can alter the speed with which claims are settled. This would give rise to changes in the development factors and it would be sensible to reflect this in the estimate of future claim payments. The development factors, either calculated directly from the data, or set using expert knowledge, are always used in the same way to estimate outstanding claim payments.

For example, suppose that a new computer system is installed that speeds up the claims process substantially. If a run-off triangle method is applied to a period that spans the time both before and after the installation of the new system, the underlying development factors are likely to change when the system is installed. Because calendar years span the diagonals of the triangle, rather than horizontal lines, the effect of a change of this kind is likely to be unpredictable.

When using run-off triangles, you should always be on the look out for factors that may distort the pattern of the run-off. It may be possible to allow for distortions by adjusting the figures produced by the triangle.

The chain ladder method can also be applied to a triangle of loss ratio data rather than cumulative payments, where the loss ratio for a given development and accident year is the cumulative payment up to and including that development year divided by the total premium income in respect of the given accident year.

Question

Apply the chain ladder method to the triangle of loss ratios shown below. Estimate the ultimate loss ratio in each year, and the amount that needs to be set aside to meet future claims. The amounts of premium received in each of the years 2009 to 2012 were £1.42m, £1.64m, £1.73m and £1.82m.

Loss ratio		Development year			
		0	1	2	3
Accident year	2009	47%	63%	70%	74%
	2010	48%	62%	71%	
	2011	49%	60%		
	2012	50%			

Solution

Applying the standard chain ladder approach we get the following triangle of figures:

Loss ratio		Development Year 0	1	2	3
Accident Year	2009	47%	63%	70%	74%
	2010	48%	62%	71%	75.06%
	2011	49%	60%	67.68%	71.55%
	2012	50%	64.24%	72.46%	76.60%

So the ultimate loss ratios are the figures in the right hand column.

The amount of the outstanding claims reserve is:

$$(0.7660 - 0.50) \times 1.82 + (0.7155 - 0.60) \times 1.73 + (0.7506 - 0.71) \times 1.64 = 0.7505\text{m}$$

or £750,500.

1.6 Assumptions underlying the method

The chain ladder technique is based on the assumption that payments from each accident year will develop in the same way. In other words, the same development factors are used to project outstanding claims for each accident year. Changes in the rate at which claims emerge can only be incorporated by adjustment of the development factors.

The final assumption made when the chain ladder technique is used concerns inflation. It is assumed that weighted average past inflation will be repeated in the future. This is because claims inflation is one of the influences swept up within the projection factors.

Using the general statistical model described earlier, it can be seen that the basic chain ladder takes the form:

$$C_{ij} = r_j s_i + e_{ij}$$

This might be an unrealistic assumption, and it will be considered in greater detail in the following section. When considering inflation, it is important to bear in mind that it is claims inflation which is important. Thus, although a standard measure of overall inflation may be used, the inflation rate inherent in claims may be quite different. For example, a court decision can affect the size of claim payments. Later, we deal with claims inflation in more detail.

When you read about inflation on news websites or in newspapers, it usually refers to inflation in the price of consumer goods. However, this is just one example of inflation. Here we are interested in the rates of inflation in the cost of settling claims, and there are in fact special indices published that show the past rates of inflation for things like the cost of repairing vehicles.

2 Adjusting for inflation

2.1 The inflation adjusted chain ladder method

Dealing with past inflation

Claims inflation will affect the payments in the run-off triangle by calendar year of payment.

These were represented by the x's in our statistical model earlier.

The inflation adjusted chain ladder method works by adjusting the figures in the triangle to allow for the effects of inflation.

In the model considered here, it will be assumed that claims inflation is at the same annual rate for all claims within a particular calendar year of payment. Each calendar year of payment corresponds to a diagonal in the triangle. For an illustration, look again at Figure 1.

When adjusting for inflation, it is the payments in each calendar year which need to be considered, rather than cumulative totals. The first step is to calculate incremental payments from the cumulative totals, by differencing along each row. The same operation was performed earlier and the following figure can be compared with Figure 6.

Example

Figure 7 gives the incremental (or non-cumulative) claim payments for the data in Figure 1.

Accident Year	Development Year				
	0	1	2	3	4
2008	786	624	806	224	79
2009	904	671	940	281	
2010	995	819	1,066		
2011	1,220	922			
2012	1,182				

Figure 7: Incremental claim payments in monetary amounts

Suppose that the annual claim payments inflation rates over the 12 months up to the middle of the given year are as follows:

2009	5.1%
2010	6.4%
2011	7.3%
2012	5.4%

For simplicity, it is also assumed that payments are made in the middle of each calendar year. An index can now be calculated in order to convert all payments to mid-2012 prices.

2012 has been chosen because it is the last 'known' year.

Question

What would the index be for the years 2008-2012 if the base value is 100 for 2008?

Solution

2008	100
2009	105.1
2010	111.8
2011	120.0
2012	126.5

The payments in Figure 7 can now be adjusted using the inflation rates. Figure 8 gives the inflation adjusted incremental payment data.

For example, we have:

$$786 \times 1.051 \times 1.064 \times 1.073 \times 1.054 = 994 \text{ (or } 786 \times \frac{126.5}{100} = 994 \text{)}$$

and $1{,}220 \times 1.054 = 1{,}286$ (or $1{,}220 \times \frac{126.5}{120.0} = 1{,}286$)

Accident Year	**Development Year**				
	0	**1**	**2**	**3**	**4**
2008	**994**	**751**	**912**	**236**	**79**
2009	**1,088**	**759**	**991**	**281**	
2010	**1,125**	**863**	**1,066**		
2011	**1,286**	**922**			
2012	**1,182**				

Figure 8: Incremental claim payments at mid-2012 prices

Now it is straightforward to form a table of inflation adjusted cumulative payments to which the chain ladder technique can be applied.

Applying the chain ladder method we get the following cumulative figures:

		Development Year				
		0	1	2	3	4
Accident Year	2008	994	1,745	2,657	2,893	2,972
	2009	1,088	1,847	2,838	3,119	
	2010	1,125	1,988	3,054		
	2011	1,286	2,208			
	2012	1,182				

This gives development factors of 1.7334, 1.5321, 1.0941 and 1.0273.

The forecasts of cumulative payments at mid-2012 prices are given in Figure 9.

Accident Year	**Development Year**			
	1	**2**	**3**	**4**
2009				**3,204**
2010			**3,341**	**3,432**
2011		**3,383**	**3,701**	**3,802**
2012	**2,049**	**3,139**	**3,434**	**3,528**

Figure 9: Forecasts of cumulative claim payments at mid-2012 prices

If you're checking these figures, you might get answers that differ by 1 or 2 from the ones shown, depending on how you have rounded your intermediate figures.

Dealing with future inflation

The predictions of cumulative payments do not, however, take account of future inflation. In order to forecast the actual payments, an assumed rate of future inflation will be needed. Again, it is necessary to convert to non-cumulative data rather than the cumulative totals before adjusting these for future inflation in a similar way to that used when dealing with past inflation.

Example

In this example we will assume a constant future rate of inflation.

Applying an annual inflation rate of 10% (at 30 June) to the data in Figure 9 gives revised forecasts of the cumulative claim payments as follows:

Accident Year	**Development Year**			
	1	**2**	**3**	**4**
2009				**2,890**
2010			**3,196**	**3,306**
2011		**3,435**	**3,820**	**3,954**
2012	**2,136**	**3,455**	**3,848**	**3,986**

Figure 10: Forecasts of cumulative claim payments in monetary amounts

The reserve that needs to be held at the end of 2012 is 5,136.

Question

Reproduce the figures shown in the table and the reserve figure of 5,136.

Solution

The incremental table in mid-2012 prices is:

Incremental payments at 2012 prices		Development Year			
		1	2	3	4
Accident Year	2009				85
	2010			287	91
	2011		1,175	318	101
	2012	867	1,090	295	94

Applying the 10% future inflation rate:

Incremental payments allowing for future inflation		Development Year			
		1	2	3	4
Accident Year	2009				94
	2010			316	110
	2011		1,293	385	134
	2012	954	1,319	393	138

Accumulating using the original figures from Figure 1:

Cumulative payments in actual amounts		Development Year				
		0	1	2	3	4
Accident Year	2008	786	1,410	2,216	2,440	2,519
	2009	904	1,575	2,515	2,796	2,890
	2010	995	1,814	2,880	3,196	3,306
	2011	1,220	2,142	3,435	3,820	3,954
	2012	1,182	2,136	3,455	3,848	3,986

These figures agree with those shown in Figure 10.

The reserve is calculated as:

$$(3{,}986 - 1{,}182) + \cdots + (2{,}890 - 2{,}796) = 5{,}136$$

Question

Can we apply the method if we believe that inflation will vary in each future year?

Solution

Of course, provided that we can estimate the inflation rate for each future year. We just multiply by the appropriate future inflation factors, which will be different for each year, exactly as we did for past inflation.

Assumptions underlying the method

The key assumption underlying this method is that, for each origin year, the amount of claims paid, in real terms, in each development year is a constant proportion of the total claims, in real terms, from that origin year.

Explicit assumptions are made for both past and future claims inflation. Therefore, using the general statistical model, the inflation adjusted chain ladder method takes the form:

$$C_{ij} = r_j s_i x_{i+j} + e_{ij}$$

As before, we are also assuming that the first origin year is fully run off.

3 The average cost per claim method

This method, considers separately the two key elements of total claim amounts, ie the number of claims and the average amounts of the claims.

3.1 Description of method

This method requires development tables for both total claim amounts and claim numbers.

We normally use cumulative figures, as before.

A third development table, of the average claim amounts, is then formed by dividing the figures in the corresponding cells of the first two tables.

The next stage is the projection of figures in the average claims and number of claims tables, using either grossing-up factors or development factors.

A grossing-up factor is not very different from a development factor. A grossing-up factor gives the proportion of the ultimate claim amount that has been paid so far. Suppose that we had cumulative payment amounts of:

500 800 1000 1,100

The development factors for this row would be:

$$\frac{800}{500}=1.6 \qquad \frac{1,000}{800}=1.25 \qquad \text{and} \qquad \frac{1,100}{1,000}=1.1$$

The grossing-up factors would be calculated as follows:

$$\frac{500}{1,100}=45.45\% \qquad \frac{800}{1,100}=72.73\% \qquad \frac{1,000}{1,100}=90.91\%$$

Grossing-up factors are particularly useful when we already know the ultimate expected payout amount for a table row and are trying to calculate individual figures in the row. We will see that this is usually the case in the next two methods that we study.

Finally, the projected ultimate claims can be calculated by multiplying together for each accident year the projected figures for the average claim amounts and claim numbers.

A reserve can then be calculated by subtracting all payments to date in respect of claims relating to the data in the table. An example is given below to illustrate this process.

3.2 Application of the method

The average cost per claim method is not uniquely defined. It may therefore equally be applied on an accident year cohort to either paid or incurred claims, or on a reporting year cohort. It is, however, important to ensure that the form of the data for the number of claims corresponds to that of the total claim amounts (*ie* paid claims corresponds to the number of claims settled and incurred claims corresponds to the number reported).

Of course, it is also important only to apply the method to data for which the development is considered to be stable and hence suitable for projection into the future. It is more likely that the reporting of claims will be more stable than the settlement, so the example given below relates to incurred claims.

In general insurance, 'incurred claims' has a technical definition. In simple cases, however, it just corresponds to the intuitive meaning of claims arising in a particular period.

Further, the method is not uniquely defined as using a particular form of grossing-up factors or development factors. However, the grossing-up method is generally considered simpler and is used in its simple average form in the example that follows.

Finally, the method described above ignored any adjustment for inflation. This can, however, be done in exactly the same way as the adjustment to the basic chain ladder method to form the inflation adjusted chain ladder method (*ie* if the data being used have been adjusted for inflation, it would simply require an index for future inflation to be applied to non-cumulative projected average claim amounts before multiplying by the projected claim numbers). In practice an adjustment for inflation would normally be made.

The following example is done on the basis that the data has not, and need not be, adjusted for inflation.

You should be able to see how the method would be adjusted for inflation (and you should be prepared to answer questions that ask you to make such an adjustment).

Cumulative incurred claims data, by years of accident and reporting development

		DY						
		0	**1**	**2**	**3**	**4**	**5**	***Ult***
	1	**2,777**	**3,264**	**3,452**	**3,594**	**3,719**	**3,717**	**3,717**
	2	**3,252**	**3,804**	**3,973**	**4,231**	**4,319**		
AY	**3**	**3,725**	**4,404**	**4,779**	**4,946**			
	4	**4,521**	**5,422**	**5,676**				
	5	**5,369**	**6,142**					
	6	**5,818**						

This is a different set of data from any of the previous examples.

Question

How can the figure of 3,719 in Development Year 4 of Accident Year 1 *reduce* in the following development year?

Solution

Since we are working with incurred claims, rather than claims paid, we are looking at the amounts that we believe we have incurred as liabilities in each development year. There are a number of factors that might cause the cumulative amount incurred to decrease. For example, we might expect a reinsurance payment relating to items incurred in the previous year to be paid. If such an amount was large, it could cause the total cumulative incurred amount to decrease.

Number of reported claims, by year of accident and reporting development

		DY						
		0	1	2	3	4	5	*Ult*
	1	414	460	482	488	492	494	494
	2	453	506	526	536	539		
AY	3	494	548	572	582			
	4	530	588	615				
	5	545	605					
	6	557						

These figures are also cumulative.

Dividing each cell in the first table by the corresponding cell in the second gives the accumulated average incurred cost per claim.

Average incurred cost per claim, by year of accident and reporting development.

		DY						
		0	1	2	3	4	5	*Ult*
	1	6.708	7.096	7.162	7.365	7.559	7.524	7.524
	2	7.179	7.518	7.553	7.894	8.013		
AY	3	7.540	8.036	8.355	8.498			
	4	8.530	9.221	9.229				
	5	9.851	10.152					
	6	10.445						

These tables lead to the grossing-up factors and projected ultimate figures given in the following table (the projections are based on the, underlined, simple averages of the grossing-up factors).

Average claim amounts

		DY						
	AY	0	1	2	3	4	5	Ult
	1	6.708	7.096	7.162	7.365	7.559	7.524	7.524
		89.2%	94.3%	95.2%	97.9%	100.5%	100.0%	
	2	7.179	7.518	7.553	7.894	8.013		7.973
		90.0%	94.3%	94.7%	99.0%	100.5%		
AY	3	7.540	8.036	8.355	8.498			8.632
		87.4%	93.1%	96.8%	98.45%			
	4	8.530	9.221	9.229				9.657
		88.3%	95.5%	95.57%				
	5	9.851	10.152					10.766
		91.5%	94.3%					
	6	10.445						11.699
		89.28%						

It may not be immediately obvious what is happening here. Let's see how these figures are calculated.

Accident Year 1 is fully run off. We can express the figures for each year as a percentage of 7.524, the final figure. For example, $\dfrac{7.365}{7.524} = 0.979$.

Now look at Accident Year 2. We use the corresponding figure in Accident Year 1 (because we have already filled in the percentages here) to find a grossing-up factor for Development Year 4, *ie* 100.5%. So we can find the ultimate expected payout figure for Accident Year 2 as $\dfrac{8.013}{1.005} = 7.973$.

Now calculate the grossing-up factors for Accident Year 2 by expressing the figures in Accident Year 2 as a percentage of 7.973.

Now look at Accident Year 3. We calculate the grossing-up factor for Development Year 3 by taking the average of the two figures that we already know, *ie* $\tfrac{1}{2}(97.9 + 99.0) = 98.45\%$.

Now calculate the ultimate figure for Accident Year 3 : $\dfrac{8.498}{0.9845} = 8.632$

Use this figure to calculate grossing-up factors for the whole of Accident Year 3.

Continue through the table, using the average of the known grossing-up factors to calculate the required grossing-up factor for each development year in turn.

We end up with an ultimate claim amount for each accident year.

We now do exactly the same for the claim number table.

Question

Why did we not just complete the average cost table using the basic chain ladder method?

Solution

No reason! As we said, the ACPC method is not a uniquely defined procedure. The method described in the Core Reading is just one possibility.

Claim numbers

		DY						
		0	**1**	**2**	**3**	**4**	**5**	***Ult***
	1	**414**	**460**	**482**	**488**	**492**	**494**	**494**
		83.8%	**93.1%**	**97.6%**	**98.8%**	**99.6%**	**100.0%**	
	2	**453**	**506**	**526**	**536**	**539**		**541**
AY		**83.7%**	**93.5%**	**97.2%**	**99.1%**	**99.6%**		
	3	**494**	**548**	**572**	**582**			**588**
		84.0%	**93.2%**	**97.3%**	**98.95%**			
	4	**530**	**588**	**615**				**632**
		83.9%	**93.0%**	**97.37%**				
	5	**545**	**605**					**649**
		84.0%	**93.2%**					
	6	**557**						**664**
		83.88%						

The total ultimate loss is therefore the sum of the following projected amounts for each accident year:

AY	Average cost per claim	×	Claim Numbers	=	Projected Loss estimate
1	7.524		494		3,717
2	7.973		541		4,313
3	8.632		588		5,076
4	9.657		632		6,103
5	10.766		649		6,987
6	11.699		664		7,768
			Total Projected Loss Estimate	=	33,964

If the claims paid to date amounted to 20,334, the total reserve required would be 13,630.

Because the triangle we started with was based on *incurred claims*, we can't deduce the total paid claims from the figures in the triangle.

Question

Using an average cost per claims method based on development factors, find the outstanding claims reserve at the end of 2012 in the following example. Claims paid to date are £1,902,000.

Cumulative incurred claims (£k)		Development Year			
		0	1	2	3
Accident Year	2009	632	714	788	822
	2010	729	784	803	
	2011	800	855		
	2012	824			

Cumulative number of claims		Development Year			
		0	1	2	3
Accident Year	2009	52	60	66	70
	2010	54	63	65	
	2011	60	70		
	2012	65			

Solution

Dividing one table of values by the other, we find the average cost per claim for each accident and development year:

Average cost per claim		Development Year			
		0	1	2	3
Accident Year	2009	12.154	11.900	11.939	11.743
	2010	13.500	12.444	12.354	
	2011	13.333	12.214		
	2012	12.677			

Since we are using development factors, we can just complete the table using the usual chain ladder approach. However, because we're not dealing with total amounts here, we do not need to accumulate in this situation. The development factors of 0.9377, 0.9979 and 0.9836 give:

Average cost per claim		Development Year			
		0	1	2	3
Accident Year	2009	12.154	11.900	11.939	11.743
	2010	13.500	12.444	12.354	12.151
	2011	13.333	12.214	12.188	11.988
	2012	12.677	11.887	11.862	11.667

Completing the claim number table in the same way, using development factors 1.1627, 1.0650. and 1.0606, we get:

Cumulative number of claims		Development Year			
		0	1	2	3
Accident Year	2009	52	60	66	70
	2010	54	63	65	68.939
	2011	60	70	74.553	79.071
	2012	65	75.572	80.488	85.366

We can now calculate the total estimated payout for all claims arising during the period:

$$822 + 68.939 \times 12.151 + 79.071 \times 11.988 + 85.366 \times 11.667 = 3,604$$

But since we have already paid out 1,902, the amount outstanding is 1,702, *ie* £1,702,000.

3.3 Assumptions underlying the method

As there is no unique way of defining the method, there is no unique set of assumptions. In particular, the assumptions relating to inflation will depend on the data used.

In general terms, however, there are the assumptions that for each origin year, both the number and average amount of claims relating to each development year are constant proportions of the totals from that origin year.

Finally, it is worth noting that for the assumptions to hold for this method, it would be normal for them to also hold for a simpler method applying to total rather than average claim amounts, such as the chain ladder method.

You may then ask: 'Why then do we not just use the basic chain ladder method?'. The answer is that the totals used in the basic chain ladder method contain a combination of the patterns of the average amount and the numbers of claims. By analysing these separately, we hope to get a more accurate projection.

You may also be asking: 'If an exam question asks for the ACPC method, what approach do I take?' The answer here is that you may need to use your judgement based on the form of the data given and any instructions included in the wording of the question. It is possible that the examiners might set a question where a variety of approaches were equally acceptable. In any case, the examiners will be trying to test your understanding, not trying to catch you out.

4 Loss ratios

The ratio of incurred claims to earned premiums over a defined period is called the loss ratio.

ie $$\text{loss ratio} = \frac{\text{incurred claims}}{\text{earned premiums}}$$

Investigation of the loss ratios for each of several different origin years would normally show some consistency, provided that there have not been any distortions, and in particular no significant change in premium rates.

A good example of a 'distortion' would be the hurricane that affected the South of England in 1987. This changed the pattern of household insurance claims completely during that origin year.

An increase in premiums would affect the loss ratio directly.

The expected loss ratios will also have formed part of the derivation of the premium basis.

It is therefore logical that a loss ratio based on trends of past data, underwriters' views, or market data, could be used as a basis for an estimate of the eventual loss and hence the outstanding claims. It is, however, on its own a very crude measure due to the fluctuations that are inherent in any claims experience.

This is similar to the approach that we used earlier when we applied the chain ladder method to a triangle of loss ratios when we used the ACPC method. However, the approach here is more general, in that the ultimate loss ratios can be estimated using any method, including subjective methods involving personal judgement. However, once the ultimate estimated loss ratios have been found, they are applied to the premium figures in order to calculate outstanding claims reserve figures, just as we did before.

Such estimated loss ratios may be useful as an input to Bornhuetter-Ferguson based estimates. These are discussed in the next section.

5 The Bornhuetter-Ferguson method

5.1 Concept of the Bornhuetter-Ferguson method

The Bornhuetter-Ferguson method combines the estimated loss ratio with a projection method.

Here, 'projection method' refers to methods such as the basic chain ladder method which are based on past claim amounts and/or numbers.

It therefore improves on the crude use of a loss ratio by taking account of the information provided by the latest development pattern of the claims, whilst the addition of the loss ratio to a projection method serves to add some stability against distortions in the development pattern.

The concepts behind the method are:

- **That whatever claims have already developed in relation to a given origin year, the future development pattern will follow that experienced for other origin years.**
- **The past development for a given origin year does not necessarily provide a better clue to future claims than the more general loss ratio.**

In other words it is a compromise that combines the loss ratios with the development pattern.

5.2 Description of the method

In its simplest form the concept leads to the following approach to calculations:

1. **Determine the initial estimate of the total ultimate claims from each origin year using premiums and loss ratios.**
2. **Divide these estimates by projection factors (*f*) determined, in a normal manner, from a claims development table. These are effectively estimates of the claims that should have developed to date.**
3. **Subtract these amounts from the corresponding total ultimate claims figures to give an estimate of the amount of claims that are yet to develop.**

Clearly, the three stages could be combined and expressed as:

$$\textbf{Future claims development} = \textbf{Premium} \times \textbf{Estimated Loss Ratio} \times (1 - 1/f)$$

We can relate this formula to the explanation given above:

1 Step 1 gives you premium $\times$ expected loss ratio

2 Step 2 gives you premium $\times$ expected loss ratio $\times 1/f$

3 Step 3 gives you Step 1 minus Step 2, which is the formula given.

As the final estimate of the ultimate loss is based on observed data and an initial estimate ignoring the observations, this method could be viewed as using a Bayesian approach. Bayesian statistics is covered in Subject CS1.

In standard Bayesian work we combine data from two sources. We find an estimate of a parameter value based on a prior distribution, and combine it with an estimate based on a likelihood function. Our final estimate for the parameter value is usually a weighted average of the two estimates from these two different sources.

We are doing the same thing here. We find an estimate of the outstanding losses based on loss ratios (*ie not* from the data in the run off triangle), and use the data in the triangle as a second source of information to refine our estimate.

The analysis is not precise here because there is no single parameter that we're trying to estimate.

5.3 Application of the method

In its original form, the Bornhuetter-Ferguson method was applied to the development of incurred claims. However, it could equally be applied to the development of paid claims, using either an accident year or policy year cohort.

Further, the original projection was done using a chain ladder approach, although alternative development factors or grossing-up factors (*g*) could easily be applied instead (*ie g* would replace 1/*f* in the above expression).

The original form also made no explicit adjustment for inflation, although the method could be adjusted in a similar way to the other methods.

The example below is based on the original form of the method, but examiners would expect students to also be able to apply the method to paid claims.

The first stage is to determine the development factors, using the same method as for the chain ladder methods.

Cumulative incurred claims data, by years of accident and reporting development

		DY						
		0	**1**	**2**	**3**	**4**	**5**	***Ult***
	1	**2,866**	**3,334**	**3,503**	**3,624**	**3,719**	**3,717**	**3,717**
	2	**3,359**	**3,889**	**4,033**	**4,231**	**4,319**		
AY	**3**	**3,848**	**4,503**	**4,779**	**4,946**			
	4	**4,673**	**5,422**	**5,676**				
	5	**5,369**	**6,142**					
	6	**5,818**						
TOTAL		**25,933**	**23,290**	**17,991**	**12,801**	**8,038**	**3,717**	
TOTAL – last no:		**20,115**	**17,148**	**12,315**	**7,855**	**3,719**		
RATIO (*r*)		**1.158**	**1.049**	**1.039**	**1.023**	**0.999**	**1.000**	
DEVELOPMENT FACTOR (*f*)		**1.290**	**1.114**	**1.062**	**1.022**	**0.999**	**1.000**	

Don't get confused by the numbers in this table. The numbers in the top left-hand corner of the triangle are slightly different from the ones used in the ACPC example.

The 'TOTAL' figures have been calculated by summing the entries in each column, *eg* the second TOTAL is:

$$3,334+3,889+4,503+5,422+6,142=23,290$$

The 'TOTAL – last no' figures have been calculated by summing all but the last entry in each column, *eg* the first TOTAL – last no figure is:

$$25,933-5,818=20,115$$

The RATIO figures are what we previously referred to as the development factors. These are calculated in the usual way. Note, however, that since we are given the TOTAL and TOTAL – last no for each development year, the calculation can be simplified. For example, the ratio of DY1 to DY0 is given by:

$$\frac{23,290}{20,115}=1.158$$

What the Core Reading refers to as DEVELOPMENT FACTORS are actually *cumulative* development factors, *ie* they apply from the development year of the column they are in up to Development Year 5. For example, the figure of 1.290 is calculated as:

$$1.158 \times 1.049 \times 1.039 \times 1.023 \times 0.999 \times 1.000 = 1.290$$

So, if we were to use the Basic Chain Ladder method to estimate the cumulative claims incurred by the end of Development Year 5 in respect of Accident Year 6, we would obtain:

$$5,818 \times 1.290 = 7,505$$

Question

Confirm the figures of $f = 1.062$ and $r = 1.039$ for Development Year 3.

Solution

The figures are:

$$r = \frac{12,801}{12,315} = 1.0394$$

and $$f = 1.000 \times 0.999 \times 1.023 \times 1.039 = 1.0618 \simeq 1.062$$

Next, the expected Ultimate Loss Ratio, say 83%, is applied to the earned premium (EP) to give the initial estimate of the total ultimate loss (UL) for each accident year.

This figure of 83% (or whatever) would normally be derived from a different source from the data in the triangle.

Initial estimate of total ultimate losses, by accident year:

AY	1	2	3	4	5	6
EP	4,486	5,024	5,680	6,590	7,482	8,502
UL (0.83 EP)	3,723	4,170	4,714	5,470	6,210	7,057

Again, the figures for earned premium are 'new data' that would be derived from a separate source.

Note that in this example, the expected loss ratio has been taken as that experienced for the fully developed first accident year. This has been done due to lack of other information.

In other words the figure of 83% we are assuming has been estimated by dividing the ultimate claims incurred for Accident Year 1 (*ie* 3,717) by the earned premium for Accident Year 1 (*ie* 4,486). If we knew that the claims experience was likely to be different for the other accident years, we would use different percentages for the other years.

The next stage is the application of the development factors to the estimated ultimate losses and the addition of the incurred claims that have already been reported.

Note that the accident years are in reverse order in the following table.

Revised estimate of total ultimate losses, by accident year

AY	6	5	4	3	2	1
f	1.290	1.114	1.062	1.022	0.999	1.000
1 – 1/*f*	0.225	0.102	0.058	0.022	–0.001	0
Initial UL	7,057	6,210	5,470	4,714	4,170	3,723
Emerging liability	1,588	633	317	104	–4	0
Reported liability	5,818	6,142	5,676	4,946	4,319	3,717
Ultimate liability	7,406	6,775	5,993	5,050	4,315	3,717

The emerging liability is calculated by multiplying the initial UL by the corresponding value of $1-\frac{1}{f}$. The reported liability for a particular accident year is the last known figure in the run-off triangle for that accident year. The ultimate liability is the sum of the emerging liability and reported liability.

The total ultimate liability relating to these six accident years is, therefore, 33,256.

If the claims paid to date amounted to 20,334 (the same figure that we used before)**, the total reserve required would be 12,922.**

If you are having some difficulty getting to terms with the method, it may be helpful to look at the procedure from a slightly different point of view.

(i) The earned premium for Accident Year 2 is 5,024 (given).

(ii) The initial expected ultimate loss amount for Accident Year 2 (*ie* before combining with the development information) is 83% of this, *ie* 4,170.

(iii) If the ultimate claim amount was 4,170, then we would expect to have paid out so far $\frac{1}{0.999}\times 4{,}170 = 4{,}174$ (*ie* $\frac{1}{f}\times \text{initial UL}$).

(iv) So that would mean we would have $4{,}170-4{,}174=-4$ to pay out in the future (*ie* we expect a rebate of 4 next year).

(v) We have actually incurred so far 4,319 (from the triangle).

(vi) So we would expect to pay out $4{,}319-4=4{,}315$ in total.

This is our Bornhuetter-Ferguson estimate for the total claim payment on claims arising in Accident Year 2.

Question

Apply the same logic to the amount needed for Accident Year 3.

Solution

(i) The earned premium for Accident Year 3 is 5,680.

(ii) The initial expected ultimate loss amount for accident Year 3 is 83% of this, *ie* 4,714.

(iii) If the ultimate claim amount was 4,714, then we would expect to have paid out so far

$$\frac{1}{1.022} \times 4{,}714 = 4{,}613\,.$$

(iv) So that would mean we have $4{,}714 - 4{,}613 = 101$ to pay in the future.

(v) We have actually incurred 4,946 so far.

(vi) So we would expect to pay out $4{,}946 + 101 = 5{,}047$ in total.

This is our Bornhuetter-Ferguson estimate for the total claim payment on claims arising in Accident Year 3.

Our figure differs slightly from the 5,050 shown in the table because of rounding.

The same approach is used to get the figures for Accident Years 4, 5 and 6.

Ultimately we add up the expected outgo for the whole period, and subtract what we have already paid out.

5.4 Assumptions underlying the method

Again the assumptions depend on whether the original or an amended version of the method is being used.

The 'original' method is the simplest form, which was outlined at the start of this section.

For the original method, the underlying assumptions are the same as for the basic chain ladder method, together with the assumption that the estimated loss ratio is appropriate.

Question

Estimate the expected outstanding claims reserve at the end of 2012 for the data in the table below, using the Bornhuetter-Ferguson method. Assume an expected loss ratio of 85%, and that the total claims paid are £1,942,000.

Claims incurred (£000)		Earned premium	Development Year			
			0	1	2	3
Accident Year	2009	860	473	620	690	715
	2010	940	512	660	750	
	2011	980	611	700		
	2012	1,020	647			

Solution

First calculate the initial expected total loss as 85% of the earned premium. This gives figures of 731, 799, 833 and 867.

Now calculate the development factors for individual years in the usual way. We find that the factors are 1.2406, 1.1250, and 1.0362.

Tackling the years one at a time:

The total expected outgo for Accident Year 2009 is 715 as we are assuming that Accident Year 2009 is fully run-off.

For Accident Year 2010, the expected outgo was initially 799. On this basis we would expect to have paid out $\dfrac{799}{1.0362}=771.09$ so far. So we would have to pay out $799-771.09=27.91$ in the future. In fact we have incurred 750, so our final figure would be $750+27.91=777.91$.

For Accident Year 2011, the expected outgo was initially 833. On this basis we would expect to have paid out $\dfrac{833}{1.0362\times 1.125}=714.58$ so far. So we would have to pay out $833-714.58=118.42$ in the future. In fact we have incurred 700 so far, so our final figure should be $700+118.42=818.42$.

For Accident Year 2012, the expected outgo was initially 867. On this basis we would expect to have paid out $\dfrac{867}{1.0362\times 1.125\times 1.2406}=599.50$ so far. So we would have to pay out $867-599.50=267.5$ in the future. In fact we have incurred 647 so far, so our final figure would be $647+267.5=914.5$.

So the total payout expected is 3,225.83, *ie* £3,225,830, of which we have already paid £1,942,000. So the balance is £1,284,000.

5.5 Grossing-up factors versus development factors

Students are sometimes unclear about when to use grossing-up factors and when to use development factors. To some extent it doesn't really matter – both are valid methods of completing a run-off triangle.

The Core Reading uses development factors for the basic chain ladder method, the inflation-adjusted chain ladder method and the Bornhuetter-Ferguson method. It uses grossing-up factors for the ACPC method. If you are not sure, stick to these approaches and you will produce sensible answers. As always, if you are in any doubt, explain carefully in the exam what you are about to do before you do it.

In fact, both procedures are very similar. Using grossing-up factors as they are used in this chapter is equivalent to taking an unweighted average of the past years' experience. Using development factors in the way they are used in this chapter is equivalent to taking a weighted average of past experience, with the years with more claims being weighted more heavily.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 21 Summary

An important feature of the claims process in general insurance is reserving *ie* estimating the components of claims reserves, which include outstanding reported claims, IBNR, reopened claims and claims handling expenses.

Run-off triangles (or delay triangles) provide a method of tabulating claims data and studying the underlying statistical model.

Three methods used for projecting claims are the basic chain ladder method, the average cost per claim method and the Bornhuetter-Ferguson method.

Basic chain-ladder method

- Calculate development factors from the cumulative claims data.
- Use these development factors to project the future cumulative claims.

Basic chain-ladder method – assumptions

- The first accident year is fully run off.
- Claims in each development year are a constant proportion in monetary terms of total claims for each accident year.
- Inflation is not allowed for explicitly, rather it is allowed for implicitly as a weighted average of past inflation.

Inflation-adjusted chain ladder method

- Apply past inflation factors to *incremental* data so that all the claims data in the table is expressed in the monetary terms of the most recent accident year.
- Accumulate the data and calculate development factors.
- Use these development factors to project the future cumulative claims (note that these will still be expressed in the monetary terms of the most recent accident year).
- Disaccumulate the data to make it incremental.
- Apply future inflation assumptions to convert the outstanding claim payments into the amounts relating to each future year.

Inflation-adjusted chain ladder method – assumptions

- The first accident year is fully run off.
- Claims in each development year are a constant proportion in real terms of total claims for each accident year.
- Inflation is allowed for explicitly and we assume that both the past and future inflation assumptions are correct

Average cost per claim – method

- Divide the entry in each cell in the cumulative claims table by the entry in the corresponding cell of the claim number table. This gives the average cost per claim.
- Calculate grossing-up factors for the average claim amounts. Use these to estimate the final average for each accident year.
- Repeat the last step for the claim number table.
- For each accident year, multiply together the figures from the ACPC and claim number tables.
- Sum over all accident years to obtain the total projected loss estimate.

Average cost per claim – assumptions

- The first accident year is fully run off.
- The average cost per claim in each development year is a constant proportion in monetary terms of the ultimate average cost per claim for each accident year.
- The number of claims in each development year is a constant proportion in of the ultimate number of claims for each accident year.
- Inflation is not allowed for explicitly, rather it is allowed for implicitly as a weighted average of past inflation.

Bornhuetter-Ferguson – method

- Decide on the amount of the loss ratio.
- Calculate development factors (as in BCL method).
- Calculate the *cumulative* development factors *f*.

For each accident year that is not fully run-off:

- Multiply the earned premium by the loss ratio to obtain the initial estimate of the ultimate loss
- Use the initial estimate and the cumulative development factors to determine the expected amount paid out so far
- Use this to see how much is expected to be paid in the future (emerging liability)
- The revised estimate of the ultimate loss is the reported liability (last known figure) plus the emerging liability.

Finally:

- Sum the revised estimates of the ultimate losses for each accident year to obtain an estimate of the total liability.

Bornhuetter-Ferguson – assumptions

- The first accident year is fully run off.
- The loss ratio is correct.
- Claims in each development year are a constant proportion in monetary terms of total claims for each accident year.
- Inflation is not allowed for explicitly, rather it is allowed for implicitly as a weighted average of past inflation.

The practice questions start on the next page so that you can keep the chapter summaries together for revision purposes.

Chapter 21 Practice Questions

21.1 The table below shows cumulative claims (not adjusted for inflation) from a portfolio of insurance policies for 4 accident years.

Year of origin	Development Year 0	1	2	3
2014	2,047	3,141	3,209	3,310
2015	2,471	3,712	3,810	
2016	2,388	3,750		
2017	2,580			

It may be assumed that payments are made in the middle of a calendar year.

It is estimated that the inflation rate applicable to these data has been 5% per annum over the relevant period.

Use the inflation adjusted chain ladder method to estimate the total outstanding payments, up to the end of Development Year 3, for Accident Year 2017 in mid-2017 prices.

21.2 Write down an equation defining the statistical model assumed by each of the following methods of projecting the payments for outstanding claims:

- basic chain ladder method
- inflation adjusted chain ladder method

Define each symbol and indicate whether the value of each quantity is assumed at the outset or is estimated by the model.

21.3 What does the factor x_{i+j} represent in the inflation adjusted chain ladder model?

A The proportion (by number) of claims paid in Development Year $i+j$.

B The proportion (by amount) of claims paid in Accident Year $i+j$.

C The volume of claims for Accident Year $i+j$.

D An index of the cost of claims paid in calendar year $i+j$.

21.4 Explain in words (*ie* without using mathematical symbols) the assumptions underlying the:

(i) basic chain ladder method.

(ii) inflation adjusted chain ladder method.

(iii) average cost per claim method

(iv) Bornhuetter-Ferguson method.

21.5 The table below shows the claim payments made by a general insurer in each year for a particular type of insurance.

Claim payments made during year (£000s)		*Development Year*			
		0	*1*	*2*	*3*
Accident Year	*2012*	10	50	50	30
	2013	50	70	30	
	2014	40	30		
	2015	90			

(i) What was the total amount paid during the 2015 calendar year?

(ii) Calculate the development factors from development years 1 to 2 for each of the 2012 and 2013 accident years.

(iii) Give three reasons why it may not be appropriate to use the basic chain ladder method to project the claim payments for this portfolio, using figures from the table to support your comments.

21.6 The following table shows incremental claims relating to the accident years 2007, 2008 and 2009. It is assumed that claims are fully run-off by the end of Development Year 2. Estimate total outstanding claims using the chain-ladder technique, ignoring inflation.

Exam style

Incremental claims		*Development Year*		
		0	*1*	*2*
Accident Year	*2007*	2,587	1,091	251
	2008	2,053	1,298	
	2009	3,190		

[7]

21.7 Exam style

The cumulative claims paid each year under a certain cohort of insurance policies are recorded in the table below, for accident years 2010, 2011, 2012 and 2013.

	Development Year			
Accident Year	*0*	*1*	*2*	*3*
2010	2,457	4,196	4,969	5,010
2011	2,648	4,715	5,561	
2012	3,084	5,315		
2013	3,341			

(i) Calculate the development factors under the basic chain ladder technique and state the assumptions underlying the use of this method. [4]

(ii) The rate of claims inflation over these years, measured over the 12 months to the middle of each year, is given in the table below.

2011	2.1%
2012	10.5%
2013	3.2%

Calculate the development factors under the inflation-adjusted chain ladder technique and state the assumptions underlying the use of this method. [6]

(iii) Based on the development factors calculated in parts (i) and (ii), calculate the fitted values under these two models and comment on how these compare with the actual values. [7]

[Total 17]

21.8 Exam style

The table below shows the payments, in £'000s, made in successive development years in respect of a particular class of general insurance business. It may be assumed that all claims are fully settled by the end of Development Year 3 and that all payments are made in the middle of a calendar year.

Year of origin	*Development Year*			
	0	*1*	*2*	*3*
2012	342	82	68	37
2013	359	90	73	
2014	481	120		
2015	591			

Use the inflation adjusted chain ladder method to estimate the amount of the reserve required at the end of 2015 to pay the outstanding amounts in respect of claims from 2013, 2014 and 2015. You should assume that past inflation has been at the rate of 5% per annum and that future inflation will also be at 5% per annum. [10]

21.9 Exam style

The tables below show the cumulative cost of incurred claims and the number of claims reported each year for a certain cohort of insurance policies. The claims are assumed to be fully run-off at the end of Development Year 2.

Cumulative cost of incurred claims:

	Development Year		
Accident Year	*0*	*1*	*2*
0	288	634	893
1	465	980	
2	773		

The numbers of claims reported in each year are:

	Development Year		
Accident Year	*0*	*1*	*2*
0	110	85	55
1	167	113	
2	285		

Given that the total amount paid in claims to date, relating to accident years 0, 1 and 2, is £2,750, calculate the outstanding claims reserve using the average cost per claim method.

[11]

21.10 Exam style

The following table shows cumulative incurred claims data, by year of accident and reporting development, for a portfolio of domestic household insurance policies:

Cumulative incurred claims (£000)		*Development Year*			
		0	*1*	*2*	*3*
Accident Year	*2011*	829	917	978	1,020
	2012	926	987	1,053	
	2013	997	1,098		
	2014	1,021			

The corresponding cumulative number of reported claims, by year of accident and reporting development, are as follows:

Cumulative number of reported claims		*Development Year*			
		0	*1*	*2*	*3*
Accident Year	*2011*	63	68	70	74
	2012	65	69	72	
	2013	71	76		
	2014	70			

Use the average cost per claim method with simple average grossing up factors to calculate an estimate of the outstanding claim amount for these policies for claims arising during these accident years. The claims paid to date are £3,640,000. State any assumptions used. [11]

21.11 An insurance company has paid the following claim amounts (in £000s):

Exam style

		Development Year				
		1	2	3	4	5
Accident Year	1	2,800	1,400	987	322	57
	2	3,260	2,004	1,017	421	
	3	3,854	1,978	857		
	4	3,722	2,114			
	5	4,627				

The earned premium in each year is 6,727 for Accident Year 1, 8,289 for Accident Year 2, 9,627 for Accident Year 3, 9,928 for Accident Year 4 and 10,004 for Accident Year 5.

Apply the Bornhuetter-Ferguson method to estimate the amount of claims yet to be paid, stating any assumptions that you make. [8]

21.12 The table below shows the cumulative costs of incurred claims. The claims are assumed to be fully run-off by the end of Development Year 2.

Exam style

£000s	*Development Year*		
Accident Year	*0*	*1*	*2*
2011	2,670	3,290	4,310
2012	2,850	3,420	
2013	3,030		

Annual premiums written were:

Year	Premiums (£000s)
2011	5,390
2012	5,600
2013	6,030

The ultimate loss ratio has been estimated at 80% and the total amount of claims paid to date is £5,720,000. Calculate the outstanding claims reserve using the Bornhuetter-Ferguson method. [6]

Chapter 21 Solutions

21.1 First we disaccumulate the amounts. This gives the following table:

INCREMENTAL CLAIM PAYMENTS – NOMINAL		Development Year			
		0	1	2	3
Origin Year	2014	2,047	1,094	68	101
	2015	2,471	1,241	98	
	2016	2,388	1,362		
	2017	2,580			

Now we convert to 2017 prices, using a past inflation rate of 5%:

INCREMENTAL CLAIM PAYMENTS – REAL		Development Year			
		0	1	2	3
Origin Year	2014	2,369.6584	1,206.135	71.4	101
	2015	2,724.2775	1,303.05	98	
	2016	2,507.40	1,362		
	2017	2,580			

Now we find the *cumulative* claim payments in real terms:

CUMULATIVE CLAIM PAYMENTS – REAL		Development Year			
		0	1	2	3
Origin Year	2014	2,369.6584	3,575.7934	3,647.1934	3,748.1934
	2015	2,724.2775	4,027.3275	4,125.3275	
	2016	2,507.40	3,869.4		
	2017	2,580			

Calculating the development factors in the usual way:

$$\lambda_1 = \frac{3575.7934 + 4027.3275 + 3869.4}{2369.6584 + 2724.2775 + 2507.4} = 1.509277$$

$$\lambda_2 = \frac{3647.1934 + 4125.3275}{3575.7934 + 4027.3275} = 1.022280$$

$$\lambda_3 = \frac{3748.1934}{3647.1934} = 1.027693$$

We now complete the triangle (still working in 2017 values):

CUMULATIVE CLAIM PAYMENTS – REAL		Development Year			
		0	1	2	3
Origin Year	2014	2,369.6584	3,575.7934	3,647.1934	3,748.1934
	2015	2,724.2775	4,027.3275	4,125.3275	**4,239.57**
	2016	2,507.40	3869.4	**3,955.61**	**4,065.15**
	2017	2,580	**3,893.93**	**3,980.70**	**4,090.93**

So the total amount of outstanding payments for 2017 at mid-2017 prices is:

$$4{,}090.93 - 2{,}580 = 1{,}510.93 \text{ or } 1{,}511 \text{ (4 SF)}$$

Note that this question asked us to find the outstanding amount in 2017 prices. If we had been asked to find the outstanding amount in nominal (actual) prices, we would have had to apply three further stages – disaccumulating, adjusting for future inflation and finally totalling.

21.2

Method	Model	Data items	Assumed values	Estimated values	Not required
Basic chain ladder	$C_{ij} = s_i r_j + e_{ij}$	C_{ij}		s_i , r_j	e_{ij}
Inflation adjusted chain ladder	$C_{ij} = s_i r_j x_{i+j} + e_{ij}$	C_{ij}	x_{i+j}	s_i , r_j	e_{ij}

The symbols are defined as follows:

s_i is a measure of the "volume" of claims relating to origin year i.

r_j is a factor for development year j, representing the proportion of claims paid in that development year.

x_{i+j} is an adjustment factor relating to calendar years

e_{ij} is a random statistical error with mean zero that represents the difference between actual and expected results.

21.3 x_{i+j} is a factor that is applied to all the entries along a particular diagonal (*ie* calendar year). In other words, x_{i+j} is an index of the cost of claims paid in calendar year $i+j$. So the answer is D.

21.4 (i) ***Basic chain ladder assumptions***

- Payments from each origin year will develop in the same way in monetary terms.
- Weighted average past inflation will be repeated in the future.

- The first year is fully run-off.

(ii) ***Inflation-adjusted chain ladder assumptions***

- Payments from each origin year will develop in the same way in real terms.
- Rates of past and future claims inflation are appropriate.
- The first year is fully run-off.

(iii) ***ACPC assumptions***

- For each origin year, the number of claims in each development year is a constant proportion of the total number of claims from that origin year.
- For each origin year, the average incurred cost per claim in monetary terms in each development year, is a constant proportion of the total claims incurred in monetary terns for that origin year.
- The first year is fully run-off.

(iv) ***BF assumptions***

- Payments from each origin year will develop in the same way.
- Weighted average past inflation will be repeated in the future.
- The first year is fully run-off.
- The estimated loss ratio is appropriate.

21.5 (i) The payments made during 2015 correspond to the figures in the longest diagonal, which total:

$$90+30+30+30=180$$

So the total amount paid during 2015 was £180,000.

(ii) The development factor for Accident Year 2012, Development Year 2 is:

$$\frac{10+50+50}{10+50}=\frac{110}{60}=1.8\dot{3}$$

The development factor for Accident Year 2013, Development Year 2 is:

$$\frac{50+70+30}{50+70}=\frac{150}{120}=1.25$$

(iii) Reasons why the basic chain ladder method may not be appropriate here include:

1. The first accident year may not be fully run-off.

2. The development ratios do not appear to be constant for each origin year. This is illustrated by the ratios calculated in part (ii).

3. The total payment amounts are relatively small (*eg* a total of £180,000 paid in 2015). There may not be enough policies for the statistical model to apply, since the figures may be dominated by the random errors e_{ij}.

4. The basic chain ladder does not explicitly model inflation but assumes that a weighted average of past inflation will apply in the future. If inflation for calendar years 2012 to 2015 has varied significantly, a chain ladder calculation will not give reliable results. The figures in the table are too erratic (*eg* the values in Development Year 0 vary from £10,000 to £90,000) to judge whether this is the case.

21.6 First of all the claim data must be accumulated to form the table below:

Cumulative claims		*Development year*	
Accident Year	*0*	*1*	*2*
2007	2587	3678	3929
2008	2053	3351	
2009	3190		

[1]

Then the development factors should be calculated:

Development factor for DY1 = $\dfrac{3678+3351}{2587+2053}=1.514871$ [1]

Development factor for DY2 = $\dfrac{3929}{3678}=1.068244$ [1]

The lower half of the run-off triangle can now be completed:

Cumulative claims		*Development year*	
Accident Year	*0*	*1*	*2*
2007	2587	3678	3929
2008	2053	3351	**3579.68**
2009	3190	**4832.44**	**5162.22**

[3]

So the estimates amount of outstanding claims is:

$$(3579.7-3351)+(5162.2-3190)=2201 \ (4 \ SF)$$ [1]

21.7 (i) ***BCL development factors and assumptions***

The development factors are:

$$d_1=\frac{4,196+4,715+5,315}{2,457+2,648+3,084}=1.737208$$

$$d_2=\frac{4,969+5,561}{4,196+4,715}=1.181686$$

$$d_3=\frac{5,010}{4,969}=1.008251$$ [2]

The assumptions underlying the method are:

- Claim amounts are fully run off by the end of Development Year 3 [½]
- Payments from each origin year will develop in the same way in monetary terms. [½]
- Weighted average past inflation will be repeated in the future. [1]

(ii) ***Inflation-adjusted chain ladder development factors and assumptions***

First we need to disaccumulate the figures to obtain incremental figures. If we do this we obtain the following results:

	Development year			
Accident year	*0*	*1*	*2*	*3*
2010	2,457	1,739	773	41
2011	2,648	2,067	846	
2012	3,084	2,231		
2013	3,341			

We now inflate the incremental figures to get real values (in mid-2001 terms):

	Development year			
Accident year	*0*	*1*	*2*	*3*
2010	2860.70	1983.09	797.74	41
2011	3019.67	2133.14	846	
2012	3182.69	2231		
2013	3341			

[2]

We can now find the cumulative figures in real terms:

	Development year			
Accident year	*0*	*1*	*2*	*3*
2010	2860.70	4843.79	5641.53	5682.53
2011	3019.67	5152.81	5998.81	
2012	3182.69	5413.69		
2013	3341			

So the development factors are now:

$$d_1 = \frac{4,843.79 + 5,152.81 + 5,413.69}{2,860.70 + 3,019.67 + 3,182.69} = 1.700340$$

$$d_2 = \frac{5,641.53 + 5,998.81}{4,843.79 + 5,152.81} = 1.164429$$

$$d_3 = \frac{5,682.53}{5,641.53} = 1.007268$$ [2]

The assumptions underlying the inflation adjusted chain ladder method are:

- the claim amounts are fully run off by the end of Development Year 3 [½]
- Payments from each origin year will develop in the same way in real terms. [½]
- Rates of past and future claims inflation are appropriate. [1]

(iii) ***Fitted values***

We now use the development factors to find the fitted claim amounts for past years. Using the development factors for the basic chain ladder method we get:

Accident Year	*Development Year*			
	0	*1*	*2*	*3*
2010	2,457	4,268.32	5,043.81	5,085.43
2011	2,648	4,600.13	5,435.90	
2012	3,084	5,357.52		
2013	3,341			

[1]

So the fitted incremental payments are:

Accident Year	*Development Year*			
	0	1	2	3
2010	2,457	1,811	776	42
2011	2,648	1,952	836	
2012	3,084	2,274		
2013	3,341			

[1]

If we compare the fitted and the actual incremental payments, we get the following:

Accident Year	*Development Year*			
	0	*1*	*2*	*3*
Fitted 2010	2,457	1,811.32	775.49	41.62
Actual 2010	2,457	1,739	773	41
A – F	--	**–72.32**	**–2.49**	**–0.62**
Fitted 2011	2,648	1,952.13	835.78	
Actual 2011	2,648	2,067	846	
A – F	--	**114.87**	**10.22**	

Fitted 2012	3,084	2,273.55
Actual 2012	3,084	2231
A – F	--	**–42.55**
Fitted 2013	3,341	
Actual 2013	3,341	
A – F	--	

[2]

We can now repeat the process with the inflation-adjusted chain ladder method. Using the inflation adjusted development factors on the real data amounts for Development Year 0, we get:

Accident Year	*Development Year* 0	*1*	*2*	*3*
2010	2,860.70	4,864.17	5,663.98	5,705.15
2011	3,019.67	5,134.47	5,978.73	
2012	3,182.69	5,411.65		
2013	3,341			

[1]

If we now disaccumulate and compare with the actual incremental (real) values:

Accident Year	*Development Year* 0	*1*	*2*	*3*
Fitted 2010	2,860.70	2,003.47	799.81	41.16
Actual 2010	2,860.70	1,983.09	797.74	41
A – F	--	–20.38	–2.08	–0.16
Fitted 2011	3,019.67	2,114.80	844.26	
Actual 2011	3,019.67	2,133.14	846	
A – F	--	18.34	1.74	

Fitted 2012	3,182.69	2,228.96
Actual 2012	3,182.69	2231
A – F	--	2.04

Fitted 2013	3,341
Actual 2013	3,341
A – F	--

The inflation adjusted method appears to fit the original data better, given that the residuals are smaller in this case. However, there are not many categories of data, and it is not clear whether either method will provide satisfactory results in the future. [2]

21.8 Adjust for past inflation (*ie* obtain figures in mid 2015 prices):

Incremental claims paid mid 95 prices (£*k*)		Development year			
		0	1	2	3
Accident year	2012	395.908	90.405	71.4	37
	2013	395.798	94.5	73	
	2014	505.05	120		
	2015	591			

[2]

Accumulate the data and project figures forward using the basic chain ladder:

Cumulative claims paid mid 2015 prices (£*k*)		Development year			
		0	1	2	3
Accident year	2012	395.908	486.313	557.713	594.713
	2013	395.798	490.298	563.298	**600.668**
	2014	505.05	625.05	**717.469**	**765.067**
	2015	591	**729.961**	**837.892**	**893.48**

[2]

The development factors are 1.235129, 1.147858 and 1.066342. [2]

Disaccumulate the data:

Incremental claims paid mid 2015 prices(£*k*)		Development year			
		0	1	2	3
Accident year	2012				
	2013				**37.3705**
	2014			**92.4189**	**47.5986**
	2015		**138.961**	**107.931**	**55.5878**

[1]

Adjust for future inflation (*ie* obtain future amounts actually paid):

Incremental claims paid – adjusted (£*k*)		Development year			
		0	1	2	3
Accident year	2012				
	2013				39.239
	2014			97.040	52.477
	2015		145.91	118.99	64.345

[2]

Totalling these up the reserve that needs to be held is 518. Since the figures are in thousands, the reserve is £518,000. [1]

21.9 First we need the cumulative numbers of claims reported:

Cumulative claims reported	*Development Year*		
Accident Year	*0*	*1*	*2*
0	110	195	250
1	167	280	
2	285		

[1]

We now divide the cumulative cost by the cumulative claim numbers to get the average cost per claim:

Cumulative ACPC	*Development Year*		
Accident Year	*0*	*1*	*2*
0	2.61818	3.25128	3.572
1	2.78443	3.5	
2	2.71228		

[1]

Assuming that Accident Year 0 is fully run off, we divide the numbers in the top row by 3.572 to get the percentages 73.297%, 91.021%, 100%. [1]

So the ultimate average cost per claim figure for Accident Year 1 is:

$$\frac{3.5}{0.91021} = 3.84525$$ [1]

We can now calculate the percentage figures for Accident Year 1 as 72.412%, 91.021%.

For accident Year 2, we take the average of the two previous figures for Development Year 0:

$$\tfrac{1}{2}(72.412+73.297)=72.855\%$$ [1]

So the ultimate figure for Accident Year 2 is:

$$\frac{2.71228}{0.72854}=3.72286$$ [1]

Using the same approach for the claim number figures, we get:

Cumulative claim numbers	*Development Year*		
Accident Year	*0*	*1*	*2*
0	110 44%	195 78%	250 100%
1	167 46.521%	280 78%	
2	285 45.261%		

These give ultimate claim number values of 250, 358.974 and 629.685. [3]

So the total expected ultimate loss will be:

$$250\times 3.572+358.974\times 3.84525+629.685\times 3.72286=4{,}617.58$$ [1]

Since the claims paid to date are 2,750, the outstanding claims reserve is 1,868 (4 SF). [1]

21.10 Dividing each cell in the first table by the corresponding cell in the second table gives the cumulative average incurred cost per claim, by year of accident and reporting development:

Cumulative average incurred cost per claim	*Development year*				
		0	1	2	3
Accident year	2011	13.159	13.485	13.971	13.784
	2012	14.246	14.304	14.625	
	2013	14.042	14.447		
	2014	14.586			

[2]

Completing the ultimate cumulative number of claims reported table and the ultimate cumulative average incurred cost per claim table using the chain ladder technique with simple grossing up factors gives:

Cumulative number of reported claims	*Development year*				
Accident year		0	1	2	3
	2011	63	68	70	74
	2012	65	69	72	76.114
	2013	71	76		83.267
	2014	70			82.095

[1½]

The simple average grossing up factors used are:

Development year 2 to 3: 94.595%

Development year 1 to 3: 91.273% (average of 91.892% and 90.653%)

Development year 0 to 3: 85.267% (average of 85.135%, 85.398% and 85.268%)

[1½]

Cumulative average incurred cost per claim	*Development year*				
Accident year		0	1	2	3
	2011	13.159	13.485	13.971	13.784
	2012	14.246	14.304	14.625	14.429
	2013	14.042	14.447		14.669
	2014	14.586			15.093

[1½]

The simple average grossing up factors used are:

Development year 2 to 3: 101.361%

Development year 1 to 3: 98.487% (average of 97.834% and 99.139%)

Development year 0 to 3: 96.642% (average of 95.465%, 98.736% and 95.725%)

[1½]

Therefore, the ultimate claims incurred amount from accident years 2011 to 2014 is:

$$1{,}020 + (76.114 \times 14.429) + (83.267 \times 14.669) + (82.095 \times 15.093) = 4{,}579$$ [1]

The claims paid to date (from accident years 2011 to 2014) amount to 3,640, resulting in a total outstanding claim amount of 939, *ie* £939,000. [½]

Assumptions:

- Claims incurred in the first accident year are fully run off. [½]
- For each accident year, the number of reported claims in each development year, is a constant proportion of the total for that accident year. [½]
- For each accident year, the average incurred cost per claim in monetary terms in each development year, is a constant proportion of the total claims incurred in monetary terns for that accident year. [½]

21.11 We need the cumulative claims data:

		Development year				
		1	2	3	4	5
Accident year	1	2,800	4,200	5,187	5,509	5,566
	2	3,260	5,264	6,281	6,702	
	3	3,854	5,832	6,689		
	4	3,722	5,836			
	5	4,627				

The ultimate loss ratio is $\dfrac{5,566}{6,727} = 0.827412$. [1]

Next we calculate the expected end of year figures (the initial ultimate liability):

$2: \quad 0.827412 \times 8,289 = 6,858.4$

$3: \quad 0.827412 \times 9,627 = 7,965.5$

$4: \quad 0.827412 \times 9,928 = 8,214.5$ [1]

$5: \quad 0.827412 \times 10,004 = 8,277.4$

The development factors are:

Year 4 to Year 5 $\dfrac{5,566}{5,509} = 1.010347$

Year 3 to Year 4 $\dfrac{5,509 + 6,702}{5,187 + 6,281} = 1.064789$

Year 2 to Year 3 $\dfrac{5,187 + 6,281 + 6,689}{4,200 + 5,264 + 5,832} = 1.187042$

Year 1 to Year 2 $\dfrac{4,200 + 5,264 + 5,832 + 5,836}{2,800 + 3,260 + 3,854 + 3,722} = 1.549721$ [1]

The emerging liabilities for each year are:

$$2: \quad 6,858.4\times\left(1-\frac{1}{1.010347}\right)=70.2$$

$$3: \quad 7,965.5\times\left(1-\frac{1}{1.010347\times1.064789}\right)=561.3$$

[3]

$$4: \quad 8,214.5\times\left(1-\frac{1}{1.010347\times1.064789\times1.187042}\right)=1,782$$

$$5: \quad 8,277.4\times\left(1-\frac{1}{1.010347\times1.064789\times1.187042\times1.549721}\right)=4,094.9$$

Given that these are claims paid (rather than incurred), we don't need to calculate the revised ultimate liability to get the reserve we can just total up the emerging liabilities:

$70.2+561.3+1,782+4,094.9=6,508$ (4 SF), *ie* £6,508,000 [1]

The assumptions are:

- Payments from each origin year will develop in the same way.
- Weighted average past inflation will be repeated in the future.
- The first year is fully run-off.
- The estimated loss ratio is appropriate. [1]

21.12 First we calculate the expected end of year figures (the initial ultimate liability):

$$2011: \quad 0.8\times5,390=4,312$$

$$2012: \quad 0.8\times5,600=4,480$$

$$2013: \quad 0.8\times6,030=4,824$$

[1]

Next we calculate the development factors:

$$\frac{3,290+3,420}{2,670+2,850}=\frac{6,710}{5,520}=1.215580$$

$$\frac{4,310}{3,290}=1.310030$$

[1]

The emerging liabilities for each year are:

$$2011: \quad 4,312\times\left(1-\tfrac{1}{1}\right)=0$$

$$2012: \quad 4,480\times\left(1-\frac{1}{1.310030}\right)=1,060.2$$

[2]

$$2013: \quad 4,824\times\left(1-\frac{1}{1.215580\times1.310030}\right)=1,794.7$$

So the ultimate liabilities for each year are:

$$2011: \quad 4{,}310 + 0 = 4{,}310$$

$$2012: \quad 3{,}420 + 1{,}060.2 = 4{,}480.2$$ [1]

$$2013: \quad 3{,}030 + 1{,}794.7 = 4{,}824.7$$

Thus the total ultimate liability is:

$$4{,}310 + 4{,}480.2 + 4824.7 = 13{,}614.9$$ [½]

Therefore the outstanding claims reserve is $13{,}614.9 - 5{,}720 = 7{,}895$ (4 SF)

ie £7,895,000 (4 SF) [½]

End of Part 4

What next?

1. Briefly **review** the key areas of Part 4 and/or re-read the **summaries** at the end of Chapters 18 to 21.
2. Ensure you have attempted some of the **Practice Questions** at the end of each chapter in Part 4. If you don't have time to do them all, you could save the remainder for use as part of your revision.
3. Attempt **Assignment X4**.

And finally ...

Good luck!

Subject CM2: Assignment X1

2019 Examinations

Time allowed: 2¾ hours

Instructions to the candidate

1. *Please:*

   – *attempt all of the questions, as far as possible under exam conditions*

   – ***begin your answer to each question on a new page***

   – ***leave at least 2cm margin on all borders***

   – *write in black ink using a medium-sized nib because we will be unable to mark illegible scripts*

   – *note that assignment marking is not included in the price of the course materials. Please purchase Series Marking or a Marking Voucher before submitting your script.*

   – *note that we only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2019 exams.*

2. *Please **do not**:*

   – *use headed paper*

   – *use highlighting in your script.*

At the end of the assignment

If your script is being marked by ActEd, please follow the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables and an electronic calculator.

Submission for marking

You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed on the summary page at the back of this pack and on our website at **www.ActEd.co.uk**.

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. *It is your responsibility to ensure that scripts reach ActEd in good time*. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

When submitting your script, please:

- complete the cover sheet, including the checklist
- scan your script, cover sheet (and Marking Voucher if applicable) and save as a pdf document, then email it to: ActEdMarking@bpp.com
- **do not submit a photograph of your script**
- **do not include the question paper in the scan.**

In addition, please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* 'CM2 Assignment X1 No. 12345', inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF **is less than 4 MB** in size.
- Do not protect the PDF in any way (otherwise the marker cannot return the script to ActEd, which causes delays).
- Please include the 'feedback from marker' sheet when scanning.
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

Subject CM2: Assignment X1

2019 Examinations

Please complete the following information:

Name:

ActEd Student Number (see Note below):

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting it will help us to process your scripts quickly. If you do not know your ActEd Student Number, please email us at ActEd@bpp.com.

Your ActEd Student Number is not the same as your IFoA Actuarial Reference Number or ARN.

Number of following pages: _______

Please put a tick in this box if you have solutions and a cross if you do not: [ ]

Please tick here if you are allowed extra time or other special conditions in the profession's exams (if you wish to share this information): [ ]

Time to do assignment
(see Note below): _____ hrs _____ mins

Under exam conditions
(delete as applicable): yes / nearly / no

Note: If you take more than 2¾ hours, you should indicate how much you completed within this exam time so that the marker can provide useful feedback on your progress.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total	
__ / 6	__ / 4	__ / 9	__ / 12	__ / 10	__ / 6	__ / 7	__ / 10	__ / 9	__ / 7	__ / 80	=_____%

Grade: A B C D E **Marker's initials:** _________

Please tick the following checklist so that your script can be marked quickly. Have you:

[ ] Checked that you are using the latest version of the assignments, *ie* 2019 for the sessions leading to the 2019 exams?

[ ] Written your full name in the box above?

[ ] Completed your ActEd Student Number in the box above?

[ ] Recorded your attempt conditions?

[ ] Numbered all pages of your script (excluding this cover sheet)?

[ ] Written the total number of pages (excluding the cover sheet) in the space above?

[ ] Included your Marking Voucher or ordered Series X Marking?

Please follow the instructions on the previous page when submitting your script for marking.

Feedback from marker

Notes on marker's section

The main objective of marking is to provide specific advice on how to improve your chances of success in the exam. The most useful aspect of the marking is the comments the marker makes throughout the script, however you will also be given a percentage score and the band into which that score falls. Each assignment tests only part of the course and hence does not give a complete indication of your likely overall success in the exam. However it provides a good indicator of your understanding of the material tested and the progress you are making with your studies:

A = Excellent progress B = Good progress C = Average progress
D = Below average progress E = Well below average progress

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

X1.1 (i) *Technical analysis* is the study of chart patterns of various asset prices. Explain whether this can be used to an investor's advantage, if the Efficient Markets Hypothesis (EMH) holds. [2]

(ii) *Insider trading* is illegal in the UK stock market. Explain what this suggests about the EMH. [2]

(iii) *Fundamental analysis* includes the analysis of balance sheets, consideration of company strategy, the environment in which the company operates *etc*. Explain how this relates to the EMH. [2]

[Total 6]

X1.2 (i) An investor has the utility function $U(w) = -\exp\left(-\frac{w}{100}\right)$.

Determine whether the investor exhibits increasing, constant or decreasing absolute and relative risk aversion. [2]

(ii) The investor has an initial wealth of $1{,}000$ and is offered a gamble with a payoff described by a random variable:

$$X = \begin{cases} +100 & \text{with probability } 0.5 \\ -50 & \text{with probability } 0.5 \end{cases}$$

Find the investor's certainty equivalent of this gamble. [2]

[Total 4]

X1.3 Two assets are available to investors. Asset B is a risk-free investment that returns 1%, and the return on Asset A is given by:

$$R_A = \begin{cases} -1\% & \text{probability } 0.5 \\ 3\% & \text{probability } 0.5 \end{cases}$$

(i) Explain why Asset B must be second-order stochastically dominant over Asset A in terms of investors and utility functions. [2]

(ii) Verify numerically the second-order stochastic dominance expressed in part (i). [2½]

(iii) What can be said about dominance if Asset A offers instead a return of:

(a) $$R_A = \begin{cases} -1\% & \text{probability } 0.5 \\ 4\% & \text{probability } 0.5 \end{cases}$$ [2½]

(b) $$R_A = \begin{cases} -1\% & \text{probability } 0.5 \\ 1\% & \text{probability } 0.5 \end{cases}$$ [1]

(iv) Can an asset that allows the possibility of a return less than 1% ever dominate Asset B? [1]

[Total 9]

X1.4 An investor is trying to choose between the investments whose distributions of returns are described below:

Investment A: 0.4 probability that it will return 10%

0.2 probability that it will return 15%

0.4 probability that it will return 20%

Investment B: 0.25 probability that it will return 10%

0.70 probability that it will return 15%

0.05 probability that it will return 40%

Investment C: A uniform distribution on the range 10% to 20%

Calculate the following for each investment:

(i) expected return [1½]

(ii) variance of return [3½]

(iii) semi-variance [3½]

(iv) expected shortfall below 12% [2½]

(v) shortfall probability below 15%. [1]

[Total 12]

X1.5 The annual rates of interest from a particular investment, in which part of an insurance company's funds is invested, are independently and identically distributed. Each year, the distribution of $(1+i_t)$, where i_t is the rate of interest earned in year *t*, is log-normal with parameters μ and σ^2.

i_t has mean value 0.07 and standard deviation 0.02, the parameter $\mu = 0.06748$ and $\sigma^2 = 0.0003493$.

(i) The insurance company has liabilities of £1m to meet in one year from now. It currently has assets of £950,000. Assets can either be invested in the risky investment described above or in an investment which has a guaranteed return of 5% per annum effective. Find, to two decimal places, the probability that the insurance company will be unable to meet its liabilities if:

(a) All assets are invested in the investment with the guaranteed return.

(b) 85% of assets are invested in the investment which does not have the guaranteed return and 15% of assets are invested in the asset with the guaranteed return. [7]

(ii) Determine the variance of return from the portfolios in (i)(a) and (i)(b) above. [3]

[Total 10]

X1.6 £1,000 is invested for 10 years. In any year the yield on the investment will be 4% with probability 0.4, 6% with probability 0.2 and 8% with probability 0.4 and is independent of the yield in any other year.

(i) Calculate the mean accumulation at the end of 10 years. [2]

(ii) Calculate the standard deviation of the accumulation at the end of 10 years. [4]
[Total 6]

X1.7 (i) Explain the following terms in the context of mean-variance portfolio theory:

(a) opportunity set

(b) efficient frontier for a portfolio of risky assets

(c) indifference curves

(d) optimal portfolio [5]

(ii) Describe, using a sketch, the effect on the efficient frontier of introducing a risk-free asset that can be bought or sold in unlimited quantities. [2]
[Total 7]

X1.8 Assets A and B have the following distributions of returns in various states:

State	Asset A	Asset B	Probability
1	10%	–12%	0.1
2	8%	0%	0.2
3	6%	3%	0.3
4	4%	16%	0.4

(i) Calculate the correlation coefficient between the returns on asset A and asset B and comment on your answer. [4]

(ii) An investor is going to set up a portfolio consisting entirely of assets A and B. Calculate the proportion of assets that should be invested in asset A to obtain the portfolio with the smallest possible variance. [2]

(iii) Assume that the means and the variances of the returns on assets A and B remain unchanged, but that the correlation ρ_{AB} between assets A and B does change. The investor decides to hold 80% of their wealth in asset A and 20% in asset B. Calculate the range of values of ρ_{AB} such that the portfolio has a smaller variance than if they were to invest everything in asset A. [4]
[Total 10]

X1.9 (i) State the assumptions of mean-variance portfolio theory (MVPT). [3]

(ii) You are given the choice of only two assets, A and B. The expected returns and variances of return of the two assets are:

$E_A = 13\%$ $E_B = 5\%$

$V_A = 36\%\%$ $V_B = 4\%\%$

Find the equation of the efficient frontier in (E, σ) space in the special case where the returns on Assets A and B are perfectly correlated. Comment on your result. [3]

(iii) State, giving the relevant equations, how your approach in (ii) would be modified if there were more than two assets? [*Numerical calculations are not required*.] [3]

[Total 9]

X1.10 (i) State the principal theme of behavioural finance. State the assumption of expected utility theory that is challenged by this theme. [2]

(ii) Outline what is meant by prospect theory. [2]

(iii) Briefly describe the behavioural finance theme categorising the behaviour of the people in each of the following cases.

(a) Short-term interest rates have remained at historically low levels for several years now and there is no compelling reason for them to change. However, Saver A continues to expect them to increase anytime soon.

(b) Investor B has been investing their cash with Bank X for the last five years and has received an annual interest rate of 5%. Bank Y has the same type of account with exactly the same interest rate, and is offering a £250 incentive for customers to switch accounts. Investor B is reluctant to switch banks. [3]

[Total 7]

Subject CM2: Assignment X2

2019 Examinations

Time allowed: 2¾ hours

Instructions to the candidate

1. *Please:*

   – *attempt all of the questions, as far as possible under exam conditions*

   – ***begin your answer to each question on a new page***

   – ***leave at least 2cm margin on all borders***

   – *write in black ink using a medium-sized nib because we will be unable to mark illegible scripts*

   – *note that assignment marking is not included in the price of the course materials. Please purchase Series Marking or a Marking Voucher before submitting your script.*

   – *note that we only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2019 exams.*

2. *Please **do not**:*

   – *use headed paper*

   – *use highlighting in your script.*

At the end of the assignment

If your script is being marked by ActEd, please follow the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables and an electronic calculator.

Submission for marking

You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed on the summary page at the back of this pack and on our website at **www.ActEd.co.uk**.

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. *It is your responsibility to ensure that scripts reach ActEd in good time*. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

When submitting your script, please:

- complete the cover sheet, including the checklist
- scan your script, cover sheet (and Marking Voucher if applicable) and save as a pdf document, then email it to: ActEdMarking@bpp.com
- **do not submit a photograph of your script**
- **do not include the question paper in the scan.**

In addition, please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* 'CM2 Assignment X2 No. 12345', inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF **is less than 4 MB** in size.
- Do not protect the PDF in any way (otherwise the marker cannot return the script to ActEd, which causes delays).
- Please include the 'feedback from marker' sheet when scanning.
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

Subject CM2: Assignment X2

2019 Examinations

Please complete the following information:

Name:

ActEd Student Number (see Note below):

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting it will help us to process your scripts quickly. If you do not know your ActEd Student Number, please email us at ActEd@bpp.com.

Your ActEd Student Number is not the same as your IFoA Actuarial Reference Number or ARN.

Number of following pages: _______

Please put a tick in this box if you have solutions and a cross if you do not: ☐

Please tick here if you are allowed extra time or other special conditions in the profession's exams (if you wish to share this information): ☐

Time to do assignment (see Note below): _____ hrs _____ mins

Under exam conditions (delete as applicable): yes / nearly / no

Note: If you take more than 2¾ hours, you should indicate how much you completed within this exam time so that the marker can provide useful feedback on your progress.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Total	
__ / 8	__ / 10	__ / 8	__ / 9	__ / 12	__ / 14	__ / 8	__ / 6	__ / 5	__ / 80	=_____%

Grade: A B C D E **Marker's initials:** _________

Please tick the following checklist so that your script can be marked quickly. Have you:

[ ] Checked that you are using the latest version of the assignments, *ie* 2019 for the sessions leading to the 2019 exams?
[ ] Written your full name in the box above?
[ ] Completed your ActEd Student Number in the box above?
[ ] Recorded your attempt conditions?
[ ] Numbered all pages of your script (excluding this cover sheet)?
[ ] Written the total number of pages (excluding the cover sheet) in the space above?
[ ] Included your Marking Voucher or ordered Series X Marking?
[ ] Rated your X1 marker at **www.ActEd.co.uk/marking**?

Please follow the instructions on the previous page when submitting your script for marking.

Feedback from marker

Notes on marker's section

The main objective of marking is to provide specific advice on how to improve your chances of success in the exam. The most useful aspect of the marking is the comments the marker makes throughout the script, however you will also be given a percentage score and the band into which that score falls. Each assignment tests only part of the course and hence does not give a complete indication of your likely overall success in the exam. However it provides a good indicator of your understanding of the material tested and the progress you are making with your studies:

A = Excellent progress B = Good progress C = Average progress
D = Below average progress E = Well below average progress

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

X2.1 (i) Explain what is meant by the multifactor model. You should define any notation you use. [4]

(ii) Briefly describe three different types of factor that can be used in a multifactor model. [4]

[Total 8]

X2.2 (i) Explain what the separation theorem implies about optimal investment strategies. [2]

(ii) Explain why an individual investor wouldn't hold the market portfolio as part of their investment portfolio in practice. [2]

You are given the following historical information for a share in Company ABC and for a portfolio of 100 shares.

	Return (% *pa*)	Standard deviation of return (% *pa*)	beta
ABC	8.5	20	0.7
Portfolio	10.5	16	1.1

(iii) Use these results to derive the expected return on the market portfolio and the risk-free rate of return assuming the CAPM applies. [3]

A student has commented that ABC's lower return and higher standard deviation, relative to the 100-share portfolio, contradicts the predictions of the CAPM.

(iv) Discuss the student's comment. [3]

[Total 10]

X2.3 (i) Use Taylor's formula to derive Ito's Lemma for a function $f(X_t)$. [2]

An oil trader uses the following model for the short-term behaviour of the oil price X_t, measured in terms of US $100 per barrel:

$$X_t = 0.05\,t + 0.10\,B_t$$

where B_t is a standard Brownian motion.

(ii) Use Ito's Lemma for a function $f(B_t, t)$ to derive the stochastic differential equation for X_t. [3]

A bank offers an exotic derivative whose value is given by:

$$G(X_t) = X_t^2$$

(iii) Use Ito's Lemma to derive the stochastic differential equation for the exotic derivative. [3]

[Total 8]

X2.4 Let B_t be a standard Brownian motion, and let F_t be its natural filtration.

(i) Derive the conditional expectations $E\left[B_t^2 \middle| F_s\right]$ and $E\left[B_t^4 \middle| F_s\right]$, where $s \leq t$.

You may assume that the third and fourth moments of a random variable with distribution $N\left(0,\sigma^2\right)$ are 0 and $3\sigma^4$ respectively. [4]

(ii) Hence construct a martingale out of B_t^4. [5]

[Total 9]

X2.5 An investment banker wishes to model exchange rate movements between US Dollars and Euros as a geometric Brownian motion. Suppose they decide to use the following stochastic differential equation for this purpose:

$$dX_t = (r_d - r_e)X_t dt + \sigma X_t dB_t$$

where:

- X_t represents the value of US Dollars in terms of Euros
- r_d and r_e are the (constant) short-term interest rates in the US and the Eurozone respectively
- B_t is a standard Brownian motion.

(i) Explain why, in economic terms, the value of the US Dollar is likely to increase when $r_d > r_e$. [2]

(ii) By considering the function $f(X_t) = \log X_t$, use Ito's Lemma to solve the above stochastic differential equation for X_t. [5]

(iii) Let $G_t = \dfrac{1}{X_t}$ denote the value of the Euro in terms of US Dollars. Derive the stochastic differential equation for G_t and comment on your answer. [5]

[Total 12]

X2.6 (i) (a) State the defining properties of standard Brownian motion B_t.

(b) Write down the probability density function, for an increment over a time lag $t-s$, of general Brownian motion $W_t = \sigma B_t + \mu t$. [5]

(ii) By first obtaining the stochastic differential equation for the function $f(S_t) = \log S_t$, solve the stochastic differential equation defining geometric Brownian motion:

$$dS_t = \mu S_t \, dt + \sigma S_t \, dB_t \qquad [5]$$

(iii) S_t, the price of a share at time *t*, is modelled as geometric Brownian motion. If μ = 20% *pa* and σ = 10% *pa*, calculate the probability that the share price will exceed 110 in six months' time given that its current price is 100. [4]

[Total 14]

X2.7 (i) Describe what is meant by the lognormal model of security prices. [2]

(ii) If X_t is defined to be the deviation of the log of the security price S_t from its trend value, show that changes in X_t over a time interval *h* are stationary. [2]

(iii) Derive expressions for the mean and the variance of the security price S_t. [4]

[Total 8]

X2.8 The stochastic differential equation that implies that the price of an asset at time t, S_t, follows a geometric Brownian motion, is:

$$dS_t = S_t\,\{\mu\, dt + \sigma\, dZ_t\}$$

where Z_t is a standard Brownian motion process.

Describe and discuss the plausibility of the assumptions behind this equation when it is used as a model of share prices. [6]

X2.9 Consider an investment market in which:

- the risk-free rate of return on Treasury bills is 4% *pa*
- the expected return on the market as a whole is 8% *pa*
- the standard deviation of the return on the market as a whole is 30% *pa*
- the assumptions of the capital asset pricing model (CAPM) hold.

(i) Consider an efficient portfolio Z that consists entirely of Treasury bills and non-dividend-paying shares, there being no other types of investment. If Z yields an expected return of 7% *pa*, determine its beta. [1]

(ii) Calculate the standard deviation of returns for Portfolio Z. [2]

(iii) Split the total standard deviation for Portfolio Z into the amounts attributable to systematic risk and specific risk. [1]

(iv) Calculate the market value of Portfolio Z assuming that its constituent securities are expected to realise a total sum of \$100 in one period from now. [1]

[Total 5]

Subject CM2: Assignment X3

2019 Examinations

Time allowed: 3¼ hours

Instructions to the candidate

1. *Please:*
   - *attempt all of the questions, as far as possible under exam conditions*
   - ***begin your answer to each question on a new page***
   - ***leave at least 2cm margin on all borders***
   - *write in black ink using a medium-sized nib because we will be unable to mark illegible scripts*
   - *note that assignment marking is not included in the price of the course materials. Please purchase Series Marking or a Marking Voucher before submitting your script.*
   - *note that we only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2019 exams.*
2. *Please **do not**:*
   - *use headed paper*
   - *use highlighting in your script.*

At the end of the assignment

If your script is being marked by ActEd, please follow the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables and an electronic calculator.

Submission for marking

You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed on the summary page at the back of this pack and on our website at **www.ActEd.co.uk**.

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. *It is your responsibility to ensure that scripts reach ActEd in good time*. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

When submitting your script, please:

- complete the cover sheet, including the checklist
- scan your script, cover sheet (and Marking Voucher if applicable) and save as a pdf document, then email it to: ActEdMarking@bpp.com
- **do not submit a photograph of your script**
- **do not include the question paper in the scan.**

In addition, please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* 'CM2 Assignment X3 No. 12345', inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF **is less than 4 MB** in size.
- Do not protect the PDF in any way (otherwise the marker cannot return the script to ActEd, which causes delays).
- Please include the 'feedback from marker' sheet when scanning.
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

Subject CM2: Assignment X3

2019 Examinations

Please complete the following information:

Name:

ActEd Student Number (see Note below):

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting it will help us to process your scripts quickly. If you do not know your ActEd Student Number, please email us at ActEd@bpp.com.

Your ActEd Student Number is not the same as your IFoA Actuarial Reference Number or ARN.

Number of following pages: _______

Please put a tick in this box if you have solutions and a cross if you do not: ☐

Please tick here if you are allowed extra time or other special conditions in the profession's exams (if you wish to share this information): ☐

Time to do assignment
(see Note below): _____ hrs _____ mins

Under exam conditions
(delete as applicable): yes / nearly / no

Note: If you take more than 3¼ hours, you should indicate how much you completed within this exam time so that the marker can provide useful feedback on your progress.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	**Total**	
__ / 10	__ / 6	__ / 10	__ / 12	__ / 9	__ / 12	__ / 9	__ / 13	__ / 11	__ / 8	__ / 100	=_____%

Grade: A B C D E **Marker's initials:** _________

Please tick the following checklist so that your script can be marked quickly. Have you:

[ ] Checked that you are using the latest version of the assignments, *ie* 2019 for the sessions leading to the 2019 exams?
[ ] Written your full name in the box above?
[ ] Completed your ActEd Student Number in the box above?
[ ] Recorded your attempt conditions?
[ ] Numbered all pages of your script (excluding this cover sheet)?
[ ] Written the total number of pages (excluding the cover sheet) in the space above?
[ ] Included your Marking Voucher or ordered Series X Marking?
[ ] Rated your X2 marker at **www.ActEd.co.uk/marking**?

Please follow the instructions on the previous page when submitting your script for marking.

Feedback from marker

Notes on marker's section

The main objective of marking is to provide specific advice on how to improve your chances of success in the exam. The most useful aspect of the marking is the comments the marker makes throughout the script, however you will also be given a percentage score and the band into which that score falls. Each assignment tests only part of the course and hence does not give a complete indication of your likely overall success in the exam. However it provides a good indicator of your understanding of the material tested and the progress you are making with your studies:

A = Excellent progress B = Good progress C = Average progress
D = Below average progress E = Well below average progress

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

X3.1 (i) In the context of a non-dividend-paying security, define the Greeks (in both words and using formulae) and state whether each has a positive or negative value for a call option and a put option. [5]

(ii) Consider a 30-day, at-the-money call option on a non-dividend-paying share currently valued at £8. The volatility of the share is 30% and the continuously compounded risk-free rate of interest is 4%.

The outputs of a computer model used to value the option are:

Option value	28.7p	Theta	-0.499pday^{-1}
Delta	0.532	Vega	$0.91\text{p}\%^{-1}$
Gamma	0.00578p^{-1}	Rho	$0.33\text{p}\%^{-1}$

After one day the share price increases by 50p, volatility is reassessed to be 35% *pa* and the risk-free rate moves to 3.5% *pa*. Estimate the new price of the option. [5]

[Total 10]

X3.2 (i) Consider a call option and a put option on a dividend-paying security, each with the same term and exercise price. By considering the put-call parity relationship or otherwise, state the value of n such that:

$$\Delta_c = \Delta_p + n$$

(Δ_c is the delta for the call option and Δ_p is the delta for the put option.) [1]

(ii) Derive similar relationships for the other five Greeks. [3]

(iii) Hence, or otherwise, decide whether or not the following relationship holds:

$$r\rho_c + q\lambda_c + (T-t)\theta_c = r\rho_p + q\lambda_p + (T-t)\theta_p \qquad [2]$$

[Total 6]

X3.3 A non-dividend-paying stock has a current price of £100. In any unit of time the price of the stock is expected to increase by 10% or decrease by 5%. The continuously compounded risk-free interest rate is 4% per unit of time.

A European call option is written with a strike price of £103 and is exercisable after two units of time, at t = 2.

Establish, using a binomial tree, the replicating portfolio for the option at the start and end of the first unit of time, *ie* at t = 0, 1. Hence, calculate the value of the option at t = 0. [10]

X3.4 (i) Explain what is meant by a "replicating portfolio". [2]

(ii) Explain what is meant by a "risk-neutral probability measure" and state mathematically what it implies about the pricing of derivatives relative to the price of the underlying asset. [2]

(iii) Consider a one-period model of a non-dividend paying-stock, currently priced at S_0 and which may move up or down to give $S_1 = S_0 u$ or $S_1 = S_0 d$. Consider a derivative that pays c_u or c_d following an up or down event. The risk-free rate of return (continuously compounded) is r.

(a) Use a replicating portfolio to derive an equation for the price of the derivative at time $t = 0$.

(b) Hence find the price of a derivative whose payoff is defined as $|S_1 - S_0|$, assuming $d < 1$ and $u > 1$.

(c) Explain how to synthesise the derivative in (iii)(b) from simpler options. [8]

[Total 12]

X3.5 Using the Black-Scholes formula for the value of a European call option on a non-dividend-paying stock, show that the call price, c, tends to the maximum of $S - Ke^{-r(T-t)}$ and zero (depending on the strike price) as σ tends to zero. [9]

X3.6 The price of a non-dividend-paying stock at time 1, S_1, is related to the price at time 0, S_0, as follows:

$$S_1 = \begin{cases} S_0 u & \text{with probability } p \\ S_0 d & \text{with probability } 1-p \end{cases}$$

The continuously compounded rate of return on a risk-free asset is r.

(i) (a) Determine the replicating portfolio for a European call option written on the stock that expires at time 1 and has a strike price of k, where $dS_0 < k < uS_0$. You should give expressions for the number of units for each constituent in the portfolio.

(b) Use your expressions in (i)(a) to find a formula for the price of the European call option.

(c) Use put-call parity to derive a formula for the price of the corresponding European put option, with the same strike price and strike date.

(d) Show that the price of the European call option in (i)(b) can be written as the discounted expected payoff under a probability measure Q. Hence find an expression for the probability, q, of an upward move in the stock price under Q.

[7]

(ii) Explain the relationship between the probability measure Q in (ii) and the real-world probability measure P. Explain what relationship you would expect q and p to have if all investors are:

(a) risk-averse

(b) risk-seeking

(c) risk-neutral. [5]

[Total 12]

X3.7 (i) State the put-call parity relationship for a non-dividend-paying share with value S_t. [1]

(ii) Using the result in part (i) and the Black-Scholes formula for the value of a European call option on a non-dividend-paying share as given in the Tables, derive an expression for p_t in terms of K, r, $T-t$, d_1 and d_2. [2]

(iii) The underlying share pays no dividends and has a current value of £20 and a volatility of 0.3. An investor who has £100 to invest has a choice between investing in either a one-year zero-coupon bond (redeemable at par) with a current market value of £94.18 or in one-year put options with a strike price of £17.50. If the investor chooses to allocate all of their money to the options, how many can they buy?

[Ignore tax and investment expenses and assume that the bond market and the options market are both arbitrage-free. Assume that the option price is quoted to the nearest penny.] [6]

[Total 9]

X3.8 The price S_t of a particular share follows a geometric random walk:

$$S_t = S_{t-1} Z_t$$

where $\{Z_t\}$ is a sequence of independent, identically distributed random variables:

$$Z_t = \begin{cases} 1.1 & \text{with probability 0.6} \\ 0.95 & \text{with probability 0.4} \end{cases}$$

and t denotes the time in months.

A 1-month European call option is available on the share with a strike price of £10.50. The current market price of the share is £10. No dividends are to be paid over the next 6 months. An annualised risk-free force of interest of 4% is available.

(i) Find the expected payoff of the call option. [1]

(ii) Construct a replicating portfolio for the derivative out of shares and cash, and hence find the fair price of the derivative. [3]

(iii) Hence state how many options need to be bought (or sold) per share, in order to construct a risk-free portfolio out of shares and these options. [1]

(iv) Describe quantitatively the arbitrage opportunity that would arise if the price of the option in the market was equal to the discounted value of the expected payoff. [4]

(v) Another derivative is available on this share. It gives the purchaser the right (but not the obligation) to buy two shares at a price of £12.00 each, or sell one share at a price of £9.00, both in 4 months' time. Determine the possible payoffs for this derivative, and hence find the fair price of this derivative using risk-neutral valuation. [4]

[Total 13]

X3.9 An investment bank has developed a new exotic derivative, which will pay an amount equal to the share price at maturity multiplied by the share price at maturity less one dollar. Let T be the maturity date of the derivative and r be the risk-free force of interest and assume that the Black-Scholes analysis applies.

(i) Use risk-neutral valuation to derive the pricing formula for this derivative at time $t<T$, based on a share that pays no dividends. [7]

(ii) (a) Derive the corresponding formula for the delta of the derivative.

(b) Derive a condition for the range of values for the current share price for which delta is positive and comment on what your answer suggests for derivatives of this type with differing terms.

(c) Derive the corresponding formula for the gamma of the derivative and comment on the sign of gamma. [4]

[Total 11]

X3.10 The price S_t of a share that pays no dividends follows a geometric Brownian motion:

$$dS_t = S_t\left(\mu dt + \sigma dZ_t\right)$$

where Z_t is a standard Brownian motion. A derivative is available on this share that can only be exercised at time T. The price of the derivative at time t, $f\left(t,S_t\right)$, depends on the time and the current share price. A cash bond is also available that offers a risk-free rate of return of r (continuously compounded). The price of the bond is B_t. You wish to set up a replicating portfolio for the derivative made out of shares and cash, so that:

$$\phi_t S_t + \psi_t B_t = f\left(t,S_t\right) \qquad (*)$$

(i) Write down the differential equation that is satisfied by B_t. [1]

(ii) What does it mean for the portfolio to be self-financing? Give a differential equation that must be satisfied by the portfolio considered above in order that this is the case. [2]

(iii) What does it mean for a process to be previsible? [1]

(iv) By applying Ito's Lemma to the right-hand side of equation (*) and using your answers to parts (i) and (ii), deduce that:

$$\phi_t = \frac{\partial}{\partial S_t} f\left(t, S_t\right)$$ [4]

[Total 8]

Subject CM2: Assignment X4

2019 Examinations

Time allowed: 3¼ hours

Instructions to the candidate

1. *Please:*

   – *attempt all of the questions, as far as possible under exam conditions*

   – ***begin your answer to each question on a new page***

   – ***leave at least 2cm margin on all borders***

   – *write in black ink using a medium-sized nib because we will be unable to mark illegible scripts*

   – *note that assignment marking is not included in the price of the course materials. Please purchase Series Marking or a Marking Voucher before submitting your script.*

   – *note that we only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2019 exams.*

2. *Please **do not**:*

   – *use headed paper*

   – *use highlighting in your script.*

At the end of the assignment

If your script is being marked by ActEd, please follow the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables and an electronic calculator.

Submission for marking

You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed on the summary page at the back of this pack and on our website at **www.ActEd.co.uk**.

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. *It is your responsibility to ensure that scripts reach ActEd in good time*. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

When submitting your script, please:

- complete the cover sheet, including the checklist
- scan your script, cover sheet (and Marking Voucher if applicable) and save as a pdf document, then email it to: ActEdMarking@bpp.com
- **do not submit a photograph of your script**
- **do not include the question paper in the scan.**

In addition, please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* ‘CM2 Assignment X4 No. 12345’, inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF **is less than 4 MB** in size.
- Do not protect the PDF in any way (otherwise the marker cannot return the script to ActEd, which causes delays).
- Please include the ‘feedback from marker’ sheet when scanning.
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

Subject CM2: Assignment X4

2019 Examinations

Please complete the following information:

Name:

ActEd Student Number (see Note below):

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting it will help us to process your scripts quickly. If you do not know your ActEd Student Number, please email us at ActEd@bpp.com.

Your ActEd Student Number is not the same as your IFoA Actuarial Reference Number or ARN.

Number of following pages: _______

Please put a tick in this box if you have solutions and a cross if you do not: [ ]

Please tick here if you are allowed extra time or other special conditions in the profession's exams (if you wish to share this information): [ ]

Time to do assignment (see Note below): _____ hrs _____ mins

Under exam conditions (delete as applicable): yes / nearly / no

Note: If you take more than 3¼ hours, you should indicate how much you completed within this exam time so that the marker can provide useful feedback on your progress.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Total	
__ / 11	_ / 9	__ / 14	__ / 13	_ / 4	_ / 6	_ / 4	_ / 6	_ / 8	_ / 6	_ / 9	__ / 10	___ / 100	=_____%

Grade: A B C D E **Marker's initials:** ________

Please tick the following checklist so that your script can be marked quickly. Have you:

[ ] Checked that you are using the latest version of the assignments, *ie* 2019 for the sessions leading to the 2019 exams?
[ ] Written your full name in the box above?
[ ] Completed your ActEd Student Number in the box above?
[ ] Recorded your attempt conditions?
[ ] Numbered all pages of your script (excluding this cover sheet)?
[ ] Written the total number of pages (excluding the cover sheet) in the space above?
[ ] Included your Marking Voucher or ordered Series X Marking?
[ ] Rated your X3 marker at **www.ActEd.co.uk/marking**?

Please follow the instructions on the previous page when submitting your script for marking.

Feedback from marker

Notes on marker's section

The main objective of marking is to provide specific advice on how to improve your chances of success in the exam. The most useful aspect of the marking is the comments the marker makes throughout the script, however you will also be given a percentage score and the band into which that score falls. Each assignment tests only part of the course and hence does not give a complete indication of your likely overall success in the exam. However it provides a good indicator of your understanding of the material tested and the progress you are making with your studies:

A = Excellent progress B = Good progress C = Average progress
D = Below average progress E = Well below average progress

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

X4.1 An analyst is using the Merton model, together with the following information, to value the five-year zero-coupon bonds (ZCBs) issued by a company:

- The nominal value of ZCBs issued is \$100 million.
- The company's shares have a market capitalisation of \$118.46 million.
- The volatility of the company's underlying assets has been estimated to be 25% *pa*.
- The five-year risk-free force of interest is 5% *pa*.

(i) Calculate the price per \$100 nominal of a five-year risk-free ZCB. [1]

(ii) Using your answer to (i) to obtain an initial estimate, and then applying linear interpolation, estimate the value of the company's assets (to the nearest \$10,000) and hence show that the value of its ZCBs is \$76.47 million. [5]

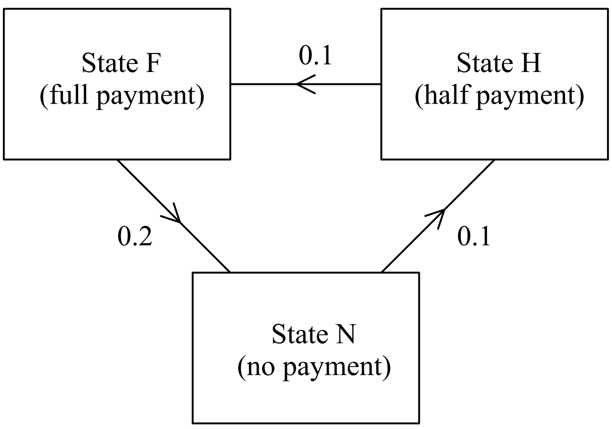
(iii) (a) State the formula for the delta of a European call option based on the Black-Scholes formula (assuming no dividends) and use it to derive a formula for the delta of the ZCBs with respect to the value of the company's assets.

(b) Estimate the numerical value of delta using your calculations in part (ii) and use it to estimate the new value of the ZCBs following a \$10 million fall in the value of the company's assets.

(c) The actual value of the ZCBs following a \$10 million fall in the value of the company's assets is \$76.16 million. Give a possible reason for the discrepancy between your estimated value of the ZCBs and the actual value. [5]

[Total 11]

X4.2 A bank is using a three-state discrete-time Markov chain model to value its bond portfolio.



On 1 January each year the bank assigns each of its client companies to one of the following categories:

- State F: The bank expects to receive any payments due that year in full.
- State H: The bank expects to receive only 50% of any payments due that year.
- State N: The bank expects to receive no payments from the company that year.

The diagram shows the risk-neutral probabilities that each company will move from its current rating level to another level at the time of each review. These probabilities are independent of the company's previous ratings and the behaviour of other companies.

Let $p_{ij}(0,t)$ denote the probability that a company initially in State i will be in State j t years later.

(i) Calculate $p_{Fj}(0,t)$ for $j=F,H$ and $t=1,2,3$. [3]

The bank is considering purchasing at par a 3-year bond issued by a company currently rated as F. Under the terms of the bond, interest of 10% of the face value of the bond will be paid at the end of each year, and the bond will be redeemed at par at the end of the 3 years.

The annual effective yields on 1-year, 2-year and 3-year government bonds are all 5%.

(ii) (a) Calculate the risk-neutral expected present value of the payments from the bond per £100 face value.

(b) Comment on your answer in (ii)(a). [3]

After negotiations, the bank agrees to purchase the bonds at a price of £95.20.

(iii) Calculate the credit spread for this bond. [3]

[Total 9]

X4.3 In the Vasicek model, the spot rate of interest is governed by the stochastic differential equation:

$$dr_t = a\left(b - r_t\right)dt + \sigma\, dB_t$$

where B_t is a standard Brownian motion and a, $b > 0$ are constants.

(i) A stochastic process $\{U_t : t \geq 0\}$ is defined by $U_t = e^{at} r_t$.

(a) Derive an equation for dU_t.

(b) Hence solve the equation to find U_t.

(c) Hence show that:

$$r_t = b + \left(r_0 - b\right)e^{-at} + \sigma \int_0^t e^{a(s-t)} dB_s \qquad [5]$$

(ii) Determine the probability distribution of r_t and the limiting distribution for large t. [4]

(iii) Derive, in the case where $s < t$, the conditional expectation $E\left[r_t \middle| F_s\right]$, where $\left\{F_s : s \geq 0\right\}$ is the filtration generated by the Brownian motion B_s. [5]

[Total 14]

X4.4 (i) Describe briefly the Vasicek one-factor model of interest rates and its key statistical properties. [4]

(ii) According to a particular parameterisation of this model, the instantaneous forward rate applicable at a fixed time $T > t$ implied by the market prices at time t is found to be:

$$f(t,T) = r_t e^{-\alpha\tau} + r_\infty(1-e^{-\alpha\tau}) + k(1-e^{-\alpha\tau})e^{-\alpha\tau}$$

where $\tau = T-t$ and $\alpha > 0$.

Show that, if a humped curve is required for $f(t,T)$, the parameter values must satisfy the condition $k > |r_\infty - r_t|$. [5]

Here a "humped curve" means one where the value of the function for some intermediate values of τ exceeds the values for both $\tau = 0$ and $\tau = \infty$. In other words, there will be a maximum value for some positive value of τ.

(iii) Describe briefly the main advantages and limitations of the Vasicek model. [4]

[Total 13]

X4.5 Claims occur according to a compound Poisson process at a rate of ¼ claims per year. Individual claim amounts, X, have probability function:

$$P(X = 50) = 0.8$$

$$P(X = 100) = 0.2$$

The insurer charges a premium at the beginning of each year using a 20% loading factor. The insurer's surplus at time t is $U(t)$. Find $P[U(2) < 0]$ if the insurer starts at time 0 with a surplus of 100. [4]

X4.6 Claims arrive in a Poisson process at rate λ, and $N(t)$ is the number of claims arriving by time t. The claim sizes are independent random variables $X_1, X_2, \ldots$ with mean μ, independent of the arrivals process. The initial surplus is u and the premium loading factor is θ.

(i) (a) Give an expression for the surplus $U(t)$ at time t.

(b) Define the probability of ruin with initial surplus u, $\Psi(u)$, and sketch a realisation of the surplus process that shows a ruin event.

(c) State the value of $\Psi(u)$ when $\theta = 0$. [4]

(ii) The unit of currency is changed so that one unit of the old currency is worth the same as 2.5 units of the new currency.

Determine a relationship between $\Psi(u)$ in (i)(b) and the probability of ruin for the new process. [2]

[Total 6]

X4.7 The general form of a run-off triangle can be expressed as:

Accident Year, i	*Development Year, j* 0	1	2	3	4	5
0	$C_{0,0}$	$C_{1,0}$	$C_{2,0}$	$C_{3,0}$	$C_{4,0}$	$C_{5,0}$
1	$C_{0,1}$	$C_{1,1}$	$C_{2,1}$	$C_{3,1}$	$C_{4,1}$	
2	$C_{0,2}$	$C_{1,2}$	$C_{2,2}$	$C_{3,2}$		
3	$C_{0,3}$	$C_{1,3}$	$C_{2,3}$			
4	$C_{0,4}$	$C_{1,4}$				
5	$C_{0,5}$					

Define a model for each *incremental* entry, C_{ij}, in general terms and explain each element of the formula. [4]

X4.8 Claims arrive in a Poisson process at rate λ. Individual claim amounts are all exactly 100. The insurer applies a premium loading factor of 20%.

(i) (a) Show that the adjustment coefficient, R, satisfies:

$$e^{100R} - 120R - 1 = 0$$

(b) By approximating e^{100R} with a series expansion up to terms in R^3, obtain an approximate value of R. [4]

(ii) Determine the minimum initial capital such that the probability of ruin is at most 0.05. [2]

[Total 6]

X4.9 The table below shows the payments made in each development year in respect of an insurer's claims for fire damage for the three most recent calendar years. You may assume that all claims are paid in the middle of each year.

Claim payments made during year (£'000)		Development year 0	1	2
Accident year	2010	830	940	150
	2011	850	920	
	2012	1,120		

The rate of claims inflation over these years, measured over the 12 months to the middle of each year are given below:

Annual claim inflation rate (past)	
-2011	2%
-2012	2.5%

Estimated annual claim inflation rate (future)	
-2013	3%
-2014	3%

Use the inflation adjusted chain ladder method to estimate the total amount outstanding for future claims arising from accident years 2011 and 2012. [8]

X4.10 Cumulative claims incurred on a motor insurance account are as follows:

Cumulative claims incurred (£'000)		Development year		
		0	1	2
Policy year	2010	1,417	1,923	2,101
	2011	1,701	2,140	
	2012	1,582		

The data have already been adjusted for inflation. Annual premiums written in 2012 were £3,073,000 and the ultimate loss ratio has been estimated as 92%. Claims paid to date for policy year 2012 are £441,000, and claims are assumed to be fully run-off by the end of Development year 2.

Estimate the outstanding claims to be paid arising from policies written in 2012 **only**, using the Bornhuetter-Ferguson technique. [6]

X4.11 The following table gives the cumulative incurred claims data, by years of accident and reporting development for a portfolio of motor insurance policies:

Cumulative incurred claims (£'000)		Development year		
		0	1	2
Accident year	2010	252	375	438
	2011	230	343	
	2012	208		

Number of reported claims		Development year		
		0	1	2
Accident year	2010	56	74	87
	2011	49	65	
	2012	44		

(i) Given that the total claims paid to date are £950,000 for Accident years 2010 to 2012 calculate the outstanding claims reserve for this cohort using the average cost per claim method with grossing-up factors. [7]

(ii) State the assumptions that underlie your result. [2]

[Total 9]

X4.12 Aggregate claims on a general insurance company's portfolio form a compound Poisson process with parameter λ.

Individual claims have an exponential distribution with mean 100. The company applies a 20% premium loading. The insurer effects proportional reinsurance with a retained proportion of α. The reinsurer applies a 30% premium loading.

(i) Calculate the minimum value of α such that the insurer's net income is greater than the expected net claims. [2]

(ii) Hence, show that the direct insurer's adjustment coefficient, R, satisfies:

$$R = \frac{1-3\alpha}{100\alpha - 1{,}300\alpha^2}$$ [4]

(iii) By differentiating the result from (ii), show that $\alpha = 0.6257$ maximises the adjustment coefficient and calculate the corresponding optimal value of R.
You may assume that the turning point is a maximum. [4]

[Total 10]

For the session leading to the April 2019 exams – CS1 & CM2 Subjects

Marking vouchers

Subjects	Assignments	Mocks
CS1	6 March 2019	13 March 2019
CM2	20 March 2019	27 March 2019

Series X and Y Assignments

Subjects	Assignment	Recommended submission date	Final deadline date
CS1	**X1**	**5 December 2018**	9 January 2019
CM2		**19 December 2018**	23 January 2019
CS1	**X2**	**19 December 2019**	23 January 2019
CS1	**Y1**	**2 January 2019**	30 January 2019
CM2	**X2**	**9 January 2019**	6 February 2019
CM2	**Y1**	**16 January 2019**	13 February 2019
CS1	**X3**	**16 January 2019**	13 February 2019
CM2		**30 January 2019**	27 February 2019
CS1	**X4**	**30 January 2019**	27 February 2019
CS1	**Y2**	**13 February 2019**	6 March 2019
CM2	**X4**	**13 February 2019**	13 March 2019
CM2	**Y2**	**27 February 2019**	20 March 2019

Mock Exams

Subjects	Recommended submission date	Final deadline date
CS1 (Paper A/B)	**27 February 2019**	13 March 2019
CM2 (Paper A/B)	**13 March 2019**	27 March 2019

We encourage you to work to the recommended submission dates where possible.
If you submit your mock on the final deadline date you are likely to receive your script back less than a week before your exam.

For the session leading to the September 2019 exams – CS1 & CM2 Subjects

Marking vouchers

Subjects	Assignments	Mocks
CS1	21 August 2019	28 August 2019
CM2	28 August 2019	4 September 2019

Series X and Y Assignments

Subjects	Assignment	Recommended submission date	Final deadline date
CS1	**X1**	**22 May 2019**	17 July 2019
CM2		**29 May 2019**	24 July 2019
CS1	**X2**	**5 June 2019**	24 July 2019
CM2		**12 June 2019**	31 July 2019
CS1	**Y1**	**19 June 2019**	31 July 2019
CM2		**26 June 2019**	7 August 2019
CS1	**X3**	**3 July 2019**	7 August 2019
CM2		**10 July 2019**	14 August 2019
CS1	**X4**	**17 July 2019**	14 August 2019
CM2		**24 July 2019**	21 August 2019
CS1	**Y2**	**31 July 2019**	21 August 2019
CM2		**7 August 2019**	28 August 2019

Mock Exams

Subjects	Recommended submission date	Final deadline date
CS1 (Paper A/B)	**14 August 2019**	28 August 2019
CM2 (Paper A/B)	**21 August 2019**	4 September 2019

We encourage you to work to the recommended submission dates where possible.

If you submit your mock on the final deadline date you are likely to receive your script back less than a week before your exam.

Assignment X1 Solutions

Markers: This document sets out one approach to solving each of the questions. Please give credit for other valid approaches.

Note that some of the numerical answers are sensitive to rounding.

Solution X1.1

Course reference: The relevance of technical analysis, fundamental analysis and insider trading to the different forms of market efficiency is discussed in Chapter 1.

(i) ***Technical analysis***

This is based purely on historical market data, *ie* prices and trading volumes. If the market is efficient in the weak sense, then technical analysis will be of no benefit. [1]

Conversely, if it is possible to generate excess risk-adjusted returns based purely on historical market data, then this would suggest that the market is weak form inefficient (and hence also semi-strong and strong form inefficient). [1]

[Total 2]

(ii) ***Insider trading***

Insider trading is based on information that is not publicly available and which will therefore not be built into the prices in a market that is only semi-strong form efficient. If the market is not efficient in the strong sense, then insider trading would enable the investor to generate excess risk-adjusted returns. [1]

Making insider trading illegal aims to remove this advantage from those with access to inside information, and so suggests that some markets may not be strong form efficient. [1]

[Total 2]

(iii) ***Fundamental analysis***

The information mentioned here is publicly available. [½]

So if the market is semi-strong form efficient, fundamental analysis will be of no benefit in helping the investor to identify mispriced securities, which can then be traded to generate excess risk-adjusted returns. [1]

Fundamental analysis can add value only if the market is semi-strong form inefficient. [½]

[Total 2]

Solution X1.2

Course reference: The risk aversion measures and their relationships to certainty equivalence are described in Chapter 2.

(i) ***Risk aversion measures***

Absolute risk aversion is measured by:

$$A(w) = \frac{-U''(w)}{U'(w)}$$ [¼]

In this case:

$$U(w) = -\exp\left(-\frac{w}{100}\right)$$

$$U'(w) = \frac{1}{100}\exp\left(-\frac{w}{100}\right)$$

$$U''(w) = \frac{-1}{10{,}000}\exp\left(-\frac{w}{100}\right)$$ [½]

So:

$$A(w) = \frac{1}{100}$$ [¼]

The investor exhibits constant absolute risk aversion. [¼]

Relative risk aversion is measured by:

$$R(w) = -w\frac{U''(w)}{U'(w)} = w \times A(w)$$ [¼]

So in this case:

$$R(w) = \frac{w}{100}$$ [¼]

The investor exhibits increasing relative risk aversion. [¼]

[Total 2]

(ii) ***Find the certainty equivalent***

The certainty equivalent of the gamble, c_X, is found from the equation:

$$U(1000 + c_X) = E[U(1000 + X)]$$ [½]

This can be written as:

$$-\exp\left(-\frac{(c_X+1000)}{100}\right)=-\frac{1}{2}e^{-1100/100}-\frac{1}{2}e^{-950/100}$$ [½]

Therefore, the certainty equivalent is:

$$c_X=-100\ln\left(\frac{1}{2}e^{-1100/100}+\frac{1}{2}e^{-950/100}\right)-1000=-0.83$$ [1]

[Total 2]

Solution X1.3

Course reference: Stochastic dominance is discussed in Chapter 3.

(i) ***Why Asset B is second-order stochastically dominant***

Both assets offer an expected return of 1%. [½]

Asset A has some risk while Asset B is risk-free. [½]

If an asset is second-order stochastically dominant over another asset, then every risk-averse investor will prefer that asset. [½]

This must be the case here, as every risk-averse investor will prefer the risk-free asset, Asset B. [½]

[Total 2]

(ii) ***Verifying second-order stochastic dominance numerically***

We can check second-order stochastic dominance by using the table below.

Return %	F_A	F_B	$\sum F_A$	$\sum F_B$
–1	0.5	0	0.5	0
0	0.5	0	1	0
1	0.5	1	1.5	1
2	0.5	1	2	2
3	1	1	3	3

[1½, less ½ for each incorrect column]

*Strictly speaking, second-order stochastic dominance is defined in terms of the integral of the cumulative probability distribution function. However, simply summing the cumulative probabilities will work for a discrete distribution provided that the probabilities are summed in **equally sized** steps.*

Asset B is second-order dominant since:

$$\sum F_B(x) \le \sum F_A(x) \text{ for all } x$$

and $$\sum F_B(x) < \sum F_A(x) \text{ for } x = -1,0,1$$ [1]

[Total 2½]

(iii)(a) ***What can be said about dominance***

We can check first-order and second-order dominance by using the table below.

Return %	F_A	F_B	$\sum F_A$	$\sum F_B$
–1	0.5	0	0.5	0
0	0.5	0	1	0
1	0.5	1	1.5	1
2	0.5	1	2	2
3	0.5	1	2.5	3
4	1	1	3.5	4

[1½, less ½ for each incorrect column]

From the table we see that Asset B is not first-order or second-order dominant. [1]

[Total 2½]

(iii)(b) ***What can be said about dominance***

We can check first-order dominance by using the table below.

Return %	F_A	F_B
–1	0.5	0
0	0.5	0
1	1	1

[¼ for each correct column]

Asset B is therefore first-order dominant (and therefore also second-order dominant). [½]

[Total 1]

(iv) ***Can Asset B be dominated?***

No. If an Asset D offers even the smallest possibility of a return of $x\% < 1\%$ then $F_D(x) > F_B(x)$ and so Asset D can never dominate Asset B. [1]

Solution X1.4

Course reference: The measures of investment risk covered by this question are discussed in Chapter 4.

(i) ***Expected return***

Investment A $0.4 \times 0.1 + 0.2 \times 0.15 + 0.4 \times 0.2 = 15\%$ [½]

Investment B $0.25 \times 0.1 + 0.7 \times 0.15 + 0.05 \times 0.4 = 15\%$ [½]

Investment C $\dfrac{0.1+0.2}{2} = 15\%$ [½]

[Total 1½]

(ii) ***Variance***

Investment A $0.4 \times 0.05^2 + 0.4 \times (-0.05)^2 = 0.002$ or $20\%\%$ [1]

Investment B $0.25 \times 0.05^2 + 0.05 \times (-0.25)^2 = 0.00375$ or $37.5\%\%$ [1]

Investment C $\displaystyle\int_{0.1}^{0.2} 10\,(0.15-x)^2\,dx$ [½]

$$=10\left[0.15^2 x+\frac{x^3}{3}-0.15x^2\right]_{0.1}^{0.2}$$

[½]

$= 10 \times (0.0011666 - 0.0010833)$

$= 0.000833$ or $8.33\%\%$ [½]

[Total 3½]

Or using the formula $Var[X]=\dfrac{1}{12}(b-a)^2$ *from page 13 of the Tables, we get:*

$$\frac{1}{12}(20-10)^2=8.33\%\%$$

Or using $Var[X]=E(X^2)-[E(X)]^2$ *we have* $E(X)=15\%$ *from part (i) and:*

$$E(X^2)=\int_{0.1}^{0.2} 10x^2\,dx=10\left[\frac{x^3}{3}\right]_{0.1}^{0.2}=0.023333=233.33\%\%$$

(iii) ***Semi-variance***

Investment A $0.4 \times 0.05^2 = 0.001$ or $10\%\%$ [1]

Investment B $0.25 \times 0.05^2 = 0.000625$ or $6.25\%\%$ [1]

Investment C $\displaystyle\int_{0.1}^{0.15}(0.15-x)^2 10\,dx$ [½]

$$= 10\left[0.15^2 x + \frac{x^3}{3} - 0.15x^2\right]_{0.1}^{0.15}$$ [½]

$= 10 \times (0.0011250 - 0.0010833)$

$= 0.000417$ or $4.17\%\%$ [½]

[Total 3½]

Or, since the distributions of returns for A and C are symmetrical, their semi-variances will be half the variance.

(iv) ***Expected shortfall below 12%***

Investment A $0.4 \times 0.02 = 0.008$ or 0.8% [½]

Investment B $0.25 \times 0.02 = 0.005$ or 0.5% [½]

Investment C $\displaystyle\int_{0.1}^{0.12}(0.12-x)10\,dx$ [½]

$$= 10\left[0.12x - \frac{x^2}{2}\right]_{0.1}^{0.12}$$ [½]

$= 10 \times (0.0072 - 0.0070)$

$= 0.002$ or 0.2% [½]

[Total 2½]

The last one can also be found by noting that there is a probability of 0.2 of a shortfall and in this case the average shortfall will be 1%. So we get $0.2 \times 1\% = 0.2\%$.

(v) ***Shortfall probability below 15%***

Investment A 0.4 [¼]

Investment B 0.25 [¼]

Investment C 0.50 [½]

[Total 1]

Solution X1.5

Course reference: The lognormal stochastic investment return model is covered in Chapter 5.

(i) ***Probability of not meeting liabilities***

(a) ***All assets with guaranteed return***

If the assets are invested in the investment with a guaranteed 5% return then the accumulated value of the assets after one year will be:

$$950{,}000 \times 1.05 = £997{,}500 \quad [1]$$

The company will therefore be unable to meet its liabilities, *ie* the probability is 1. [1]

(b) ***Assets split 85% : 15%***

The accumulated value of the 15% of assets invested in the guaranteed investment is:

$$0.15 \times 950{,}000 \times 1.05 = £149{,}625 \quad [1]$$

For the company to be unable to meet its liabilities, the remaining investment must be worth less than $1{,}000{,}000 - 149{,}625 = £850{,}375$ one year from now. [1]

We therefore require:

$$\begin{aligned} P(0.85 \times 950{,}000(1+i_1) < 850{,}375) &= P\left(\log(1+i_1) < \log\left(\frac{850{,}375}{0.85 \times 950{,}000}\right)\right) \\ &= P(\log(1+i_1) < \log 1.053096) \\ &= P\left(\frac{\log(1+i_1) - \mu}{\sigma} < \frac{\log 1.053096 - \mu}{\sigma}\right) \quad [1] \end{aligned}$$

Since $\dfrac{\log(1+i_1)-\mu}{\sigma} \sim N(0,1)$, $\mu = 0.06748$ and $\sigma^2 = 0.0003493$, this is equivalent to:

$$\begin{aligned} P\left(Z < \frac{\log 1.053096 - 0.06748}{\sqrt{0.0003493}}\right) &= P(Z < -0.842) = 1 - P(Z < 0.842) \\ &= 1 - 0.80 = 0.2 \quad [2] \end{aligned}$$

(ii) ***Variance of return***

The return from the portfolio in (i)(a) is guaranteed to be 5% and so the variance of return is 0. [1]

The return from the portfolio in (i)(b) is:

$$0.15 \times 0.05 + 0.85 \times i_1 \quad [1]$$

The variance of return is therefore:

$$0 + 0.85^2 \times 0.02^2 = 0.000289$$ [1]

Solution X1.6

Course reference: The "variable" stochastic interest rate model is covered in Chapter 5.

(i) ***Mean***

Let:

$$S_{10} = \prod_{t=1}^{10} (1 + i_t)$$

where i_t $(t = 1, 2, \ldots, 10)$ are independent and identically distributed random variables.

The mean of the rate of return in each year, j, is:

$$E[i_t] = j = 0.04 \times 0.4 + 0.06 \times 0.2 + 0.08 \times 0.4 = 0.06$$ [1]

So:

$$E[1{,}000 S_{10}] = 1{,}000(1+j)^{10} = 1{,}000(1.06)^{10} = £1{,}790.85$$ [1]

(ii) ***Standard deviation***

The variance of the rate of return in each year, s^2, is:

$$\begin{aligned} \text{var}[i_t] = s^2 &= 0.04^2 \times 0.4 + 0.06^2 \times 0.2 + 0.08^2 \times 0.4 - 0.06^2 \\ &= 0.00032 \end{aligned}$$ [2]

The variance of the accumulation is:

$$\begin{aligned} \text{var}(1{,}000 S_{10}) &= 1{,}000^2 \,\text{var}[S_{10}] \\ &= 1{,}000^2 \left[\left((1+j)^2 + s^2 \right)^{10} - \left((1+j)^{10} \right)^2 \right] \\ &= 1{,}000^2 \left[\left(1.06^2 + 0.00032 \right)^{10} - (1.06)^{20} \right] \\ &= 1{,}000^2 \times 0.095633^2 \end{aligned}$$ [1½]

So the standard deviation is £95.63. [½]

Solution X1.7

Course reference: The portfolio theory terms covered by this question are introduced in Chapter 6.

(i)(a) ***Opportunity set***

The opportunity set is the set of all the possible combinations of mean and variance that are available to the investor by suitable choice of investment portfolios. [1]

(i)(b) ***Efficient frontier for a portfolio of risky assets***

The efficient frontier for a portfolio of risky assets is the set of efficient portfolios consisting solely of risky assets – *ie* those that yield an uncertain investment outcome. An efficient portfolio is one such that no other portfolio offers a higher expected return for the same (or lower) variance of investment return or equivalently a lower variance of investment return for the same (or higher) expected return. [1]

The efficient frontier turns out to be the upper "half" of the envelope curve (*ie* above the point of minimum variance) around the opportunity set. This is because investors are assumed to prefer more to less – and therefore to like a higher expected return – and to be risk-averse – and so dislike variance of returns. [½]

(i)(c) ***Indifference curves***

An indifference curve is a locus of combinations of points in expected return-variance space that yield the same expected utility to the investor and hence between which the investor is indifferent. They slope upwards for a risk-averse investor, because the investor requires an additional expected return to compensate for any additional risk (*ie* variance) if the expected utility is to remain unchanged. [1]

(i)(d) ***Optimal portfolio***

Within the context of mean-variance portfolio theory, the optimal portfolio is the portfolio that maximises the expected utility of the investor as a function of the mean and variance of investment return over a single time period. [1]

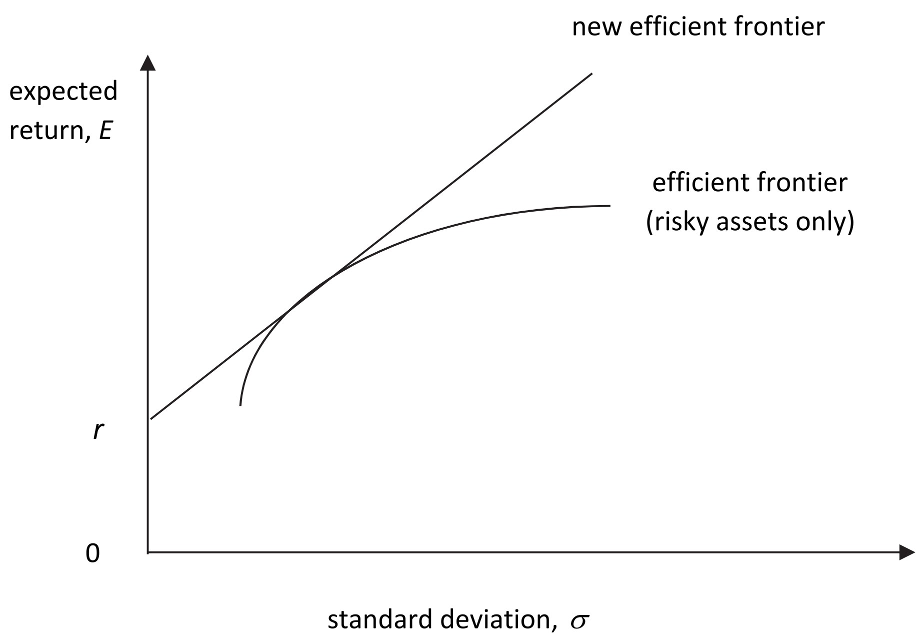
It arises where an indifference curve is tangential to the efficient frontier. [½]

[Total 5]

(ii) ***Efficient frontier with risk-free asset***

The effect on the efficient frontier of introducing a risk-free asset that can be bought or sold in unlimited quantities is to make the efficient frontier a straight line in mean-standard deviation space. (In mean-variance space it remains parabolic – but will be different to before.) [½]

In both cases, the efficient frontier intersects the expected return axis at the (certain) risk-free rate of return and is tangential to the efficient frontier of risky assets only. [½]



[1 for correct diagram]
[Total 2]

Solution X1.8

Course reference: The calculation of the variance of a portfolio is covered in Chapter 6.

(i) ***Calculate the correlation coefficient***

The correlation coefficient ρ_{AB} is defined as:

$$\rho_{AB} = \frac{Cov(A,B)}{\sigma_A \sigma_B}$$

where:

$$Cov(A,B) = E[AB] - E[A]E[B]$$

Using the data given:

$$E[A] = 6.0\% \qquad E[B] = 6.1\% \qquad E[AB] = 19\%\% \qquad [1]$$

So: $Cov(A,B) = 19 - 6.0 \times 6.1 = -17.6\%\%$ [½]

Also, using the data given:

$$\sigma_A = 2\%$$

$$\sigma_B = 9.0714\% \qquad [1]$$

So: $\rho_{AB} = \dfrac{-17.6}{2 \times 9.0714} = -0.970$ [½]

This shows that the returns are very strongly negatively correlated, so a poor return on one asset is very likely to be accompanied by a favourable return on the other. [1]

[Total 4]

(ii) ***Calculate the proportion in Asset A***

The portfolio variance is given by:

$$V_P = x_A^2 \sigma_A^2 + \left(1 - x_A\right)^2 \sigma_B^2 + 2x_A\left(1 - x_A\right)Cov(A,B) \qquad [½]$$

where x_A is the portfolio proportion invested in Asset A. Differentiating this expression with respect to x_A and setting it equal to zero gives the following condition for a minimum:

$$\frac{\partial V_P}{\partial x_A} = 2x_A\sigma_A^2 - 2\left(1 - x_A\right)\sigma_B^2 + 2\left(1 - 2x_A\right)Cov(A,B) = 0 \qquad [½]$$

Cancelling out the 2's and rearranging then gives:

$$x_A = \frac{\sigma_B^2 - Cov(A,B)}{\sigma_A^2 + \sigma_B^2 - 2Cov(A,B)}$$ [½]

If we substitute in the values calculated in (i) we obtain the value of x_A that minimises the portfolio variance:

$$x_A = \frac{9.0714^2 + 17.6}{2^2 + 9.0714^2 + 2 \times 17.6} = 0.8222$$ [½]

ie 82.22% of the portfolio should be invested in Asset A.

[Total 2]

(iii) ***Calculate the range of values of the correlation coefficient***

As before, the portfolio variance is given by:

$$V_P = x_A^2 \sigma_A^2 + \left(1 - x_A\right)^2 \sigma_B^2 + 2x_A\left(1 - x_A\right)Cov(A,B)$$ [1]

If we substitute in all the actual values given – including $x_A = 0.8$ – except for the correlation coefficient ρ_{AB}, then this becomes:

$$V_P = 0.8^2 \times 2^2 + 0.2^2 \times 9.0714^2 + 2 \times 0.8 \times 0.2 \times 2 \times 9.0714 \times \rho_{AB}$$ [1]

This simplifies to:

$$V_P = 5.85161 + 5.805696\rho_{AB}$$ [½]

We require that this portfolio variance be less than the variance of Asset A, *ie*:

$$5.85161 + 5.805696\rho_{AB} < 2^2$$ [½]

This in turn requires that:

$$\rho_{AB} < -0.319$$

So the portfolio will have a lower variance provided that ρ_{AB} lies in the range $(-1, -0.319)$. [1]

[Total 4]

Solution X1.9

Course reference: The assumptions underlying portfolio theory are listed in Chapter 6, in which the Core Reading also defines the efficient frontier and outlines how it may be derived using the Lagrangian approach.

(i) ***MVPT assumptions***

- All expected returns and variances of return are known for each individual asset, and covariances of return are known for each pair of assets. [½]
- Investors make decisions purely on the basis of the expected return and variance of return of their portfolio. [½]
- Investors are non-satiated, *ie* they prefer more to less. [½]
- Investors are risk-averse. [½]
- There is a single one-step time horizon. [½]
- Assets may be held in any amounts. [½]
- There are no taxes or transaction costs. [½]

[Maximum 3]

(ii) ***Equation of efficient frontier***

The expected return and variance of the portfolio are:

$$E = x_A E_A + (1 - x_A) E_B$$

$$\Leftrightarrow \quad x_A = \frac{E - E_B}{E_A - E_B} \qquad [½]$$

and $$\sigma^2 = x_A{}^2 \sigma_A{}^2 + x_B{}^2 \sigma_B{}^2 + 2 x_A x_B \sigma_A \sigma_B \rho_{AB}$$

$$= \left(x_A \sigma_A + x_B \sigma_B \right)^2 \quad \text{since } \rho_{AB} = 1 \qquad [1]$$

This gives:

$$\sigma = x_A \sigma_A + x_B \sigma_B$$

$$= \frac{E - E_B}{E_A - E_B} \sigma_A + \frac{E_A - E}{E_A - E_B} \sigma_B = \frac{6}{8}(E - 5) + \frac{2}{8}(13 - E)$$

$$= \frac{1}{2} E - \frac{1}{2} \qquad [1]$$

This is a straight line, which is what we would expect since the returns on A and B are perfectly correlated. The actual return R_B on Asset B will be a linear function of R_A, the actual return on Asset A. So both the mean and the standard deviation will be linearly related to the asset proportions. [½]

[Total 3]

(iii) ***Method with more than two assets***

Here we need to minimise the portfolio variance subject to the two constraints of specifying the expected portfolio return and having a fully-invested portfolio. To do this, we use a Lagrangian function:

$$W = \sum_i \sum_j x_i x_j C_{ij} - \lambda(\sum_i x_i E_i - E) - \mu(\sum_i x_i - 1)$$ [½]

We can differentiate this with respect to x_k, μ and λ:

$$\frac{\partial W}{\partial x_k} = 2\sum_j x_j C_{jk} - \lambda E_k - \mu\,, \quad k = 1,\ldots,n$$

$$\frac{\partial W}{\partial \lambda} = E - \sum_i x_i E_i$$

$$\frac{dW}{d\mu} = 1 - \sum_i x_i$$ [1]

Setting these to zero gives $n+2$ simultaneous equations in $n+2$ unknowns. [½]

We then solve these and substitute the solutions for the x_i's as functions of the portfolio mean E back into the expression for the variance above. [½]

This gives us the efficient frontier in (E,V) space. [½]

[Total 3]

Solution X1.10

Course reference: Behavioural finance appears in Chapter 3 of the Course Notes.

(i) ***Principal theme of behavioural finance***

The field of behavioural finance looks at how a variety of mental biases and decision-making errors affect financial decisions. It relates to the psychology that underlies and drives financial decision-making behaviour. [1]

Behavioural finance challenges the assumption of expected utility theory that decisions are made on the basis of maximising the expected value of utility under the investor's particular beliefs about the probability of different outcomes. [1]

[Total 2]

(ii) ***Prospect theory***

This is a theory of how individuals make decisions when faced with risk and uncertainty. It replaces the conventional risk-averse / decreasing marginal utility theory based on total wealth with a concept of value defined in terms of gains and losses relative to a reference point. [1]

Individuals are assumed to be risk-averse when facing gains relative to the reference point, but risk-seeking when facing losses relative to the reference point. [½]

This generates utility curves with a point of inflexion at the chosen reference point. [½]
[Total 2]

(iii) ***Describe the behavioural finance theme***

(a) This is an example of *anchoring and adjustment*. [½]

The saver's expectations with regard to future short-term interest rates are anchored by past experience of higher interest rates. Consequently, they expect interest rates to rise back to average historical levels even when there is no evidence that they are likely to do so. [1]

(b) This is an example of *status quo bias*. [½]

The investor is preferring to stick with their current situation, even though the alternative is more favourable. This behaviour can be driven by the fear of incurring a loss when switching to a new bank, even though there's no rational reason why Bank Y should be any more risky than Bank X. [1]
[Total 3]

Assignment X2 Solutions

Markers: This document sets out one approach to solving each of the questions. Please give credit for other valid approaches.

Solution X2.1

Course reference: Multifactor models are covered in Chapter 7.

(i) ***Multifactor model***

The multifactor model attempts to explain returns on assets by relating them to a series of n factors known as indices:

$R_i = a_i + b_{i,1}I_1 + \ldots + b_{i,n}I_n + c_i$ where: [1]

- a_i, c_i are the constant and random parts of the return, specific to asset i [½]
- $I_1, I_2, \ldots, I_n$ are the (changes in the) n factors explaining the returns on all the stocks [½]
- $b_{i,k}$ is the sensitivity of the return on stock i to factor k [½]
- $E[c_i] = 0$ [½]
- $cov[c_i, c_j] = 0$ for all $i \neq j$ [½]
- $cov[c_i, I_k] = 0$ for all stocks and indices. [½]

[Total 4]

(ii) ***Three types of factors***

1. *Macroeconomic* – the factors will include some macroeconomic variables such as interest rates, inflation, economic growth and exchange rates. [1]

2. *Fundamental* – the factors will be company specifics such as P/E ratios, liquidity ratios and gearing levels. [1]

3. *Statistical* – the factors do not necessarily have a meaningful interpretation. This is because they are derived from historical data, using techniques such as principal components analysis to identify the most appropriate factors. [2]

[Total 4]

Solution X2.2

Course reference: The capital asset pricing model (CAPM) is outlined in Chapter 8.

(i) ***What the separation theorem implies about optimal investment strategies***

The separation theorem states that, under the assumptions of CAPM, the optimal combination of risky assets for an investor can be determined without any knowledge of their preferences towards risk and return. [1]

This combination is the *market portfolio*, which consists of all risky assets held in proportion to their market capitalisation. [1]

[Total 2]

(ii) ***Why an investor wouldn't hold the market portfolio in practice***

In practice, the investor won't hold the market portfolio because:

- It is impossible to identify exactly the constituents of the market portfolio, as in principle it contains *all* risky assets including shares, property, bonds, commodities and human capital, many of which are never traded and have no market values. [1]

- Even if it were possible to identify exactly the constituents of the market portfolio, many of the different risky assets required (*eg* human capital) cannot be invested in, in practice. [1]

[Total 2]

(iii) ***Derive expected return on market portfolio and risk-free rate of return***

According to the CAPM *security market line*, the expected return on any portfolio, efficient or otherwise, including a single risky stock is given by:

$$E_p = r + \beta_p (E_m - r)$$

where:

- r is the risk-free rate of return
- E_m is the expected return on the market portfolio
- $\beta_p = \dfrac{Cov(R_p, R_m)}{Var(R_m)}$ [1]

For ABC shares we have:

(1) $8.5 = r + 0.7 \times (E_m - r) = 0.3r + 0.7E_m$ [½]

whereas for the 100-share portfolio:

(2) $10.5 = r + 1.1 \times (E_m - r) = -0.1r + 1.1E_m$ [½]

(1) + 3×(2) gives:

$40 = 4 \times E_m, \quad \Rightarrow \quad E_m = 10\%$ [½]

Substituting this back into (1) gives:

$8.5 = 0.3r + 0.7 \times 10, \quad \Rightarrow \quad r = 5\%$ [½]

[Total 3]

(iv) ***Discuss the student's comment***

The student's reasoning is flawed because according to CAPM, what counts in the determination of expected returns for inefficient portfolios such as we have here is β and not the standard deviation. [1]

The student believes that the higher standard deviation of ABC (not surprisingly higher than the standard deviation of the well-diversified, 100-stock portfolio) justifies a higher return for ABC. [1]

In fact, ABC has a lower β than that of the 100-share portfolio and so, r and E_m being the same for both, the return on ABC should be lower. [1]

[Total 3]

Solution X2.3

Course reference: Ito's Lemma is stated in Chapter 10 and also appears on page 46 in the Tables.

(i) ***Derive Ito's Lemma***

Ito's Lemma states that if $f(X_t)$ is a twice continuously differentiable function with respect to X_t and $dX_t = a_t\,dt + b_t\,dB_t$ (where a_t and b_t are F_t-measurable), then $f(X_t)$ satisfies:

$$df(X_t) = \left[a_t f'(X_t) + ½\, b_t^2 f''(X_t)\right]dt + b_t f'(X_t)\,dB_t$$ [½]

This can be derived as follows using Taylor's formula and ignoring higher than second-order terms:

$$df(X_t) = f'(X_t)dX_t + ½ f''(X_t)(dX_t)^2$$

$$= f'(X_t)(a_t dt + b_t\,dB_t) + ½ f''(X_t)(a_t\,dt + b_t\,dB_t)^2$$ [½]

Since:

- $(dt)^2 = 0$
- $dB_t \times dt = 0$
- $(dB_t)^2 = dt$

from the multiplication grid for increments in the Core Reading:

$$df(X_t) = f'(X_t)(a_t dt + b_t\,dB_t) + ½ f''(X_t) b_t^2 dt$$

$$= \left[a_t f'(X_t) + ½\, b_t^2 f''(X_t)\right]dt + b_t f'(X_t)\,dB_t$$ [1]

[Total 2]

(ii) ***Use Ito's Lemma to derive the stochastic differential equation for*** X_t

Here X_t is a function of B_t and time *t*:

$$X_t = f(B_t, t) = 0.05t + 0.10B_t$$

So:

$$\frac{\partial f}{\partial t} = 0.05 \qquad [½]$$

$$\frac{\partial f}{\partial B_t} = 0.10 \qquad [½]$$

$$\frac{\partial^2 f}{\partial B_t^2} = 0 \qquad [½]$$

Note that as X_t *is a direct function of standard Brownian motion* B_t*, we need to derive the drift and volatility terms* $a(t, X_t)$ *and* $b(t, X_t)$ *(using the notation on page 46 in the Tables) from the stochastic differential equation for* B_t*. As we can write:*

$$dB_t = 0 \times dt + 1 \times dB_t$$

We have:

$$a(t, X_t) = 0 \qquad [¼]$$

$$b(t, X_t) = 1 \qquad [¼]$$

So, substituting all of these terms into Ito's Lemma from page 46 in the *Tables* gives:

$$dX_t = df(B_t, t) = \left[0 \times 0.10 + ½ \times 1^2 \times 0 + 0.05\right]dt + 1 \times 0.10\, dB_t$$

$$= 0.05dt + 0.10\, dB_t \qquad [1]$$

[Total 3]

Alternatively, using a Taylor Series expansion, we have:

$$df(B_t, t) = \frac{df}{\partial t}dt + \frac{\partial f}{\partial B_t}dB_t + ½\frac{\partial^2 f}{\partial B_t^2}dB_t^2 \qquad [½]$$

where the three partial derivatives are the same as above and so:

$$df(B_t, t) = 0.05\, dt + 0.10\, dB_t + ½ \times 0\, dB_t^2$$

ie $dX_t = 0.05\, dt + 0.10\, dB_t$ *[1]*

(iii) ***Derive the stochastic differential equation for the exotic derivative***

The value of the exotic derivative is given by:

$$G(X_t) = X_t^2$$

So:

$$\frac{\partial G}{\partial t} = 0 \qquad [½]$$

$$\frac{\partial G}{\partial X_t} = 2X_t \qquad [½]$$

$$\frac{\partial^2 G}{\partial X_t^2} = 2 \qquad [½]$$

And from the SDE for X_t (*and working in terms of the notation on page 46 in the Tables*) we have:

$$a(t, X_t) = 0.05 \qquad [¼]$$

$$b(t, X_t) = 0.10 \qquad [¼]$$

So, substituting these terms into Ito's Lemma from page 46 in the *Tables* gives:

$$dG(X_t) = \left[0.05 \times 2X_t + ½ \times 0.10^2 \times 2 + 0\right]dt + 0.10 \times 2X_t dB_t$$

$$= \left(0.10X_t + 0.10^2\right)dt + 0.20X_t dB_t \qquad [1]$$

[Total 3]

Alternatively, using a Taylor Series expansion, we have:

$$dG\left(X_t\right) = \frac{\partial G}{\partial X_t} dX_t + ½ \frac{\partial^2 G}{\partial X_t^2} dX_t^2$$

Note that there is no term in time t here, as G is an explicit function of X_t *only and not time t.*

Here the partial derivatives are the same as above and so:

$$dG\left(X_t, t\right) = 2X_t dX_t + ½ \times 2\, dX_t^2$$

Substituting in for dX_t *from part (i) gives:*

$$dG\left(X_t, t\right) = 2X_t\left(0.05dt + 0.10dB_t\right) + ½ \times 2 \times \left(0.05dt + 0.10dB_t\right)^2$$

Next multiplying the increments in the right-hand bracket and remembering that $(dt)^2 = 0$*,* $dB_t dt = 0$ *and* $(dB_t)^2 = dt$*, gives:*

$$dG(X_t, t) = 2X_t(0.05dt + 0.10dB_t) + ½ \times 2 \times 0.10^2 dt$$

ie $$dG(X_t, t) = (0.10X_t + 0.10^2)dt + 0.20X_t dB_t$$

Solution X2.4

Course reference: Chapter 9 covers the important properties of Brownian motion and martingales.

(i) ***Conditional expectations of B_t^2 and B_t^4***

We can evaluate these expectations by splitting B_t into the (independent) past and future components at time B_t *ie* by writing $B_t = B_s + (B_t - B_s)$. We also need to remember that the value of B_s is known at time s and that $B_t - B_s \sim N(0, t-s)$. We then get:

$$E[B_t^2 | F_s] = E[\{B_s + (B_t - B_s)\}^2 | F_s]$$

$$= E[B_s^2 + 2B_s(B_t - B_s) + (B_t - B_s)^2 | F_s]$$

$$= B_s^2 + 2B_s E[(B_t - B_s)] + E[(B_t - B_s)^2] \qquad [1]$$

We lose the filtration F_s since we have independent increments.

Now:

$$E[(B_t - B_s)^2] = \text{var}(B_t - B_s) + \{E[B_t - B_s]\}^2 = t - s + 0 = t - s$$

So:

$$E[B_t^2 | F_s] = B_s^2 + 2B_s \times 0 + (t-s)$$

$$= B_s^2 + t - s \qquad [1]$$

Similarly:

$$E[B_t^4 | F_s] = E[\{B_s + (B_t - B_s)\}^4 | F_s]$$

$$= E\left[B_s^4 + 4B_s^3(B_t - B_s) + 6B_s^2(B_t - B_s)^2 + 4B_s(B_t - B_s)^3 + (B_t - B_s)^4 | F_s\right]$$

[1]

Evaluating the expectations:

$$E[B_t^4 \mid F_s] = B_s^4 + 0 + 6B_s^2 E[(B_t - B_s)^2] + 0 + E[(B_t - B_s)^4]$$

$$= B_s^4 + 6(t-s)B_s^2 + 3(t-s)^2$$ [1]

[Total 4]

NB The very last term in the above is derived using the information given in the question.

(ii) ***Constructing a martingale based on B_t^4***

To find a martingale we can re-express the result we've just derived so that we have an expression involving B_t and t on the LHS and an identical expression involving B_s and s on the RHS.

The result based on B_t^4 can be written:

$$E[B_t^4 \mid F_s] = B_s^4 + 6tB_s^2 - 6sB_s^2 + 3(t-s)^2$$ [½]

We have also shown that:

$$E[B_t^2 \mid F_s] = B_s^2 + t - s$$

or $B_s^2 = E[B_t^2 \mid F_s] - (t-s)$ [½]

This allows us to express the $6tB_s^2$ term purely in terms of t to get:

$$E[B_t^4 \mid F_s] = B_s^4 + 6t\{E[B_t^2 \mid F_s] - (t-s)\} - 6sB_s^2 + 3(t-s)^2$$ [1]

Taking the $6t\,E[B_t^2 \mid F_s]$ over to the LHS gives:

$$E[B_t^4 - 6tB_t^2 \mid F_s] = B_s^4 - 6t(t-s) - 6sB_s^2 + 3(t-s)^2$$

$$= B_s^4 - 6sB_s^2 - 3t^2 + 3s^2$$ [1]

Taking the $-3t^2$ over to the LHS gives:

$$E[B_t^4 - 6tB_t^2 + 3t^2 \mid F_s] = B_s^4 - 6sB_s^2 + 3s^2$$

Since the functions take the same form on both sides, we have shown that $B_t^4 - 6tB_t^2 + 3t^2$ is a martingale. [1]

Strictly speaking, we should also check that the absolute expectation is bounded for all values of *t*, *ie* that $E[|B_t^4 - 6tB_t^2 + 3t^2|] < \infty$

To show this:

$$E\left[\left|B_t^4 - 6tB_t^2 + 3t^2\right|\right] \leq E\left[\left|B_t^4\right| + \left|6tB_t^2\right| + \left|3t^2\right|\right]$$

$$= E\left[B_t^4\right] + 6tE\left[B_t^2\right] + 3t^2$$

$$= 3t^2 + 6t \times t + 3t^2$$

$$= 12t^2 < \infty$$

[1]

[Total 5]

Solution X2.5

Course reference: Ito's Lemma is stated in Chapter 10 and also appears on page 46 in the Tables.

(i) ***Explain why the value of the US Dollar is likely to increase***

If $r_d > r_e$, then the short-term interest rate is higher in the US than in the Eurozone. This means that "hot money" is likely to flow from Euros to US Dollars in search of the higher interest rate. [1]

The consequent higher demand for US Dollars (and higher supply of Euros) on the currency market will lead the value of US Dollars to rise against the Euro. [1]

[Total 2]

(ii) ***Use Ito's Lemma to solve the stochastic differential equation (SDE)***

Note that we cannot integrate the stochastic differential equation for X_t directly, as to do so would involve the following Ito integral: $\int_0^t \sigma X_t dB_t$.

As X_t is being modelled as a geometric Brownian motion, we can solve its SDE by applying Ito's Lemma to:

$$f(X_t) = \log(X_t)$$ [½]

So:

$$\frac{\partial f}{\partial X_t} = \frac{1}{X_t}$$ [½]

$$\frac{\partial^2 f}{\partial X_t^2} = \frac{-1}{X_t^2}$$ [½]

Note that $\frac{\partial f}{\partial t} = 0$, as $f(X_t)$ does not depend directly on time t.

From the SDE for X_t (and working in terms of the notation on page 46 in the *Tables*) we have:

$$a(t, X_t) = (r_d - r_e)X_t \qquad [½]$$

$$b(t, X_t) = \sigma X_t \qquad [½]$$

So, substituting these terms into Ito's Lemma from page 46 in the *Tables* gives:

$$df(X_t) = \left[(r_d - r_e)X_t \frac{1}{X_t} - ½\sigma^2 X_t^2 \frac{1}{X_t^2} \right] dt + \sigma X_t \frac{1}{X_t} dB_t$$

$$= (r_d - r_e - ½\sigma^2)dt + \sigma dB_t \qquad [1]$$

Alternatively, using a Taylor Series expansion, we have:

$$df\left(X_t\right) = \frac{\partial f}{\partial X_t} dX_t + ½ \frac{\partial^2 f}{\partial X_t^2} dX_t^2 \,]$$

where the two partial derivatives are the same as above and so:

$$df\left(X_t\right) = \frac{1}{X_t} dX_t - ½ \frac{1}{X_t^2} dX_t^2$$

Substituting in the expression for dX_t given in the question and remembering that $(dt)^2 = 0$, $dB_t dt = 0$ and $(dB_t)^2 = dt$, gives:

$$df\left(X_t\right) = \frac{1}{X_t}\left((r_d - r_e)X_t dt + \sigma X_t dB_t\right) - ½ \frac{1}{X_t^2}\left(\sigma^2 X_t^2 dt\right)$$

$$= \left(r_d - r_e - ½\sigma^2\right)dt + \sigma dB_t$$

This can be integrated directly as follows:

$$\int_0^t df(X_s) = (r_d - r_e - ½\sigma^2)\int_0^t ds + \sigma \int_0^t dB_s$$

$$\Rightarrow \qquad f(X_t) - f(X_0) = (r_d - r_e - ½\sigma^2)(t - 0) + \sigma(B_t - B_0) \qquad [½]$$

Hence, remembering that $f(X_t) = \log(X_t)$ and $B_0 = 0$, this becomes:

$$\log(X_t) - \log(X_0) = (r_d - r_e - ½\sigma^2)t + \sigma B_t$$

ie $\qquad X_t = X_0 \exp\{(r_d - r_e - ½\sigma^2)t + \sigma B_t\}$ [1]

[Total 5]

(iii) ***Derive the stochastic differential equation for G_t***

Here the function we need to "differentiate" using Ito's Lemma is:

$$G_t = \frac{1}{X_t}$$

So:

$$\frac{\partial G_t}{\partial X_t} = \frac{-1}{X_t^2} \qquad [½]$$

$$\frac{\partial^2 G_t}{\partial X_t^2} = \frac{2}{X_t^3} \qquad [½]$$

And from the SDE for X_t (and working in terms of the notation on page 46 in the *Tables*) we again have:

$$a(t, X_t) = (r_d - r_e)X_t \qquad [½]$$

$$b(t, X_t) = \sigma X_t \qquad [½]$$

So, substituting these terms into Ito's Lemma from page 46 in the *Tables* gives:

$$dG_t(X_t) = \left[(r_d - r_e)X_t \frac{-1}{X_t^2} + ½\sigma^2 X_t^2 \frac{2}{X_t^3}\right]dt + \sigma X_t \frac{-1}{X_t^2} dB_t$$

$$= (\sigma^2 + r_e - r_d)\frac{1}{X_t}dt - \sigma \frac{1}{X_t} dB_t \qquad [1]$$

Or, remembering that $G_t = \dfrac{1}{X_t}$, this becomes:

$$dG_t = (\sigma^2 + r_e - r_d)\,G dt - \sigma G dB_t \qquad [1]$$

So, G_t, which represents the value of the Euro in terms of US Dollars, is also a geometric Brownian motion, but with drift $(\sigma^2 + r_e - r_d)$ and volatility parameter, $-\sigma$.

[1]
[Total 5]

Alternatively, using a Taylor Series expansion, we have:

$$dG_t(X_t) = \frac{\partial G_t}{\partial X_t} dX_t + ½ \frac{\partial^2 G_t}{\partial X_t^2} dX_t^2$$

where the two partial derivatives are the same as above. So, substituting in the expression for dX_t *given in the question and remembering that* $(dt)^2 = 0$*,* $dB_t dt = 0$ *and* $(dB_t)^2 = dt$*, gives:*

$$dG_t(X_t) = \frac{-1}{X_t^2}\left((r_d - r_e)X_t dt + \sigma X_t dB_t\right) + ½ \frac{2}{X_t^3}\left(\sigma^2 X_t^2 dt\right)$$

$$= \frac{1}{X_t}\left((\sigma^2 + r_e - r_d)\, dt - \sigma dB_t\right)$$

ie $$dG_t(X_t) = \left(\sigma^2 + r_e - r_d\right) G_t dt - \sigma G_t dB_t$$

Solution X2.6

Course reference: The defining properties of standard Brownian motion are stated in Chapter 9, which also describes geometric Brownian motion. Ito's Lemma is stated in Chapter 10 and on page 46 in the Tables.

(i)(a) ***Defining properties of standard Brownian motion***

Standard Brownian motion is a continuous-time stochastic process with state space $\mathbb{R}$. [½]

It has:

- $B_0 = 0$ [½]
- independent increments [½]
- stationary increments [½]
- continuous sample paths [½]
- The distribution of the increments $B_t - B_s$ $(0 \le s < t)$ is given by:

$$B_t - B_s \sim N(0, t-s)$$ [½]

[Total 3]

(i)(b) ***Probability density for general Brownian motion***

The distribution of the increments for the general process is given by:

$$W_t - W_s \sim N[\mu(t-s), \sigma^2(t-s)]$$ [1]

which has probability density function:

$$\frac{1}{\sqrt{2\pi\sigma^2(t-s)}}\exp\left[-\frac{\{x-\mu(t-s)\}^2}{2\sigma^2(t-s)}\right], \ -\infty<x<\infty$$ [1]

[Total 2]

(ii) ***Solving the stochastic differential equation***

The question tells us to consider the function $f(S_t)=\log S_t$.

Now:

- $\frac{\partial f}{\partial t}=0$ [½]
- $\frac{\partial f}{\partial S_t}=\frac{1}{S_t}$ [½]
- $\frac{\partial^2 f}{\partial S_t^2}=\frac{-1}{S_t^2}$ [½]

and from the stochastic differential equation for S_t given in the question, the drift function is $a=\mu S_t$ and the volatility function is $b=\sigma S_t$.

Substituting these five terms in Ito's Lemma from page 46 in the *Tables* gives:

$$df=\left[\mu S_t\frac{1}{S_t}-½\sigma^2 S_t^2\frac{1}{S_t^2}+0\right]dt+\sigma S_t\frac{1}{S_t}dB_t$$

$$=\left(\mu-½\sigma^2\right)dt+\sigma dB_t$$ [1½]

Alternatively, by ***Taylor's theorem*** *(see page 3 in the Tables), the stochastic differential equation for this process is:*

$$d(\log S_t)=\frac{1}{S_t}dS_t+\frac{1}{2}\left(-\frac{1}{S_t^2}\right)(dS_t)^2$$

Substituting in for dS_t *and simplifying:*

$$d(\log S_t)=\frac{1}{S_t}(\mu S_t\,dt+\sigma S_t\,dB_t)-\frac{1}{2S_t^2}(\mu S_t\,dt+\sigma S_t\,dB_t)^2$$

Now, from the multiplication grid for increments in the Core Reading:

- $(dt)^2=0$
- $dB_t\,dt=0$
- $(dB_t)^2=dt$

So:

$$d(\log S_t) = \mu\,dt + \sigma\,dB_t - \tfrac{1}{2}\sigma^2\,dt$$
$$= (\mu - \tfrac{1}{2}\sigma^2)dt + \sigma\,dB_t$$

Integrating this equation between limits of $s=0$ and $s=t$, we get:

$$\int_0^t d\log S_s = (\mu - \tfrac{1}{2}\sigma^2)\int_0^t ds + \sigma\int_0^t dB_s \qquad [1]$$

$$\Rightarrow \qquad \log S_t - \log S_0 = (\mu - \tfrac{1}{2}\sigma^2)t + \sigma B_t$$

$$\Rightarrow \qquad S_t = S_0\,e^{(\mu - ½\sigma^2)t + \sigma B_t} \qquad [1]$$

[Total 5]

(iii) ***Probability that the share price will exceed 110 (at the end of the period)***

We need to calculate:

$$P(S_{6/12} > 110 \mid S_0 = 100) \qquad [½]$$

Using the result from (ii), this gives:

$$P\left(\frac{S_{1/2}}{S_0} > \frac{11}{10}\right) = P\left(e^{\sigma B_{1/2} + \frac{1}{2}(\mu - ½\sigma^2)} > \frac{11}{10}\right) \qquad [1]$$

This simplifies to:

$$P\left(\sigma B_{1/2} + \tfrac{1}{2}(\mu - ½\sigma^2) > \log\frac{11}{10}\right) = P\left(0.1B_{1/2} + \tfrac{1}{2}(0.2 - ½\times 0.1^2) > \log\frac{11}{10}\right)$$
$$= P\left(B_{1/2} > -0.022\right) \qquad [1]$$

Since $B_{1/2} \sim N\left(0, \tfrac{1}{2}\right)$, this is:

$$1 - \Phi\left(\frac{-0.022 - 0}{\sqrt{1/2}}\right) = \Phi\left(0.031\right) = 0.512 \quad \text{or } 51.2\% \qquad [1½]$$

[Total 4]

Solution X2.7

Course reference: The properties of the lognormal model are discussed in Chapter 11.

(i) ***Describe the lognormal model***

The continuous-time lognormal model of security prices assumes that log prices form a random walk in continuous-time. [½]

If S_t denotes the market price of an investment, then the model states that, for $T > t$, the log returns are modelled by the following distribution:

$$\log(S_T) - \log(S_t) \sim N\left[\mu(T-t), \sigma^2(T-t)\right] \quad [1]$$

where:

- μ is the parameter associated with the *drift*
- σ^2 is the parameter associated with the *volatility*. [½]

[Total 2]

(ii) ***Show that changes in X_t are stationary***

By definition:

$$X_t = \log S_t - \mu t \quad [½]$$

Thus:

$$\begin{aligned} X_{t+h} - X_t &= \log S_{t+h} - \mu(t+h) - \log S_t + \mu t \\ &= \log S_{t+h} - \log S_t - \mu h \\ &\sim N[0, \sigma^2 h] \end{aligned} \quad [1]$$

Since the distribution of $X_{t+h} - X_t$ depends only on σ, which is constant, and the time interval h, it must be stationary. [½]

[Total 2]

(iii) ***Mean and variance of share price***

Starting with the distribution of the log of the returns:

$$\log(S_T) - \log(S_t) \sim N\left[\mu(T-t), \sigma^2(T-t)\right]$$

the share price S_T is lognormally distributed with parameters $\mu(T-t) + \log(S_t)$ and $\sigma^2(T-t)$. [1]

From the *Tables* we have:

$$E\left[S_T\right]=\exp\left(\mu(T-t)+\log\left(S_t\right)+\tfrac{1}{2}\sigma^2\left(T-t\right)\right)$$
$$=S_t\exp\left(\mu(T-t)+\tfrac{1}{2}\sigma^2\left(T-t\right)\right)$$

[1½]

and:

$$Var\left(S_T\right)=\exp\left(2\mu(T-t)+2\log\left(S_t\right)+\sigma^2\left(T-t\right)\right)\left(\exp\left(\sigma^2\left(T-t\right)\right)-1\right)$$
$$=S_t^2\exp\left(2\mu(T-t)+\sigma^2\left(T-t\right)\right)\left(\exp\left(\sigma^2\left(T-t\right)\right)-1\right)$$

[1½]

[Total 4]

Solution X2.8

Course reference: Geometric Brownian motion is the same process as the lognormal model. The properties of the lognormal model and its suitability for modelling share prices are discussed in Chapter 11.

This model is a continuous-time random walk. Graphs of share prices do appear to have this form, with the price changing by a small 'random' amount from day to day. [½]

The RHS contains an S_t factor, which implies that prices changes are proportional to the current price. This is plausible since we would expect price movements to be based on percentage changes, not absolute changes. [½]

Empirical evidence suggests that the volatility parameter σ may not be constant over time, as estimates of volatility from past data are critically dependent on the time period chosen for the data and how often the estimate is re-parameterised. [1]

It can also be argued that the drift parameter μ may not be constant over time, as it is likely to vary with the level of bond yields. [1]

The underlying Brownian motion has normal increments. However, studies have shown that the distribution of log-share price increments has fatter tails and is more peaked than a normal distribution. [1]

Brownian motion also assumes continuous sample paths. However, share prices often exhibit sudden jumps both upwards and downwards. [1]

Brownian motion assumes independent increments. However the empirical evidence suggests a degree of dependence between the increments. [1]

In particular, there is some evidence of mean reversion, *ie* negative serial correlation, in the long term, although this is based largely on a small number of market crashes. [½]

In addition, daily movements appear to be subject to "momentum" effects, *ie* they are positively correlated. [½]

[Maximum 6]

Solution X2.9

Course reference: Chapter 8 describes the capital asset pricing model (CAPM).

(i) ***Beta***

According to the capital asset pricing model:

$$E_z = r + \beta_z(E_m - r)$$ [½]

where:

- E_z is the expected return on Portfolio Z, *ie* 7% or 0.07
- r is the risk-free rate of return, *ie 4% or 0.04*
- E_m is the expected return on the market as a whole, *ie* 8% or 0.08

Thus:

$$0.07 = 0.04 + \beta_z \times (0.08 - 0.04)$$

From which:

$$\beta_z = 0.75$$ [½]

[Total 1]

(ii) ***Standard deviation***

As both Portfolio Z is efficient *and* the market includes a risk-free security, the efficient frontier (here, the capital market line) is a straight line and the returns on any efficient portfolio are perfectly correlated with those of the market. Hence, $\rho_{zm} = 1$. [1]

Recalling the definition of beta:

$$\beta_z = \frac{\rho_{zm}\sigma_z\sigma_m}{\sigma_m^2}$$ [½]

We can substitute in the relevant values to obtain:

$$0.75 = \frac{1 \times \sigma_z \times 0.3}{0.3^2}$$

ie $\sigma_z = 0.225$ or 22.5% *pa* [½]

[Total 2]

Alternatively, you can note that, since Z is efficient and the assumptions of CAPM apply, Portfolio Z must consist of 25% risk-free asset and 75% market portfolio.

ie $R_Z = 0.25r + 0.75R_m$ (*)

$\Rightarrow$ $Var[R_Z] = 0.75^2 V_m$

$\Rightarrow$ $\sigma_Z = 0.75\sigma_m = 0.75 \times 30\% = 22.5\% \ pa$

(iii) ***Systematic and specific risk***

Portfolio Z is efficient. Consequently it can have no specific risk – this having been diversified away. Hence, the entire standard deviation of 22.5% is attributable entirely to systematic risk. (This is immediately apparent from equation (*) above, since the only source of randomness is R_m.) [1]

(iv) ***Market value of Portfolio Z***

If Portfolio Z makes a single payment of $100 in one period's time, then its market value must be given by:

$PV(\mathrm{Z}) = 100v$ @ $i = 7\%$

ie $PV(\mathrm{Z}) = \$93.46$ [1]

Assignment X3 Solutions

Markers: This document sets out one approach to solving each of the questions. Please give credit for other valid approaches.

Note that some of the numerical answers are sensitive to rounding.

Solution X3.1

Course reference: The Greeks are discussed in Chapter 13.

(i) ***The Greeks***

Delta measures the sensitivity of the derivative price to small changes in the price of the underlying security *S*, all else being equal, *ie*:

$$\Delta = \frac{\partial f}{\partial S}$$

It is positive for a call option and negative for a put option. [1]

Gamma measures the *curvature* or convexity of the relationship between the derivative price and the price of the underlying security, all else being equal, *ie*:

$$\Gamma = \frac{\partial \Delta}{\partial S},$$

or: $$\Gamma = \frac{\partial^2 f}{\partial S^2}$$

It has a positive value for both put and call options. [1]

Vega measures the sensitivity of the derivative price to small changes in the assumed level of volatility (σ) of the underlying security price, all else being equal, *ie*:

$$\nu = \frac{\partial f}{\partial \sigma}$$

Vega is positive for both options. [1]

Rho measures the sensitivity of the derivative price to small changes in the risk-free rate of interest *r*, all else being equal, *ie*:

$$\rho = \frac{\partial f}{\partial r}$$

It is positive for a call option and negative for a put option. [1]

Theta measures the rate of change of the derivative price with respect to time, t. Here t measures time since the start of the contract and not the outstanding time to expiry.

$$\theta = \frac{\partial f}{\partial t}$$

Theta is generally negative for vanilla options. [1]

[Total 5]

NB Theta may be positive for a deeply in-the-money European put option.

(ii) ***Estimate the new price of option***

Because delta measures the rate of increase in the option price with changes in the share price, we expect that if the share price increases by 50p, then the option price will increase by:

$$50 \times 0.532 = 26.6p\ .$$

Using the same logic for all the other Greeks, we see that the change in the option price is:

$$df = f' - f \ = \Delta dS + \tfrac{1}{2}\Gamma dS^2 + V d\sigma + \rho dr + \theta dt \qquad [2]$$

where S, σ, r and t are defined as in (i) above.

The new option price f can be estimated from the original theoretical value, f, using the relationship:

$$f' = f + df \qquad [1]$$

Thus, substituting in the relevant values gives:

$$\begin{aligned} df' &= 0.532 \times 50 + \tfrac{1}{2} \times 0.00578 \times 50^2 + 0.91 \times 5 + 0.33 \times (-\tfrac{1}{2}) - 0.499 \times 1 \\ &= 37.7 \end{aligned} \qquad [1]$$

So, the new option price is approximately:

$$28.7 + 37.7 = 66.4p \qquad [1]$$

[Total 5]

Solution X3.2

Course reference: Put-call parity is described in Chapter 12 and the formula also appears on page 47 in the Tables. The Greeks are introduced in Chapter 13.

(i) ***Value for*** Δ

The put-call parity relationship is given in the *Tables* on page 47:

$$c_t + Ke^{-r(T-t)} = p_t + S_t e^{-q(T-t)}$$

Differentiating once with respect to S_t, we obtain:

$$\Delta_c = \Delta_p + e^{-q(T-t)} \Rightarrow n = e^{-q(T-t)}$$ [1]

(ii) ***Derive relationships for the other five Greeks***

Starting from the result in (i), we differentiate once again with respect to S_t, to obtain:

$$\Gamma_c = \Gamma_p$$ [½]

Starting from the original put-call parity relationship, we differentiate once with respect to r to obtain:

$$\rho_c - (T-t)Ke^{-r(T-t)} = \rho_p$$ [½]

Starting from the original put-call parity relationship, we differentiate once with respect to q to obtain:

$$\lambda_c = \lambda_p - (T-t)S_t e^{-q(T-t)}$$ [½]

Starting from the original put-call parity relationship, we differentiate once with respect to σ to obtain:

$$V_c = V_p$$ [½]

Starting from the original put-call parity relationship, we differentiate once with respect to t to obtain:

$$\theta_c + rKe^{-r(T-t)} = \theta_p + qS_t e^{-q(T-t)}$$ [1]

[Total 3]

(iii) ***Decide whether the relationship holds***

Using the relationships derived in (ii) we see that:

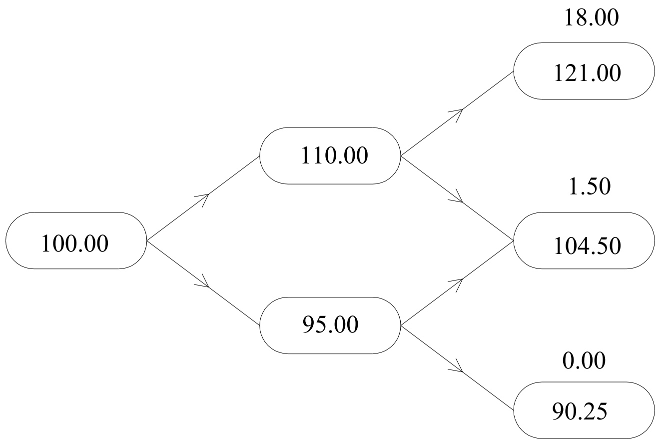
$$\begin{aligned} r\rho_c + q\lambda_c + (T-t)\theta_c &= r(T-t)Ke^{-r(T-t)} + r\rho_p + q\lambda_p - q(T-t)S_t e^{-q(T-t)} \\ &\quad + (T-t)\theta_p + (T-t)qS_t e^{-q(T-t)} - (T-t)rKe^{-r(T-t)} \\ &= r\rho_p + q\lambda_p + (T-t)\theta_p \end{aligned}$$

Therefore the relationship holds. [2]

Solution X3.3

Course reference: Valuing derivatives using binomial trees is described in Chapter 14.

The binomial tree looks as follows, with share prices within the boxes and the option payoffs at expiry above the right-hand boxes.



[1 for correct tree]

Given that an up-movement occurs over the first time period, a portfolio of ϕ_{11} shares and ψ_{11} cash set up a time 1 will replicate the value of the call option at time 2 if:

$$S_{11}\,u\,\phi_{11} + \psi_{11}\,e^{r} = c_{21}$$

and $$S_{11}\,d\,\phi_{11} + \psi_{11}\,e^{r} = c_{22}$$

Substituting in the relevant numerical values then gives:

$$121.00\,\phi_{11} + \psi_{11}\,e^{0.04} = 18 \quad \text{and} \quad 104.50\,\phi_{11} + \psi_{11}\,e^{0.04} = 1.5$$ [1]

Solving these equations simultaneously gives:

$$\phi_{11} = 1 \quad \text{and} \quad \psi_{11} = -98.961$$ [1]

Thus, the value of the replicating portfolio – and hence the call option – at the upper node at time 1 is given by:

$$\phi_{11}\,S_{11} + \psi_{11} = 1 \times 110 - 98.961 = 11.039$$ [1]

Given that a down-movement occurs over the first time period, a portfolio of ϕ_{12} shares and ψ_{12} cash set up at time 1 will replicate the value of the call option at time 2 if:

$$104.50\,\phi_{12} + \psi_{12}\,e^{0.04} = 1.5$$

and $$90.25\,\phi_{12} + \psi_{12}\,e^{0.04} = 0$$ [1]

Solving these equations simultaneously gives:

$$\phi_{12} = 0.10526 \quad \text{and} \quad \psi_{12} = -9.1275$$ [1]

Thus, the value of the replicating portfolio – and hence the call option – at the lower node at time 1 is given by:

$$\phi_{12}\,S_{12}+\psi_{12} \;=\; 0.10526\times 95-9.1275 \;=\; 0.872$$ [1]

Finally, a portfolio of ϕ_{01} shares and ψ_{01} cash set up a time 0 will replicate the value of the call option at time 1 if:

$$110.00\,\phi_{01}+\psi_{01}\,e^{0.04} \;=\; 11.039$$

and $$95.00\,\phi_{01}+\psi_{01}\,e^{0.04} \;=\; 0.872$$ [1]

Solving these equations simultaneously gives:

$$\phi_{01} \;=\; 0.6777$$

$$\psi_{01} \;=\; -61.023$$ [1]

Thus, the value of the replicating portfolio – and hence the call option – at time 0 is given by:

$$\phi_{01}\,S_{01}+\psi_{01} \;=\; 0.6777\times 100-61.023 \;=\; 6.75$$ [1]

[Total 10]

Solution X3.4

Course reference: Replicating portfolios, the risk-neutral probability measure and the use of one-period binomial trees are all introduced in Chapter 14.

(i) *Replicating portfolio*

A *replicating portfolio* is one that reproduces the payoffs of a derivative for all possible outcomes of the value of the underlying asset. [½]

It is usually constructed out of a suitable combination of the underlying asset and risk-free cash. [½]

In an arbitrage-free world it must have the same value as the derivative it replicates. [1]

[Total 2]

(ii) *Risk-neutral probability measure*

A *risk-neutral probability measure* is a synthetic set of probabilities under which the expected values of the relevant assets accumulate at the risk-free rate of interest, *ie* at the rate of return that would apply if all investors were risk-neutral. [1]

If the risk-neutral probability measure for the underlying asset is Q and the risk-free interest rate (continuously compounded) is r, then the value at time t of a derivative with payoff X_T at time T is:

$$V_t = e^{-r(T-t)}E_Q[X_T \mid F_t]$$

This can be used to calculate a derivative price in an arbitrage-free world. [1]

[Total 2]

(iii)(a) ***Derive an equation for the price***

Let's set up a replicating portfolio at the start of the period consisting of ϕ units of the underlying share and ψ units of cash. [½]

Equating the values of the portfolio at the end of the period to the corresponding derivative payoffs leads to the equations:

$$\phi S_0 u + \psi e^r = c_u$$

$$\phi S_0 d + \psi e^r = c_d$$ [1]

We can solve these equations by subtracting to find ϕ, and by first multiplying the equations by d and u respectively then subtracting to find ψ. This gives:

$$\phi = \frac{c_u - c_d}{S_0(u-d)} \quad \text{and} \quad \psi = \frac{uc_d - dc_u}{e^r(u-d)}$$ [1]

The value at the start of the period of the replicating portfolio is $\phi S_0 + \psi$. So the value of the derivative (in an arbitrage-free world) must be:

$$c = \phi S_0 + \psi = \frac{c_u - c_d}{S_0(u-d)} S_0 + \frac{uc_d - dc_u}{e^r(u-d)} = \frac{c_u - c_d}{(u-d)} + \frac{uc_d - dc_u}{e^r(u-d)}$$ [1½]

Note that this can be written in an alternative form involving risk-neutral probabilities:

$$c = e^{-r}\left\{\frac{e^r - d}{u-d} c_u + \frac{u - e^r}{u-d} c_d\right\}$$

or $\quad c = e^{-r}\{qc_u + (1-q)c_d\}$ *where* $q = \dfrac{e^r - d}{u-d}$

(iii)(b) ***Price of derivative***

For this derivative the payoffs are:

$$c_u = |S_0 u - S_0| = S_0(u-1)$$

and $\quad c_d = |S_0 d - S_0| = S_0(1-d)$ [1]

The replicating portfolio consists of ϕ shares, where:

$$\phi = \frac{c_u - c_d}{S_0(u-d)} = \frac{S_0(u-1) - S_0(1-d)}{S_0(u-d)} = \left(\frac{u+d-2}{u-d}\right)$$

and an amount ψ in cash, where:

$$\psi = \frac{uc_d - dc_u}{e^r(u-d)} = \frac{uS_0(1-d) - dS_0(u-1)}{e^r(u-d)} = \left(\frac{u+d-2ud}{u-d}\right)S_0e^{-r}$$

So we find that:

$$c = \psi S_0 + \phi = \left(\frac{u+d-2}{u-d}\right)S_0 + \left(\frac{u+d-2ud}{u-d}\right)S_0e^{-r} \qquad [1]$$

Alternatively, using the formula derived in part (iii)(a):

$$c = e^{-r}\left\{\frac{e^r - d}{u-d}(u-1) + \frac{u-e^r}{u-d}(1-d)\right\}S_0$$

(iii)(c) ***How to synthesise the derivative***

This derivative pays the absolute value of the difference between S_1, the stock price at time 1, and the original stock price S_0.

Consider a portfolio consisting of a call option and a put option, both with exercise price equal to S_0. [1]

If $S_1 > S_0$ the payoff from the call option will be the difference $S_1 - S_0 = |S_1 - S_0|$ and the put will have zero payoff. [½]

If $S_1 < S_0$ the payoff from the put option will be the difference $S_0 - S_1 = |S_1 - S_0|$ and the call will have zero payoff. [½]

So this portfolio would reproduce the payoffs.

[Total 8]

Solution X3.5

Course reference: The Black-Scholes formula is stated in Chapter 15.

The Black-Scholes formula is given on page 47 of the *Tables* (it is the Garman-Kohlhagen formula with $q=0$) as:

$$c = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

where:

$$d_1 = \frac{\ln(S/K) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T-t}$$

Consider $d_1 = \dfrac{\ln(S/K)+r(T-t)}{\sigma\sqrt{T-t}} + ½\sigma\sqrt{T-t}$. As $\sigma \to 0$, the second term vanishes.

The numerator in the first term does not depend on σ. So, as $\sigma \to 0$:

(1) if $\ln(S/K)+r(T-t)>0$ then $d_1 \to +\infty$ [1]

(2) if $\ln(S/K)+r(T-t)<0$ then the first term $d_1 \to -\infty$ [1]

(3) if $\ln(S/K)+r(T-t)=0$ then the first term $d_1 \to 0$ [1]

The behaviour of d_2 is the same as the behaviour of d_1 in all three cases. [½]

By taking the exponential of each of the above three conditions and rearranging, we see that they can be rewritten as:

(1) if $S-Ke^{-r(T-t)}>0$ then $\Phi(d_1)\to 1$ and $\Phi(d_2)\to 1$ [1]

(2) if $S-Ke^{-r(T-t)}<0$ then $\Phi(d_1)\to 0$ and $\Phi(d_2)\to 0$ [1]

(3) if $S-Ke^{-r(T-t)}=0$ then $\Phi(d_1)\to ½$ and $\Phi(d_2)\to ½$ [1]

So, the Black-Scholes formula $c = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$ implies that:

(1) if $S-Ke^{-r(T-t)}>0$ then $c \to S-Ke^{-r(T-t)}$ [½]

(2) if $S-Ke^{-r(T-t)}<0$ then $c \to 0$ [½]

(3) if $S-Ke^{-r(T-t)}=0$ then $c \to ½\left(S-Ke^{-r(T-t)}\right)$ [½]

However, given the condition $S-Ke^{-r(T-t)}=0$ we can see that in case (3) we actually have $c \to 0$

So it follows that $c \to \max\left(S-Ke^{-r(T-t)},0\right)$ as required. [1]

[Total 9]

Solution X3.6

Course reference: The valuation of derivatives using binomial trees is discussed in detail in Chapter 14.

(i)(a) ***Replicating portfolio of European call option***

We have two possibilities for the price at time 1:

$$S_1 = \begin{cases} S_0 u & \textit{if the price goes up} \\ S_0 d & \textit{if the price goes down} \end{cases}$$

We can hold an amount ϕ *of the stock, and amount* ψ *of cash at time 0 with the intention of replicating a derivative whose payoff at time 1 is* c_u *if the stock price goes up and* c_d *if the stock price goes down.*

We equate the value of the portfolio at time 1 with the payoffs on the call option at time 1 and solve the simultaneous equations:

$$\phi S_0 u + \psi e^r = c_u$$

$$\phi S_0 d + \psi e^r = c_d \qquad [1]$$

giving $$\psi = e^{-r}\left(\frac{c_d u - c_u d}{u-d}\right) \qquad [½]$$

and $$\phi = \frac{c_u - c_d}{S_0(u-d)} \qquad [½]$$

By the no-arbitrage principle, the value of this portfolio at time 0, V_0*, must also be the value of the derivative contract at that time.*

Finally, we are actually asked to replicate a European call with strike price of k. This implies that $c_u = uS_0 - k$ and $c_d = 0$ since we are told that $dS_0 < k < uS_0$. Substituting in our expressions for ϕ and ψ gives:

$$\phi = \frac{S_0 u - k}{S_0(u-d)} \qquad [½]$$

and $$\psi = e^{-r}\left(\frac{-(S_0 u - k)d}{u-d}\right) = -e^{-r}\left(\frac{S_0 u - k}{u-d}\right)d \qquad [½]$$

[Total 3]

Note that we could have alternatively worked with the explicit expressions for the payoff from the start. In this question it wouldn't have really mattered; perhaps the maths is slightly easier just dealing specifically with the call option, although at least with c_u and c_d you get answers that should be familiar from Core Reading. In addition, if the question had asked later for the corresponding expressions for a put, then the method presented here would have saved you a considerable amount of time. It often pays to read all the parts of a question before starting it.

(i)(b) ***Price of European call option***

By the no-arbitrage principle, the price of the European call option is given by:

$$V_0 = \phi S_0 + \psi \qquad [½]$$

Substituting the expressions for ϕ and ψ from (i)(a) into this gives:

$$V_0 = \phi S_0 + \psi = \frac{S_0 u - k}{S_0(u-d)} S_0 - e^{-r}\left(\frac{(S_0 u - k)d}{u-d}\right) = e^{-r}(S_0 u - k)\left(\frac{e^r - d}{u-d}\right)$$ [1]

[Total 1½]

(i)(c) ***Price of European put option using put-call parity***

Put-call parity says that for European calls and puts with the same strike price and date:

$$p_0 + S_0 = c_0 + ke^{-r}$$ [½]

So:

$$p_0 = c_0 + ke^{-r} - S_0$$

$$= e^{-r}(S_0 u - k)\left(\frac{e^r - d}{u-d}\right) + ke^{-r} - S_0$$

$$= \frac{e^{-r}}{u-d}\left[\cancel{S_0 ue^r} - S_0 ud - ke^r + \cancel{kd} + ku - \cancel{kd} - \cancel{S_0 ue^r} + S_0 de^r\right]$$

ie $$= e^{-r}(S_0 d - k)\left(\frac{e^r - u}{u-d}\right)$$ [1]

[Total 1½]

(i)(d) ***Price of European call under risk-neutral measure Q***

As in part (i(a)) we could work with the general case using c_u and c_d but it's quicker to use the explicit expressions because $c_d = 0$ simplifies matters.

We want to show that:

$$V_0 = \phi S_0 + \psi = e^{-r} E_Q(C \mid F_0) = e^{-r}\left(q(S_0 u - k)\right)$$

where *C* is the call option payoff at time 1 and *q* is the required (risk-neutral) probability of an upward stock price movement. [½]

Now, comparing the pricing formula for the call option in (i)(b) to:

$$e^{-r} E_Q(C) = e^{-r}\left(q(S_0 u - k)\right)$$

gives us:

$$q = \frac{e^r - d}{u-d}$$ [½]

[Total 1]

(ii) ***Risk-neutral versus real-world probabilities***

The probability measure *Q* was constructed in part (ii) so that the value of the derivative at time 0 was the discounted value, at the continuously compounded risk-free rate, of its expected payoff at time 1 using the risk-neutral probabilities q and $1-q$. [1]

Under the probability measure *Q* investors are therefore assumed to be risk-neutral, *ie* they demand no extra return from the stock even though it has higher risk (variance) than the cash. [½]

Under the real-world probability measure *P*, we would value the same derivative at time 0 as the discounted value (using a real-world risk discount rate) of its expected payoff at time 1, using the real-world probabilities p and $1-p$. [1]

The relationship of *Q* to the real-world probability measure *P* will depend on the preferences of investors. [½]

If, as is generally considered to be the case, investors are risk-averse, then the actual real-world probability *p* must be greater than the risk-neutral probability *q*. This must be so because the actual expected return must be higher than the risk-free rate to compensate them for the risk. [1]

If the investors are actually risk-neutral in the real world, then $p=q$. [½]

While if they are risk-seeking, then $p<q$. [½]

[Total 5]

Solution X3.7

Course reference: Valuing options using the Black-Scholes pricing formula is described in Chapter 15.

(i) ***Put-call parity relationship***

$$c_t + Ke^{-r(T-t)} = p_t + S_t \quad [1]$$

(ii) ***Derive a formula for the value of a put option***

Using put-call parity, we have:

$$\begin{aligned} p_t &= Ke^{-r(T-t)} - S_t + c \\ &= Ke^{-r(T-t)} - S_t + S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \end{aligned} \quad [1]$$

This can be simplified to:

$$\begin{aligned} p_t &= Ke^{-r(T-t)}(1-\Phi(d_2)) - S_t(1-\Phi(d_1)) \\ &= Ke^{-r(T-t)}\Phi(-d_2) - S_t\Phi(-d_1) \end{aligned} \quad [1]$$

[Total 2]

Notice that the bottom line answer is in the Tables and so you definitely need to show all your working here.

(iii) ***How many options can be bought?***

Working in pence, we have:

$$S_t = 2000\,,\ \sigma = 0.3\,,\ K = 1750$$

The one-year zero-coupon bond has a current value of £94.18. So:

$$e^{-r} = 0.9418 \Rightarrow r \approx 6\%\ pa \qquad [1]$$

We can now use this to calculate d_1 and d_2:

$$d_1 = \frac{\log\frac{2000}{1750} + (0.06 + \frac{0.3^2}{2})}{0.3} = 0.7951 \qquad [1]$$

$$d_2 = d_1 - 0.3 = 0.4951 \qquad [1]$$

So, the price of the put option is:

$$p_t = 1750e^{-0.06}\Phi(-0.4951) - 2000\Phi(-0.7951)$$

$$= 1750e^{-0.06} \times 0.3103 - 2000 \times 0.2133$$

$$= 84.79\,p \qquad [2]$$

For £100, the investor can purchase 117 options for 85 pence each, with 55 pence left over. [1]

[Total 6]

Solution X3.8

Course reference: The valuation of derivatives using binomial trees is discussed in detail in Chapter 14.

(i) ***Expected option payoff***

Given the strike price of £10.50, the payoff is £0.50 if the share price goes up, otherwise it is zero. So:

$$E[C_1] = 0.6 \times 0.5 + 0.4 \times 0 = £0.30 \qquad [1]$$

(ii) ***Replicating portfolio***

If we let ϕ be the number of shares held at time 0 and ψ be the amount of cash, then for replication we require:

$$11\phi + e^{0.04/12}\psi = 0.50$$ [½]

$$9.5\phi + e^{0.04/12}\psi = 0$$ [½]

Solving these equations gives:

$$\phi = \tfrac{1}{3} = 0.333$$ [½]

and $$\psi = -9.5 \times \frac{1}{3} \times e^{-0.04/12} = -3.156$$ [½]

The fair price is therefore:

$$10\phi + \psi = £0.18$$ [1]

[Total 3]

(iii) ***Number of options to construct risk-free portfolio***

Since one derivative can be replicated by ϕ shares and ψ cash, as we saw in part (ii), one derivative and $-\phi$ shares, or equivalently, $-\dfrac{1}{\phi}$ derivatives and one share must replicate the cash, *ie* they will form a risk-free portfolio.

We therefore need to sell $\dfrac{1}{\phi} = 3$ call options for every share owned to construct a risk-free portfolio. [1]

(iv) ***Arbitrage opportunity***

The discounted value of the expected payoff is $0.3e^{-0.04/12} = £0.29900$. [½]

This is greater than the fair price found in Part (ii), so we need to *sell* the options. [½]

By part (iii), a portfolio consisting of 1 share and -3 options will be risk-free. [½]

This portfolio will cost $10 - 3 \times 0.29900 = £9.103$. [1]

We borrow this amount in the cash market to have zero initial expenditure. [½]

At the end of the month you would owe $e^{0.04/12} \times 9.103 = £9.13$ … [½]

… but have a portfolio worth either:

$11 - 3 \times 0.5 = £9.50$ if the share price goes up [½]

or $9.50 - 3 \times 0 = £9.50$ if the share price goes down. [½]

Note the portfolio is constructed to be risk-free and hence the payoffs must be equal.

Either way, you make a (risk-free) profit of $9.50 - 9.13 = £0.37$ per share. [½]

[Maximum 4]

If your arbitrage portfolio involved selling one call option, the profit would be £0.12 per option. It is also possible to take the profit up-front and have a portfolio that exactly breaks even at the end of the month.

(v) ***Risk-neutral valuation of option***

The general risk-neutral valuation formula for the price at time *t* of a derivative that pays *X* at time *T* is:

$$V_t = e^{-r(T-t)} E_Q\left[X \middle| F_t\right]$$ [½]

where *r* is the risk-free force of interest and *Q* is the risk-neutral measure.

Here the binomial tree is recombining and:

$$q = \frac{e^r - d}{u - d} = \frac{e^{0.04/12} - 0.95}{1.1 - 0.95} = 0.35559$$ [½]

There are 5 possible share prices at time 4:

(1) 14.641

(2) 12.6445

(3) 10.92025

(4) 9.431125

(5) 8.1450625 [Maximum 1]

These occur with risk-neutral probabilities $\binom{4}{i} q^i (1-q)^{4-i}$:

(1) $q_1 = 0.015989$

(2) $q_2 = 0.11590$

(3) $q_3 = 0.31505$

(4) $q_4 = 0.380622$

(5) $q_5 = 0.172442$ [Maximum 1]

The corresponding payoffs for the derivative are:

(1) $X_1 = 2\times(14.641-12) = 5.282$ (exercise and buy 2 shares at £12)

(2) $X_2 = 2\times(12.6445-12) = 1.289$ (exercise and buy 2 shares at £12)

(3) $X_3 = 0$ (do not exercise)

(4) $X_4 = 0$ (do not exercise)

(5) $X_5 = 1\times(9-8.145063) = 0.854937$ (exercise and sell 1 share at £9)

[Maximum 1]

It follows that the fair price of the derivative is:

$$V_0 = e^{-4\times 0.04/12}\left(\sum_{i=1}^{5} X_i q_i\right) = £0.38$$ [½]

[Maximum 4]

Solution X3.9

Course references: The risk-neutral valuation of derivatives in continuous time is introduced in Chapter 17. The Greeks are discussed in Chapter 13.

(i) ***Risk-neutral pricing formula for the derivative***

The general risk-neutral formula for pricing a derivative at time $t < T$ is:

$$V_t = e^{-r(T-t)} E_Q\left[X_T \middle| F_t\right]$$ [½]

In this instance, the payoff function of the derivative is:

$$X_T = S_T(S_T - 1) = S_T^2 - S_T$$

So, the price will be given by:

$$V_t = e^{-r(T-t)} E_Q\left[S_T^2 - S_T \middle| F_t\right]$$ [1]

Given that the share price at maturity can take any value from zero upwards, this can be evaluated as:

(1) $$V_t = e^{-r(T-t)}\left[\int_0^\infty S_T^2\, f(S_T|S_t)\,dS_T - \int_0^\infty S_T\, f(S_T|S_t)\,dS_T\right]$$ [1]

where:

- $f(S_T|S_t)$ is the probability density function for S_T given S_t

- The assumption of independent increments underlying Brownian motion means we don't need to condition on the full past history of the share price.

Under the risk-neutral measure Q, the share price follows geometric Brownian motion with drift r and so:

$$\log S_T \sim N\left[\log S_t + \left(r - ½\sigma^2\right)(T-t), \sigma^2(T-t)\right]$$ [½]

So, using the formula for the truncated moments of a lognormal distribution on page 18 in the *Tables,* (with $k=2$ for the first integral and $k=1$ for the second integral) (1) above becomes:

$$V_t = e^{-r(T-t)}\left[\begin{array}{l}\left(e^{2(\log S_t + (r-½\sigma^2)(T-t)) + ½\times 4\sigma^2(T-t)}\right)\{\Phi(U_2)-\Phi(L_2)\} \\ -\left(e^{\log S_t + (r-½\sigma^2)(T-t) + ½\sigma^2(T-t)}\right)\{\Phi(U_1)-\Phi(L_1)\}\end{array}\right]$$

$$= e^{-r(T-t)}\left[S_t^2\left(e^{(2r+\sigma^2)(T-t)}\right)\{\Phi(U_2)-\Phi(L_2)\} - S_t\left(e^{r(T-t)}\right)\{\Phi(U_1)-\Phi(L_1)\}\right]$$

$$= S_t^2\left(e^{(r+\sigma^2)(T-t)}\right)\{\Phi(U_2)-\Phi(L_2)\} - S_t\{\Phi(U_1)-\Phi(L_1)\}$$

[1½]

Now, also from page 18 in the *Tables*:

- $U_2 = \dfrac{\log(\infty) - \log S_t - (r-½\sigma^2)(T-t)}{\sigma\sqrt{T-t}} - 2\sigma\sqrt{T-t} = \infty$ [¼]

- $L_2 = \dfrac{\log(0) - \log S_t - (r-½\sigma^2)(T-t)}{\sigma\sqrt{T-t}} - 2\sigma\sqrt{T-t} = -\infty$ [¼]

- $U_1 = \dfrac{\log(\infty) - \log S_t - (r-½\sigma^2)(T-t)}{\sigma\sqrt{T-t}} - \sigma\sqrt{T-t} = \infty$ [¼]

- $L_1 = \dfrac{\log(0) - \log S_t - (r-½\sigma^2)(T-t)}{\sigma\sqrt{T-t}} - \sigma\sqrt{T-t} = -\infty$ [¼]

So:

- $\Phi(U_2) = 1$
- $\Phi(L_2) = 0$
- $\Phi(U_1) = 1$
- $\Phi(L_1) = 0$

[½]

Hence:

$$V_t = S_t^2\left(e^{(r+\sigma^2)(T-t)}\right)\{1-0\} - S_t\{1-0\}$$

$$= S_t^2 e^{(r+\sigma^2)(T-t)} - S_t$$ [1]

[Total 7]

Alternatively, as the lognormal distribution isn't actually truncated here, you could instead use the formula for the (untruncated) moments of the lognormal distribution on page 14 in the Tables, with $r=2$ for the first integral and $r=1$ for the second integral. So:

$$V_t = e^{-r(T-t)}\left[\int_0^\infty S_T^2\, f(S_T\,|S_t)\,dS_T - \int_0^\infty S_T\, f(S_T\,|S_t)\,dS_T\right]$$

$$= e^{-r(T-t)}\left[\begin{array}{l}\left(e^{2(\log S_t + (r-½\sigma^2)(T-t)) + ½\times 4\sigma^2(T-t)}\right)\\ -\left(e^{\log S_t + (r-½\sigma^2)(T-t) + ½\sigma^2(T-t)}\right)\end{array}\right]$$ *[2]*

$$= e^{-r(T-t)}\left[S_t^2 e^{(2r+\sigma^2)(T-t)} - S_t e^{r(T-t)}\right]$$

$$= S_t^2 e^{(r+\sigma^2)(T-t)} - S_t$$ *[2]*

(ii)(a) ***Formula for the delta of the derivative***

Recall that for a derivative with price f_t delta is defined as:

$$\Delta = \frac{\partial f_t}{\partial S_t}$$

So, here:

$$\Delta = 2S_t\, e^{(r+\sigma^2)(T-t)} - 1$$ [1]

(ii)(b) ***Range of values for the current share price for which delta is positive***

Delta will be positive if:

$$2S_t\, e^{(r+\sigma^2)(T-t)} - 1 > 0$$

$$S_t > ½e^{-(r+\sigma^2)(T-t)}$$ [½]

As the term on the right-hand side decreases as T increases, this indicates that delta is more likely to be positive for a derivative of this type with a longer outstanding term. [½] (ii)(c) ***Derive formula for the gamma of the derivative***

Recall that:

$$\Gamma = \frac{\partial \Delta}{\partial S_t}$$

So for this derivative:

$$\Gamma = 2e^{(r+\sigma^2)(T-t)}$$ [1]

So, gamma is always positive. [½]

This reflects the fact that the value of the derivative depends directly upon the square of the share price and hence is a convex function of the share price. [½]

[Total 4]

Solution X3.10

Course references: Self-financing and previsible are defined in Chapter 16. Ito's Lemma is stated in Chapter 10 and also on page 46 in the Tables. Note also that the result in (iv)(b) is the Black-Scholes PDE from Chapter 15 (and page 46 in the Tables).

(i) ***Differential equation***

$$dB_t = rB_t dt$$ [1]

(ii) ***Self-financing portfolio***

The changes in the value of a self-financing portfolio are due purely to the changes in the prices of the constituent assets, and not due to injections or withdrawals of money into or out of the portfolio. [1]

If a portfolio of shares and cash has value V_t, *ie* $V_t = \phi_t S_t + \psi_t B_t$ then it will be self-financing if and only if:

$$dV_t = \phi_t dS_t + \psi_t dB_t$$ [1]

[Total 2]

(iii) ***Previsible***

A process is previsible if its value at time *t* can be deduced from the information that is known up to but not including time *t*. [1]

(iv) ***Deduce that*** $\boldsymbol{\phi_t = \frac{\partial}{\partial S_t} f(t, S_t)}$

Starting from $\phi_t S_t + \psi_t B_t = f(t, S_t)$ we get:

(1) $d(\phi_t S_t + \psi_t B_t) = df(t, S_t)$ [½]

Assuming the portfolio is self-financing, the left-hand side of (1) must be $\phi_t dS_t + \psi_t dB_t$. [½]

Given that $dS_t = S_t\left(\mu dt + \sigma dZ_t\right)$ and $dB_t = rB_t dt$, this is equal to:

$$\phi_t S_t\left(\mu dt + \sigma dZ_t\right) + \psi_t rB_t dt$$ [½]

Applying Ito's Lemma to the right-hand side of (1) we get:

$$df\left(t, S_t\right) = \left[\frac{\partial f}{\partial t} + \mu S_t \frac{\partial f}{\partial S_t} + \tfrac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 f}{\partial S_t^2}\right]dt + \sigma S_t \frac{\partial f}{\partial S_t} dZ_t$$ [1]

Alternatively, using a Taylor Series expansion, the right-hand side of (1) is equal to:

$$df\left(t, S_t\right) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial S_t}dS_t + \tfrac{1}{2}\frac{\partial^2 f}{\partial S_t^2}\left(dS_t\right)^2$$

Again using $dS_t = S_t\left(\mu dt + \sigma dZ_t\right)$ *and* $dB_t = rB_t dt$ *, and using the 2-by-2 multiplication table for increments given in the Core Reading to note that* $\left(dS_t\right)^2 = \sigma^2 S_t{}^2 dt$ *, this can be written as:*

$$df\left(t, S_t\right) = \left[\frac{\partial f}{\partial t} + \mu S_t \frac{\partial f}{\partial S_t} + \tfrac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 f}{\partial S_t^2}\right]dt + \sigma S_t \frac{\partial f}{\partial S_t} dZ_t$$

So, equating the two sides of (1) gives:

$$\phi_t S_t\left(\mu dt + \sigma dZ_t\right) + \psi_t rB_t dt = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial S_t}S_t\left(\mu dt + \sigma dZ_t\right) + \tfrac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 f}{\partial S_t^2}dt$$ [½]

Comparing the dZ_t terms we must have:

$$\phi_t \sigma S_t = \sigma \frac{\partial f}{\partial S_t} S_t$$ [½]

and therefore:

$$\phi_t = \frac{\partial f}{\partial S_t}\left(t, S_t\right)$$ [½]

[Total 4]

Assignment X4 Solutions

Markers: This document sets out one approach to solving each of the questions. Please give credit for other valid approaches.

Note that some of the numerical answers are sensitive to rounding.

Solution X4.1

Course reference: The Merton model for valuing risky corporate debt is described in Chapter 19.

(i) ***Price of risk-free zero-coupon bond***

Price = $100\,e^{-5\times 0.05} = \$77.88$ [1]

(ii) ***Value of company's assets and ZCBs***

Adding together the value of the company shares and the risk-free ZCBs gives a starting value for the value of its assets of:

$$F_t = 118.46 + 77.88 = \$196.34\text{m}$$ [½]

Recall the Merton model for the current value of the shares:

$$E_t = F_t\Phi(d_1) - L\,e^{-r(T-t)}\Phi(d_2),$$

where $d_1 = \dfrac{\log\left(\dfrac{F_t}{L}\right) + \left(r + ½\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$, $d_2 = d_1 - \sigma\sqrt{T-t}$ [½]

So, substituting the starting value of 196.34 for F_t into this formula, together with:

- the outstanding term of the ZCBs, $T-t=5$
- the volatility of the company's assets, $\sigma_F = 0.25$
- the risk-free force of interest, $r = 0.05$
- the nominal value of the ZCBs, $L = 100$

gives:

$$d_1 = \frac{\log\left(\dfrac{196.34}{100}\right) + \left(0.05 + ½\times 0.25^2\right)\times 5}{0.25\sqrt{5}} = 1.9336$$ [½]

$$d_2 = 1.9336 - 0.25\sqrt{5} = 1.3746$$ [½]

and hence:

$$E_t = 196.34 \times \underbrace{\Phi(1.9336)}_{0.97342} - 100 \times e^{-0.05\times 5} \times \underbrace{\Phi(1.3746)}_{0.91537} = \$119.832\text{m}$$ [½]

This is (about 1.37) *greater* than the observed market capitalisation and delta (the change in the share price with respect to the total asset value) is *positive*. So, try a *lower* estimate for the total asset value. [½]

Note also that as the option represented by the shares is very "in-the-money" (as F_t is much greater than $L = 100$), delta is close to one.

So, in order to reduce the market capitalisation by about 1.37, we need to try a value of F_t a bit more than 1.37 less than 196.34. So, we next try $F_t = 194.90$. This gives:

$$d_1 = \frac{\log\left(\frac{194.90}{100}\right) + \left(0.05 + ½ \times 0.25^2\right) \times 5}{0.25\sqrt{5}} = 1.9205$$

$$d_2 = 1.9205 - 0.25\sqrt{5} = 1.3614$$

and hence:

$$E_t = 194.90 \times \underbrace{\Phi(1.9205)}_{0.97260} - 100 \times e^{-0.05\times 5} \times \underbrace{\Phi(1.3614)}_{0.91331} = \$118.431\text{m}$$

This is just below the market capitalisation of \$118.46 (and delta is close to one), so let's try an asset value that is slightly higher, *eg* $F_t = 194.95$. This gives:

$$d_1 = \frac{\log\left(\frac{194.95}{100}\right) + \left(0.05 + ½ \times 0.25^2\right) \times 5}{0.25\sqrt{5}} = 1.9209$$

$$d_2 = 1.9209 - 0.25\sqrt{5} = 1.3619$$

and hence:

$$E_t = 194.95 \times \underbrace{\Phi(1.9209)}_{0.97263} - 100 \times e^{-0.05\times 5} \times \underbrace{\Phi(1.3619)}_{0.91338} = \$118.479\text{m}$$

Finally, linear interpolation between 194.90 and 194.95 gives:

$$F_t = 194.90 + 0.05 \times \frac{118.46 - 118.431}{118.479 - 118.431} = \$194.93\text{m}$$

[2½ for correct value for F_t via appropriate interpolation]

Alternatively, trying $F_t = 195.00$ *gives:*

$$d_1 = \frac{\log\left(\dfrac{195.00}{100}\right) + \left(0.05 + ½ \times 0.25^2\right) \times 5}{0.25\sqrt{5}} = 1.9214$$

$$d_2 = 1.9214 - 0.25\sqrt{5} = 1.3624$$

and hence:

$$E_t = 195.00 \times \underbrace{\Phi(1.9214)}_{0.97266} - 100 \times e^{-0.05\times5} \times \underbrace{\Phi(1.3624)}_{0.91346} = \$118.528\text{m}$$

So, linear interpolation between 194.90 and 195.00 gives:

$$F_t = 194.90 + 0.10 \times \frac{118.46 - 118.431}{118.528 - 118.431} = \$194.93\text{m}$$

[2½ for correct value for F_t *via appropriate interpolation]*

Hence, the value of the company's ZCBs must be:

$$B_t = 194.93 - 118.46 = \$76.47\text{m}$$ [½]

[Maximum 5]

(iii)(a) ***Delta of the ZCBs***

If the underlying asset pays no dividends then, according to the Black-Scholes formula:

$$\Delta_{call} = \Phi(d_1)$$ [½]

In the Merton model:

$$F_t = E_t + B_t$$ [½]

Differentiating this equation partially with respect to F_t gives:

$$1 = \Delta_E + \Delta_B$$

ie $\Delta_B = 1 - \Delta_E$ [½]

So, recalling that the Merton model suggests that the company's shares are effectively a European call option on its underlying assets, we have:

$$\Delta_B = 1 - \Phi(d_1)$$ [½]

[Total 2]

(iii)(b) ***Numerical value of delta***

$$\Delta_B = 1 - \Phi(d_1)$$

So, using $\Phi(d_1) \approx 0.97262$ from part (ii) we have:

$$\Delta_B = 1 - 0.97262 = 0.02738$$ [½]

Markers: Please accept any value for Δ_B between:

$$\Delta_B = 1 - 0.9726 = 0.0274$$

and $\Delta_B = 1 - 0.9727 = 0.0273$

Estimated fall in the value of the ZCBs

If the value of the company's assets falls by \$10 million, the value of the ZCBs will therefore fall by approximately:

$$10 \times \Delta_B = 10 \times 0.02738 = \$0.2738\text{m}$$ [1]

to:

$$B_t' = 76.47 - 0.2738 = \$76.196\text{m}$$ [½]

[Total 2]

(iii)(c) ***Reason for discrepancy between estimate and actual value***

The difference between the estimated value and the actual value stems from the fact that delta is not constant but varies with the change in the value of the company's assets. [1]

In other words, the difference arises because gamma is non-zero.

Alternatively, the values may differ because the values of one or more of the other parameters (*eg* the volatility) may have changed. [1]

[Maximum 1]

In fact, partially differentiating the formula for delta gives:

$$\Gamma_B = \frac{\partial^2 F_t}{\partial E_t^2} = -\frac{\phi(d_1)}{F_t \sigma \sqrt{T-t}} < 0$$

*which tells us that delta decreases as the value of the company's assets increases and conversely that it **increases** with a **fall** in the value of the company's assets. So, in this instance, assuming delta is constant means that the fall in the value of the ZCBs is under-estimated. Hence, the estimated ZCB value is greater than the actual value.*

Solution X4.2

Course reference: Multiple state models of credit risk are discussed in Chapter 19. Note, however, that this is a discrete-time model in which changes only occur once a year. So we need to think of this as a Markov chain, rather than as a jump process (like the Jarrow-Lando-Turnbull model).

(i) ***Transition probabilities***

For time $t=1$, we can infer the probabilities directly from the diagram:

$$p_{FF}(0,1)=1-0.2=0.8\,,\quad p_{FH}(0,1)=0 \qquad [½]$$

For the later times, we can use the fact that the t-step transition probabilities are the entries in the matrix P^t, where P is the one-step transition probability matrix, which is:

$$P=\begin{array}{c c} & \begin{array}{ccc} F & H & N \end{array} \\ \begin{array}{c} F \\ H \\ N \end{array} & \begin{bmatrix} 0.8 & 0 & 0.2 \\ 0.1 & 0.9 & 0 \\ 0 & 0.1 & 0.9 \end{bmatrix} \end{array} \qquad [½]$$

We find that:

$$P^2=\begin{bmatrix} 0.8 & 0 & 0.2 \\ 0.1 & 0.9 & 0 \\ 0 & 0.1 & 0.9 \end{bmatrix}\begin{bmatrix} 0.8 & 0 & 0.2 \\ 0.1 & 0.9 & 0 \\ 0 & 0.1 & 0.9 \end{bmatrix}=\begin{bmatrix} 0.64 & 0.02 & 0.34 \\ 0.17 & 0.81 & 0.02 \\ 0.01 & 0.18 & 0.81 \end{bmatrix} \qquad [½]$$

So: $p_{FF}(0,2)=0.64\,,\quad p_{FH}(0,2)=0.02$ [½]

An alternative method is to add up the probabilities for each possible path.

For time 3:

$$P^3=P^2P=\begin{bmatrix} 0.64 & 0.02 & 0.34 \\ 0.17 & 0.81 & 0.02 \\ 0.01 & 0.18 & 0.81 \end{bmatrix}\begin{bmatrix} 0.8 & 0 & 0.2 \\ 0.1 & 0.9 & 0 \\ 0 & 0.1 & 0.9 \end{bmatrix}=\begin{bmatrix} 0.514 & 0.052 & 0.434 \\ 0.217 & 0.731 & 0.052 \\ 0.026 & 0.243 & 0.731 \end{bmatrix} \qquad [½]$$

So: $p_{FF}(0,3)=0.514\,,\quad p_{FH}(0,3)=0.052$ [½]

[Total 3]

(ii)(a) ***Calculate the present value***

The payments due from the bond in State F are £10 at the end of years 1 and 2, and £110 at the end of year 3.

The payments due from the bond in State H are £5 at the end of years 1 and 2, and £55 at the end of year 3.

The payments due from the bond in State N are £0 at the end of each year.

Hence, using the probabilities calculated in part (i), the risk-neutral expected amounts of the payments are:

Time 1: $10p_{FF}(0,1)+5p_{FH}(0,1)=10\times 0.8+5\times 0=8$ [½]

Time 2: $10p_{FF}(0,2)+5p_{FH}(0,2)=10\times 0.64+5\times 0.02=6.5$ [½]

Time 3: $110p_{FF}(0,3)+55p_{FH}(0,3)=110\times 0.514+55\times 0.052=59.4$ [½]

We can calculate the risk-neutral expected present value by discounting these payments using the yields on the government bonds (which can be considered to be default-free). This gives:

$$\frac{8}{1.05}+\frac{6.5}{1.05^2}+\frac{59.4}{1.05^3}=64.83$$ [½]

(ii)(b) ***Comment***

The risk-neutral present value calculated in part (ii)(a) is the fair value of the bonds (allowing for the possibility of default). [½]

So, if the bank is considering purchasing the bonds at par (*ie* at £100), it will be paying far too much (according to the model). [½]

[Total 3]

(iii) ***Calculate the credit spread***

We first need to find the rate of return that the bank will earn on the bond if the company makes all the payments in full. This is found from the equation:

$$\frac{10}{1+i}+\frac{10}{(1+i)^2}+\frac{110}{(1+i)^3}=95.2$$ [1]

By trial and improvement, we find that the solution is 12%. [1]

So the credit spread is the difference between this and the yield on a corresponding default-free bond, *ie* $12\%-5\%=7\%$. [1]

[Total 3]

Solution X4.3

Course reference: The Vasicek model is introduced in Chapter 18.

(i)(a) ***Derive an equation for dU_t***

Starting with Taylor's theorem:

$$dU_t = \frac{\partial U_t}{\partial t}dt + \frac{\partial U_t}{\partial r_t}dr_t + \frac{\partial^2 U_t}{\partial {r_t}^2}(dr_t)^2 \qquad [½]$$

Since $U_t = e^{at}r_t$:

$$\frac{\partial U_t}{\partial t} = ae^{at}r_t \qquad \frac{\partial U_t}{\partial r_t} = e^{at} \qquad \frac{\partial^2 U_t}{\partial {r_t}^2} = 0 \qquad [½]$$

Substituting these values, and the expression for dr_t, into Taylor's theorem gives:

$$dU_t = ae^{at}r_t dt + e^{at}\left(a\left(b - r_t\right)dt + \sigma\, dB_t\right)$$
$$= abe^{at}dt + \sigma e^{at}dB_t \qquad [1]$$

Note that this approach is equivalent to applying the "product rule". (It is not obvious that this is allowed in stochastic calculus, and indeed it isn't in general. However, it is legitimate to use it in this situation, where one of the factors in the product is deterministic).

Alternatively, if we use Ito's Lemma, the three partial derivatives are the same as above and the drift and volatility functions from the stochastic differential equation for r_t are $a(b - r_t)$ and σ respectively. So, substituting these five items into Ito's Lemma from page 46 in the Tables gives:

$$dU_t = \left[a\left(b - r_t\right)e^{at} + ½ \times 0 \times \sigma^2 + ae^{at}r_t\right]dt + \sigma e^{at}dB_t$$
$$= ab\,e^{at}dt + \sigma e^{at}dB_t$$

(i)(b) ***Solve the equation***

We integrate both sides from 0 to t :

$$U_t - U_0 = \int_0^t abe^{as}ds + \sigma\int_0^t e^{as}dB_s \qquad [1]$$

$$\Rightarrow \qquad U_t = U_0 + b\left(e^{at} - 1\right) + \sigma\int_0^t e^{as}dB_s \qquad [1]$$

(i)(c) ***Show that***

Expressing the previous expression for U_t in terms of r_t, we have:

$$e^{at}r_t = r_0 + b\left(e^{at} - 1\right) + \sigma\int_0^t e^{as}dB_s$$

$$\Rightarrow \qquad r_t = b + e^{-at}\left(r_0 - b\right) + \sigma\int_0^t e^{a(s-t)}dB_s \qquad [1]$$

[Total 5]

(ii) ***Probability distribution of r_t***

$$dB_s \sim N(0, ds)$$

$$\Rightarrow \qquad \sigma e^{a(s-t)}dB_s \sim N\left(0, \sigma^2 e^{2a(s-t)}ds\right)$$

$$\Rightarrow \qquad \int_0^t \sigma e^{a(s-t)}dB_s \sim N\left(0, \int_0^t \sigma^2 e^{2a(s-t)}ds\right) \qquad [1]$$

In the last line here just treat the integral as a sum, and therefore use the standard results for adding up independent normal random variables, ie their mean and variance just add.

You can then evaluate this last integral, and you need to take the other non-random terms into account, which just change the mean, ie shift the distribution.

The distribution of r_t is given by:

$$r_t \sim N\left(b + e^{-at}\left(r_0 - b\right), \int_0^t \sigma^2 e^{2a(s-t)}ds\right) = N\left(b + e^{-at}\left(r_0 - b\right), \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)\right) \qquad [2]$$

As $t \to \infty$, $e^{-2at} \to 0$, so we get:

$$r_t \sim N\left(b, \frac{\sigma^2}{2a}\right) \qquad [1]$$

[Total 4]

(iii) ***Derive the conditional expectation***

We have:

$$E\left[r_t \mid F_s\right] = E\left[b + e^{-at}\left(r_0 - b\right) + \int_0^t \sigma e^{a(u-t)} dB_u \mid F_s\right]$$

$$= b + e^{-at}\left(r_0 - b\right) + E\left[\int_0^t \sigma e^{a(u-t)} dB_u \mid F_s\right] \qquad [1]$$

The integral in the last expression is an Ito integral, which suggests that it might be a martingale. In fact, it won't be a martingale itself because it involves the variable t in the integrand. However, if the factor e^{-at} is removed, then the remaining part will be a martingale. If you didn't spot this, you could still continue as follows.

$$E\left[r_t \mid F_s\right] = b + e^{-at}\left(r_0 - b\right) + E\left[\int_0^s \sigma e^{a(u-t)} dB_u \mid F_s\right] + E\left[\int_s^t \sigma e^{a(u-t)} dB_u \mid F_s\right] \qquad [1]$$

What we've done here, is to split the integral up into two parts. In the first part, from 0 to s, the random increments are all known, since we are conditioning on F_s. We can therefore treat that integral as a constant and take it outside the expectation. In the second integral, from s to t, the increments are all independent of the past (ie the period before s) and so the conditioning can be dropped. This leaves the expectation of such as integral, which is always zero, since the increments dB_u themselves have zero mean.

$$E\left[r_t \mid F_s\right] = b + e^{-at}\left(r_0 - b\right) + \int_0^s \sigma e^{a(u-t)} dB_u \qquad [1]$$

We can now relate this to r_s which we know from (i)(c):

$$E\left[r_t \mid F_s\right] = b + e^{-at}\left(r_0 - b\right) + e^{-a(t-s)} \int_0^s \sigma e^{a(u-s)} dB_u \qquad [1]$$

using the result from part (i)(c) with *s* replaced by *u* and *t* replaced by *s*.

$$E\left[r_t \mid F_s\right] = b + e^{-at}\left(r_0 - b\right) + e^{a(s-t)}\left[r_s - b - \left(r_0 - b\right)e^{-as}\right]$$

$$= b\left(1 - e^{a(s-t)}\right) + e^{a(s-t)} r_s$$

[1]

[Total 5]

Solution X4.4

Course reference: The Vasicek model is introduced in Chapter 18.

(i) ***The Vasicek model and its statistical properties***

This is a model used for modelling the short-rate of interest $r(t)$. [½]

It assumes that $r(t)$ has the dynamics of an Ito process (in fact, an Ornstein-Uhlenbeck process) under the risk-neutral probability measure Q. [1]

The Vasicek model assumes the model $dr(t) = \alpha[\mu - r(t)]dt + \sigma d\tilde{W}(t)$, where $\tilde{W}(t)$ is standard Brownian motion. [1]

The movements in the interest rate are therefore normally distributed and the parameter σ controls the volatility. [½]

$r(t)$ can take negative values. [½]

The parameter α is chosen to be positive, so that $r(t)$ is mean-reverting to the constant value μ.
[½]
[Total 4]

(ii) ***Condition for a humped curve***

If the graph is humped, there will be a local maximum for some positive value of τ.

Let $f(\tau) = f(t,T)$, so that the function given is:

$$f(\tau) = r_t e^{-\alpha\tau} + r_\infty(1 - e^{-\alpha\tau}) + k(1 - e^{-\alpha\tau})e^{-\alpha\tau}$$
$$= r_\infty + (r_t - r_\infty + k)e^{-\alpha\tau} - ke^{-2\alpha\tau}$$

Differentiating with respect to τ:

$$f'(\tau) = -(r_t - r_\infty + k)\alpha e^{-\alpha\tau} + 2k\alpha e^{-2\alpha\tau}$$
$$= [-(r_t - r_\infty + k) + 2ke^{-\alpha\tau}]\alpha e^{-\alpha\tau} \qquad [2]$$

For this to equal zero, the expression in square brackets must also equal zero:

ie $$e^{-\alpha\tau} = \frac{r_t - r_\infty + k}{2k} \qquad [1]$$

This means that there is at most one stationary point. Now, since $0<\tau<\infty$ and $\alpha>0$, then $e^{-\alpha\tau}$ is in the range $(0,1)$. So there will be a hump if:

$$0<\frac{r_t-r_\infty+k}{2k}<1$$ [1]

or $$0<r_t-r_\infty+k<2k$$

(k must be positive – otherwise it would be a "dip", rather than a hump.)

If we subtract k 's, we see that this condition is equivalent to:

$$-k<r_t-r_\infty<k$$

ie $$k>\left|r_\infty-r_t\right|$$ [1]

[Total 5]

(iii) ***Advantages and limitations***

Advantages

The mean-reverting property is consistent with real-world observations. [½]

The model is mathematically tractable. [½]

It provides an arbitrage-free model of short-term interest rates. [½]

It allows a wide range of possible yield curve shapes (upward-sloping, downward-sloping, humped) depending on the parameter values chosen. [½]

Limitations

The model has the disadvantage that it allows r_t to take negative values, which is not common in the real world. [½]

The model assumes a constant volatility, whereas the actual volatility of short-term interest rates appears to vary through time. [½]

It can be difficult to calibrate the model to past date and to the current yield curve as there are only three parameters. [½]

Since the model incorporates only one source of randomness:

- it is not flexible enough to price a wide range of derivatives, *eg* those whose payoffs depend on more than one interest rate [½]

- it produces perfectly correlated movements in interest rates and hence in zero-coupon bond prices. [½]

[Maximum 4]

Solution X4.5

The premium is:

$$c = (1+\theta)E(S) = 1.2 \times \lambda E(X) = 1.2 \times ¼ \times 60 = 18$$ [½]

Hence:

$$U(2) = 100 + 2 \times 18 - S(2) = 136 - S(2)$$ [½]

So the probability of ruin is:

$$P[U(2) < 0] = P[136 - S(2) < 0] = P[S(2) > 136]$$ [½]

Considering $P[S(2) < 136]$, remembering that $N(2) \sim Poi(2 \times ¼)$, we have:

Number of claims	Amount of claim(s)	Probability
0 claims	0	$(e^{-0.5}) = 0.60653$
1 claim	50	$(0.5e^{-0.5}) \times 0.8 = 0.24261$
	100	$(0.5e^{-0.5}) \times 0.2 = 0.06065$
2 claims	50, 50	$(\frac{0.5^2}{2} e^{-0.5}) \times 0.8^2 = 0.04852$

[½ for each probability = 2]

Hence:

$$P[S(2) < 136] = 0.9583 \quad \Rightarrow \quad P[U(2) < 0] = 0.0417$$ [½]

[Total 4]

Solution X4.6

(i)(a) ***Surplus***

The surplus process $U(t)$ is given by:

$$U(t) = u + ct - S(t)$$

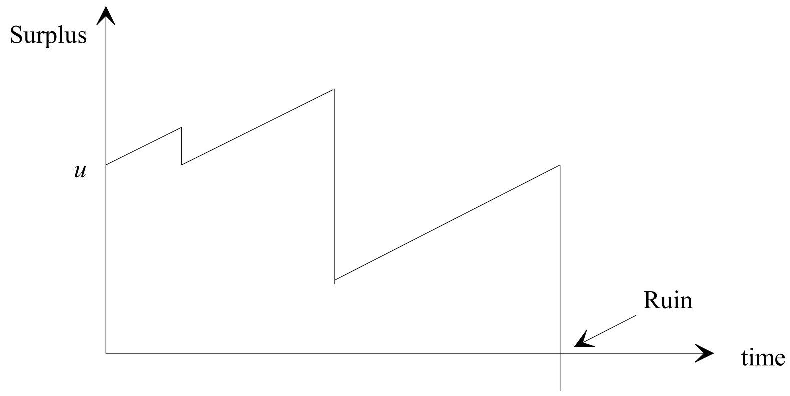
where c, the rate of premium income per unit time, is $c = (1+\theta)\lambda E[X]$ and $S(t)$, the aggregate claim amount, is $S(t) = X_1 + X_2 + \cdots + X_{N(t)}$. [1]

(i)(b) ***Probability of ruin***

The probability of ruin is given by:

$$\Psi(u) = P\left[U(t) < 0, t > 0\right]$$ [1]

The diagram showing a ruin event is as follows:



[1]

(i)(c) ***State probability of ultimate ruin***

$\Psi(u) = 1$ [1]

[Total 4]

(ii) ***Change of currency***

Let the surplus process under the new currency be $\tilde{U}(t)$. A surplus of 100 in the old currency is the same as a surplus of 250 in the new currency. So $\tilde{U}(t) = 2.5U(t)$. The probability of ruin, $\Psi(u)$, will be the same under the new currency since:

$$P\left(\tilde{U}(t) < 0\right) = P\left(2.5U(t) < 0\right) = P\left(U(t) < 0\right)$$ [2]

Solution X4.7

Each incremental entry, C_{ij}, in the run-off triangle can be expressed in general terms as:

$$C_{ij} = r_j \, s_i \, x_{i+j} + e_{ij}$$ [1]

where:

- r_j is the development factor for Development Year j, representing the proportion of claim payments in year j. Each r_j is independent of the Accident Year i. [1]
- s_i is a parameter for Accident Year i representing exposure (*eg* number of claims or claim amount in respect of Accident Year *i*). [1]
- x_{i+j} is a parameter varying by calendar year (*eg* a measure of inflation). [½]
- e_{ij} is an error term. [½]

[Total 4]

Solution X4.8

(i)(a) ***Adjustment coefficient equation***

The adjustment coefficient is the unique positive root, R , of the equation:

$$\lambda + cr = \lambda M_X(r) \quad \text{equation (1)}$$ [½]

We have $P(X=100)=1$, hence:

$$E(X)=100 \quad \text{and} \quad M_X(t)=e^{100t}$$ [½]

$$\Rightarrow \quad c=(1+\theta)\lambda E(X)=120\lambda$$ [½]

Substituting these into equation (1) gives:

$$\lambda + 120\lambda R = \lambda e^{100R} \quad \Rightarrow \quad 1+120R=e^{100R}$$ [½]

(i)(b) ***Adjustment coefficient***

Using the series expansion from Page 2 of the *Tables*:

$$1+120R \simeq \left(1+100R+\frac{(100R)^2}{2!}+\frac{(100R)^3}{3!}\right)$$ [½]

$$0=R\left(-20+5{,}000R+\frac{500{,}000}{3}R^2\right)$$ [½]

Solving this gives $R=0, 0.00357, -0.0336$. [½]

Since the adjustment coefficient is the positive root, $R=0.00357$. [½]

[Total 4]

(ii) ***Minimum initial capital***

Using Lundberg's inequality:

$$e^{-RU}<0.05 \quad [1]$$

Using the value of R from Part (i)(b):

$$-RU<\ln 0.05 \quad \Rightarrow \quad RU>-\ln 0.05 \quad \Rightarrow \quad U>-\frac{\ln 0.05}{0.00357}=838$$ [1]

[Total 2]

Solution X4.9

Adjusting the incremental data for past inflation (*ie* change the figures to mid-2012 money terms):

Claim payments in mid-2012 money terms (£'000)		Development year		
		0	1	2
Accident year	2010	$830 \times 1.02 \times 1.025$ **867.765**	940×1.025 **963.5**	150
	2011	850×1.025 **871.25**	920	
	2012	1,120		

[1 mark for top left, ½ mark for other two bold entries = total 2]

Next, we accumulate, find the ratios and use the basic chain ladder to project the values:

×2.083084 ×1.081911

Claim payments in mid-2012 money terms (£'000)		Development year		
		0	1	2
Accident year	2010	867.765	1,831.265	1,981.265
	2011	871.25	1,791.25	**1,937.97**
	2012	1,120	**2,333.05**	**2,524.16**

[1 mark for 1st ratio, ½ mark for 2nd ratio, ½ mark for each bold entry $=$ total 3]

Finally, we need incremental data again, so we can adjust for future inflation (*ie* calculate the actual money to be paid in each future year):

Actual money paid (£'000)		Development year		
		0	1	2
Accident year	2010			
	2011			146.72×1.03 **151.12**
	2012		$1{,}213.05 \times 1.03$ **1,249.44**	191.11×1.03^2 **202.75**

[1 mark for bottom right, ½ mark for other two bold entries = total 2]

So the estimated total future amount for outstanding claims is:

$$151.12 + 1{,}249.44 + 202.75 = 1{,}603.31 \simeq £1{,}603{,}000$$ [1]

[Total 8]

Rounding will affect the accuracy of students' answers. Our general guidelines are to round the final figure to the same degree of accuracy as the figures given in the question but to keep several more significant figures throughout the workings. In this case the final answer should be correct to at least 3 SF and the workings correct to at least 4 SF. For reference, our workings are based on non-rounded figures, ie storing non-rounded figures in the calculator memory.

Solution X4.10

The development factors are:

$$\text{DY 1} \rightarrow \text{DY 2} \qquad \frac{2{,}101}{1{,}923} = 1.09256 \qquad [1]$$

$$\text{DY 0} \rightarrow \text{DY 1} \qquad \frac{1{,}923 + 2{,}140}{1{,}417 + 1{,}701} = 1.30308 \qquad [1]$$

The initial estimate of the ultimate loss incurred for Policy Year 2012 is given by:

$$(\text{ultimate loss ratio}) \times (\text{earned premium}) = 0.92 \times £3{,}073 = £2{,}827.16 \qquad [1]$$

Next we calculate the emerging liability for Policy Year 2012:

$$(\text{intial ultimate loss}) \times (1 - \tfrac{1}{f}) = £2{,}827.16 \times \left(1 - \frac{1}{1.09256 \times 1.30308}\right)$$

$$= £841.37 \qquad [1]$$

In the formula above, f is the cumulative development factor (*ie* the product of all the development factors from the last known payment to the end).

So the revised estimate of the ultimate loss incurred for policies written in 2012 is:

$$£1{,}582 + £841.37 = £2{,}423.37 \qquad [1]$$

Alternatively, we can calculate the expected claims incurred to date as $£2{,}827.16 \times \dfrac{1}{1.09256 \times 1.30308} = £1{,}985.79$. *We have actually incurred £1,582. The actual is less than the expected figure by £403.79. Adjusting the initial estimate of the ultimate loss incurred by this gives a revised estimate of the ultimate loss incurred of* $£2{,}827.16 - £403.79 = £2{,}423.37$.

Since all our figures are in 000's, we have £2,423,370. Now the figures given in the table are claims incurred (rather than claims paid). So we need to use the fact that for policy year 2012 we have paid £441,000. Therefore, the amount left to pay is:

$$£2{,}423{,}370 - £441{,}000 \simeq £1{,}982{,}000 \text{ (4 SF)} \qquad [1]$$

[Total 6]

Solution X4.11

(i) ***ACPC***

First divide each number in the first table by the corresponding entry in the second table to obtain the average cost incurred per claim reported:

Average cost per claim (£'000)		Development year		
		0	1	2
Accident year	2010	252 ÷ 56 = **4.5**	375 ÷ 74 = **5.0676**	438 ÷ 87 = **5.0345**
	2011	230 ÷ 49 = **4.6939**	343 ÷ 65 = **5.2769**	
	2012	208 ÷ 44 = **4.7273**		

[1]

Then calculate the grossing-up factors for the average cost incurred per claim reported:

Average cost per claim (£'000)		Development year		
		0	1	2
Accident year	2010	4.5 **89.384%**	5.0676 **100.657%**	5.0345 **100%**
	2011	4.6939 **89.536%**	5.2769 **100.657%**	5.2769 ÷ 1.00657 **5.2425**
	2012	4.7273 **89.460%**		4.7273 ÷ 0.89460 **5.2843**

[½ mark for each of the shaded figures = total 2]

Next calculate the grossing-up factors for the number of reported claims:

Number of reported claims		Development year		
		0	1	2
Accident year	2010	56 **64.368%**	74 **85.057%**	87 **100%**
	2011	49 **64.120%**	65 **85.057%**	65 ÷ 0.85057 **76.419**
	2012	44 **64.244%**		44 ÷ 0.64244 **68.489**

[½ mark for each of the shaded figures = total 2]

The projected ultimate incurred loss for each accident year is then obtained by multiplying the ultimate figures for the average cost per claim and the number of reported claims.

Accident Year	ACPC (£'000)	Number of reported claims	Projected incurred claims (£'000)
2010	5.0345	87	438
2011	5.2425	76.419	400.6
2012	5.2843	68.489	361.9
		Total	**1,201** (4 SF)

[1]

The total claims paid is £950,000. So the estimated outstanding claims reserve is approximately:

$$£1,201,000 - £950,000 = £251,000$$ [1]

[Total 7]

(ii) ***Assumptions***

- The 2010 accident year is fully run-off.
- For each origin year, the numbers of claims reported in each development year are constant proportions of the total number of claims reported from that accident year.
- For each origin year, the average claim amounts incurred in each development year are constant proportions of the total average claim amount incurred from that accident year.

[Total 2: subtract 1 mark per error or omission]

Solution X4.12

(i) ***Minimum value of α***

The net premium is given by:

$$c_{net} = (1+\theta)E(S) - (1+\xi)E(S_R)$$

$$= (1+\theta)\lambda E(X) - (1+\xi)\lambda E(Z)$$ [½]

$$= 1.2\lambda \times 100 - 1.3\lambda \times 100(1-\alpha)$$

$$= 130\alpha\lambda - 10\lambda$$ [½]

The expected net claims received by the insurer are given by:

$$E(S_I) = \lambda E(Y) = \lambda\alpha E(X) = 100\alpha\lambda$$ [½]

Hence:

$$130\alpha\lambda - 10\lambda > 100\alpha\lambda \quad \Rightarrow \quad 30\alpha > 10 \quad \Rightarrow \quad \alpha > \tfrac{1}{3}$$ [½]

[Total 2]

(ii) ***Adjustment coefficient***

The adjustment coefficient is the unique positive root, R, of the equation:

$$\lambda + c_{net}R = \lambda M_Y(R)$$ equation (1) [½]

From Part (i)(a) we have $c_{net} = 130\alpha\lambda - 10\lambda$. Also:

$$M_Y(R) = E[e^{RY}] = E[e^{R\alpha X}] = M_X(\alpha R) = (1-100\alpha R)^{-1}$$ [1]

Substituting into equation (1) gives:

$$\lambda + (130\alpha\lambda - 10\lambda)R = \lambda(1-100\alpha R)^{-1}$$ [1]

Solving this:

$$1 + (130\alpha - 10)R = (1-100\alpha R)^{-1}$$

$$\Rightarrow \quad 1 = (1 + 130\alpha R - 10R)(1-100\alpha R)$$

$$\Rightarrow \quad 1 = 1 - 100\alpha R + 130\alpha R - 13{,}000\alpha^2 R^2 - 10R + 1{,}000\alpha R^2$$

$$\Rightarrow \quad 0 = 30\alpha R - 13{,}000\alpha^2 R^2 - 10R + 1{,}000\alpha R^2$$

$$\Rightarrow \quad 0 = 10R(3\alpha - 1{,}300\alpha^2 R - 1 + 100\alpha R)$$

$$\Rightarrow \quad R = 0 \text{ or } \frac{1-3\alpha}{100\alpha - 1{,}300\alpha^2}$$ [1]

Since R is the unique positive root, we have $R = \frac{1-3\alpha}{100\alpha - 1{,}300\alpha^2}$. [½]

[Total 4]

(iii) ***Maximum value of the adjustment coefficient***

In order to maximise R we need to differentiate the above equation with respect to α. Using the quotient rule, we get:

$$\frac{dR}{d\alpha} = \frac{(100\alpha - 1{,}300\alpha^2)(-3) - (1-3\alpha)(100 - 2{,}600\alpha)}{(100\alpha - 1{,}300\alpha^2)^2}$$ [1]

$$= \frac{-300\alpha + 3{,}900\alpha^2 - 100 + 2{,}600\alpha + 300\alpha - 7{,}800\alpha^2}{(100\alpha - 1{,}300\alpha^2)^2}$$

$$= \frac{2{,}600\alpha - 3{,}900\alpha^2 - 100}{(100\alpha - 1{,}300\alpha^2)^2}$$ [1]

Setting this equal to zero, we get:

$$39\alpha^2 - 26\alpha + 1 = 0 \quad \Rightarrow \quad \alpha = \frac{26 \pm \sqrt{26^2 - 4 \times 39}}{2 \times 39} = 0.6257 \text{ or } 0.0410$$ [1]

But from (i) we know that $\alpha > \frac{1}{3}$, so we have $\alpha = 0.6257$. [½]

Markers please note that students can also substitute the given value of α *in to show that the derivative is zero.*

Substituting this value back into the equation in Part (ii), we get $R = 0.0019649$. [½]

[Total 4]