PORTFOLIO MANAGEMENT

MODERN PORTFOLIO THEORY I

PORTFOLIO MANAGEMENT

Table of contents

1. The risk / return framework

What happens if individuals' current income is not sufficient to cover their consumption? They will need to borrow the difference. On the other hand, when individuals' current income exceeds their consumption, they tend to save the excess. This imbalance creates a market; instead of putting their savings under their mattresses, individuals can give up immediate possession of their savings for a higher level of future consumption by lending their savings. This is called *investment.*

The **required rate of return** is what investors who lend their savings will demand in order to compensates them for the *time*, the expected rate of *inflation*, and the *uncertainty* of the return.

Time: The rate of exchange between *future consumption* and *current consumption* is called the *real risk-free rate.* This rate of exchange is also sometimes referred to as *the pure time value of money*.

Inflation: The real risk-free rate of interest does only compensate the investor for the passage of time. In absence of inflation and uncertainty of the returns the required rate of return and the real risk-free rate of interest would be the same. However, historically, inflation has almost never been null. Therefore the investor will need to account for it if he does not want his purchasing power to decline over time.

One of the problems is obviously that we do not know what the future inflation will be. Therefore the best we can do is to estimate what the future inflation will be; we talk about the *expected inflation.*

The pure rate of interest increased by the expected rate of inflation is called the *nominal risk-free rate.*

Uncertainty: If the future payment form the investment is not certain, the investor will demand a *risk-premium* to reward him for taking this additional risk.

Adding the risk-premium to the nominal risk-free rate yields what we defined above as the required rate of return.

Before we have a closer look at the different components of the required rate of return we need to understand how return and risk are typically measured.

1.1 Return and measures of return

1.1.1 Holding period return

The most common measure of return is the **holding period return**, also called rate of return over a given period. For an asset paying no dividend or coupon, such as gold, the rate of return equals the percentage change in the price of the asset:

$$R_{t-1,t} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

where $R_{t-1,t}$ is the return of the asset over the time period going from time t-1 to t, P_{t-1} is the price of the asset at time t–1, and P_t is the price of the asset at time t.

Example:
An investor buys one ounce of gold at time t=0 for EUR 350 and sells it at time t=1 at EUR 400. Over the period, the investor's return is:

$$R_{0,1} = \frac{400-350}{350} = 0.1429 = 14.29\%$$

However, most financial assets have intermediate cash flows taking the form of dividends or coupon payments. If the return on these assets is computed immediately after the dividend or coupon payment, the return equals:

$$R_{t-1,t} = \frac{D_t + P_t - P_{t-1}}{P_{t-1}}$$

where D_t is the dividend paid at time t.

Example:
An investor buys a stock at time t=0 for EUR 100; at time t=1 a dividend of EUR 10 is paid; at the same time, the stock is priced at EUR 105.
Over the period, the investor's return is:

$$R_{0,1} = \frac{10+105-100}{100} = 0.15 = 15\%$$

It is a common practice to assume that one time period is one year. In most cases, however, payments are made during the time period; for instance, quarterly dividend payments are common in the USA. The problem is how to deal with these intermediate cash flows. The easiest way to measure returns in the presence of intermediate cash flows is to assume that these payments are re-invested at a given rate. Then the above formula changes to:

$$R_{t-1,t} = \frac{D_\tau \cdot (1+R^*_{\tau,t})^{t-\tau} + P_t - P_{t-1}}{P_{t-1}}$$

where D_τ is the dividend paid at time τ and $R^*_{\tau,t}$ is the annualised rate of return used for the reinvestment, from time τ to time *t*. Generally, $R^*_{\tau,t}$ is set equal to the risk-free rate for the period under consideration.

Example:
An investor buys a stock at time t=0 for EUR 100; at time t=0.5 a dividend of EUR 10 is paid, which is reinvested at a risk-free rate of 4% p.a. for the rest of the period; at time t=1, the stock is priced at EUR 105. Over the period, the investor's return is:

$$R_{0,1} = \frac{10 \cdot 1.04^{0.5} + 105 - 100}{100} = 0.152 = 15.2\%$$

Since we are computing the return on a stock, it may be more appropriate (although sometimes cumbersome) to consider that the dividend payments are directly reinvested into the asset itself. Then the holding period return could be computed as:

$$R_{t-1,t} = \frac{D_\tau \cdot \frac{P_t}{P_\tau} + P_t - P_{t-1}}{P_{t-1}}$$

where P_τ is the price of the asset at time τ.

These formulas can be generalised when any number k of intermediate payments are made:

$$R_{t-1,t} = \frac{P_t - P_{t-1} + \sum_{j=1}^{k} D_{\tau_j} \cdot \left(1 + R^*_{\tau_j,t}\right)}{P_{t-1}}$$

where τ_j is the time of the j[th] dividend or coupon payment, such that $t-1 \le \tau_j \le t$, and $R^*_{\tau,t}$ is a risk-free rate for the considered period (τ_j to t) if the reinvestment is done in a risk-free asset, or $R^*_{\tau_j,t} = \frac{P_t}{P_{\tau_j}} - 1$ if time τ_j dividend is reinvested in the same asset.

1.1.2 Arithmetic versus geometric average of holding period returns

An investor will typically hold assets over more than one time period and he will be probably interested in computing the average return per period on his investment. Take for instance an investment horizon (the holding period) of two years. Now, the investor wants to compute the average yearly return. The first and intuitive approach is to take the arithmetic average of the holding period returns over the period considered, i.e. the sum of the holding period returns divided by the number of compounding periods in the holding period:

$$\overline{R}^{(a)}_{0,T} = \frac{1}{T} \cdot \sum_{t=1}^{T} R_{t-1,t}$$

where

$R_{t-1,t}$	holding period returns
T	number of compounding periods in the holding period

The following example shows that this is not the appropriate method.

Example:
Let there be three stocks A, B, C, held for two time periods; the ends of period prices are:

	t=0	Period 1		Period 2	
	Price	**Price**	**Period return**	**Price**	**Period return**
A	EUR 100	EUR 110	10%	EUR 121	10%
B	EUR 100	EUR 150	50%	EUR 121	-19.3%
C	EUR 100	EUR 200	100%	EUR 121	-39.5%

It is clear that the three assets have yielded the same return over the two periods since the beginning and end of period values are identical.

Yet, the arithmetic mean $\frac{1}{2}\cdot\left(\frac{P_1-P_0}{P_0}+\frac{P_2-P_1}{P_1}\right)$ gives:

$$\text{for A}: \overline{R}^{(a)}_{0,2}=\frac{1}{2}\cdot\left(\frac{110-100}{100}+\frac{121-110}{110}\right)=\frac{(10\%+10\%)}{2}=10\%$$

$$\text{for B}: \overline{R}^{(a)}_{0,2}=\frac{1}{2}\cdot\left(\frac{150-100}{100}+\frac{121-150}{150}\right)=\frac{(50\%-19.3\%)}{2}=15.33\%$$

$$\text{for C}: \overline{R}^{(a)}_{0,2}=\frac{1}{2}\cdot\left(\frac{200-100}{100}+\frac{121-200}{200}\right)=\frac{(100\%-39.5\%)}{2}=30.25\%$$

This seems to indicate that C has better performed which is obviously not true.

The appropriate way to average holding period returns is to take the **geometric average of the holding period returns** over the period under consideration:

$$\begin{aligned}
R^{(g)}_{0,T} &= \sqrt[T]{(1+R_{0,1})\cdot(1+R_{1,2})\cdot\ldots\cdot(1+R_{T-1,T})}-1\\
&= \sqrt[T]{\left(\frac{P_1}{P_0}\right)\cdot\left(\frac{P_2}{P_1}\right)\cdot\ldots\cdot\left(\frac{P_T}{P_{T-1}}\right)}-1\\
&= \sqrt[T]{\left(\frac{P_T}{P_0}\right)}-1\\
&= \sqrt[T]{(1+R_{0,T})}-1\\
&= (1+R_{0,T})^{\frac{1}{T}}-1
\end{aligned}$$

This is because holding period returns are multiplicative, but not additive.

Example:
Using the same data as our previous example, the above equation yields for the three stocks:

$$\text{for A}: \overline{R}^{(g)}_{0,2}=\sqrt{\left(\frac{P_1}{P_0}\right)\cdot\left(\frac{P_2}{P_1}\right)}-1=\sqrt{\left(\frac{110}{100}\right)\cdot\left(\frac{121}{110}\right)}-1=\sqrt{1.10\cdot1.10}-1=10\%$$

$$\text{for B}: \overline{R}^{(g)}_{0,2}=\sqrt{\left(\frac{P_1}{P_0}\right)\cdot\left(\frac{P_2}{P_1}\right)}-1=\sqrt{\left(\frac{150}{100}\right)\cdot\left(\frac{121}{150}\right)}-1=\sqrt{1.50\cdot0.81}-1=10\%$$

$$\text{for C}: \overline{R}^{(g)}_{0,2}=\sqrt{\left(\frac{P_1}{P_0}\right)\cdot\left(\frac{P_2}{P_1}\right)}-1=\sqrt{\left(\frac{200}{100}\right)\cdot\left(\frac{121}{200}\right)}-1=\sqrt{2.00\cdot0.605}-1=10\%$$

We should note that a consequence of geometric averaging is that a given percentage market increase followed by the same percentage decrease does not lead to a zero average return over the two periods!

Example:
During the first year, a stock price increases by $R_{0,1}$=10%. The second year, the stock price decreases by 10%, i.e. $R_{1,2}$=–10%.
The average annual return during the two years is

$$\overline{R}_{0,2}^{(g)} = \sqrt[2]{(1+0.10)\cdot(1-0.10)} - 1 = \sqrt{0.99} - 1 = 0.995 - 1 = -0.5\%$$

1.1.3 Time value of money: compounding and discounting

1.1.3.1 Compounding period equal to the holding period

As we have seen receiving EUR 1 today is worth more than receiving EUR 1 tomorrow. Both EUR are linked through the concept of interest and we can write:

$$1 + R_{t-1,t} = \frac{P_t}{P_{t-1}}$$

or

$$\left(1 + R_{t-1,t}\right) \cdot P_{t-1} = P_t$$

which is the fundamental equation of the time value of money, as it defines the future value at time t of an amount P_{t-1} invested at a rate $R_{t-1,t}$ over one period:

$$\text{Future value} = \text{Present value} \cdot (1 + \text{Interest rate})$$

The term $(1 + R_{t-1,t})$ is generally called the **capitalisation factor** for the period t-1 to t. We can also write

$$P_{t-1} = \frac{1}{\left(1 + R_{t-1,t}\right)} \cdot P_t$$

that is

$$\text{Present value} = \frac{1}{(1+\text{Interest rate})} \cdot \text{Future value}$$

The above equation defines the present value of an amount P_t to be received at the end of a given period if the rate of return over that period is $R_{t-1,t}$.

The term $\frac{1}{(1 + R_{t-1,t})}$ is generally called the **discount factor** for the period from t-1 to t.

1.1.3.2 Compounding period shorter than the holding period

So far, we have been considering returns over a single period (the holding period), at the end of which interest was calculated and added to the principal. But what happens if the holding period differs from the **compounding period**, that is the period at the end of which interest is calculated and added to the principal amount?

Let us first consider the case of a compounding period that is shorter than the holding period. For instance, we could have two interest payments (at time 1 and 2) during the holding period, as illustrated below:

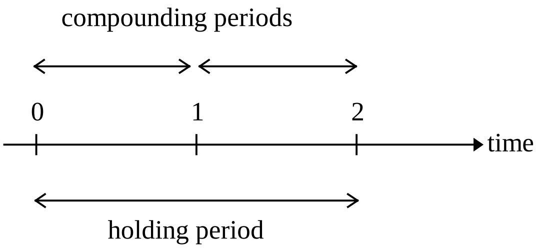


Figure 1-1: Holding period longer than compounding period

In such a case, the principal amount invested at time 0 will earn interest income at time 1 (equal to $R_{0,1}$) and another interest payment at time 2 (equal to $R_{1,2}$). But the interest payment received at time 1 (equal to $R_{0,1}$) can be reinvested from time 1 to time 2, which creates what is called **compound interest**, or interest on interest (equal to $R_{0,1} \cdot R_{1,2}$), to be received at time 2. This has to be considered in the holding period return calculation. We have:

$$1+R_{0,2}=(1+R_{0,1})\cdot(1+R_{1,2})=1+R_{0,1}+R_{1,2}+\left(R_{0,1}\cdot R_{1,2}\right)$$

In fact, neglecting compounding is equivalent to setting the ($R_{0,1} \cdot R_{1,2}$) term equal to zero.

Example:
If an investor deposits EUR 100 on his bank account at a 10% annual rate, he will receive EUR 10 at the end of the first year. At the end of the second year, he will receive again EUR 10, corresponding to the interest paid on the principal amount, plus an extra amount of EUR 1 corresponding to the second year interests on the first year interest (10% on EUR 10).
Thus, over two years, the holding period return is (100+10+10+1)/100–1=21%. If we neglect the compound interest, the holding period return is (100+10+10)/100–1=20%.

More generally, if $R_{t,\, t+1}$ is the rate of return to be paid from time t to time t+1 (one period), and if the proceeds from one period can be reinvested immediately, the effective rate of return from time 0 to time T (over T periods) is given by the product of the individual period returns, that is:

$$1+R_{0,T}=\left(1+R_{0,1}\right)\cdot\left(1+R_{1,2}\right)\cdot\left(1+R_{2,3}\right)\cdot\ldots\cdot\left(1+R_{T-1,T}\right)$$

Example:
A deposit of EUR 100 on a bank account earns interest at the rate of 7% the first year, 9% the second year, and 10% the third year. Interest amounts are credited annually, at the end of each year, and are immediately considered for the following year's interest computation.
The value on the account at the end of the third year will be 100·1.07·1.09·1.10= EUR 128.29. The effective rate of return over the three years is therefore 28.29%.

A special case is the situation when all rates are equal, that is, $R_{0,1}=R_{1,2}=\ldots=R_{T-1,T}$. In such a case, we have:

$$1+R_{0,T}=\left(1+R_{0,1}\right)^{T}$$

The following table shows the final value of an initial amount of EUR 100 invested at a simple versus a compound rate, for various rates and various holding periods.

Time (Years)	Simple Rate	Compound Rate	Simple Rate	Compound Rate	Simple Rate	Compound Rate
Rate	*2%*	*2%*	*7%*	*7%*	*10%*	*10%*
1	102.00	102.00	107.00	107.00	110.00	110.00
2	104.00	104.04	114.00	114.49	120.00	121.00
3	106.00	106.12	121.00	122.50	130.00	133.10
4	108.00	108.24	128.00	131.08	140.00	146.41
5	110.00	110.41	135.00	140.26	150.00	161.05
6	112.00	112.62	142.00	150.07	160.00	177.16
7	114.00	114.87	149.00	160.58	170.00	194.87
8	116.00	117.17	156.00	171.82	180.00	214.36
9	118.00	119.51	163.00	183.85	190.00	235.79
10	120.00	121.90	170.00	196.72	200.00	259.37
15	130.00	134.59	205.00	275.90	250.00	417.72
20	140.00	148.59	240.00	386.97	300.00	672.75

Table 1-1: Impact of compounding

It is easy to see that neglecting compound interest can cause big errors, particularly in calculations carried out over long periods with large interest rates.

1.1.3.3 Compounding period longer than the holding period

Let us now consider the case of a compounding period that is longer than the holding period. For instance, the holding period is τ days while the compounding period is one year, as illustrated below:

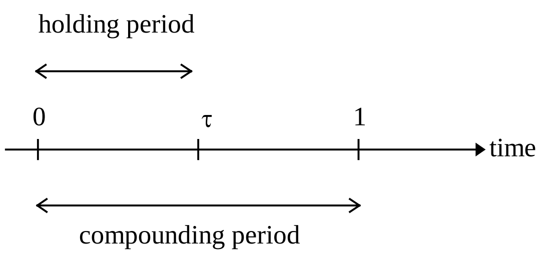


Figure 1-2: Holding period shorter than compounding period

In such a case, the principal amount invested at time 0 will earn interest income at time 1 (equal to $R_{0,1}$). But what is the rate of return earned on the investment from time 0 to time τ? We have:

$$\left(1+R_{0,\tau}\right)=\left(1+R_{0,1}\right)^{\tau}$$

where τ is measured relatively to the total period (time 0 to time 1) length. Let us illustrate this.

Example:
Let us again think of a stock purchased at time 0 at EUR 100 whose time 1 value (a year later) is EUR 105 and suppose ownership of this stock has entitled its owner to four (quarterly) dividends of EUR 2.5 which have been reinvested at a risk-free annual rate of return of 5%, the return on the stock would have been:

$$R_{0,1} = \frac{2.5\cdot(1+0.05)^{270/360}}{100} + \frac{2.5\cdot(1+0.05)^{180/360}}{100} + \frac{2.5\cdot(1+0.05)^{90/360}}{100} + \frac{2.5\cdot(1+0.05)^{0}}{100} + \frac{105-100}{100}$$
$$= 15.18\%$$

Here, the τ's were measured as the number of days before time t payment is made.

1.1.3.4 Continuously compounded returns

Let us examine what happens if we decide to *compound more and more often*. For instance, what is the impact on the effective rate of return ($R_{0,1}^{eff}$) of compounding twice in the holding period (at a rate of $\frac{R_{0,1}^{nom.}}{2}$) rather than once at a rate $R_{0,1}^{nom.}$? The annual effective rate is determined with the following equation[1]:

$$1+R_{0.1}^{eff} = \left(1+\frac{R_{0,1}^{nom}}{2}\right)^2$$

More generally, if we increase the frequency of payments and decide to pay interest m times a year at a rate $\frac{R_{0,1}^{nom.}}{m}$, the annual effective rate of return is determined as follows:

$$1+R_{0,1}^{eff} = \left(1+\frac{R_{0,1}^{nom}}{m}\right)^m$$

As m increases, the quantity $\left(1+\frac{R_{0,1}^{nom}}{m}\right)^m$ tends to the exponential of $R_{0,1}^{nom.}$ and we have:

$$1+R_{0,1}^{eff} = \lim_{m\to\infty}\left(1+\frac{R_{0,1}^{nom}}{m}\right)^m = e^{R_{0,1}^{nom.}}$$

with e=2.71828.

At the limit, we can derive the following general formula for **continuously compounded return** or **instantaneous return**, i.e. the return over an infinitesimal (i.e. as short as possible) period that we will denote by a lower case letter:

$$r = \lim_{m\to\infty} R_{0,1}^{nom} = \ln\left(1+R_{0,1}^{eff}\right)$$

Thus, for each discrete time rate (or simple rate), there is a continuous time rate that is defined by the above equation. But one should not forget that the continuously compounded rate is only an *approximation* of the discrete rate valid for an infinitesimal time period. In fact, for a small difference in price (which is normally the case if the time period is small), using the fact that

1 Instantaneous compounding will lead to a higher future value: as interest is paid continuously, there is more interest on interest.

$$\ln\left(\frac{x}{y}\right)=\ln(x)-\ln(y)$$

and that

$$d\left(\ln(x)\right)=\frac{dx}{x}$$

one has:

$$\ln\left(1+R_{t-1,t}\right)=\ln\left(\frac{P_t}{P_{t-1}}\right)=\ln\left(P_t\right)-\ln\left(P_{t-1}\right)\cong d\left(\ln(P)\right)=\frac{dP}{P}\cong\frac{P_t-P_{t-1}}{P_{t-1}}=r_{t-1,t}$$

where d denotes the differential. In other words, the difference between the natural logs of asset prices is the measure of the percentage change in the asset price. For instance, the continuously compounded return of a stock just after the dividend payment is given by:

$$r_{t-1,t}=\ln\left(\frac{P_t+D_t}{P_{t-1}}\right)$$

This measure would be exact only if the differences were very small.

The convenience of this method (as we will see later) justifies its utilisation, especially for *short period returns* (daily, weekly). However one has to remember that it is only an *approximation.* For longer holding periods (which imply generally larger returns) the error can be substantial as shown in the table below.

Price in t=1 (base in t=0 is 100)	Holding Period Return	Continuously Compounded Return
50	–50%	ln(0.50) = –69.3%
80	–20%	ln(0.80) = –22.3%
90	–10%	ln(0.90) = –10.5%
95	–5%	ln(0.95) = –5.1%
97	–3%	ln(0.97) = –3.1%
99	–1%	ln(0.99) = –1.0%
100	0%	ln(1.00) = 0.0%
101	1%	ln(1.01) = 1.0%
103	3%	ln(1.03) = 2.9%
105	5%	ln(1.05) = 4.9%
110	10%	ln(1.10) = 9.5%
120	20%	ln(1.20) = 18.2%
150	50%	ln(1.50) = 40.5%

Table 1-2: Holding period returns vs. continuously compounded returns

The approximation error increases as the return is increasing, as illustrated by the following figure.

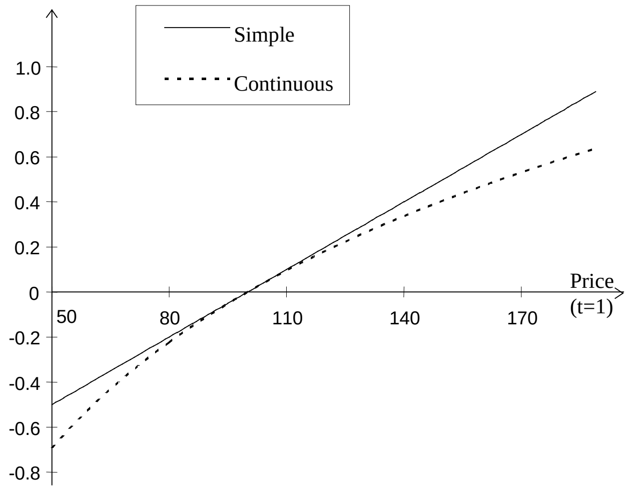


Figure 1-3: Approximating simple returns by continuously compounded returns

If using continuously compounded returns creates approximation errors, why should we use them? Let us examine what happens if we decide to compound returns over two infinitesimal periods. The continuously compounded rate over these two periods, denoted r_2, is equal to

$$r_2 = \ln\left(1+R_{0,1}\right)^2 = 2\cdot\ln\left(1+R_{0,1}\right) = 2\cdot r$$

More generally, using continuous time, we assume that compounding takes place at every moment in time. As a consequence of the standard property of logarithm, we thus have:

$$\text{Future value} = \text{Actual value}\cdot e^{\text{Time}\,\cdot\,\text{Instantaneous interest rate}}$$

We see that the continuously compounded rate of return over N periods is simply N times the continuously compounded rate of return. Thus, *while simple returns are multiplicative, continuously compounded returns are additive.* This makes using continuously compounded returns easy to use for discounting and compounding.

1.1.3.5 Averaging continuously compounded return

Averaging continuously compounded returns is simple. As we already saw, continuously compounded returns are additive. Thus, we can use the *arithmetic average of the continuously compounded returns* over the period considered.

$$\overline{r} = \overline{r}_{0,T}^{(a)} = \frac{1}{T}\cdot\sum_{t=1}^{T} r_{t-1,t}$$

Example:
Let there be three stocks A, B, C, held for two time periods; the prices at the end of each period are the same as in our previous example.

	t=0	Period 1		Period 2	
	Price	**Price**	**Continuous Compounded return**	**Price**	**Continuous Compounded return**
A	EUR 100	EUR 110	9.53%	EUR 121	9.53%
B	EUR 100	EUR 150	40.54%	EUR 121	–21.48%
C	EUR 100	EUR 200	69.31%	EUR 121	–50.25%

It is clear that the three assets have yielded the same return over the two periods since the beginning and end of period values are identical. In fact, using continuously compounded returns, we have:

$$\begin{aligned}\bar{r}_{0,2} &= \frac{1}{2}\left(r_{0,1} + r_{1,2}\right) = \frac{1}{2}\left(\ln\left(\frac{P_1}{P_0}\right) + \ln\left(\frac{P_2}{P_1}\right)\right) \\ &= \frac{1}{2}\left(\ln(P_1) - \ln(P_0) + \ln(P_2) - \ln(P_1)\right) \\ &= \frac{1}{2}\left(\ln(P_2) - \ln(P_0)\right) = \frac{1}{2}\ln\left(\frac{P_2}{P_0}\right)\end{aligned}$$

When taking the numbers of the previous section, this gives:

$$\text{for A}: \frac{1}{2}\cdot\left(\ln\left(\frac{110}{100}\right) + \ln\left(\frac{121}{110}\right)\right) = \frac{1}{2}\cdot(9.53\% + 9.53\%) = 9.53\%$$

$$\text{for B}: \frac{1}{2}\cdot\left(\ln\left(\frac{150}{100}\right) + \ln\left(\frac{121}{150}\right)\right) = \frac{1}{2}\cdot(40.54\% - 21.48\%) = 9.53\%$$

$$\text{for C}: \frac{1}{2}\cdot\left(\ln\left(\frac{200}{100}\right) + \ln\left(\frac{121}{200}\right)\right) = \frac{1}{2}\cdot(69.31\% - 50.25\%) = 9.53\%$$

9.53% is the average continuously compounded return over periods 1 and 2 for stocks A, B and C. We can prove it as follows:

$$e^{0.0953} - 1 = 10\%$$

$$100\cdot(1+10\%)^2 = 121$$

1.1.4 *Annualisation of returns*

In some cases the period considered is smaller than one year, for instance daily, monthly or quarterly. Nevertheless, returns are generally compared on an annual basis. For this reason, returns have to be annualised. There is again a difference in computation between holding period and continuously compounded returns.

1.1.4.1 *Annualising holding period returns*

Consider a time period of τ days over which a simple return R_τ has been obtained. We want to express the rate R_τ as an annualised simple rate of return; then the following formula can be used:

$$R_{an} = \left(1 + R_\tau\right)^{360/\tau} - 1$$

Example:
Let there be a stock worth EUR 100 on 31st of December and EUR 110 on 31st of March. The annualised simple return obtained for this stock is:

$$R_{an} = \left(\frac{110}{100}\right)^{360/90} - 1 = 46.41\%$$

Note that the convention of 360 days versus 365 or the effective number of days varies from one country to another.

1.1.4.2 *Annualising continuously compounded returns*

Since continuously compounded returns are additive, if r_τ is the continuously compounded rate of return earned over a period of τ days, the corresponding annualised return is:

$$r_{an} = \frac{360}{\tau} \cdot r_\tau$$

Example:
Let there be a stock worth EUR 100 on 31st of December and EUR 110 on 31st of March. The annualised continuously compounded return obtained for this stock is:

$$r_{an} = \frac{360}{90} \cdot \ln\left(\frac{110}{100}\right) = 38.12\%$$

Note again the difference with our previous example result due to the approximation when computing continuously compounded returns and the fact that we are using large numbers as examples.

1.2 Risk

The previous sections have made clear how to determine the holding period return from past data, i.e. once the results of a given investment are known (ex post). But interesting questions arise at the moment in time when the choice of investment is to be made. At this moment, that is, *ex ante*, returns are not known[2]. One can talk of estimates or prospects, but these concepts are difficult to describe precisely. The key step taken by modern portfolio theory, and most of modern finance, is to *describe ex-ante returns in probabilistic terms*, i.e. to view holding period returns as random variables and to compute an *expected return* denoted E(R) or E(r). Moreover, the notion of an expected return has to be considered in pair with the corresponding risk. In order to be able to quantify risk, we will proceed with a quick review of probability concepts.

1.2.1 *Probability concepts*

A **sample space** (also called an **event space**) will be defined as the set of all possible outcomes (or possible ‘states of nature’) of the random variable under observation. With

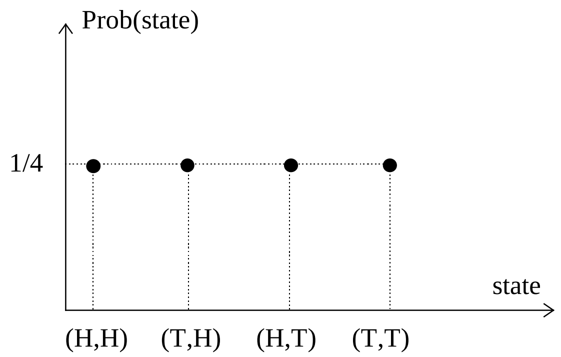
2 Otherwise every investor would simply invest his/her entire wealth in the one asset paying the highest return.

every state s of the sample space, we can associate a number denoted π_S and called the **probability** of state s[3].

> **Example:**
> When one flips a coin twice, the sample space consists of 4 outcomes which are given by {(head, head), (tail, tail), (head, tail), (tail, head)}. One can easily verify that it includes all the possible states. To each state, we can associate a probability of 1/4, as each state is equally probable.

Then, we can define a **probability distribution function** as a function p(s) that associates with state s the probability π_S of being in that state.

> **Example:**
> If we take our previous example (flipping a coin twice), the probability distribution function will be a discontinuous function represented as follows:



Let us try to apply these concepts to returns. To assign probabilities to our states, we first need to define the states themselves. When considering financial assets in each of the possible state of nature, the asset under study will take a different value and the return on the asset will be affected correspondingly. Thus, one has to outline several possible scenarios for the future depending on how precise one wants to be and how much information is available, each scenario associated with a certain probability. Formally, the states of nature should be defined in such a way as to cover all the relevant possibilities, so that their probabilities sum up to 1.

1.2.1.1 Probability trees

A good way of representing individual outcomes are event trees[4]. Often, binomial trees are used, where at time t+1, only two states of nature are considered possible, given the state of nature at time t. They are particularly suitable for states of nature that follow one another in time.

For instance, let there be an investor who on January 1, is considering what the return of the SMI might be over the next twelve months. Suppose his assessment is as follows: there is a 50-50 chance of gaining or losing 10%. If the SMI is at 7000, this can be represented by the following tree:

---

3 Intuitively, the probability π_S gives the ‘number of chances’ out of 100% to have state s if we generate randomly an ‘event’ from the sample space. From this interpretation, it follows that p(s) must be between 0 and 1 (as no event can have a negative probability, nor more than 100% chances to occur). Furthermore, the sum of probabilities for all event in the sample space should be 1 (as all the possible events should be in the sample space, we have 100% chances to get an event of the sample space if we generate one randomly).

4 This type of modelling approach is also very often used for the pricing of derivative securities in the absence of closed form solution (for instance American options on dividend paying stocks).

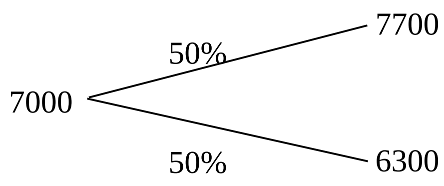


Figure 1-4: A simple binomial tree

This evaluation is relatively rough, since there are only two possible outcomes after one year. Let us refine the estimates by assuming that there is a 50-50 chance to have a period return of –5% or +5% until July and –5% or +5% from July until December. This can be represented by:

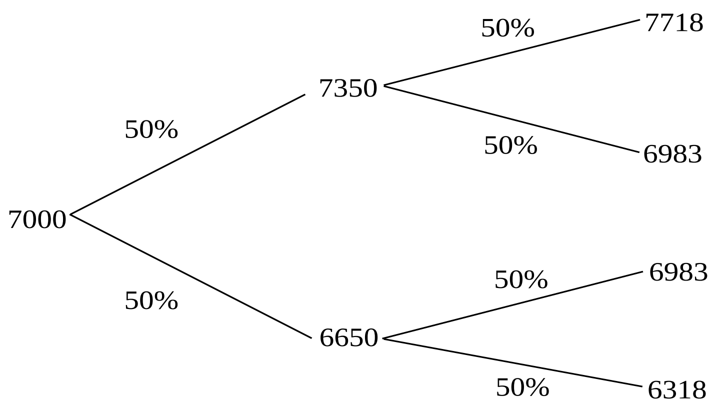


Figure 1-5: A two-period binomial tree

Now we already have four possible returns for the SMI for the current year. This process can be continued indefinitely. If we take monthly estimates, we will have 2^{12}=4'096 possible outcomes and so forth. Hence, this type of representation is a quite powerful tool.

In Figure 1-5, there are four outcomes, but only three possible values the SMI can take. This comes from the fact that the returns chosen for the two time periods are the same. This nevertheless illustrates that there will typically be more values close to the middle of the range of possible outcomes than towards the extremes.

Distribution trees do not need to be binomial. For instance, it is possible that at every instant t, the price of the asset either goes up, remains the same or goes down. This type of tree is often referred to as a multinomial tree. At the most general level, the number of possible alternatives is not countable and probabilities are represented by **continuous distributions**. We will present such a concept in the next section.

1.2.1.2 Probability distributions

A possible way of representing an infinite number of individual outcomes is to group them into categories. For instance, let there be an investor considering a one-year investment in the UK stock market, and who is interested in predicting the return on the UK market over the next year. The only available information is the list of the last 50 years of annual returns on a market index (FTSE All-Share Index).

1959	43.40%	1969	-15.19%	1979	4.35%	1989	30.01%	1999	21.25%
1960	-4.71%	1970	-7.52%	1980	27.07%	1990	-14.31%	2000	-7.97%
1961	-2.52%	1971	41.93%	1981	7.24%	1991	15.06%	2001	-15.41%
1962	-1.81%	1972	12.82%	1982	22.07%	1992	14.83%	2002	-24.97%
1963	10.60%	1973	-31.36%	1983	23.10%	1993	23.35%	2003	16.57%
1964	-10.00%	1974	-55.34%	1984	26.02%	1994	-9.55%	2004	9.28%
1965	6.73%	1975	136.33%	1985	15.18%	1995	18.48%	2005	-1.10%
1966	-9.31%	1976	-3.87%	1986	22.34%	1996	11.71%	2006	35.03%
1967	28.98%	1977	41.18%	1987	4.16%	1997	19.73%	2007	2.03%
1968	43.36%	1978	2.65%	1988	6.48%	1998	10.91%	2008	-32.78%

Table 1-3: List of the last 50 years returns

From these, the investor can count the number of returns below zero percent and the number of returns greater than zero percent, and divide the numbers by fifty. This gives 33/50=0.66 chances out of one to have a return greater than zero, and 17/50=0.34 chances out of one to have a return lower than zero. This defines a (very simple) **discrete distribution** that we can write as:

$$\begin{cases} \text{Prob}(R \leq 0\%)=0.34 \\ \text{Prob}(R>0\%)=0.66 \end{cases}$$

Graphically, probability distributions are generally portrayed as **histograms**, with possible outcomes represented on the horizontal axis and probabilities on the vertical axis[5]. Thus, the box associated with $R \leq 0\%$ will have an area of 0.34 (or 34% of all boxes total surface), and the box associated with $R > 0\%$ will have an area of 0.66 (or 66% of all boxes total surface).

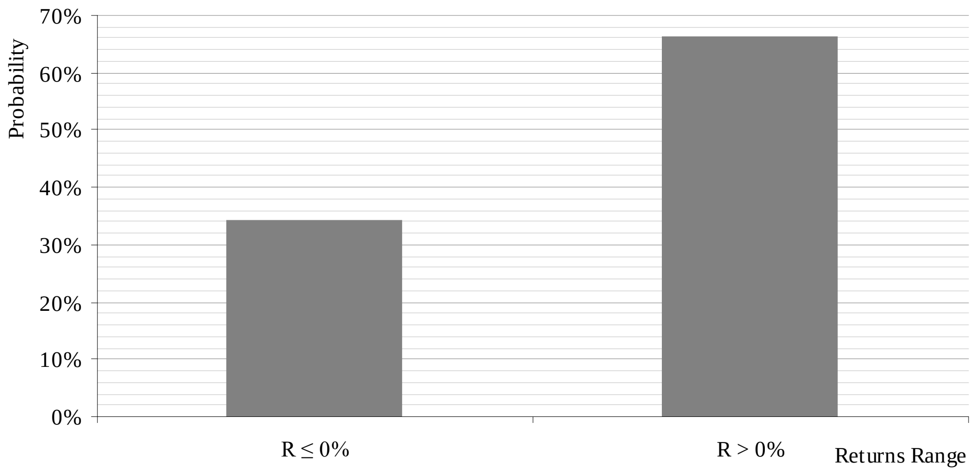


Figure 1-6: A first probability distribution

This information is useful, but in order to make more precise inferences, we have to divide the data into narrower ranges. For instance, we could consider the following ranges: less than -15%, between–15% and 0%, between 0% and 15%, between 15% and 30%, and over 30%.

Applying the same methodology would yield the following distribution:

5 Note that the probabilities are represented by the area of the boxes, not the height. Otherwise, the picture would be distorted when intervals are of different width.

$$\begin{cases} \text{Prob}(R \leq -15\%)=0.12 \\ \text{Prob}(-15\%<R \leq 0\%)=0.22 \\ \text{Prob}(0\%<R \leq 15\%)=0.26 \\ \text{Prob}(15\%<R \leq 30\%)=0.26 \\ \text{Prob}(R > 30\%)=0.14 \end{cases}$$

this can also be represented as a histogram:

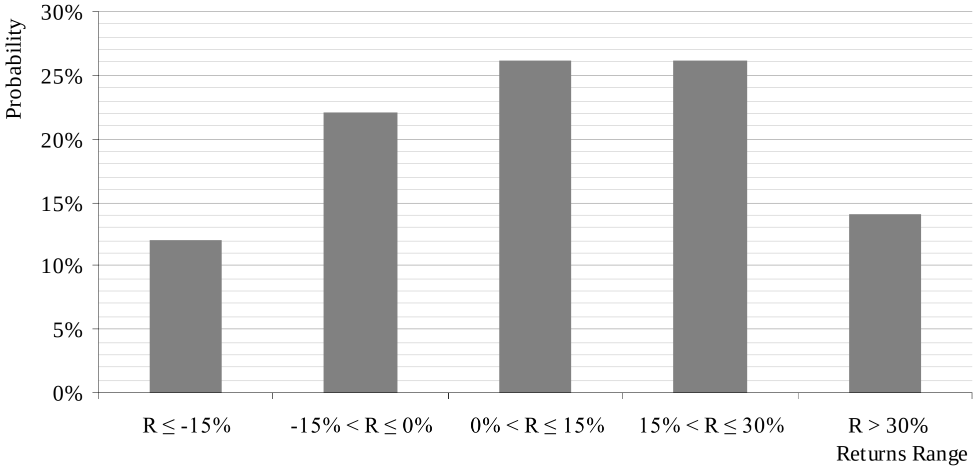


Figure 1-7: A second probability distribution

If we still wish to make more precise inferences, we would have to consider narrower return intervals. This refinement can go on and on. If categories chosen are small enough, it is possible to attribute one probability to every return. At the limit, under the condition that the number of outcomes is large enough[6], we obtain a **continuous probability distribution**, represented as a curve[7].

6 Note that in our example, we are in fact limited by the number of observations (50). To consider narrower intervals, we may have to augment the sample by extending the measurement period (!) or to consider shorter time periods (for instance, using monthly returns over the same period would provide us $50 \cdot 12=600$ observations).

7 In Figure Figure 1-8 it has been plotted the probability distribution function of a normal distribution. But there are many other different examples of continuous distributions.

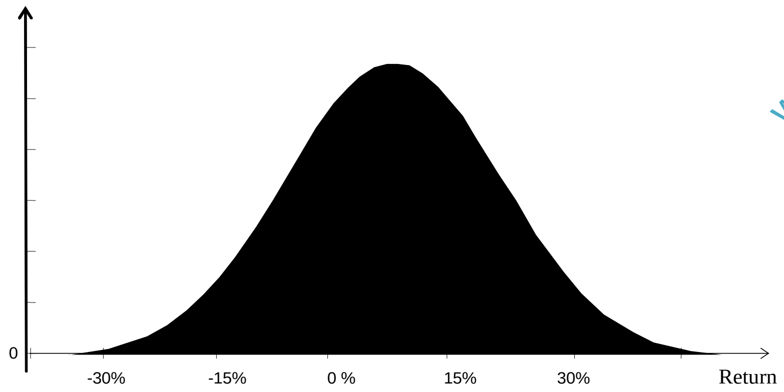


Figure 1-8: Limiting case: a continuous distribution

The equation of this curve is called the **probability density** function of the distribution. It is written as a function of the return.

When we constructed the histogram, we said that the probabilities of the return to be in a given interval were represented by the (relative) surface associated with the given interval. The principle can be applied to continuous distributions. The probability of achieving a given return R^* will then be estimated as the area under the probability density curve in a very small interval around R^*. More generally, the probability for the return R to be lower than a given value R^* will be given by the area under the curve from $-\infty$ to R^*. It is called the **integral of the probability density** from $-\infty$ to R^* and it corresponds to the shaded area in the following figure.

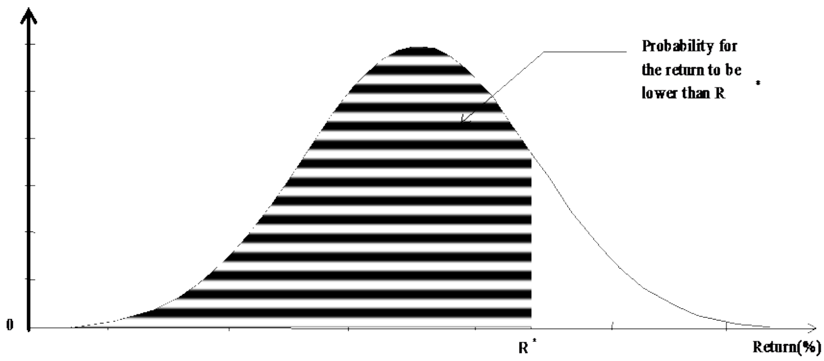


Figure 1-9: Probability distribution and upper bound

Similarly, the non-shaded area represents the probability for the return to be higher than R^*, that is, one minus the probability for the return to be lower than R^*. Thus, an important property of the continuous distributions is that the total area under their curve is bounded and equal to 1.

Thus, the probability of the random variable R taking a higher value than the bound R^* is

$$\Pr ob\left[R > R^*\right] = 1 - \Pr ob\left[R \le R^*\right]$$

The above relationship is based on the fact that the total sum of the probabilities of all outcomes is 100%, hence the probability of an event taking place equals 100% minus the probability that it does not take place.

More formally we can define the integral of the probability density as the cumulative distribution function. This function takes values ranging from 0 to 1, or 0 % to 100 %. The figure below depicts the cumulative distribution function of a standard normal distribution.

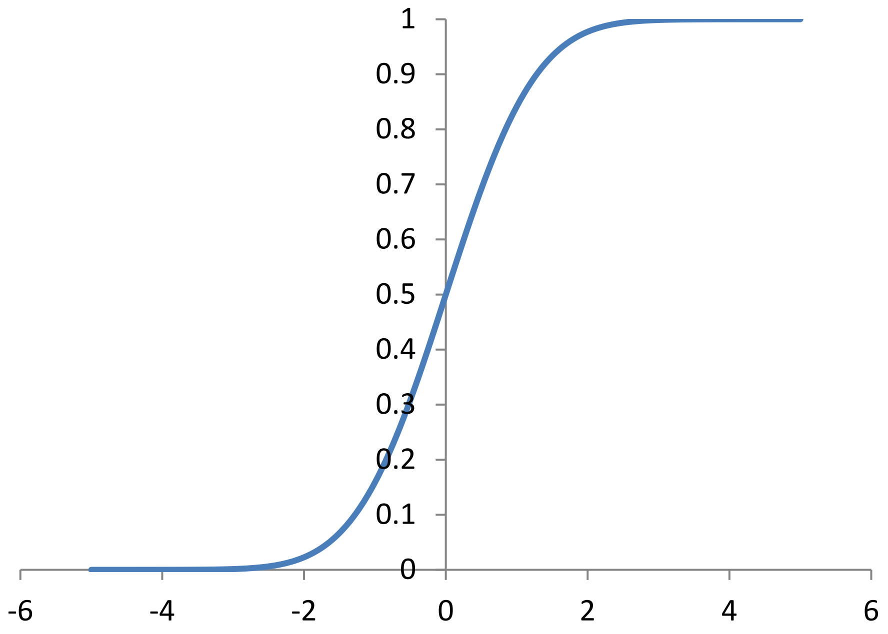


Figure 1-10: Cumulative distribution function

1.2.1.3 Measures of central tendency and dispersion of a return distribution

In need of summary measures to represent such complex objects as return distributions, analysts like to describe probability distributions using two parameters: the central tendency of returns and the dispersion of returns.

The central tendency of a distribution can be described by three measures:

- The **mean** is the expected value of all possible outcomes. It is the sum of all the possible outcomes weighted by their respective probabilities.
- The **median** is the value that has a 50-50 chance of being too high or too low.
- The **mode** is the observation that appears the most frequently. There can be several modes (in this case we have a multimodal distribution). Graphically, it is the highest point of the graph.

Let us illustrate these three values. The first distribution of the next figure is an unimodal symmetric distribution, the mean, the median and the mode are identical. The second and third distributions are asymmetric.

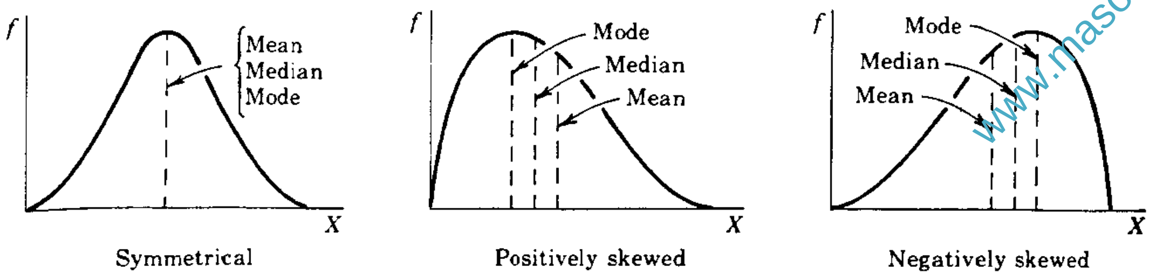


Figure 1-11: Mean, median, and mode of various continuous distributions

There are several ways to measure the dispersion of a probability distribution:

- The *range of possible outcomes* describes the set of possible values to be taken by the variable at hand; of particular interest are the minimum and the maximum of all possible outcomes. The typical lower bound for the price of a security is 0, since both stocks and bonds have liabilities limited to the stock price. Some securities typically have an upper bound, for example, bonds held until expiration, while others such as stocks do not.
- *The standard deviation* of returns and *the variance* are the most common measures of dispersion[8]. We will be discussing them presently as well as a relative measure of the dispersion, the *coefficient of variation*.

When an investor buys an asset, he must consider the risk involved. There is a risk of upward price moves as well as a risk of downward moves (however, obviously only the latter is unpleasant). The exact expected return is hardly ever achieved and the investor will probably earn more or less than expected. From this standpoint, *measuring risk involves measuring deviations from the mean*. The simplest way of doing it would be to take each state i with its probability p_i and to compute the sum of all the probability-weighted deviations from the mean,

$$\sum p_i \cdot \left(R_{\text{state i}} - E(R)\right)$$

But in the case of a symmetric distribution, the sum of these deviations will turn out to be zero. For this reason, **squared deviations** are used to measure the dispersion. The sum of the squared deviations is called the **variance**:

$$\text{Var}(R) = \sigma^2 = \sum p_i \cdot \left(R_i - E(R)\right)^2$$

Because it is more convenient to compare distances in the same dimension unit, the **standard deviation**, the square root of the variance, is most often used:

$$\sigma = \sqrt{\text{Var}(R)} = \sqrt{\sum p_i \left(R_i - E(R)\right)^2}$$

The standard deviation of the returns of an asset is often referred to as the **volatility** of the asset.

8 There are also two other moments of distribution which are not used by the MPT: the skewness is a measure of the eventual asymmetry of the probability distribution, while the kurtosis measures the importance to be attributed to extreme values (the tails) of the distribution. Leptokurtosis refers to tails that are fatter than those of the normal distribution that we will consider hereafter.

Example:
Let there be two investments which both last one time period and two possible states of nature in t=1 which each has a 50% chance of taking place.

	t=0	t=1		Simple Return	
Investment 1	EUR 100	EUR 95	EUR 115	-5%	15%
Investment 2	EUR 100	EUR 90	EUR 120	-10%	20%

The expected value in t=1 for both investments is equal to EUR 105 and their expected return is 5%, but the standard deviation is different:

$$\sigma_1 = \sqrt{0.5 \cdot (-5\% - 5\%)^2 + 0.5 \cdot (15\% - 5\%)^2} = 10\%$$

$$\sigma_2 = \sqrt{0.5 \cdot (-10\% - 5\%)^2 + 0.5 \cdot (20\% - 5\%)^2} = 15\%$$

The second investment opportunity is riskier than the first one. The higher standard deviation of the second investment indicates that the expected returns are spread wider around the mean than for the first investment. Therefore there is a higher probability of not being close to the mean.

1.2.1.4 The normal distribution

A common continuous probability distribution used in finance is the **normal distribution**. Its probability density is given by the following function[9]:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$$

where x is the value of the variable, μ is the mean of the distribution, and σ its standard deviation.

One can show that the normal distribution is not just a theoretical distribution. Let us consider the example of flipping a coin: each time one gets a tail, the player gets EUR 1 and each time a head comes up, the player must pay EUR 1. If the coin is well balanced (no cheating is possible), the player has equal chances of paying or receiving EUR 1. After the first round, he will either have won or lost EUR 1. If he won the first round, after the second round, he will either have EUR 2 or he will have broken even. If he lost the first round, after the second round he will at best break even or even have lost EUR 2. Hence after two periods, he has out of 4 possible outcomes either lost or won EUR 2 (each in one case) or broken even (in two cases). This game can go on with as many rounds as the player wants, the expected gain from the game will be zero and the standard deviation $\sqrt{n}$. This means, that it is nearly impossible to lose or win more than $3\sqrt{n}$ EUR even if there are an infinite number of rounds. When the number of games tends towards infinity, the discrete binomial distribution has so many possible outcomes that it forms a continuous range of outcomes that is the normal distribution with a mean of 0 and a standard deviation of 1.

Graphically, the normal distribution is a bell-shaped curve, and it has several important characteristics:

- It is completely characterised by its mean and its standard deviation.

9 The somewhat complex looking formula (formulated by Abraham de Moivre in 1733) is rarely used in every day applications since the necessary values are given by special numerical tables and are available in most mathematical programs. The normal distribution is in fact the continuous version of a discrete distribution: the binomial distribution. The binomial distribution is the type of distribution we had with the event trees we initially considered.

- It is symmetric around its mean; thus, the normal distribution has mean, median and mode at the same point.
- 68% of the possible values of the variable forming the normal distribution will be in the range of one standard deviation around the mean, 95% it will be within the interval of $2{\cdot}\sigma$ around the mean and with 99% it will be within $3{\cdot}\sigma$.

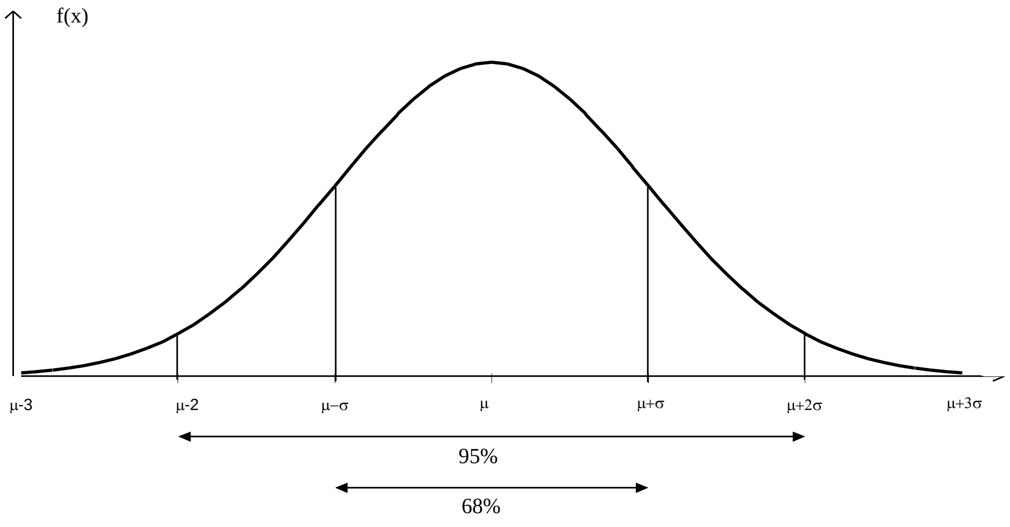


Figure 1-12: The normal distribution

These characteristics will be very useful when considering standardised variables, as we will see later.

1.2.1.5 Standardised variables

For every mean and standard deviation, the shape of the normal distribution is different. Now, each time one was to compute the probability that a **normal variable** (a variable that follows a normal distribution) is lower than a certain bound, a new integral (the surface under the curve) would have to be calculated. Fortunately this is not necessary, because it is possible to transform a normal variable such that it has a mean of zero and a standard deviation of 1. This process is called standardization.

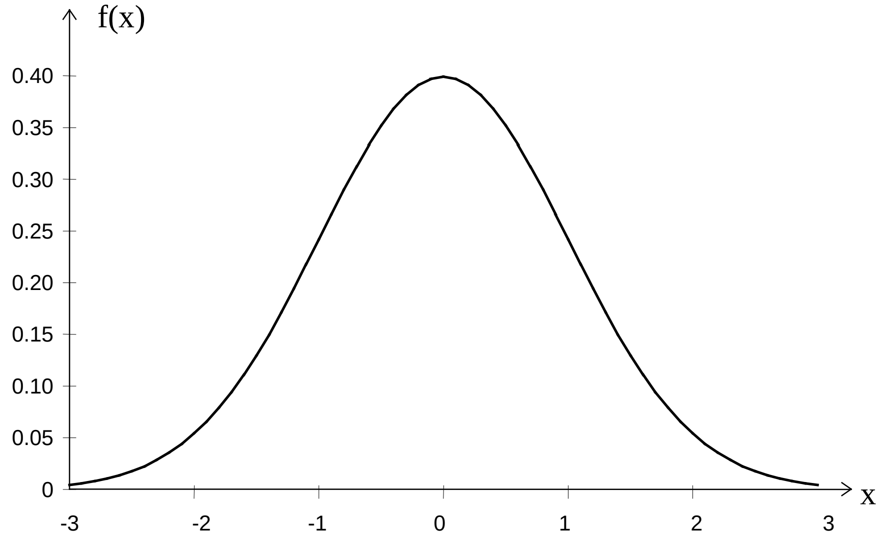


Figure 1-13: The standard normal distribution

This distribution, called the **standard normal distribution**, is obtained by transforming the random variable R by subtracting its mean, and dividing by its standard deviation:

$$U = \frac{R - \mu_R}{\sigma_R}$$

By doing so, we are able to transform any variable R with density

$$f(R) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma_R} \cdot e^{-\frac{(R-\mu_R)^2}{2 \cdot \sigma_R^2}}$$

into a standard normal variable (or **unit** variable) U with density

$$f(U) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{U^2}{2}}$$

that is, in a normal variable with mean 0 and standard deviation equal to 1. This is very useful, as there exist tables of values for the standard normal variable integral denoted N(x).

The following table lists the values of N(x) when x is positive. The table should be used with linear interpolation. For instance, if one is looking for N(0.6278), one can write:

$$\begin{aligned} N(0.6278) &= N(0.62) + 0.78 \cdot \big(N(0.63) - N(0.62)\big) \\ &= 0.7324 + 0.78 \cdot \big(0.7357 - 0.7324\big) \\ &= 0.7350 \end{aligned}$$

For negative values of x, one has to remember that N(–x)=1–N(x), as the normal distribution is symmetric around its mean (0).

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 1-4: Values for N(x) (i.e. Prob[x ≤ Z]) when Z ≥ 0

Let us illustrate this.

Example:
Let us consider a stock with an average continuously compounded return of 11.5% and a volatility of 31%, what is the probability of having a return below or equal to 0%, between 0 and 15% and above 15%?

$$\begin{aligned}\text{Prob}\left[r \le r^*\right] &= \text{Prob}\left[r \le 0\%\right]\\ &=\text{Prob}\left[\frac{r-11.5\%}{31\%} \le \frac{0\%-11.5\%}{31\%}\right]\\ &=\text{Prob}\left[\frac{r-11.5\%}{31\%} \le -0.3709\right]\end{aligned}$$

As $\dfrac{(r-11.5\%)}{31\%}$ is normally distributed with mean 0 and standard deviation 1, we have

$$u = \frac{r - 11.5\%}{31\%}$$

where 'u' is a standard normal variable. Our probability becomes

$$\text{Prob}\left[r \le 0\right] = \text{Prob}\left[u \le -0.3709\right] = 35.53\%$$

That is, that there is a 35.53% chance of losing money over the holding period. Similarly, the probability of having a return of more than 15% is:

$$\begin{aligned}\text{Prob}\left[r > 15\%\right] &= \text{Prob}\left[\frac{r - 11.5\%}{31\%} > \frac{15\% - 11.5\%}{31\%}\right] \\ &= \text{Prob}\left[u > 0.1129\right] = 1 - \text{Prob}\left[u \le 0.1129\right] \\ &= 45.50\%\end{aligned}$$

This means that the probability of having returns between 0% and 15%, is:

$$\text{prob}\left[0\% < \tilde{r} < 15\%\right] = 1 - \text{prob}\left[\tilde{r} < 0\%\right] - \text{prob}\left[\tilde{r} > 15\%\right] = 18.97\%$$

The last property uses the fact that the sum of all probabilities is by definition equal to 1.

1.2.1.6 Caveats

When applying the normal probability distribution to measure uncertain outcomes in financial analysis, one should proceed with caution:

- Probability estimates are subject to sampling errors: for instance, when using historical returns, we consider a small sample of the entire universe of historical returns, and thus, we may have wrong estimates of the future central tendency and dispersion of returns.
- The normal distribution is at the most a reasonable approximation of an asset return distribution, but certainly not a perfect model. It is an inexact model of reality.
- Stock prices do not change continuously or even necessarily by small increments.
- Many investment strategies such as those involving options or dynamic trading rules often generate non-normal return distributions.

We cannot assume that both simple and continuously compounded returns are normally distributed. Assuming normality of continuously compounded returns implies a log-normality of simple returns.

1.2.2 Computing and annualising volatility and practice

1.2.2.1 Computing volatility

We saw in the previous sections that when considering a single period model with a given number of states, the volatility of the returns was computed considering for all states the deviation of the realised return from its expected value and weighting these deviations by the state probability. That is,

$$\sigma = \sqrt{\text{Var}(R)} = \sqrt{\sum p_{\text{state i}} \cdot \left(R_{\text{state i}} - E(R)\right)^2}$$

Thus, to compute the volatility, one needs to have the probability of each state of nature.

When working with effective data, what is the available information? Generally, one observes only successive realisations of the return considered as a random variable, that is, a return at time 1, at time 2, at time 3, etc. How then can we compute the volatility?

A naïve solution consists in considering all past observations as realisations of the random variable[10]. Each past return is then considered as equally weighted, which gives the variance:

$$\mathrm{Var}(R) = \sigma^2 = \frac{1}{N-1}\sum_{t=1}^{N}(R_t - E(R_t))^2$$

and the volatility by taking its square root. This would be correct if the return were additive over time[11]. But we have seen that returns are multiplicative over time[12].

The better solution is to use continuously compounded returns, which are additive over time. Therefore, we will denote:

$$\mathrm{Var}(R) = \sigma^2 = \frac{1}{N-1}\sum_{t=1}^{N}(r_t - E(r_t))^2$$

where r_t is the continuously compounded return between time t and time t+1, computed as

$$r_t = r_{t,t+1} = \ln\left(\frac{P_{t+1}}{P_t}\right) = \ln(1 + R_{t,t+1})$$

Example:
The following table lists in its first column a set of 12 simple returns, out of which we want to compute the variance. The second column lists the corresponding continuously compounded returns.

R	r=ln(1+R)	(r–E(r))^2
0.100	0.095	0.018
–0.120	–0.128	0.008
0.030	0.030	0.005
–0.560	–0.821	0.610
0.300	0.262	0.092
0.150	0.140	0.032
0.180	0.166	0.042
–0.130	–0.139	0.010
–0.050	–0.051	0.000
–0.090	–0.094	0.003
0.020	0.020	0.004
0.040	0.039	0.006
	E(r)=–0.040	σ^2=0.0754 σ=27.46%

The third column allows us to compute the variance (by summing the elements and dividing by 11), which finally gives the volatility. Using the naive methodology, one would get a standard deviation of 21.57%.

10 This implies assuming the stationarity over time of the return generating process.

11 Nevertheless, this is often done in practice to avoid the needed additional complexity and because most popular computer software do not provide user-friendly tools to carry out the appropriate analysis. The observed systematic errors are therefore simply neglected, consciously or not.

12 That is, two consecutive periods with returns $R_{0,1}$ and $R_{1,2}$ will give a total return of $(1+R_{0,1}) \cdot (1+R_{1,2}) - 1$ and not $(R_{0,1}+R_{1,2})$.

1.2.2.2 Annualising volatility

Another practical problem often encountered in practice is that the sample from which the volatility is computed has a time unit that differs from the desired one. For instance, we could have estimated a volatility of 6% using 36 monthly returns (here the time unit is the month, and the obtained 6% is a monthly volatility), but we are interested in the annual volatility.

Assuming that the returns from which volatility is computed are independently and identically distributed (iid), the rule to be applied is the following: **volatility is proportional to the square root of time**.

$$\sigma_T = \sqrt{\frac{T}{t}}\sigma_t$$

where σ_T denotes the volatility observed over a time interval of length T. Let us illustrate this.

Example:
The monthly volatility of the ABC stock is 3%. What is its annual volatility?

$$\sigma_{1Y} = \sigma_{12M} = \sqrt{\frac{12}{1}} \cdot 3 = 10.39\%$$

as one year can be considered as 12 months.

A consequence of such a rule is that **variance is proportional to time.**

$$\sigma_T^2 = \frac{T}{t}\sigma_t^2$$

where σ_T^2 denotes the variance observed over a time interval of length T.

The mechanism above can be better understood by looking at the variance of the sum of daily log returns:

$$\operatorname{var}\left(r_1 + r_2 + r_3 + \ldots + r_{250}\right)$$

Since we have assumed that daily returns are independently and identically distributed, we can use the facts that the variance of the sum is equal to the sum of the variances and that all daily variances are equal to σ_t^2

$$\operatorname{var}\left(r_1 + r_2 + r_3 + \ldots + r_{250}\right) = \sum_{t=1}^{250} \operatorname{var}(r_t) = 250 \cdot \sigma_t^2$$

Note that we use an index ranging from 1 to 250, since there are approximately 250 trading days per year. Now from the equation above we can retrieve the standard deviation at the yearly frequency from the standard deviation at daily frequency by taking the square root

$$\sigma_T = \sqrt{250} \cdot \sigma_t$$

1.2.3 Statistical concepts

The **covariance** between the returns R_X and R_Y of two securities X and Y is defined as

$$\sigma_{X,Y} = \text{Cov}(R_X, R_Y) = E\left[\left(R_X - E(R_X)\right)\cdot\left(R_Y - E(R_Y)\right)\right]$$

where E(.) denotes the expectation operator. Intuitively, the covariance is a measure of the degree to which the two returns move together, or covary.

The **correlation coefficient** between the returns R_X and R_Y of two securities X and Y is defined as the covariance divided by the product of standard deviations

$$\rho_{X,Y} = \text{Corr}(R_X, R_Y) = \frac{\sigma_{X,Y}}{\sigma_X \cdot \sigma_Y} = \frac{E\left[\left(R_X - E(R_X)\right)\cdot\left(R_Y - E(R_Y)\right)\right]}{\sqrt{E\left[\left(R_X - E(R_X)\right)^2\right]}\cdot\sqrt{E\left[\left(R_Y - E(R_Y)\right)^2\right]}}$$

It is easy to see that when $R_X = R_Y$, that is, X = Y, the correlation coefficient equals 1, and that when $R_X = -R_Y$, that is, X = –Y, the correlation coefficient equals –1. The correlation coefficient between two random variables X and Y takes values in the interval [-1,1].

2. Portfolio theory

The recognition that the creation of an optimal portfolio is more than simply the selection of individual securities with desirable risk-return characteristics was a major step forward in the investment community a few decades ago. Indeed, before the development of the portfolio model by Markowitz, which is generally accepted as the origin of the modern portfolio theory, investors had no specific measure of the risk of a portfolio.

Markowitz[13], showed that the variance of the rate of return could be used as a measure of the risk of a portfolio.

2.1 Diversification and portfolio risk

2.1.1 Definition of a portfolio

A portfolio is a **basket of securities**. It is essentially defined by portfolio weights, that is, the proportion of the portfolio total value invested in each individual asset.

Suppose we consider potentially investing in N assets. Let x_i represent the portfolio weight (in percent, or relative terms) of asset i, i=1,2...N. Then a portfolio is fully defined by the vector of weights (x_1, x_2, ..., x_N), where, since we are talking about proportions, it must be that

$$\sum_{i=1}^{N} x_i = 1$$

Note that it is possible to have some asset weights equal to zero, that is, to say that nothing has been invested in these assets. It is also possible to have negative asset weights: this corresponds to a short sale.

> **Example:**
> Let there be a portfolio with the following weights: (0.50, 0.60,–0.10). This means that, for EUR 1'000 invested, EUR 500 will be invested in asset 1, EUR 600 in asset 2 with the difference between the initial capital (EUR 1'000) and the two long positions (EUR 1'100) being the proceeds from selling short asset 3 for a value of EUR 100.

2.1.2 Average and expected return on a portfolio

Ex post, the average return on a portfolio is the weighted average of the individual realised returns of the securities composing the portfolio:

$$\overline{R}_P = \sum_{i=1}^{N} x_i \cdot \overline{R}_i = x_1 \cdot \overline{R}_1 + x_2 \cdot \overline{R}_2 + \cdots + x_N \cdot \overline{R}_N$$

where

$\overline{R}_P$	average return on the portfolio
$\overline{R}_i$	average return on asset i
x_i	relative weight of asset i in portfolio p

13 MARKOWITZ Harry, 1952, “Portfolio Selection”, The Journal of Finance

N number of assets available.

Example:
Let there be three securities A, B, and C with realised returns of 10%, 11% and 15% over last year. A portfolio was equally invested in all three assets. What is the realised return of the portfolio over last year?

$$\bar{R}_P = 33.33\% \cdot 10\% + 33.33\% \cdot 11\% + 33.33\% \cdot 15\% = 12\%$$

Ex ante, similarly, the expected return on a portfolio is the weighted average of the individual expected returns of the securities with the proportions as weights:

$$E(R_P) = \sum_{i=1}^{N} x_i \cdot E(R_i) = x_1 \cdot E(R_1) + x_2 \cdot E(R_2) + \cdots + x_N \cdot E(R_N)$$

Example:
Let there be three securities A, B, and C with expected returns of 10%, 11% and 15% over the next year. A portfolio is equally invested in all three assets. What is the expected return of the portfolio over the next year?

$$E(R_P) = 0.333 \cdot 10\% + 0.333 \cdot 11\% + 0.333 \cdot 15\% = 12\%$$

2.1.3 Variance of a portfolio

Before we discuss the variance of a portfolio we need to understand two basic statistical concepts: covariance and correlation.

The **covariance** between the returns R_X and R_Y of two securities X and Y is defined as:

$$\sigma_{X,Y} = \mathrm{Cov}(R_X, R_Y) = E\left[(R_X - E(R_X)) \cdot (R_Y - E(R_Y))\right]$$

Intuitively, the covariance is a measure of the degree to which the two returns move together, or covary. The extent of the covariance depends on the variance of the rate of return of the individual assets as well as on the relationship between them. The covariance is an *absolute* measure of the co-movement of two securities over time.

The **correlation coefficient** between the returns R_X and R_Y of two securities X and Y is defined as the covariance divided by the product of standard deviations:

$$\rho_{X,Y} = \mathrm{Corr}(R_X, R_Y) = \frac{\sigma_{X,Y}}{\sigma_X \cdot \sigma_Y} = \frac{E\left[(R_X - E(R_X)) \cdot (R_Y - E(R_Y))\right]}{\sqrt{E\left[(R_X - E(R_X))^2\right]} \cdot \sqrt{E\left[(R_Y - E(R_Y))^2\right]}}$$

It is easy to see that when R_X=R_Y, that is, X=Y, the correlation coefficient equals 1, and that when R_X=–R_Y, that is, X=–Y, the correlation coefficient equals –1. The correlation coefficient is a standardized measure of the co-movement of the rates of return of two securities over time.

Now that we have seen the statistical concept of covariance and correlation we can consider computing the variance of a portfolio.

The variance of a portfolio is the variance of its rate of return. As we have just seen, the portfolio rate of return is a weighted average of the random rates of return of the assets in the portfolio.

Let us consider a portfolio of two assets. The variance of the portfolio can be computed as:

$$\begin{aligned}\sigma_P^2 &= E\left(R_P - \overline{R}_P\right)^2 \\ &= E\left(x_1 \cdot R_1 + x_2 \cdot R_2 - x_1 \cdot \overline{R}_1 - x_2 \cdot \overline{R}_2\right)^2 \\ &= E\left(x_1 \cdot \left(R_1 - \overline{R}_1\right) + x_2 \cdot \left(R_2 - \overline{R}_2\right)\right)^2 \\ &= E\left(x_1^2 \cdot \left(R_1 - \overline{R}_1\right)^2 + x_2^2 \cdot \left(R_2 - \overline{R}_2\right)^2 + 2 \cdot x_1 \cdot x_2 \cdot \left(R_1 - \overline{R}_1\right) \cdot \left(R_2 - \overline{R}_2\right)\right)\end{aligned}$$

where x_i is the proportion of the initial portfolio invested into asset i. As the expected value of a sum is the sum of the expectations, we have:

$$\begin{aligned}\sigma_P^2 &= x_1^2 \cdot \sigma_1^2 + x_2^2 \cdot \sigma_2^2 + 2 \cdot x_1 \cdot x_2 \cdot \sigma_{12} \\ &= x_1^2 \cdot \sigma_1^2 + x_2^2 \cdot \sigma_2^2 + 2 \cdot x_1 \cdot x_2 \cdot \rho_{12} \cdot \sigma_1 \cdot \sigma_2\end{aligned}$$

which generally differs from $\left(x_1 \cdot \sigma_1 + x_2 \cdot \sigma_2\right)^2$, as ρ_{12} generally differs from 1.

Thus, *the risk (standard deviation) of a portfolio is generally not equal to the weighted average of the standard deviations of the assets in the portfolio*, as there is an extra term depending on the correlation coefficient ρ_{12} of the assets in the portfolio. In fact, as ρ_{12} is always between –1 and +1, the risk of a portfolio can only be **smaller** ($\rho_{12} < 1$) **or equal** ($\rho_{12}=1$) to the weighted average of the standard deviations of the assets in the portfolio.

More generally, one can show that in a portfolio with N assets, we have:

$$\sigma_P^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i \cdot x_j \cdot \sigma_{i,j} = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i \cdot x_j \cdot \rho_{ij} \cdot \sigma_i \cdot \sigma_j$$

Let us illustrate this with an example.

Example:
Let us consider an equally weighted portfolio with three assets A, B, and C. We know that σ_A=15%, σ_B=18%, and σ_C=25%.
If ρ_{AB}=0.5, ρ_{AC}=0.7, and ρ_{BC}=0.55, the portfolio standard deviation is

$$\begin{aligned}\sigma_P &= \sqrt{\frac{1}{3^2} \cdot \left(0.15^2 + 0.18^2 + 0.25^2 + 2 \cdot 0.50 \cdot 0.15 \cdot 0.18 + 2 \cdot 0.70 \cdot 0.15 \cdot 0.25 + 2 \cdot 0.55 \cdot 0.18 \cdot 0.25\right)} \\ &= 16.55\%\end{aligned}$$

If ρ_{AB}=0, ρ_{AC}=0.1, and ρ_{BC}=0.8, the portfolio standard deviation is

$$\begin{aligned}\sigma_P &= \sqrt{\frac{1}{3^2} \cdot \left(0.15^2 + 0.18^2 + 0.25^2 + 2 \cdot 0.00 \cdot 0.15 \cdot 0.18 + 2 \cdot 0.10 \cdot 0.15 \cdot 0.25 + 2 \cdot 0.80 \cdot 0.18 \cdot 0.25\right)} \\ &= 14.79\%\end{aligned}$$

whereas it would be 19.33% if it were simply additive. Hence combining assets considerably reduces the risk.

The importance of correlation in the portfolio choice is further illustrated in the following figure which computes the standard deviation of a portfolio constituted of two assets A and B with expected returns of 5% and 7% and volatility of 10% and 12% respectively. The portfolio is made of 50% of each asset; therefore it has an expected return of 6%. Its standard deviation of course depends on the correlation between A and B. What is represented is how portfolio risk changes as ρ_{AB} changes.

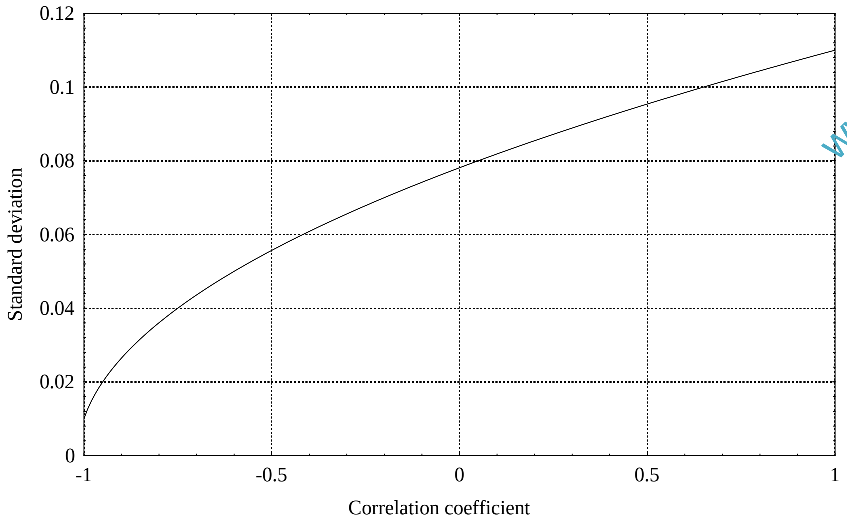


Figure 2-1: Effect of correlation on the standard deviation of a portfolio of two assets

The risk reduction made possible by combining several assets in a portfolio is referred to as **diversification**. As is clear from the graph above, the lower is the correlation between the individual assets, the larger is the benefit of diversification. If the correlation is +1, the risk of the portfolio is the weighted average of the individual assets and there is no benefit of diversifying. But we will come back to this later.

2.1.4 Risk aversion and risk premiums

One of the basic assumptions of the portfolio model is that investors are basically **risk averse** meaning that given the choice between two investments with equal rates of return, they will select the investment with the lower level of risk.

Further, in order to accept the risk attached to a given investment, that is the risk of maybe not getting the expected return, investors will ask for a compensation in the form of a surplus of expected return. This extra expected return is called the **risk premium**. The size of the risk premium depends on the quantity of risk attached to the investment and on the extra return that investors require per unit of risk taken. This risk premium per unit of risk is related to the average risk aversion of investors. Highly risk-averse investors request a large risk premium, while risk-neutral (no risk aversion) investors are willing to take risks even if there is no premium. Note that, since returns are not guaranteed, the risk premium itself is not certain: it takes the form of the expectation of an excess return (a return above the risk-free rate).

2.2 Markowitz model and efficient frontier

2.2.1 Portfolio selection

Our objective is now to define an optimal procedure for selecting a portfolio. To simplify the process, we will make the following assumptions:

The investing horizon is well defined, say one year. At date t=0 the investor selects a portfolio; at t=1 he liquidates it. Thus, this is a one period model: if the investor wants to maintain his investment over several periods, he is supposed to sell at the end of each period and reinvest the proceeds.

The investor is endowed with a certain initial capital, i.e. he has at his disposal a certain amount of money that he is willing to invest at date t=0. This amount will remain invested until t=1. At the end of the holding period, he sells all his assets and either spends the money on consumption or reinvests it in a new set of assets.

There are a given finite number N of assets (in principle, of any definition: fixed income, stocks, real estate, etc.); all assets are valued for their returns only.

Future (ex ante) asset returns are defined in probabilistic terms by their mean, their variance, and their covariance with one another, i.e. it is assumed that the typical investor only cares for the two first moments of the probability distributions on asset and portfolio returns. This can be justified either as a useful simplification, or as being fully compatible with expected utility theory under one of two possible hypotheses: quadratic utility function or normally distributed asset returns.

We need to define a criterion of portfolio selection: in other words, what is the objective of the investor? Ex post, the portfolio with the highest return is clearly the most desirable. But the game has to be played ex ante: at date 0, only the mean, variance, and covariance of returns can be estimated.

On the basis of this information, the investor can describe (compute) the probabilistic characteristics (and we know the mean and variance provides all the relevant information) of any given portfolio and he can contemplate a large number of potential portfolios. How should he rank these alternative portfolios?

2.2.2 The concept of dominance

We shall hypothesise (plausibly) that a *rational investor* will act according to the two following principles:

1) Confronted with portfolios with identical level of risk (variance), the typical investor will choose (prefer) the portfolio with the highest expected return: we say that the typical investor is not **satiated**; he prefers more to less.

2) Confronted with portfolios with identical expected return, he will choose the portfolio with the lowest variance (or, equivalently, lowest standard deviation): we say that the typical investor is risk averse; he dislikes risk per se.

These two criteria are enough to solve the problem of choice in a limited set of situations only. Take portfolios A and B: if $E(R_A) > E(R_B)$ and $\sigma_A < \sigma_B$, then clearly A is preferable to B on both dimensions (more expected return, less risk) and one could not think of any rational (mean-variance) investor preferring B to A. Portfolio A is said to **dominate** portfolio B.

Example:
The following table shows the values of two investments at time 0 and at time 1 (where two possible states of nature are equally probable).

	t=0	t=1	
Investment A	EUR 100	EUR 95	EUR 115
Investment B	EUR 100	EUR 90	EUR 110

The expected returns over the period are R_A=5% and R_B=0%, and the expected standard deviations are σ_A=10% and σ_B=10%. This is a limit case of the above dominance property: portfolio A provides a higher return than portfolio B for the same risk. Thus, A dominates B.

Unfortunately, the situation is not always as simple. In a more common situation, the investor has the opportunity to get higher expected returns, but only at the price of having to accept more risk. Clearly, no asset dominates the other. Which investment opportunity should he chose?

Example:
The following table shows the values of two investments at time 0 and at time 1 (where two possible states of nature are equally probable).

	t=0	t=1	
Investment A	EUR 100	EUR 95	EUR 115
Investment B	EUR 100	EUR 90	EUR 130

The expected returns over the period are R_A=5% and R_B=10%, and the expected standard deviations are σ_A=10% and σ_B=20%. Thus, we are in a situation of absence of dominance, as $E(R_A) < E(R_B)$ and $\sigma_A < \sigma_B$.

We have to find a way to describe the terms under which an individual may be willing to exchange risk against expected return: how large an increase in expected return might be necessary to compensate for a unit increase in risk? Here, the issue becomes more delicate: the answer might not be the same for any two individuals!

2.2.3 Indifference curves and utility level

In order to determine in a general way what the optimal portfolio for an arbitrary investor in terms of mean and variance is, a somewhat abstract tool is used: **indifference curves**. These curves are the locus of points in the mean-standard deviation plane - usually drawn with the volatility (standard deviation) on the horizontal axis and the expected return on the vertical axis - each point thus representing a particular asset or portfolio, between which the investor is indifferent. These curves are level curves (in the sense of level curves on a geographical map), the level in question being the level of satisfaction in the fulfilment of the investment objective. **Two portfolios on the same curve are equally desirable**, thus signifying that for the particular investor whose preferences are depicted the trade-off between risk and the slope of the line joining the two points appropriately represents return. The variation in risk exactly matches the extra yield he wants, thus leaving him with the same ‘utility’ level (or ‘satisfaction’).

Let us consider the following example: an investor could be indifferent between portfolios A and B (B is riskier, but it has a higher expected return). Thus, A and B would be on the same indifference curve. Portfolio C would lie on another indifference curve (as it is riskier than A but it has the same expected return, and investor would not be indifferent between A and C).

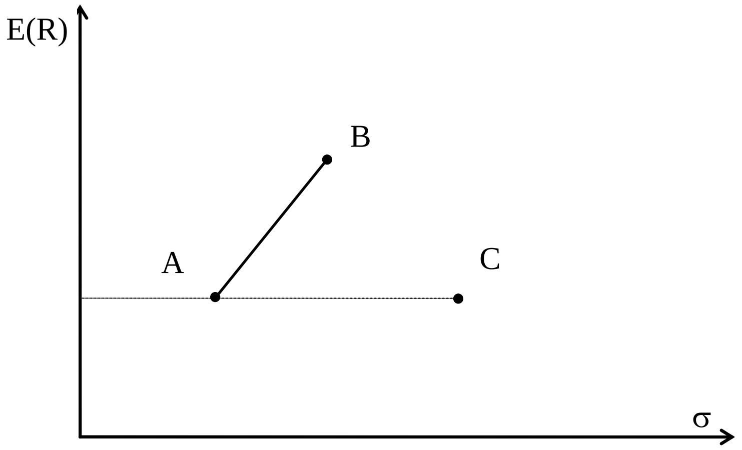


Figure 2-2: A basic example of indifference curves

Now, what do we know about indifferent curves?

First, they are **upward sloping** (from the dominance property, an increase of risk should be rewarded by an increase of expected return if the investor is risk-averse).

Second, the curves located further (higher) in the northwest direction (more towards the upper left corner) correspond to **higher levels of utility**, i.e. to more desirable portfolios in the eyes of the investor represented. Hence, in the figure below, the investor prefers A and B to C and D, but he is indifferent between A and B.

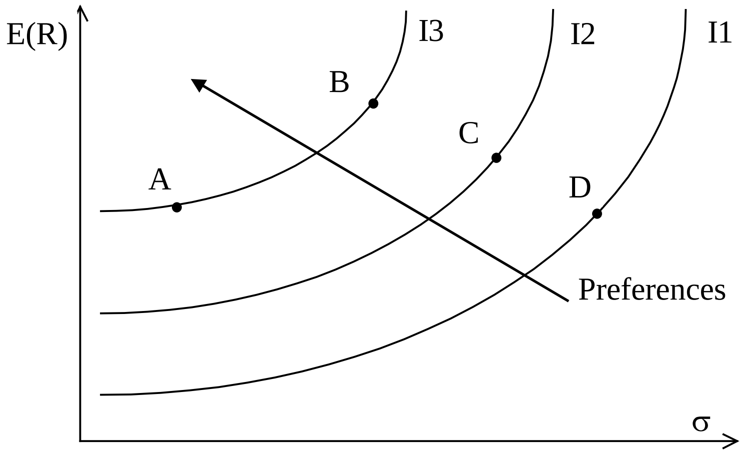


Figure 2-3: Indifference curves and utility levels

Although only three indifference curves have been plotted, the investor has an infinite number of curves. Again think of isolevel curves on a geographical map, one curve could be drawn for each level (or elevation).

Third, the slope is a measure of risk aversion: investors are risk-averse, that is, they are not ready to undertake a fair gamble (a fair gamble is a game in which there is an expected return of zero with equal chances of winning and losing). The more an investor is risk averse, the more extra expected returns, i.e. the higher risk premiums he wants for a certain level of risk. This is equivalent to saying that the indifference curve is steeper for a more risk-averse investor.

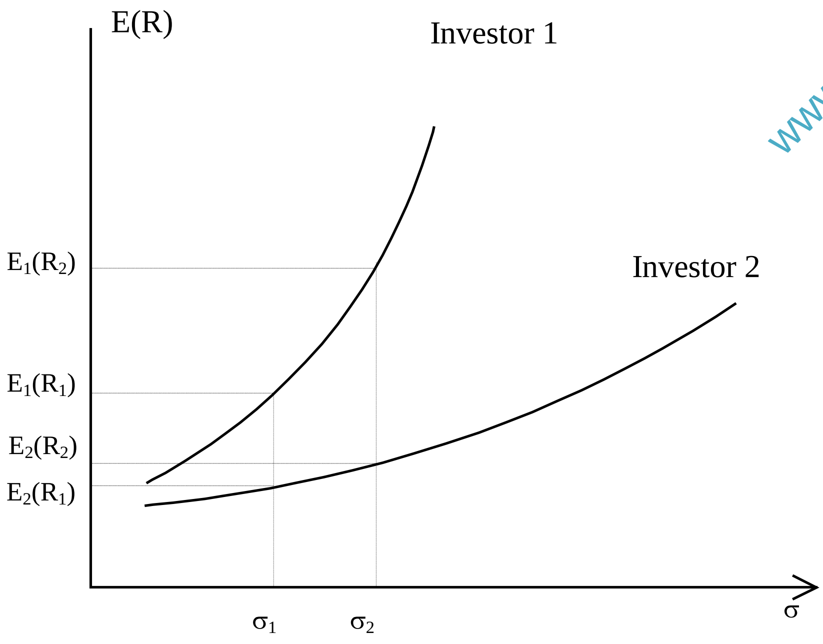


Figure 2-4: Indifference curves and risk aversion

For an increase in risk from σ_1 to σ_2, investor 2 requires a smaller increase in expected return (from $E_2(R_1)$ to $E_2(R_2)$) than investor 1 (from $E_1(R_1)$ to $E_1(R_2)$). Investor one, which is more risk-averse, will require the largest increase in expected returns in order to accept the increase in risk.

In many applications, as it is convenient to model indifference curves as a positive function of expected return and a negative function of risk, the following representation is frequently adopted: [14]

$$U = E(R) - \lambda \cdot \sigma^2$$

where:

U	'Utility level'
λ	Risk aversion coefficient. It has no economic meaning in and of itself, but it is merely an index of our aversion toward risk. The higher the coefficient, the more risk averse the investor.

According to this specification, for a given investor (and thus, a given risk aversion coefficient), it is possible to identify combinations of expected return and standard deviation that yield the same level of utility, and thus are on the same indifference curve. The following table shows several combinations based on a risk aversion coefficient of 2 that all give a utility level of 4% and 3%. Tracing a curve through all these combinations of expected return and risk creates one indifference curve that corresponds to a utility level of 4% and a second one that corresponds to a utility level of 3%.

14 However, it is not the only possible representation. Clearly, the utility should depend on other parameters (such as the wealth level, the age, etc.).

E(R)	σ	U	σ	U
4.0%	0.00%	4%	7.07%	3%
4.5%	5.00%	4%	8.66%	3%
5.0%	7.07%	4%	10.00%	3%
5.5%	8.66%	4%	11.18%	3%
6.0%	10.00%	4%	12.25%	3%
6.5%	11.18%	4%	13.23%	3%
7.0%	12.25%	4%	14.14%	3%
7.5%	13.23%	4%	15.00%	3%
8.0%	14.14%	4%	15.81%	3%
8.5%	15.00%	4%	16.58%	3%
9.0%	15.81%	4%	17.32%	3%
9.5%	16.58%	4%	18.03%	3%
10.0%	17.32%	4%	18.71%	3%
10.5%	18.03%	4%	19.36%	3%
11.0%	18.71%	4%	20.00%	3%

Table 2-1: Risk return combination with equal risk aversion (λ=2)

2.2.4 From the feasible set to the efficient frontier

Now, let us consider the set of feasible investments, that is, the set of all possible portfolios that can be formed from a set of N securities. To simplify things, let us consider N=4, that is, four risky securities labelled A, B, C, and D are available.

The **feasible investment set** takes the form shown in figure below (the construction of this is illustrated later under section 2.2.9). All the portfolios in the grey area are feasible by an adequate mix of A, B, C, and D. Thus, using a given set of N securities, we can create an infinite number of portfolios, but all of them will lie in the greyed area. A particular portfolio is Q, which is the **minimum variance portfolio**, that is, the portfolio with the *lowest possible variance* that we can create using A, B, C, and D.

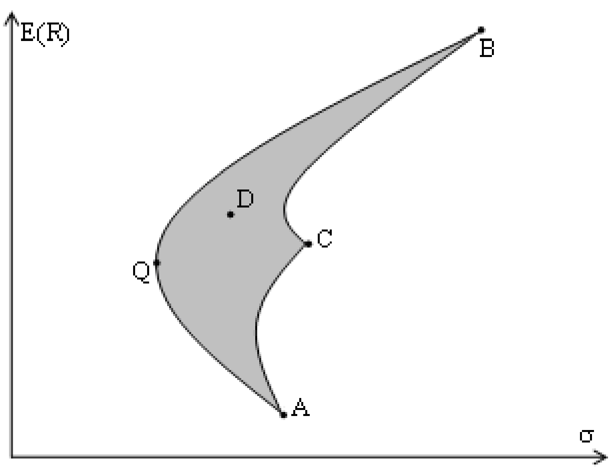


Figure 2-5: The feasible investment set

However, not all these portfolios are relevant. The set of candidate 'best' portfolios is substantially reduced thanks to the **efficient frontier theorem**. An investor will choose his or her optimal portfolio from the set of portfolios so that it:

1) Offers a minimum risk for varying levels of expected return

2) Offers a maximum expected return for varying risk levels

Let us again consider our feasible set. For any portfolio D located inside the grey area, it is possible to find a portfolio D' on the boundary with a lower risk for the same expected return.

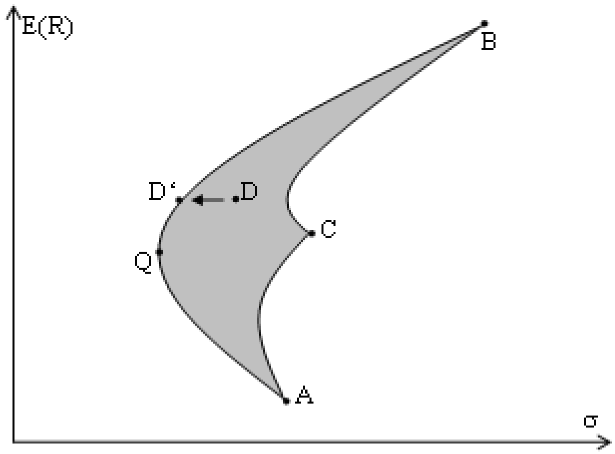


Figure 2-6: A lower risk for the same return

Thus, the rational investor should select a portfolio that lies on the boundary of the feasible set. This part of the feasible investment set is called the **minimum variance frontier**.

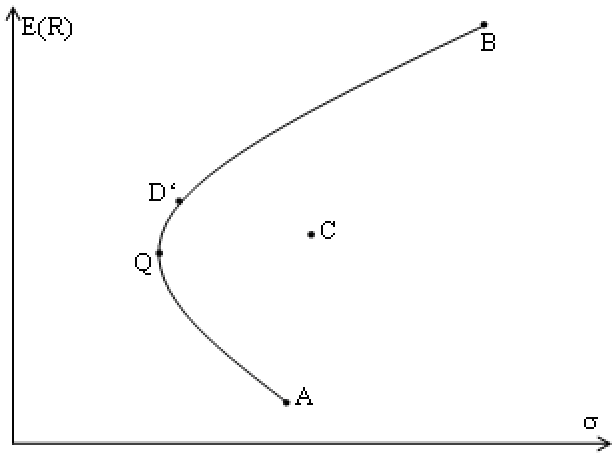


Figure 2-7: The minimum variance frontier

But actually, for any portfolio located in the lower part of the feasible set boundary, the second rule can be applied. For each portfolio we can find another portfolio on the minimum variance frontier with a higher expected return for the same risk. Thus, we have to restrict the minimum variance frontier: the rational investor should only select a portfolio that lies on the 'upper' boundary of the feasible set, from Q (the minimum variance portfolio) to B. This part of the minimum variance frontier is called the **efficient set** or **efficient frontier**.

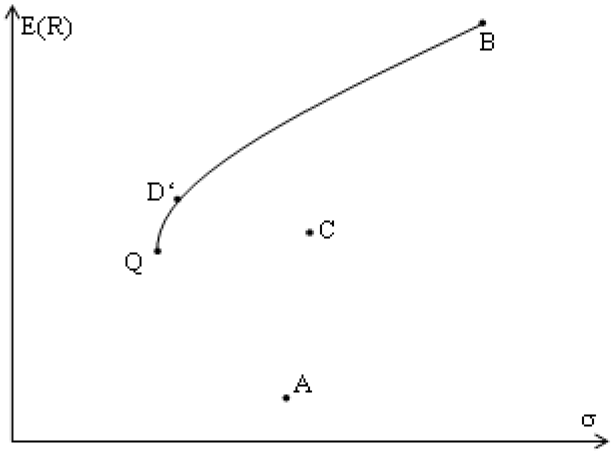


Figure 2-8: The efficient frontier

In fact, the efficient frontier can be seen as the set of rationally feasible investments. Thus, we have restricted our investment set from the total 'feasible set' to this curve. But how can we select the optimal portfolio for an investor in this curve?

2.2.5 The optimal portfolio

We have seen that the investor's preferences can be graphically represented by indifference curves.

We can plot the indifference curves and the feasible set on the same diagram (see below). The more an indifference curve is situated towards the upper-left of the figure, the more utility the investor gets (i.e. $I_3 > I_2 > I_1$). Thus, the maximum utility an investor can obtain from a set of N assets is at the tangency point of the efficient frontier and the indifference curve.

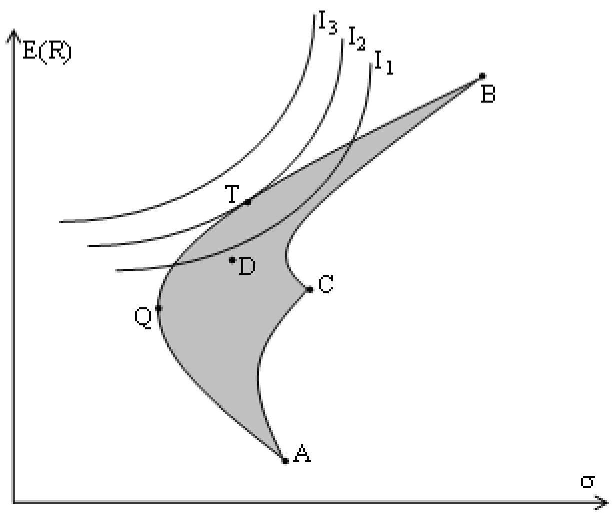


Figure 2-9: The optimal portfolio

The investor will choose the portfolio T, because it is the portfolio that gives him the highest utility. His satisfaction would be higher on I_3, but there is no feasible portfolio there. Conversely, there are an infinite number of portfolios that would yield the satisfaction of I_1, but when choosing one of these portfolios, the investor does not maximise his utility. Moreover, all but two portfolios on I_1 are inefficient.

Different investors have different risk aversions, hence differently shaped indifference curves. This means that the tangency points will vary among the different investors as shown in the following figure.

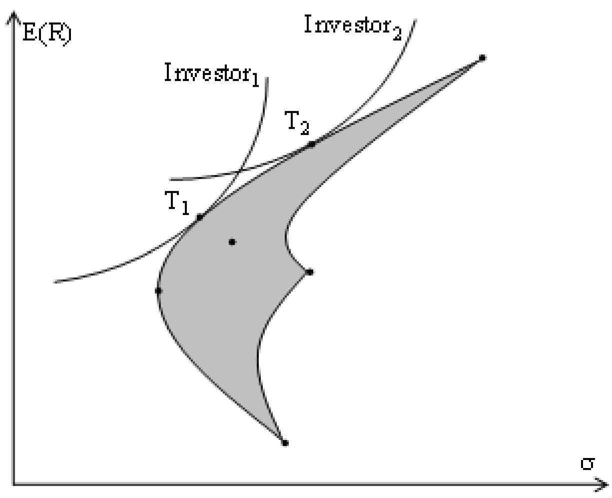


Figure 2-10: Optimal portfolios for two investors

Now that we know the essentials of the approach, we have to be more specific about the shape of the efficient frontier.

2.2.6 The efficient frontier

In this part, we will focus on the shape of the efficient frontier that one can create with different number of risky assets. We will start with a two-asset portfolio and later analyse the implication of adding more assets.

2.2.7 Two risky assets

Let us consider an investor who has the choice to invest his wealth between two risky assets (characterised by $(E(R_1), \sigma_1)$ and $(E(R_2), \sigma_2)$). If we denote by x_1 the relative amount of his wealth invested in the risky asset 1 and by $x_2=(1-x_1)$ the relative amount of his wealth invested in the risky asset 2, the expected return on the total portfolio is given by:

$$E(R_P) = x_1 \cdot E(R_1) + (1-x_1) \cdot E(R_2)$$

and the portfolio risk is:

$$\sigma_P^2 = x_1^2 \cdot \sigma_1^2 + x_2^2 \cdot \sigma_2^2 + 2 \cdot x_1 \cdot x_2 \cdot \sigma_{12}$$

we have:

$$\begin{cases} x_1 = \dfrac{E(R_P) - E(R_2)}{E(R_1) - E(R_2)} \\ \\ x_2 = 1 - x_1 = 1 - \dfrac{E(R_P) - E(R_2)}{E(R_1) - E(R_2)} = \dfrac{E(R_1) - E(R_P)}{E(R_1) - E(R_2)} \end{cases}$$

Substituting for x_1 and x_2 in the variance equation gives:

$$\sigma_P^2 = \left(\frac{E(R_P) - E(R_2)}{E(R_1) - E(R_2)}\right)^2 \cdot \sigma_1^2 + \left(\frac{E(R_1) - E(R_P)}{E(R_1) - E(R_2)}\right)^2 \cdot \sigma_2^2$$
$$+ 2 \cdot \left(\frac{(E(R_P) - E(R_2)) \cdot (E(R_1) - E(R_P))}{(E(R_1) - E(R_2))^2}\right) \cdot \sigma_{12}$$

Collecting terms in $E(R_P)$ and $(E(R_P))^2$ gives:

$$\begin{aligned} \sigma_P^2 &= \left[\frac{\sigma_1^2 + \sigma_2^2 - 2 \cdot \sigma_{12}}{(E(R_1) - E(R_2))^2}\right] \cdot (E(R_P))^2 \\ &- 2 \cdot \left[\frac{(E(R_2)) \cdot \sigma_1^2 + (E(R_1)) \cdot \sigma_2^2 - E(R_2) \cdot \sigma_{12} - E(R_1) \cdot \sigma_{12}}{(E(R_1) - E(R_2))^2}\right] \cdot E(R_P) \\ &+ \left[\frac{E(R_2)^2 \cdot \sigma_1^2 + E(R_1)^2 \cdot \sigma_2^2 - 2 \cdot E(R_1) \cdot E(R_2) \cdot \sigma_{12}}{(E(R_1) - E(R_2))^2}\right] \\ &= A \cdot (E(R_P))^2 + B \cdot E(R_P) + C \end{aligned}$$

This is the equation of a *parabola* in a (E(R), σ^2) plane[15]. Hence, the feasible investment set created by combining two risky assets (that is, 'spanned by two risky assets') is a parabola. The corresponding minimum variance portfolio can be obtained by differentiating the parabola equation with respect to $E(R_P)$. This gives:

$$\frac{d\sigma_P^2}{dE(R_P)} = 2 \cdot A \cdot E(R_P) + B$$

The first-order condition for the minimum variance portfolio equates this differential to zero, which gives:

$$E(R_P^*) = \frac{-B}{2 \cdot A}$$

where R_P^* denotes the return for the minimum variance portfolio. We can also solve for the proportions x_1 and x_2, as:

$$E(R_P^*) = x_1 \cdot E(R_1) + x_2 \cdot E(R_2) = \frac{-B}{2 \cdot A}$$

which implies:

$$x_1 = \frac{\sigma_2^2 - \sigma_{12}}{\left(\sigma_1^2 - 2 \cdot \sigma_{12} + \sigma_2^2\right)}$$

and $x_2 = 1 - x_1$.

In fact, it can be proven that our parabola is within a triangle as shown in the next figure.

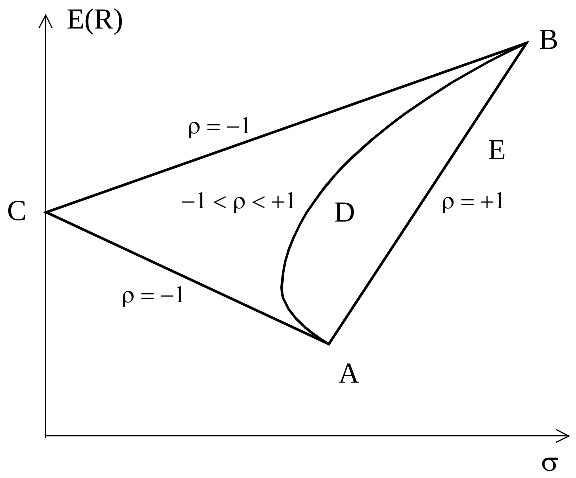


Figure 2-11: The efficient frontier (two risky assets)

and its shape depends on the correlation between the two assets:

- If the correlation is perfectly positive (ρ=+1), there is no gain from diversification. The efficient frontier is the **straight line** between the two assets with a slope equal to:

15 A parabola equation is of the form $y = A \cdot x^2 + B \cdot x + C$.

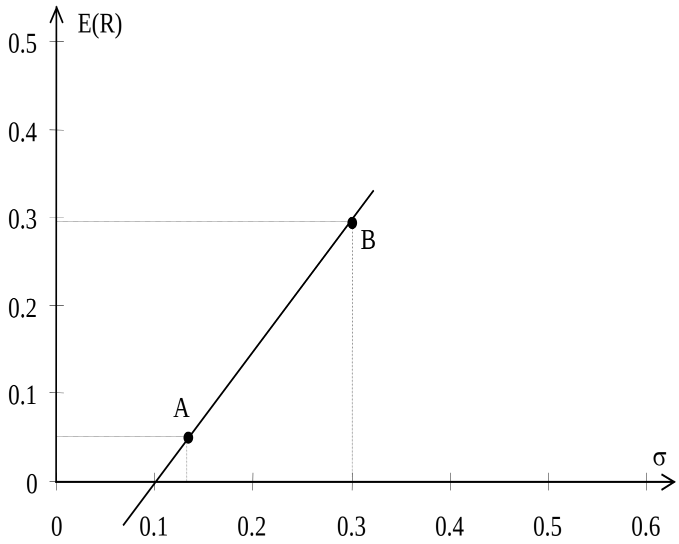
$$\frac{E(R_A) - E(R_B)}{\sigma_A - \sigma_B}$$

Note that, in this case, the standard deviation of a portfolio made of A and B is the weighted average of σ_A and σ_B. (This is the only exception to the statement made earlier about the standard deviation of a portfolio).

Example:
We want to plot the opportunity set defined by the two following risky assets: A: $E(R_A)$=5%, σ_A=13.7% and B: $E(R_B)$=29.5%, σ_B=30%, and ρ_{AB}=+1.
Using the opportunity set formula previously defined (with $\sigma_{AB}=\rho_{AB}\cdot\sigma_A\cdot\sigma_B$=0.041), we have:

$$\sigma_P^2 = 0.443\cdot(E(R_P))^2 + 0.138\cdot E(R_P) + 0.011 \cong (0.665\cdot E(R_P) + 0.104)^2$$

which gives the following figure:



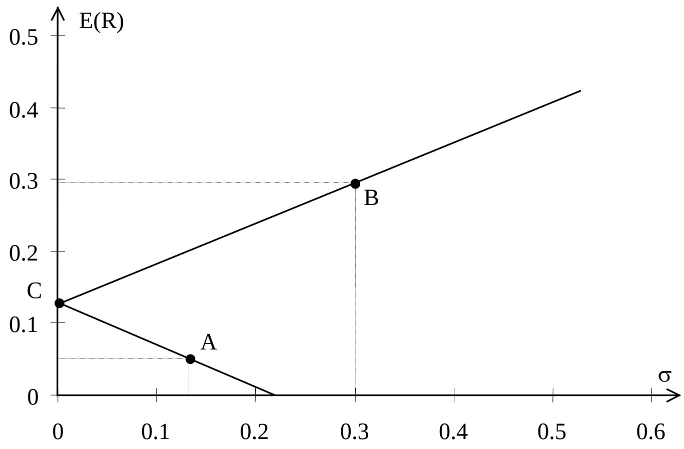
As the correlation between the two assets is +1, the minimum variance portfolio has a standard deviation of zero and a return of –15.6%.

- If the correlation is perfectly negative (ρ=–1), the benefit of diversification is the largest; it is in fact possible to create a risk-free portfolio having a positive return (point C). ACB is the minimum variance set. Only BC belongs to the efficient set.

Example:
We want to plot the opportunity set defined by the two following risky assets: A: $E(R_A)$=5%, σ_A=13.7% and B: $E(R_B)$=29.5%, σ_B=30%, and ρ_{AB}=–1.
Using the opportunity set formula previously defined (with $\sigma_{AB}=\rho_{AB}\cdot\sigma_A\cdot\sigma_B$=–0.041), we have:

$$\sigma_P^2 = 3.181\cdot(E(R_P))^2 - 0.807\cdot E(R_P) + 0.051$$

which gives the following figure:



As the correlation between the two assets is –1, the minimum variance portfolio C has no variance at all and offers an expected return of 12.7%.

If $-1 < \rho < +1$, the efficient frontier is defined by the following equation (defining a parabola):

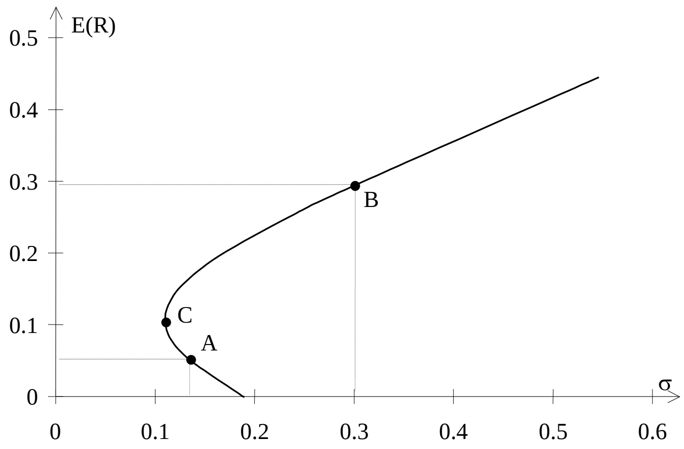
$$\sigma_P = \sqrt{x_A^2 \cdot \sigma_A^2 + (1 - x_A)^2 \cdot \sigma_B^2 + 2 \cdot x_A \cdot (1 - x_A) \cdot \rho \cdot \sigma_A \cdot \sigma_B}$$

where x_A is the proportion invested in A, $1 - x_A$ in B. This is the general case. It justifies our using the 'parabola' as the standard shape for the efficient frontier in the previous sections.

Example:
We want to plot the opportunity set defined by the two following risky assets: A: $E(R_A)=5\%$, $\sigma_A=13.7\%$ and B: $E(R_B)=29.5\%$, $\sigma_B=30\%$, and $\rho_{AB}=-0.243$.
Using the opportunity set formula previously defined (with $\sigma_{AB}=\rho_{AB}\cdot\sigma_A\cdot\sigma_B=-0.00999$), we have:

$$\sigma_P^2 = 2.145 \cdot (E(R_P))^2 - 0.449 \cdot E(R_P) + 0.036$$

which gives the following figure:



The minimum variance portfolio has a standard deviation of 11.18% and a return of 10.47%.

2.2.8 Three risky assets

Now, let there be three assets A, B and C. Unless C is perfectly correlated with one of the two assets, there will be a gain from adding one asset. The increase in diversification will make the curve shift to the left. This is illustrated in the figure below, where the line going through X is the old efficient frontier.

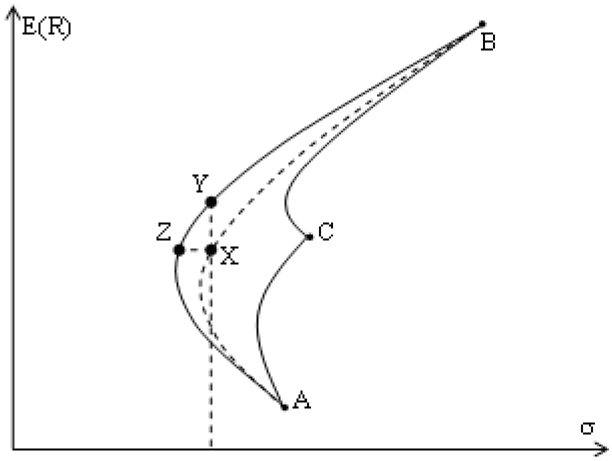


Figure 2-12: The increase in diversification when adding an asset

This leftward shift in efficient portfolios will increase the utility of the investors since it allows them to get higher returns for the same risk (moving from X to Y), or to have a lower risk for the same return (moving from X to Z). **This means that the three-asset portfolio dominates the two-asset portfolio in any case.**

Note that it is possible for the efficient frontiers representing two asset combinations to overlap each other. While the efficient frontiers of AB, AC and BC did not overlap in the above figure, it may not always be the case. However the three asset frontier, will always envelope the two asset portfolios on the left.

2.2.9 Four risky assets

Now, let us consider four risky assets A, B, C and D. The additional asset causes the efficient frontier to shift further to the left. The feasible set consists of all the asset combinations bounded by the efficient frontiers of various asset combinations. This is illustrated in the figure below and the same figure was produced under section 2.2.4 with the feasible set shaded in grey and it was also used in the following figures. On the right hand side the two asset efficient frontiers determine the boundary and are represented by full lines. The curves AB, AD, BD and CD are entirely included within the feasible set and are shown dashed.

The minimum variance frontier is determined by the combinations of all the 4 assets and shown by the bold curve which also bounds the feasible set on the left hand side. The point Q represents the asset combination on the efficient frontier with the minimum variance.

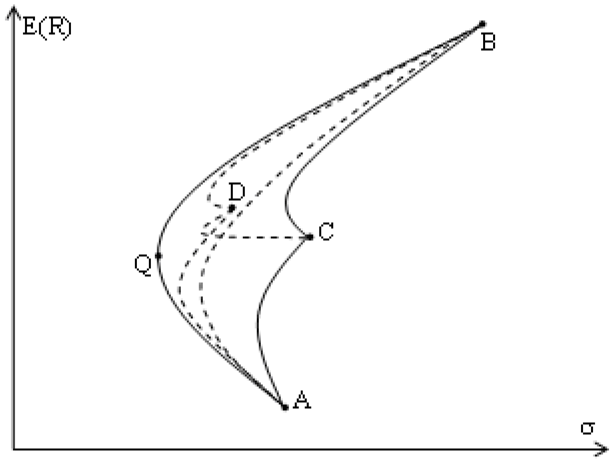


Figure 2-13: Four risky assets

The four asset efficient frontier envelopes, the two and three asset efficient frontiers, on the left signifying that a **four asset portfolio is more efficient than a three asset portfolio.**

The four asset efficient frontier is just a simplification of an efficient frontier of n risky assets.

2.2.10 N risky assets

The number of additional assets can be increased arbitrarily. However, the more assets there are the less each asset will add to the diversification possibilities. This means that the shape of the efficient frontier will not change much if we have 100 or 101 assets in our portfolio. In theory, the risk decrease will asymptotically tend towards zero with $N \to \infty$.

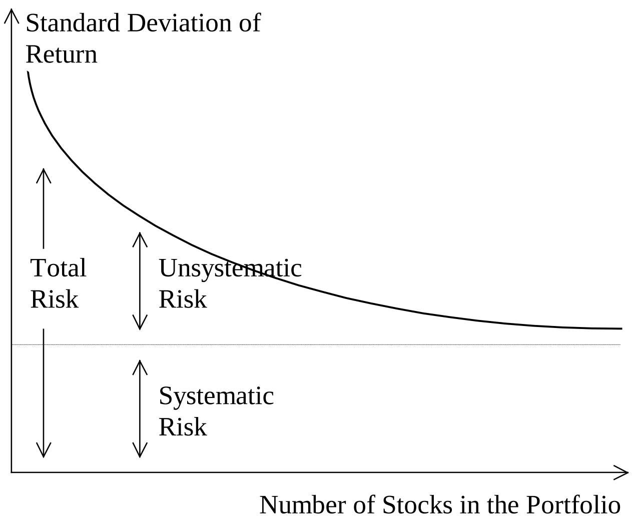


Figure 2-14: Diversification effect

2.2.11 The four steps of Markowitz's approach

In its classic article Markowitz exposed that investors should not choose portfolios that maximise expected return because this criterion by itself ignores the principle of diversification, but should rather consider variances of return in order to select the portfolio with the highest expected return for a given level of variance.

Markowitz's approach of portfolio selection takes place in four steps.

1) First, the investor specifies the set of assets he wants to take into consideration as well as his horizon of investment.

2) Second, security analysis is conducted, which here takes the specific form of trying to define the expected returns, volatility and correlation of the assets considered.

3) The third step is to compute the efficient set using the data calculated in the second step. If a risk-free asset is used, the efficient set is a line, otherwise, it will be a curve.

4) The final step is to determine the optimal portfolio for the particular investor considered.

The advantage of this approach is that step one to three are independent of the investor considered[16] and have to be made once. Only step four has to be considered for each investor. However, the process is still long: if we have N assets considered, we have to compute N expected returns, N volatilities, and $N \cdot (N-1)$ correlations[17] to compute the efficient frontier.

16 For investors with the same time horizon.

17 In fact, only half of this number, as $\rho_{XY}=\rho_{YX}$.

3. Capital Asset Pricing Model

This chapter continues where the last chapter ended, Markowitz's efficient frontier. The capital market theory expands the portfolio theory and develops a model, the **capital asset pricing model** (CAPM) for pricing risky assets. This implies that CAPM not only applies to stock pricing, but in theory, to all risky securities such as corporate bonds and other investments such as real estate.

3.1 Major assumptions

The capital market theory is based on a set of simplifying assumptions of which the main ones are:

- All investors are mean-variance optimisers, which means that they all select their portfolio in the manner described by the MPT.
- All investors have homogeneous (i.e. identical) expectations[18]. This means that their views on the available assets are represented by the same vector of expected returns and the same matrix of return variances and covariances: they use the same input list. This restrictive assumption follows from the Efficient Markets Hypothesis, which states that all relevant information is instantaneously reflected in asset prices and thus known to all market participants.

There are a few additional assumptions, some of which are stated below:

- Markets are perfect: there are no arbitrage opportunities, no transaction costs, no bid-ask spreads, assets exist in unlimited quantity and are infinitely divisible. All assets are publicly traded.
- There are no short selling restrictions.
- All investors can borrow and lend at the same risk-free rate.
- All investors have the same holding period. The model does not account for what happens after the period ends.
- There is a large number of investors. Each investor has a small individual wealth, hence no amount of buying/selling by an individual investor can affect the market price: i.e. investors are considered to be price takers.

We will see later that several of these assumptions can be relaxed, thus making the model more applicable to the real world, without changing the main implication or conclusions drawn from the model.

18 A version oft he CAPM with heterogeneous expectations also holds see Lintner 1969 Journal of Financial and Quantitative Analysis.

3.1.1 *Risk-free asset*

One of the most important factors that allowed the portfolio theory to develop into the capital market theory is the introduction of the risk-free asset.

A risk-free asset has a certain payoff. That is to say that the expected return is always the actual return[19]. Hence, if an investor buys the asset in t=0, he is certain of its value in t=1.

Since there is no risk associated with the risk-free asset, one of its key characteristics is that its standard deviation is zero. This implies that the correlation of R_F, the rate of return on the risk-free asset, with any risky asset is null.

We will now see what happens to the efficient frontier if you introduce a risk-free asset.

3.1.2 *One risky and one risk-free asset*

Let us consider an investor who has the choice to invest his wealth between a risky asset[20] (characterised by $(E(R_1), \sigma_1)$) and the risk-free asset (characterised by $(R_F, \sigma_F=0)$). If we denote by x_1 the relative amount of his wealth invested in the risky asset and by $x_2=(1-x_1)$ the relative amount of his wealth invested in the risk-free asset, the expected return on the total portfolio is given by:

$$E(R_P) = x_1 \cdot E(R_1) + (1 - x_1) \cdot R_F$$

and the portfolio risk is:

$$\sigma_P^2 = x_1^2 \cdot \sigma_1^2 + x_2^2 \cdot \sigma_F^2 + 2 \cdot x_1 \cdot x_2 \cdot \rho_{1F} \cdot \sigma_1 \cdot \sigma_F = x_1^2 \cdot \sigma_1^2$$

Thus, in this particular instance, the risk on the portfolio is simply proportional to the proportion of initial wealth invested in the risky asset.

Example:
For instance, if we consider a risky asset with an expected return of 10% and a volatility of 20%, and a risk-free rate of 4%, the portfolio return would be given by:

$$E(R_P) = R_F + x_1 \cdot [E(R_1) - R_F] = 4\% + x_1 \cdot 6\%$$

and its risk by:

---

19 In fact, there are several problems when considering a risk-free asset:

- The fact that the asset is supposed to be risk-free means not only that the returns are foreseeable, but also that there is no default risk. Therefore, only domestic government bonds are considered to be risk-free.
- Another problem is the time horizon, even a default risk free government bond has some price risk due to fluctuations in interest rates, hence only T-bonds that expire in T=1 are truly risk free.
- Yet another problem is the fact that the risk-free asset must be a zero-coupon bond, since otherwise there is an interest rate risk on the reinvestment of the coupons. In MPT, this problem is avoided since it is a one period model (consequently, no intermediary coupons can be paid), but in reality the investor has to buy a zero coupon T-bond that has the exact time to maturity as his/her time horizon. Any other asset is not risk-free even if the borrower is AAA or the bond expires with a one day difference.
- We ignore the problems caused by inflation, unless otherwise stated, the cash flows are always real.

20 which might also be interpreted as a portfolio.

$$\sigma_P = x_1 \cdot 20\%$$

From there, we have:

$$x_1 = \frac{\sigma_P}{\sigma_1}$$

which can be replaced in the return equation to give:

$$E(R_P) = \left(\frac{\sigma_P}{\sigma_1}\right) \cdot E(R_1) + (1 - x_1) \cdot R_F = R_F + \left(\frac{E(R_1) - R_F}{\sigma_1}\right) \cdot \sigma_P$$

The set of all possible investments, which in this case corresponds with the efficient frontier, is the straight line joining the two points representing the two assets taken under consideration. Some authors call it the **Capital Allocation Line** (CAL):

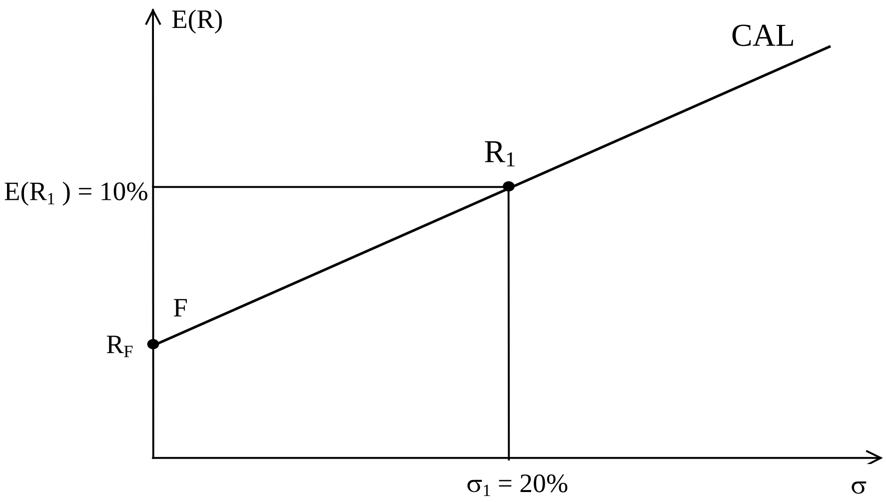


Figure 3-1: The Capital Allocation Line (CAL)

The CAL has four segments:

1) F is the point where the investor only holds the risk-free asset (x_1=0), hence the standard deviation is zero.

2) The segment from F to R_1 is the locus of all portfolios, which are at the same time long in the risky and the risk-free asset ($0 \leq x_1 \leq 1$).

3) In R_1 all the portfolio is invested in the risky asset (x_1=1).

4) Beyond R_1, the share of the risky asset is of over 100% of the wealth of the portfolio (that is, $x_1 > 1$, $x_2 < 0$). This means that the investor borrows at the risk-free rate in order to buy risky assets.

The slope of the CAL, as can be seen on the graph, equals the return-risk ratio.

3.1.3 N risky assets + one risk-free asset

The investor now has the opportunity to invest part of his wealth in a risk free asset and the rest in a combination of risky assets. Recalling what we said in the previous sections, the investor can select a portfolio on any linear combination of the risk free asset and any risky portfolio on the minimum variance frontier. This implies that the efficient frontier in this more general case is the straight line from R_F which is tangent to the parabola representing the efficient frontier when there are only risky assets. Though the shape of the feasible set with N risky assets could be different (it may have more protrusions on the right), the basic parabolic shape (which determines the efficient frontier) will be similar. Hence we continue to show the feasible set with 4 risky assets in the illustrations.

The efficient frontier is represented below:

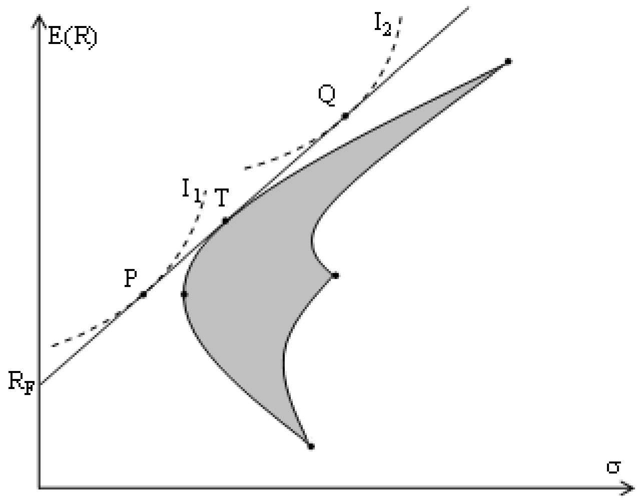


Figure 3-2: N risky assets + one risk-free asset

The investor is going to choose a portfolio on the efficient frontier according to his risk aversion. The totally risk averse investor will put all his wealth in R_F. Note that unlike the case with N risky assets, there is only *one tangency portfolio*. Hence, the risk/return ratio remains the same for all investors. Thus, the tangency portfolio T maximizes the slope of the line connecting the risk-free rate to any of the efficient portfolios obtained by combining risky assets only. In other words it maximizes the risk return tradeoff, i.e. the Sharpe ratio.

Between R_F and T, the investor invests part of his wealth in the risk-free asset and part in the tangency portfolio. This situation is relatively typical for normal investors: only a fraction of the portfolio will be invested in risky bonds and stocks.

Notice that the proportion invested in this portfolio do no longer depend on the risk aversion. All investor, irrespective of their aversion to risk, hold the same proportion of risky assets in portfolio T. As we discuss later, an investor will choose her exposure to market risk by allocating part of her wealth in this portfolio and another part in the risk free asset.

In T, we have all our wealth invested in the tangency portfolio.

What do portfolios beyond T, respectively with weights of more than 100% mean? This is the case when the investor starts leveraging his portfolio, that is to say that he borrows at the risk-free rate in order to buy even more risky assets. Note that the portfolios beyond T, such as Q, needs in order to be feasible, that there is no short-selling restriction on the assets since at T the investor is fully invested in the N risky assets. However, it is very unlikely that an investor, who can buy risk-free T-bills can also sell bonds at the same price. These problems are some of those known as market imperfections.

3.1.4 Market imperfections

A series of market imperfections can exist; we will only illustrate here the short selling restrictions, and different borrowing and lending rates.

Short-selling restrictions: frequently it is not possible at all to sell short any quantity that the investor wants. In this case, the efficient frontier is the following:

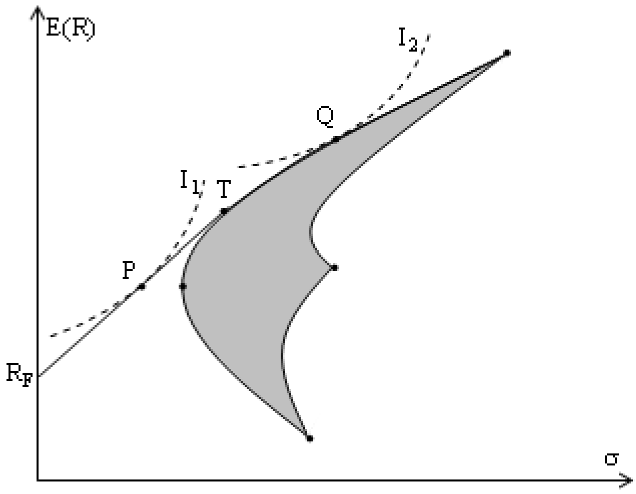


Figure 3-3: Short selling restrictions

When reaching T, the investor has to take a pure combination of the N assets on the efficient set of the N risky assets.

Different borrowing and lending rates: an investor will generally not be able to borrow at exactly the same rate as he lends. In this case the efficient frontier will not be a straight line but three segments.

The investor will have to borrow the money at a higher rate (R_{high}) than he is able to lend (R_{low}). For this reason the efficient frontier will have the following form:

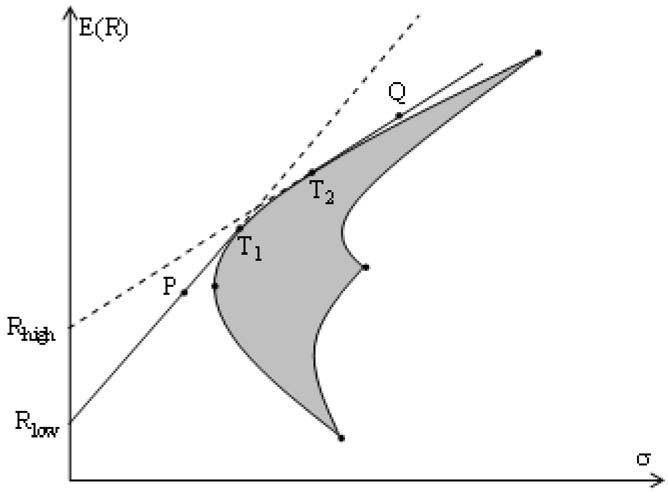


Figure 3-4: Different borrowing and lending rates

When starting with the least risky portfolios, the investor is partly invested in the risk-free asset and in the market portfolio. Hence the first segment of the efficient set is R_{low} to T_1. At T_1 the portfolio is fully invested in the risky assets. From T_1 to T_2, the investor remains fully invested in the N risky assets and the segment is the efficient set of the N assets. From T_2, the investor starts leveraging (borrowing) his portfolio. But he borrows at R_{high}. Thus the relevant straight line is the one issued from R_{high} and tangent at T_2.

3.1.5 The separation theorem

The **separation theorem (**or two funds theorem) states that the optimal combination of *risky assets* for an investor can be determined without any knowledge of his preferences toward risk and return. From the previous chapter, we know that under the assumptions stated above, the efficient frontier is a line that joins the risk-free asset and the tangency portfolio. Since we have also assumed that all investors share the same expectations, they are all confronted with the same efficient frontier, including the same tangency portfolio.

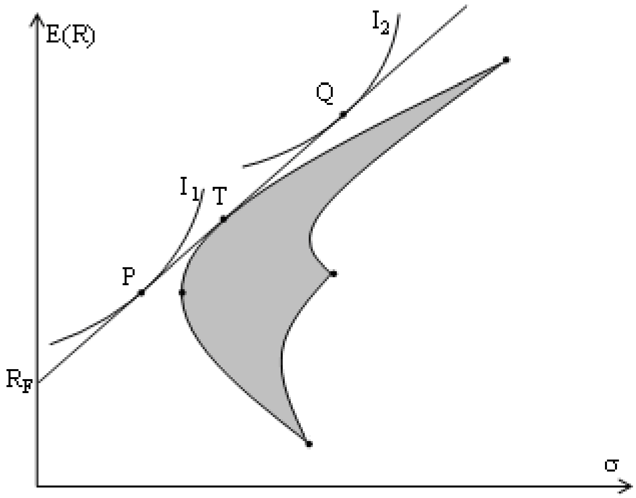


Figure 3-5: The separation theorem

Hence, without knowing the specific risk-return preferences of individual investors, we know they will all choose a portfolio on this unique efficient frontier. This means *the relevant tangency portfolio T is the same for all investors* and all investors hold different combinations of the portfolio T and of the risk-free asset R_F.

3.1.6 The market portfolio

The separation theorem implies that *all existing risky assets traded in the market must be included in the tangency portfolio*. For this reason, the tangency portfolio T, which is the optimal risky portfolio, can be called the **market portfolio** and is often represented as M.

Let us consider an asset not belonging to M. Since all investors hold the same risky portfolio, if a given asset were not a part of the market portfolio (for instance, because its risk-return characteristics are not attractive enough), nobody would hold it and there will be no demand for it. Thus, if that particular asset does exist, there will be a supply of it. Obviously an asset with some supply and no demand cannot be in equilibrium. Therefore, since the supply of this asset exceeds the demand for it, the market price of the asset will drop until the expected return on it will increase to a level that makes investment in the asset desirable. This adjustment process must go on until supply equals demand, which in this particular case must mean not only that every asset must be a part of the market portfolio, but moreover that the weighting of the asset in M is the same as the ratio of its market capitalisation to the total market capitalisation (of all existing assets).

3.2 Capital market line (CML)

The efficient frontier common to all investors is the set of all efficient portfolios. From the MPT and the assumed existence of a risk-free security, we know that all efficient portfolios are combinations of the risk-free asset and the market portfolio in different proportions. The locus of these combinations is known as the **Capital Market Line** (CML), named as such since *all rational investors have their optimal portfolio located on this line.*

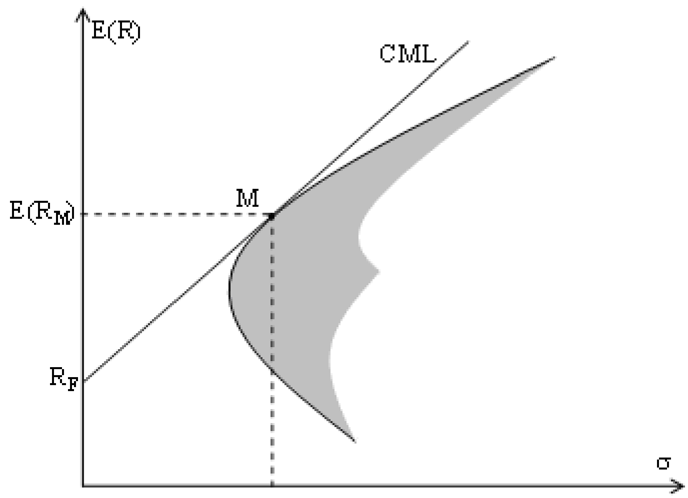


Figure 3-6: The capital market line (CML)

CML is the locus of all the possible combinations of the market portfolio and the risk free asset. Using geometry, we can prove that the slope of the CML is

$$\frac{E(R_M) - R_F}{\sigma_M}$$

(vertical distance over horizontal distance between points (0, R_F) and (σ_M, $E(R_M)$), i.e. the difference between the expected market return and the risk-free return-thus, the risk premium paid by the market-divided by the market risk). The slope represents the unitary risk premium, i.e. the *price of one unit of risk.*

The equation of the CML, which holds for all efficient portfolios, can thus be written:

$$E(R_P) = R_F + \left[\frac{E(R_M) - R_F}{\sigma_M}\right] \cdot \sigma_P$$

or

$$E(R_P) - R_F = \left[\frac{E(R_M) - R_F}{\sigma_M}\right] \cdot \sigma_P$$

i.e. the risk premium on any efficient portfolio P is the product of the (quantity of) risk of that portfolio measured by its standard deviation and the market price of risk.

This means that equilibrium expected returns on efficient portfolios depend on two factors: the reward for delaying consumption by investing rather than consuming, R_F, and the reward for taking risk which is expressed in the above equation.

Since one of the basic assumptions of the CAPM is that all investors have the same parameters, such as time horizon, information set, risk-free rate, etc., we have an equilibrium that is common to all investors. The slope of the CML is solely defined by the unitary reward for risk that is required by investors. Therefore, the slope of the CML will change if the investors become more or less risk-averse, as they will require a higher or a lower unitary risk premium. This implies that the shape of the efficient set of the N risky assets also changes. For instance, let us suppose that the investors get more risk-averse. They will then require a higher return as compensation for bearing a certain level of risk. This, in turn, will push the whole efficient frontier upwards. The tangency point, i.e. the market portfolio, will have a higher return to risk ratio and its composition will have also changed. The CML will then be tilted upward, i.e. it will have a higher slope (the equilibrium risk-free rate may change; it will probably decrease).

3.3 Security market line (SML)

The CML describes the risk-return relationship applicable to efficient portfolios. It does not apply to individual assets or non-efficient portfolios. For the latter, one has to define the relevant quantity of risk (to be multiplied by the price of risk to determine the risk premium). From the MPT, we know that the total variance of a risky portfolio depends on the variances and correlations among the individual assets included in the portfolio. Since the risky portfolio held by every investor is the market portfolio, the risk and correlation of a single asset or of a non-efficient portfolio must be evaluated in terms of its *contribution to the risk of the market portfolio*. An asset will be deemed desirable (and thus will fetch a higher price or a lower expected return) not because its total risk is low, but if it contributes negatively to the risk of the market portfolio. Conversely, securities that contribute positively (tend to increase) to the risk of the market portfolio will have to be rewarded accordingly, i.e. they will promise expected returns larger than R_M.

To derive the relation applicable to individual risky assets let us consider a particular portfolio which we denote P. It is invested with a proportion α in the market portfolio and with proportion (1-α) in an arbitrary risky asset j. Using our previous results we can compute the expected return μ_p and variance σ_p^2 of the portfolio P.

$$\begin{aligned}
\mu_p &= \alpha E\left[R_M\right] + (1-\alpha) E\left[R_j\right] \\
&= \alpha \mu_M + (1-\alpha)\mu_j \\
\sigma_p^2 &= \alpha^2 \operatorname{var}\left[R_M\right] + (1-\alpha)^2 \operatorname{var}\left[R_j\right] + 2\alpha(1-\alpha)\operatorname{cov}\left[R_j, R_M\right] \\
&= \alpha^2 \sigma_M^2 + (1-\alpha)^2 \sigma_j^2 + 2\alpha(1-\alpha)\sigma_{jM}
\end{aligned}$$

As the proportion invested in each asset varies we obtain a series of portfolio invested in asset j and in the market portfolio. The locus of points constructed using the mean and standard deviation of these portfolios is located to the right of the efficient frontier. We will call it the dominated frontier. Indeed, as the portfolios constructed are not optimally diversified they can never dominate the efficient frontier. As shown in the figure below, when $\alpha = 1$, the locus is tangent to the efficient frontier since in this case it is composed of the market portfolio only and is thus efficient.

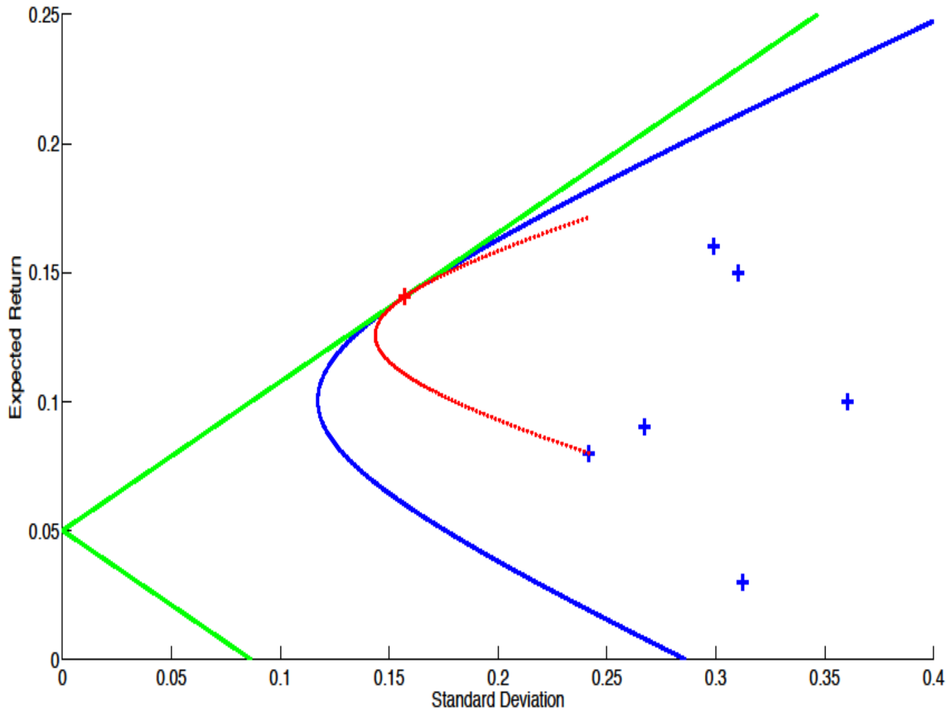


Figure 3-7: Efficient frontiers and dominated frontier

We can use this result to derive an important relation between the expected return of an arbitrary asset and its covariance with the market portfolio. The tangency at $\alpha = 1$ implies that the slope of the dominated frontier must be equal to the slope of the capital market line.

$$\left.\frac{\partial \mu_p}{\partial \sigma_p}\right|_{\alpha=1} = \frac{\mu_M - r_f}{\sigma_M}$$

We now need to explicitly obtain the partial derivative $\frac{\partial \mu_p}{\partial \sigma_p}$. We first decompose this term as follows $\frac{\partial \mu_p}{\partial \sigma_p} = \frac{\frac{\partial \mu_p}{\partial \alpha}}{\frac{\partial \sigma_p}{\partial \alpha}}$ and we obtain the next two equalities:

$$\frac{\partial \mu_p}{\partial \alpha} = \mu_M - \mu_j$$
$$\frac{\partial \sigma_p^2}{\partial \alpha} = 2\sigma_p \frac{\partial \sigma_p}{\partial \alpha} = 2\alpha\sigma_M^2 - 2(1-\alpha)\sigma_j^2 + 2(1-2\alpha)\sigma_{jM}$$

We now use the fact that when $\alpha = 1$ the standard deviation of the portfolio is equal to the standard deviation of the market, $\sigma_p = \sigma_M$, to obtain the following:

$$\frac{(\mu_M - \mu_j)\sigma_M}{\sigma_M^2 - \sigma_{jM}} = \frac{\mu_M - r_f}{\sigma_M}.$$

From this last equation we can isolate the expected return of the arbitrary risky asset i, and we obtain the central equation of the CAPM for **individual securities** or **non-efficient portfolios**:

$$\mu_i = r_f + \frac{(\mu_M - r_f)}{\sigma_M}\frac{\sigma_{iM}}{\sigma_M},$$

where the unit price of risk is given by:

$$\frac{(\mu_M - r_f)}{\sigma_M}.$$

The above equation, known as the **security market line** or SML, is often rewritten as

$$E[R_i] = r_f + (E[R_M] - r_f)\beta_i$$

where β_i (beta) is defined as follows:

$$\beta_i = \frac{\sigma_{iM}}{\sigma_M^2} = \frac{\mathrm{Cov}(R_i, R_M)}{\mathrm{Var}(R_M)}$$

A beta larger than 1 means that the individual security returns are more volatile than the returns of market portfolio. On the other hand, a beta less than 1 means that the security returns have smaller fluctuations than those of the market index. The relevant measure of risk is the asset's covariance with the market portfolio.

The CAPM asserts that the equilibrium return on an asset does not depend on the total amount of risk of that asset, as would be measured by its standard deviation or variance, but on the covariance of the asset with the market portfolio. Therefore, a risky security that is not correlated with the market (i.e. $\beta=0$) will not be expected to yield higher return than R_F. Conversely, a security with a relatively low volatility can have high expected returns simply because it strongly covaries with the market ($\beta > 1$).

Note that:

- By definition, the beta of the market portfolio is 1. Inversely, the expected return of a security with a beta of 1 is the expected rate of return on the market portfolio, $E(R_M)$.
- By definition, risk-free securities have a beta of 0. Inversely, the expected return of a security with a beta of 0 is the risk-free rate R_F.
- As we will see later, a firm can affect its beta risk through changes in the composition of its assets or through debt financing.
- Securities with negative betas (if such securities exist!) can be viewed as either hedges or insurance policies: the security is expected to do well when the market does poorly and vice-versa.
- The β of a portfolio is *the weighted average* of all the β's of the individual assets.

$$\beta_p = \sum_{i=1}^{N} w_i \beta_i$$

The SML has the following graphical representation:

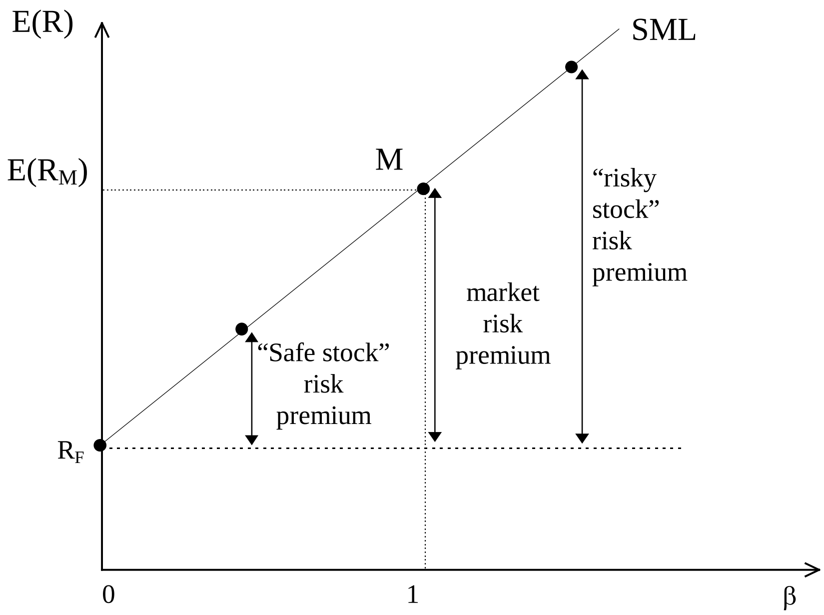


Figure 3-8: The security market line (SML)

Expected rates of returns are shown on the vertical axis, while risk measured by *beta* is shown on the horizontal axis. The slope of the SML reflects the degree of risk aversion in the economy: the greater the average investor's risk aversion, the steeper the slope of the SML, the greater the risk premium for any stock and the higher the expected (required) rate of return on stocks.

Based on the above discussion, it is easy to understand the impact of inflation and changes in the average investor's risk aversion.

Knowing that the nominal risk-free rate consists of a real inflation rate of return and an inflation premium equal to the anticipated rate of inflation, if the expected inflation rises, it produces a vertical shift in the SML.

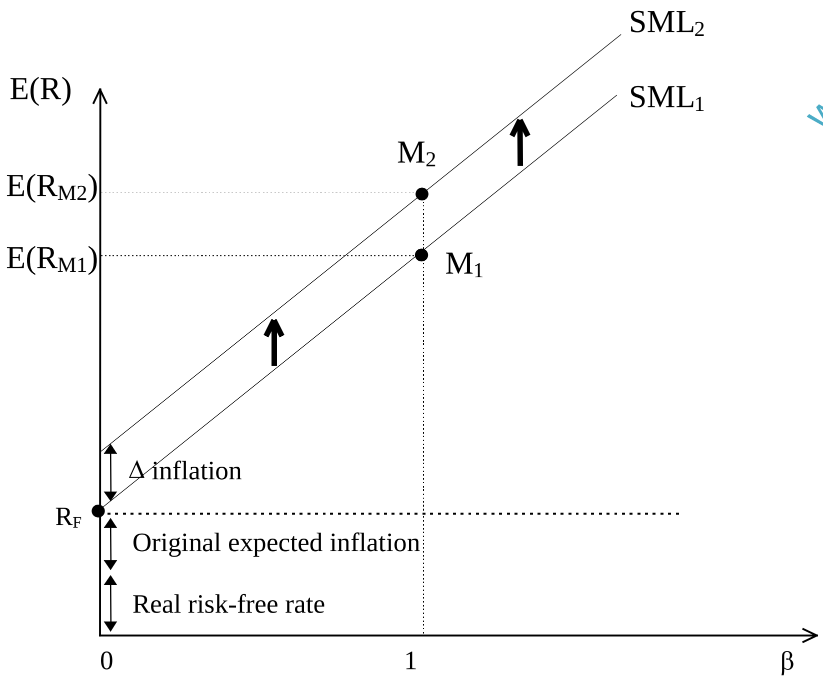


Figure 3-9: The SML and an increase in anticipated inflation

If there were no risk aversion, the SML would be horizontal. As risk aversion increases, so does the risk premium and thus the slope of the SML.

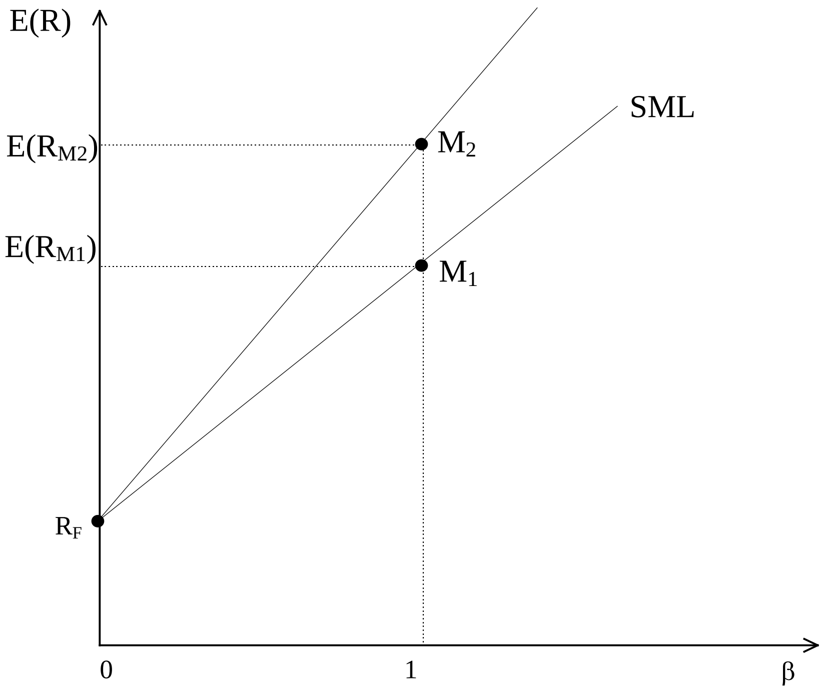


Figure 3-10: The SML and an increase in risk aversion

3.3.1 Reconciling the CML and the SML

What is the link between the SML and the CML? The CML represents the expected returns of the *efficient portfolios* as a function of their volatility measured by the standard deviations of their returns. The SML, on the other hand, graphs the expected return of an *individual asset* as a function of its sensitivity to market fluctuations. The usefulness of the SML lies in its ability to evaluate individual assets: a correctly priced asset will lie exactly on the Security Market Line.

Note that *all the efficient portfolios of the CML are also located on the SML, but the opposite is not true*. This is due to the fact that a portfolio of risky assets will have an expected return in proportion to its beta as predicted by the SML. However, unless it is a replication of the market portfolio, *every portfolio that lies on the SML need not be efficient and thus need not be located on the CML.*

We know that efficient portfolios are fully diversified (as they are a combination of the market portfolio and the risk-free asset). Let a portfolio P be efficient. The SML equation is

$$
\begin{aligned}
E(R_p) &= R_f + \left(E(R_M) - R_f\right)\cdot \beta_p \\
&= R_f + \left(E(R_M) - R_f\right)\cdot \frac{\sigma_{PM}}{\sigma_M^2} \\
&= R_f + \left(E(R_M) - R_f\right)\cdot \frac{\rho_{PM}\cdot\sigma_P\cdot\sigma_M}{\sigma_M^2}
\end{aligned}
$$

If the portfolio P is on the CML, we have either $\sigma_p = 0$, or $\rho_{PM} = 1$. If $\rho_{PM} = 1$, then

$$
\begin{aligned}
E(R_p) &= R_f + \left(E(R_M) - R_f\right)\cdot \frac{\sigma_P}{\sigma_M} \\
&= R_f + \frac{\left(E(R_M) - R_f\right)}{\sigma_M}\cdot \sigma_p
\end{aligned}
$$

which, is the CML.

3.3.2 Standard deviation versus beta as a risk measure

We are now confronted with two risk measures for a portfolio: the standard deviation (or variance) and the beta. What sort of investor rationally views the variance of returns as an appropriate measure of risk, and what sort of investor rationally views the beta of returns as a proper measure of security risk?

A rational risk-averse investor will view the variance of his portfolio’s return as the appropriate risk measure if he only holds one security. In such a case, the variance of the security becomes the variance of his portfolio’s return. On the other hand, for assets held in a diversified portfolio, the contribution of any one asset to the riskyness of the portfolio is its systematic or non-diversifiable risk. Thus, for a reasonably well-diversified portfolio, the appropriate measure of the risk of an individual asset is how the return on the asset moves relative to the returns on the market portfolio that is measured by the beta of the individual security. Thus, for investors holding a diversified portfolio, the appropriate measure of risk of an individual security is its beta.

3.3.3 Overvalued and undervalued securities

As discussed earlier, the SML can be used to detect differences in observed prices relative to theoretical prices. The mere fact that some securities are mispriced implies that the assumptions underlying the CAPM are violated. Typically, not all investors have the same knowledge of the individual stocks. Also, the data underlying their calculations are not necessarily the same. Hence, differences in valuation of stocks can appear. If the model were a true representation of reality, it would mean that the market is in disequilibrium. Nevertheless, the CAPM is based on a series of very restrictive assumptions, which make it difficult for the SML to be an exact replication of reality. However, market participants can use the concept developed above to try to detect mispriced assets.

One way of detecting mispriced assets is to compute what is often referred to as **alpha**. It is, for a given beta, the difference between the theoretical return on a given security as predicted by the SML and the return expected according to the investor's own forecast and security analysis model.

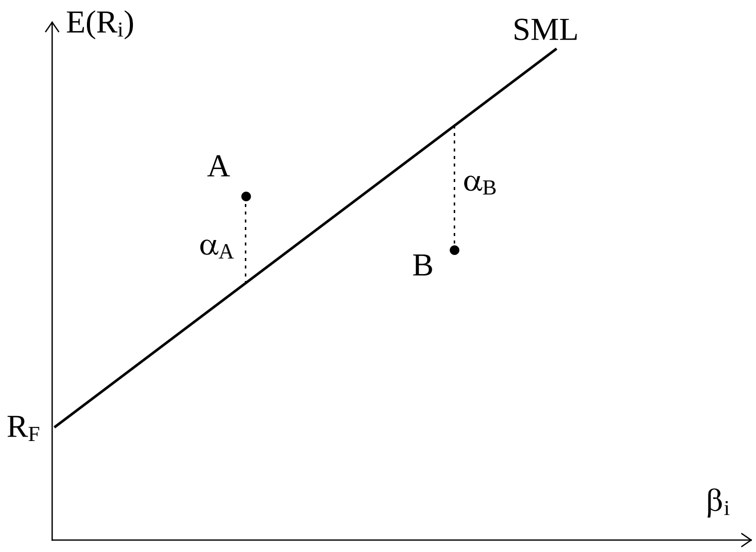


Figure 3-11: Overvalued and undervalued securities

The value of alpha is given by

$$\alpha_i = E(R_i) - E_{CAPM}(R_i)$$

When reapplying $E_{CAPM}(R_i)$ by the equation stemming from the CAPM:

$$\alpha_i = E(R_i) - [R_F + (E(R_M) - R_F) \cdot \beta_i]$$

For any alpha different from zero, the investor will consider the security to be not correctly priced. He will buy the security if α_i is positive and sell it if α_i is negative. In Figure 3-10, the asset A is underpriced whereas the security B is overpriced. According to his calculations, the investor should go short B and long A. If he buys and sells in sufficient quantity or other investors have the same outlook, the prices of undervalued assets will rise until the asset lies on the SML again. Similarly, the prices of overvalued assets will decrease until the risk-return relationship of the asset plots on the SML.

Example:
A stock with a beta of 0.8 is quoted at EUR 460. In one year, you expect its price to be EUR 590. The market return is 12%; the risk free rate is 4%. Using the CAPM check whether the stock is, correctly valued or if it is above or below the SML. Based on your calculations, would you buy the stock or sell it?
The predicted return on the stock based on the SML

$$\begin{aligned} E_{CAPM}(R) &= R_F + \beta \cdot (R_M - R_F) \\ &= 0.04 + 0.8 \cdot (0.12 - 0.04) \\ &= 0.104 = 10.4\% \end{aligned}$$

while the expected return on the stock is

$$E(R) = \frac{590 - 460}{460} = 0.283 = 28.3\%$$

The stock alpha is α=28.3%–10.4%=17.9%. The CAPM predicts a return of 10.4%, which is not consistent with your expectations of 28.3%. Thus, if your expectations are correct, the stock is undervalued by the market (its alpha is positive and it is above the SML) and you should buy it.

3.4 The zero-beta CAPM

This version of the CAPM relaxes the assumption that all investors can borrow or lend at the same risk-free interest rate. If the borrowing rate differs from the lending rate, investors will not all have the same tangency portfolio.

In this context, Black[21] developed a model of the CAPM with restricted borrowing. His model rests on the following properties of the mean-variance criterion:

- A portfolio constructed by a combination of other efficient portfolios is itself on the efficient frontier.
- Every efficient portfolio P has a corresponding portfolio on the dominated segment of the mean-variance curve with which the efficient portfolio is *uncorrelated*. This corresponding portfolio is called the **companion portfolio**, or the **zero-beta portfolio** of the efficient portfolio and is denoted Z(P).

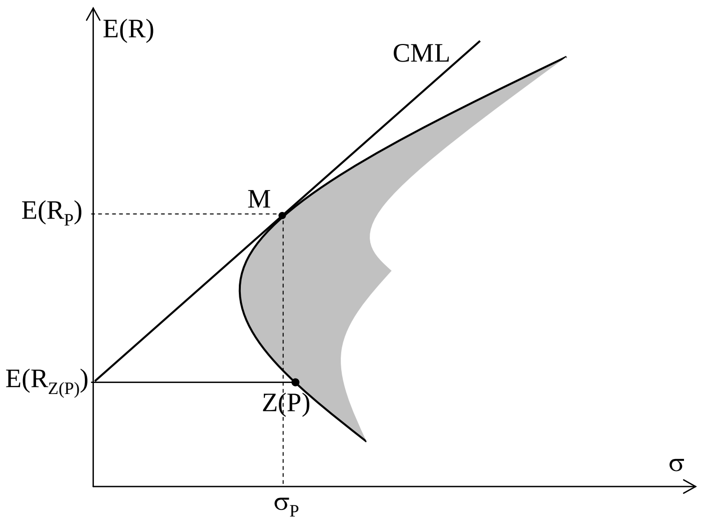


Figure 3-12: Example of a zero-beta portfolio

21 See BLACK Fischer, 1972, “Capital Market Equilibrium with restricted borrowing", Journal of Business.

- The expected rate of return of any portfolio is a linear function of the form:

$$E(R_i) = E(R_Q) + [E(R_P) - E(R_Q)] \cdot \frac{Cov(R_i, R_P) - Cov(R_P, R_P)}{\sigma_P^2 - Cov(R_P, R_Q)}$$

where P and Q are portfolio located on the efficient frontier. All investors will invest in portfolios according to their risk aversion. If all holders of portfolios have efficient portfolios, the aggregate portfolio will be efficient. We know that under normal conditions, the market portfolio is efficient because it is the aggregate of all portfolios, held by all investors, all of which are efficient. With the above properties, we can find the locus of all portfolios that are not correlated with the market portfolio. It is the horizontal line that intersects the SML on the vertical axis. Since we want this zero beta portfolio to be efficient we take the feasible portfolio with the lowest variance on this horizontal line: $E(R_{Z(M)})$.

We form a portfolio made of the market portfolio M and its zero-beta companion Z(M). Since the two portfolios are uncorrelated $Cov(R_M, R_{Z(M)}) = 0$. Our equation becomes:

$$E(R_i) = E(R_{Z(M)}) + [E(R_M) - E(R_{Z(M)})] \cdot \frac{Cov(R_i, R_M)}{\sigma_M^2}$$

This formula is similar to the standard CAPM, except that R_F has been replaced by $E(R_{Z(M)})$. This equation has found a broad application since it not only allows having a SML for cases where there is no risk-free asset, but also for the realistic case where lending and borrowing cannot be done at the same rate.

The conclusion of this is that without a risk free borrowing and lending opportunity, the expected returns would be the same as they would be in a hypothetical market with borrowing and lending at the risk-free rate.

4. Index and market models

4.1 Introduction

While the Capital Asset Pricing Model (**CAPM**) was an **equilibrium theory** explaining the expected returns on assets, the **market model (MM) and index models** provide empirical descriptions of **ex-post returns** (description of return generating processes).

More specifically:

- an **index model** (also called a **factor model**) stipulates a return generating process, i.e. hypothesises a mechanism that is supposed to determine actual, observed returns. Typically, these models decompose the sources of return in two stochastic parts: one part of return arises as a compensation for a security's sensitivity to the movements of various common factors, this is the systematic part of return; the other part is fully specific (idiosyncratic) to a security. A factor model becomes an index model when the issue of factor measurement is solved by using one or several indexes to approximate the corresponding factor, most prominently the return on a market index.
- a **single-index model** or **single-factor model** often uses a market index as factor, typically a well-diversified stock index like the SPI, the S&P 500 or the Topix. In this case, one generally uses the term "**market model**" although again some authors reserve this label for a specific version of the single-index model. Of course, there is a close link between the market model and the CAPM, a market index being the natural empirical counterpart to the notion of the market portfolio. This link is explored in more detail in this chapter. Recall however that, in theory, the market portfolio includes bonds or real estate as well as stocks and that consequently a stock index like the SPI or the S&P 500 is not the appropriate approximation for the overall market. In practice, however, the market portfolio is often approximated by a stock index.

The different versions of factor or index models have following distinguishing features:

- one factor or index / several factors or indexes.
- expressed in terms of returns (R_i) or excess returns (over the risk-free rate: $R_i - R_F$) or sometimes unanticipated returns ($R_i - E(R_i)$).
- with the expected values of the factors or indexes normalised to 0 or not: example: if the factor is the rate of inflation, the variable used could be the rate of inflation itself (whose expected value is typically positive) or the deviation of the rate of inflation from its average (with the expected value of the factor thus equal to zero).
- with a constant term in the equation or not; the meaning of the constant term is affected by the point above: if factors are expressed in deviations from their mean, then (and only then) is the constant term equal to the expected return on the asset.

4.2 The single-index model and its hypothesis

There is clearly more than one factor that can explain the behaviour of the stock market. Factors such as business cycles, interest rates, inflation, technological changes, prices of commodities or unemployment are all likely to have an impact on the behaviour of a security. Instead of taking all these factors separately into account, one can hypothesise that they might all be reflected in the change in a given factor.

Factor or index models are written as statistical equations in the form of simple or multiple regressions. For this reason, the most natural way to start describing them is in the form of a standard (simple) regression equation:

$$R_i = \alpha_i + \beta_i \cdot R_{index} + \varepsilon_i$$

One factor implies a simple regression.

For a particular time period, t, the single index model can be written as:

$$R_{it} = \alpha_i + \beta_i \cdot R_{index,t} + \varepsilon_{it}$$

where R_{it} is the return on portfolio i at time t and $R_{index,t}$ the return on the index at time t. In the following discussion we will omit subscripts whenever possible.

Theoretically, a single-index model could be formulated for any conceivable definition of the unique index or factor. However, empirical investigations have shown that the best results for single index models are achieved when the index is the market itself, approximated by the return on a broad index. From the perspective of the CAPM, there is nothing surprising in this result. The single-index model using the market return as the one factor is generally referred to as the **market model**.

Alternatively, the index is sometimes defined as the random **changes** of the market, i.e. the **unexpected** part of the return on the market index or the deviation of the rate of return on the market index from its average; one may also talk of the unanticipated market return. In that case, the constant term in the regression equation is the expected return on security i. **In the following section, we will develop the model in terms of total returns**. You will notice that the equation representing the market model can easily be transformed from one form to another. The coefficients of the independent variables remain the same. Only the constant term in the equation changes.

The single index model (market model) can be written as:

$$R_{it} = \alpha_i + \beta_i \cdot R_{Mt} + \varepsilon_{it}$$

This equation implies that there are three components to the return on a particular asset i:

- α_i is the non-stochastic part of the return on asset i. This is the expected return on the asset if the market return is zero. Indeed

$$R_{Mt} = 0 \quad \rightarrow \quad R_{it} = \alpha_i + \varepsilon_{it}$$

and thus, $E(R_{it}) = \alpha_i$.

- $\beta_i \cdot R_{Mt}$ is the portion of the return on asset i which depends upon changes in the market return. β_i is a measure of the sensitivity of the return on asset i to changes in the return on the market index. This implies $\Delta R_{it} = \beta_i \cdot \Delta R_{Mt}$.

- ε_{it} is the random element of the return specific to asset i and to date t. It is also called the idiosyncratic, or residual or firm-specific return, which means that it is the part of the asset return not explained by the index (individual responsiveness to the index is captured by the weight β_i).

In addition, the idiosyncratic returns should obey the classical linear regression model assumptions:

- idiosyncratic returns have zero expected value and constant variance for all observations, that is,

$$E(\varepsilon_i) = 0 \qquad \text{and} \qquad E(\varepsilon_i^2) = \sigma_{\varepsilon_i}^2$$

- idiosyncratic returns are statistically independent across firms, that is, the covariance of ε_i and ε_j is zero for all distinct i and j:

$$\sigma_{\varepsilon_i \varepsilon_j} = 0 \ \forall i \neq j$$

  The latter assumption is crucial as it represents the key assumption underlying a factor model: what is common to all assets is their sensitivity to the variations in the market return and everything else is absolutely **specific to each individual asset**.

- idiosyncratic returns are normally distributed.

As a corollary to these assumptions, we implicitly assume that the idiosyncratic returns are independent of the market returns, and therefore, uncorrelated with the market returns. This means that

$$\begin{aligned} Cov(\varepsilon_i, R_M) &= E\left[\left(\varepsilon_i - E(\varepsilon_i)\right)\cdot\left(R_M - E(R_M)\right)\right] \\ &= E\left[\varepsilon_i \cdot \left(R_M - E(R_M)\right)\right] \\ &= 0 \end{aligned}$$

The above assumptions will be used throughout the discussion in this chapter. If these assumptions are violated, the use of the market model may be inappropriate.

One should also note that the market model can be expressed in terms of expectations:

$$E(R_i) = \alpha_i + \beta_i \cdot E(R_M)$$

or

$$\alpha_i = E(R_i) - \beta_i \cdot R_M$$

as the random error term is always assumed to have a zero mean.

In the following figure, the above equation is represented by the (regression) line:

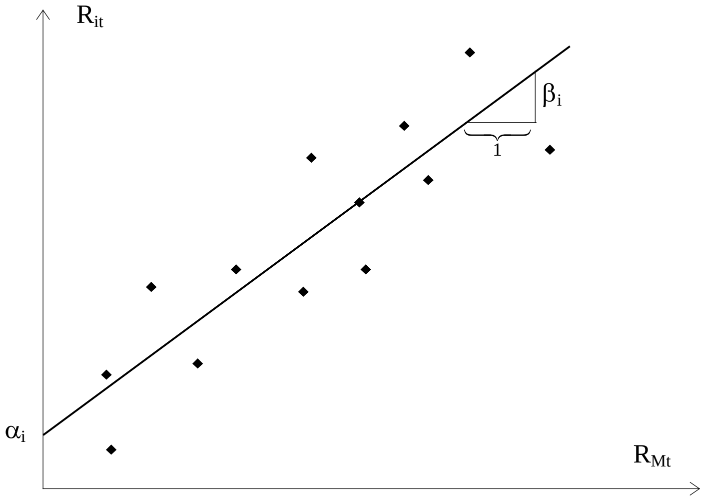


Figure 4-1: Single index model regression estimates

The figure plots the asset returns versus the market returns (R_{Mt}, R_{it}). The intercept (α_i) and the slope (β_i) are chosen so as to minimise the sum of the squared deviations from the regression line.

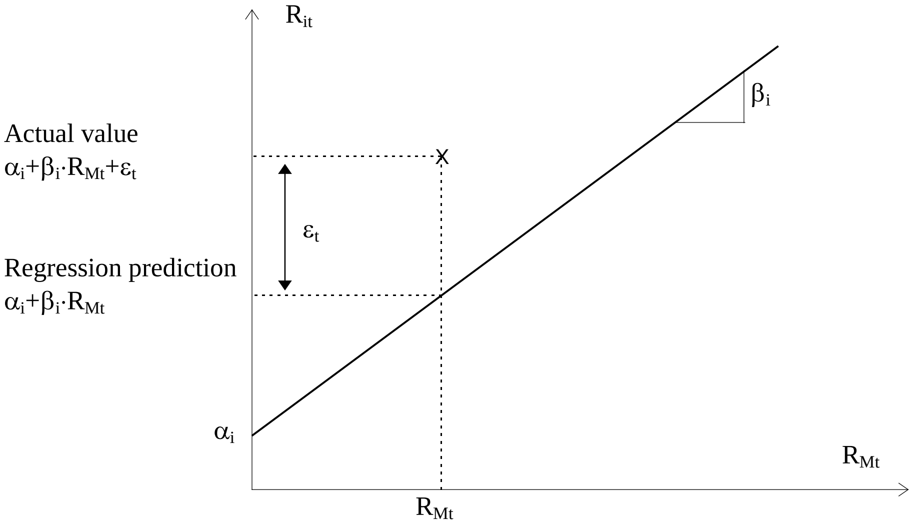


Figure 4-2: Simple regression estimates and residuals

If we take a closer look at one observation, we can see that the component of R_{it} that is explained by the regression model is $\alpha_i + \beta_i \cdot R_{Mt}$, while the unexplained component is represented by the disturbance term ε_t.

Example:
If α_i = 2% and β_i = 1.5, then:

R_{Mt}	ε_{it}	R_{it}
10%	3%	2% + 15% + 3% = 20%
–6%	4%	2% – 9% + 4% = –3%
0%	2%	2% + 2% = 4%

As already noted above, α_i is the expected return on security i if the market return is zero. It is thus the constant term of the regression. The coefficient β_i on the other hand, measures the sensitivity of the asset returns to changes in market returns. In the diagram, it is the slope of the regression line.

4.3 Decomposing variance into systematic and diversifiable risk

4.3.1 In the case of a single security

In the single index model, the return on security i is given by:

$$R_{it} = \alpha_i + \beta_i \cdot R_{Mt} + \varepsilon_{it}$$

Taking expectations, and recalling that i) the expected value of a sum of random variables is the sum of the expected values, ii) α_i and β_i are constants by construction (thus $E(\alpha_i) = \alpha_i$ and $E(\beta_i) = \beta_i$) and iii) $E(\varepsilon_i)$ is zero:

$$E(R_i) = \alpha_i + \beta_i \cdot E(R_M)$$

The variance of the returns on security i is given by:

$$\begin{aligned}\sigma_i^2 &= E\left(R_i - E(R_i)\right)^2 \\ &= E\left(\alpha_i + \beta_i \cdot R_M + \varepsilon_i - \alpha_i - \beta_i \cdot E(R_M)\right)^2 \\ &= E\left(\beta_i \cdot \left(R_M - E(R_M)\right) + \varepsilon_i\right)^2\end{aligned}$$

Squaring terms in the parenthesis and taking expectations gives

$$\sigma_i^2 = \beta_i^2 \cdot E\left(R_M - E\left(R_M\right)\right)^2 + 2 \cdot \beta_i \cdot E\left[\varepsilon_i \cdot \left(R_M - E\left(R_M\right)\right)\right] + E(\varepsilon_i^2)$$

By definition, idiosyncratic returns are independent of the market returns. Thus, we have:

$$\sigma_i^2 = \beta_i^2 \cdot E\left(R_M - E\left(R_M\right)\right)^2 + E(\varepsilon_i^2)$$

This equation tells us that the contribution of the variance of R_M to that of R_i depends on the slope coefficient β_i. It can be rewritten as:

$$\sigma_i^2 = \underbrace{\beta_i^2 \cdot \sigma_M^2}_{\substack{\text{market} \\ \text{risk}}} + \underbrace{\sigma_{\varepsilon_i}^2}_{\substack{\text{residual} \\ \text{risk}}}$$

where σ_i^2 is the total variance of the asset returns, $\beta_i^2 \cdot \sigma_M^2$ is its market or systematic risk (also called **explained variance**) and $\sigma_{\varepsilon_i}^2$ is its idiosyncratic or residual or unsystematic risk, (also labelled "diversifiable" risk for reasons that will be clear later on but can easily be anticipated) or **unexplained variance**.

Example:
You have the following information about the German market:

Asset i	β_i	σ_i
DAX index	1.00	0.0260
E.ON	1.10	0.0363
Deutsche Bank	1.69	0.0585

From there, we want to find the idiosyncratic risk of E.ON and Credit Suisse returns.
For E.ON, we have

$$\sigma_{\varepsilon_1} = \sqrt{\sigma_1^2 - \beta_1^2 \cdot \sigma_M^2} = \sqrt{0.0363^2 - 1.10^2 \cdot 0.0260^2} \approx 0.0224$$

And for Deutsche Bank

$$\sigma_{\varepsilon_2} = \sqrt{\sigma_2^2 - \beta_2^2 \cdot \sigma_M^2} = \sqrt{0.0585^2 - 1.69^2 \cdot 0.0260^2} \approx 0.0387$$

Of course, the validity of this decomposition depends on the assumption that idiosyncratic returns are statistically independent across firms. It also requires the independence between R_M and ε_i which is a feature of a correctly specified regression equation.

In the market model context, we can also compute the covariance between two assets. Recall that the covariance between the returns of assets i and j is given by:

$$\sigma_{ij} = \mathrm{Cov}(R_i, R_j) = E\left[(R_i - E(R_i))(R_j - E(R_j))\right]$$

Substituting for R_i, R_j, $E(R_i)$, and $E(R_j)$ with the values computed above yields

$$\begin{aligned}\sigma_{ij} &= E\left[(\alpha_i + \beta_i \cdot R_M + \varepsilon_i - \alpha_i - \beta_i \cdot E(R_M)) \cdot (\alpha_j + \beta_j \cdot R_M + \varepsilon_j - \alpha_j - \beta_j \cdot E(R_M))\right] \\ &= E\left[(\beta_i \cdot (R_M - E(R_M)) + \varepsilon_i) \cdot (\beta_j \cdot (R_M - E(R_M)) + \varepsilon_j)\right]\end{aligned}$$

Multiplying the terms, we get:

$$\begin{aligned}\sigma_{ij} = {} & \beta_i \cdot \beta_j \cdot E(R_M - E(R_M))^2 + \beta_j \cdot E\left[\varepsilon_i \cdot (R_M - E(R_M))\right] \\ & + \beta_i \cdot E\left[\varepsilon_j \cdot (R_M - E(R_M))\right] + E\left[\varepsilon_i \cdot \varepsilon_j\right]\end{aligned}$$

According to the assumptions underlying the market model, the last three terms are zero. Thus, the covariance between asset i and j returns is given by:

$$\sigma_{ij} = \beta_i \cdot \beta_j \cdot E(R_M - E(R_M))^2$$

which we can rewrite as:

$$\sigma_{ij} = \beta_i \cdot \beta_j \cdot \sigma_M^2$$

Let us illustrate this result using the following example:

Example:
Here is an extract from the German market:

Asset i	β_i	σ_i
DAX index	1.00	0.0260
E.ON	1.10	0.0363
Deutsche Bank	1.69	0.0585

From there, what is the covariance and the correlation coefficient of E.ON and Deutsche Bank returns?
The covariance is given by:

$$\sigma_{12} = \beta_1 \cdot \beta_2 \cdot \sigma_M^2 = 1.10 \cdot 1.69 \cdot 0.0260^2 \approx 0.0013$$

From there, we can easily infer the correlation coefficient:

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2} = \frac{0.0013}{0.0363 \cdot 0.0585} \approx 0.59$$

4.3.2 In the case of a portfolio: implications for diversification

The equation

$$\sigma_i^2 = \beta_i^2 \cdot \sigma_M^2 + \sigma_{\varepsilon_i}^2$$

holds for portfolios as well as for individual securities. Knowing that

$$R_p = \alpha_p + \beta_p \cdot R_M + \varepsilon_p$$

and recalling that

$$R_p = \sum_{i=1}^{N} x_i \cdot R_i$$

where x_i is the weight of asset *i* in the portfolio, one gets

$$\begin{aligned} R_p &= \sum_{i=1}^{N} x_i \cdot (\alpha_i + \beta_i \cdot R_M + \varepsilon_i) \\ &= \sum_{i=1}^{N} x_i \cdot \alpha_i + R_M \cdot \sum_{i=1}^{N} x_i \cdot \beta_i + \sum_{i=1}^{N} x_i \cdot \varepsilon_i \end{aligned}$$

Clearly, the beta coefficient for a portfolio of N securities is a simple weighted average of the betas of the stocks included in the portfolio, where the weights are the relative amounts invested in each security:

$$\beta_p = \sum_{i=1}^{N} x_i \cdot \beta_i$$

From the above, we get

$$\sigma_p^2 = \beta_p^2 \cdot \sigma_M^2 + \sigma_{\varepsilon_p}^2$$

and thus, we can decompose the variance of the portfolio in the following way

$$\sigma_p^2 = \left(\sum_{i=1}^{N} x_i \beta_i \right)^2 \cdot \sigma_M^2 + \sum_{i=1}^{N} x_i^2 \sigma_{\varepsilon_i}^2$$

The last term in the above equation can be restated as follows:

$$\sigma_{\varepsilon_p}^2 = \sum_{i=1}^{N} x_i^2 \cdot \sigma_{\varepsilon_i}^2$$

This means that the residual variance of a portfolio is the weighted average of the residual variances of the securities in the portfolio. Note that, this time, in taking the average, we square the portfolio weights.

This last result means that, as an investor attempts to diversify his portfolio by increasing the number of stocks in his portfolio, he reduces the specific risk of his portfolio. since the residuals, $\sigma^2_{\varepsilon_p}$, are uncorrelated, the residual variance of the portfolio approaches zero as N gets larger and larger. However, the beta of his portfolio does not decrease since it is the weighted average of the individual betas.

4.3.3 Quality of an index model: R^2 and ρ^2

How can we tell if the market model is a good representation of reality? If one accepts the hypothesis that asset returns are linearly related to the market returns, the indicator of the explanatory power of the model is the percentage of the variation of the dependent variable (R_i) that can be explained by the variations in the independent variable (R_M), or the part of the fluctuations in returns of a specific asset that can be explained by the variations in the market return.

This indicator is defined as the **coefficient of determination**, also called **R-squared (R^2)** of the regression

$$R^2 = \frac{\text{Explained variance in } R_i}{\text{Total variance in } R_i} = \frac{\beta_i^2 \cdot \sigma_M^2}{\sigma_i^2} = \frac{\beta_i^2 \cdot \sigma_M^2}{\beta_i^2 \cdot \sigma_M^2 + \sigma_{\varepsilon_i}^2}$$

An R^2 equal to 1 would mean that 100% of the variations in the returns of an individual asset could be explained by the variations in the market return. Hence a R^2 of 0.55 means that 45% of the return cannot be explained by the model.

Note that as the unexplained variance $\sigma^2_{\varepsilon_i}$ has to be the difference between 1 and the coefficient. Thus,

$$R^2 = 1 - \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_i}$$

It is easy to show that the coefficient of determination is the square of the correlation coefficient, as

$$\rho_{iM} = \frac{\sigma_{iM}}{\sigma_i \cdot \sigma_M} = \frac{\beta_i \cdot \beta_M \cdot \sigma_M^2}{\sigma_i \cdot \sigma_M} = \frac{\beta_i \cdot \sigma_M}{\sigma_i} = \sqrt{R^2}$$

or

$$\rho^2_{iM} = R^2$$

In practice, the market model performs poorly on individual assets; typically, the variation in the returns on the market index explains less than half of the variation in the returns on an individual asset (i.e., $R^2 < 50\%$). The performance of the market model is far more satisfactory for well diversified portfolios where the model accounts for a major part of the variation in returns.

4.4 The link with the CAPM

4.4.1 *About beta (β)*

Our starting regression equation was

$$R_i = \alpha_i + \beta_i \cdot R_M + \varepsilon_i$$

This has an implication for the covariance of an asset with the market

$$\begin{aligned} Cov(R_i, R_M) &= Cov(\alpha_i + \beta_i R_M + \varepsilon_i, R_M) \\ &= \beta_i Cov(R_M, R_M) + Cov(\varepsilon_i, R_M) \end{aligned}$$

Since, by definition, the residual errors are independent of the market returns, the second term equals zero. The covariance of a single security with the market is given by

$$Cov(R_i, R_M) = \beta_i \cdot \sigma_M^2$$

This can be rewritten in the following form

$$\beta_i = \frac{Cov(R_i, R_M)}{\sigma_M^2}$$

Thus, the beta in the market model is of the same form as the beta in the CAPM. The former helps to give empirical content to the latter.

In the CAPM, we use ex ante betas related to the general abstract notion of the market portfolio, while in the market model, we have ex post estimated beta, specific to the particular index selected. Furthermore, in the CAPM, we use expected returns, while in the market model, we use realised returns. The common independent variable of the two models is the return on the market portfolio, which, in the market model (and also in the empirical versions of the CAPM), takes the form of a market index.

Clearly, this commonality indicates a direct relationship between the two models. The definition of β_i corresponds to the definition in the CAPM, provided we accept the market index as the appropriate measure of the market portfolio. The application of the market model will thus provide us with empirical estimates of the β's.

4.4.2 *Estimating the alphas (α)*[22]

The market model can be written in expectations form, keep in mind that $E(\varepsilon_i) = 0$, as:

$$E(R_i) = \alpha_i + \beta_i \cdot E(R_M)$$

Similarly, the CAPM can be expressed as:

$$E(R_i) = R_F + \beta_i \cdot \left[E(R_M) - R_F\right]$$

or

$$E(R_i) = R_F \cdot (1 - \beta_i) + \beta_i \cdot E(R_M)$$

22 Not the same α as in the previous chapter

Comparing these equations, one may conclude that if the CAPM holds (and if the index used in the market model is a good approximation), one should get the following estimates

$$\alpha_i = R_F \cdot (1 - \hat{\beta}_i)$$

Suppose our estimation of the market model yields a greater value, such as

$$\hat{\alpha}_i > R_F \cdot (1 - \hat{\beta}_i)$$

then this would suggest that over the (past) period of estimation, asset i has had an average return larger than the equilibrium return predicted by the CAPM, or that asset i was undervalued.

On this score, the particular formulation of the market model is important. Suppose we write

$$R_i^e = \alpha_i^* + \beta_i R_M^e + \varepsilon_i$$

where the (e) superscript denotes an excess return over the risk-free rate[23].

In its expectations form, the market model predicts

$$E(R_i^e) = \alpha_i^* + \beta_i \cdot E(R_M^e)$$

while the CAPM predicts

$$E(R_i^e) = \beta_i \cdot E(R_M^e)$$

Hence, for both models to be in accordance, $\alpha_i^* = 0$ must be true. In that case, observing $\alpha_i^* > 0$ would imply that asset i is undervalued. **Thus, α, is usually interpreted as an indicator of undervaluation ($\alpha_i^* > 0$) or overvaluation ($\alpha_i^* < 0$) of the asset in question**.

How useful is this approach to valuation? It would be extremely useful if $\hat{\alpha}_i$ (resp. $\hat{\alpha}_i^*$) were stable over time. Unfortunately, in practice, this is not the case.

4.4.3 Estimating the betas (β)

The β coefficient indicates the sensitivity of the asset return to changes in the index. The estimation of the β can be performed with an OLS **regression**, where α is the constant of the regression.

23 Note that we changed the notation for the constant term, because as we shall show, it is affected by this rewriting, whereas this is not the case for the other coefficients of the regression.

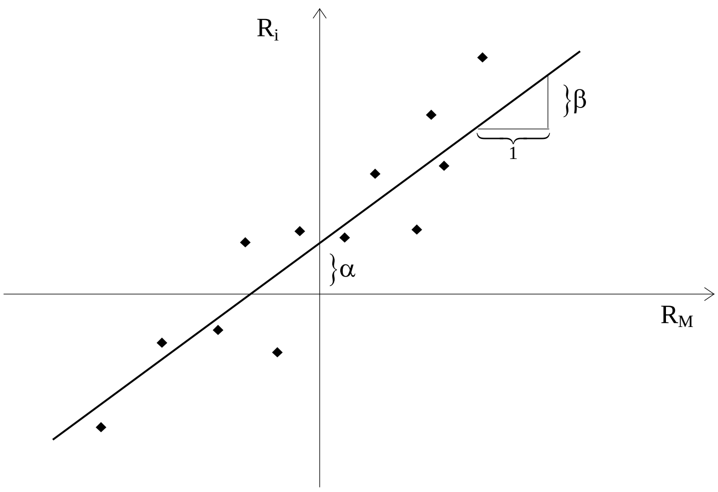


Figure 4-3: Estimating the βs with a regression

Let there be a series of returns for a single asset and for the market index. If we plot the asset returns against the market returns, we will probably get a scattered graph like the one in the figure above. We know that the points are on a line, plus/minus an error term. For this reason, we will try to draw the line that best fits our points. This is equivalent to saying that we want to minimise the sum of the squared error terms of our equation.

4.4.4 An illustration: estimating α and β and quantifying the precision

Let us illustrate all this with an example. The following table lists a set of 30 returns for a stock market index (denoted M) and for a portfolio (denoted i).

t	R_{Mt}	R_{it}	$R_{Mt}–E(R_M)$	$R_{it}–E(R_i)$	$[R_{Mt}– E(R_M)]·[R_{it}– E(R_i)]$	$[R_{Mt} – E(R_M)]^2$
1	1.9837%	–2.8920%	1.8580%	–1.5473%	–0.0287%	0.0345%
2	–0.4484%	–1.4669%	–0.5741%	–0.1222%	0.0007%	0.0033%
3	–1.8301%	–1.5515%	–1.9558%	–0.2067%	0.0040%	0.0383%
4	–1.4178%	–6.5201%	–1.5434%	–5.1754%	0.0799%	0.0238%
5	–1.7042%	–8.7879%	–1.8298%	–7.4431%	0.1362%	0.0335%
6	–4.6031%	–5.4707%	–4.7288%	–4.1260%	0.1951%	0.2236%
7	4.5779%	2.0986%	4.4522%	3.4434%	0.1533%	0.1982%
8	1.6255%	5.8628%	1.4998%	7.2076%	0.1081%	0.0225%
9	–3.9472%	–10.7743%	–4.0728%	–9.4295%	0.3840%	0.1659%
10	–3.2426%	–8.7411%	–3.3682%	–7.3963%	0.2491%	0.1135%
11	–3.3664%	–10.6126%	–3.4921%	–9.2679%	0.3236%	0.1219%
12	0.5693%	–1.4477%	0.4437%	–0.1029%	–0.0005%	0.0020%
13	3.3353%	2.5510%	3.2097%	3.8957%	0.1250%	0.1030%
14	–1.4239%	–1.7480%	–1.5495%	–0.4033%	0.0062%	0.0240%
15	–1.6261%	–1.9809%	–1.7518%	–0.6361%	0.0111%	0.0307%
16	3.7423%	8.6251%	3.6167%	9.9698%	0.3606%	0.1308%
17	–4.5111%	–7.6572%	–4.6368%	–6.3125%	0.2927%	0.2150%
18	–1.8398%	–6.0547%	–1.9655%	–4.7099%	0.0926%	0.0386%
19	3.0806%	7.5542%	2.9550%	8.8989%	0.2630%	0.0873%
20	–3.0676%	–5.2470%	–3.1933%	–3.9022%	0.1246%	0.1020%
21	2.7000%	6.2042%	2.5743%	7.5489%	0.1943%	0.0663%
22	1.3727%	0.3833%	1.2471%	1.7280%	0.0215%	0.0156%
23	3.4203%	–0.6167%	3.2946%	0.7280%	0.0240%	0.1085%
24	2.6141%	3.3802%	2.4885%	4.7250%	0.1176%	0.0619%
25	–4.7804%	–7.6248%	–4.9061%	–6.2801%	0.3081%	0.2407%
26	3.3756%	6.9062%	3.2499%	8.2510%	0.2682%	0.1056%
27	3.6740%	6.3474%	3.5484%	7.6922%	0.2729%	0.1259%
28	4.6518%	3.4624%	4.5262%	4.8071%	0.2176%	0.2049%
29	–2.5041%	–5.7860%	–2.6297%	–4.4413%	0.1168%	0.0692%
30	3.3593%	1.2625%	3.2336%	2.6073%	0.0843%	0.1046%
	$E(R_M)$ = 0.1257%	**$E(R_i)$= –1.3447%**			**Σ = 4.5061%**	**Σ = 2.8155%**

Table 4-1: Estimating alpha and beta

Using the classical regression formulae, we can estimate beta as

$$\beta_i = \frac{\sigma_{iM}}{\sigma_M^2} = \frac{\sum_{t=1}^{30}\left[\left(R_{it} - E(R_i)\right)\cdot\left(R_{Mt} - E(R_M)\right)\right]}{\sum_{t=1}^{30}\left(R_{Mt} - E(R_M)\right)^2} \approx \frac{4.5061\%}{2.8155\%} \approx 1.60$$

and alpha as

$$\alpha_i = E(R_i) - \beta_i \cdot E(R_M) \approx -1.55\%$$

Graphically, the situation can be represented as follows

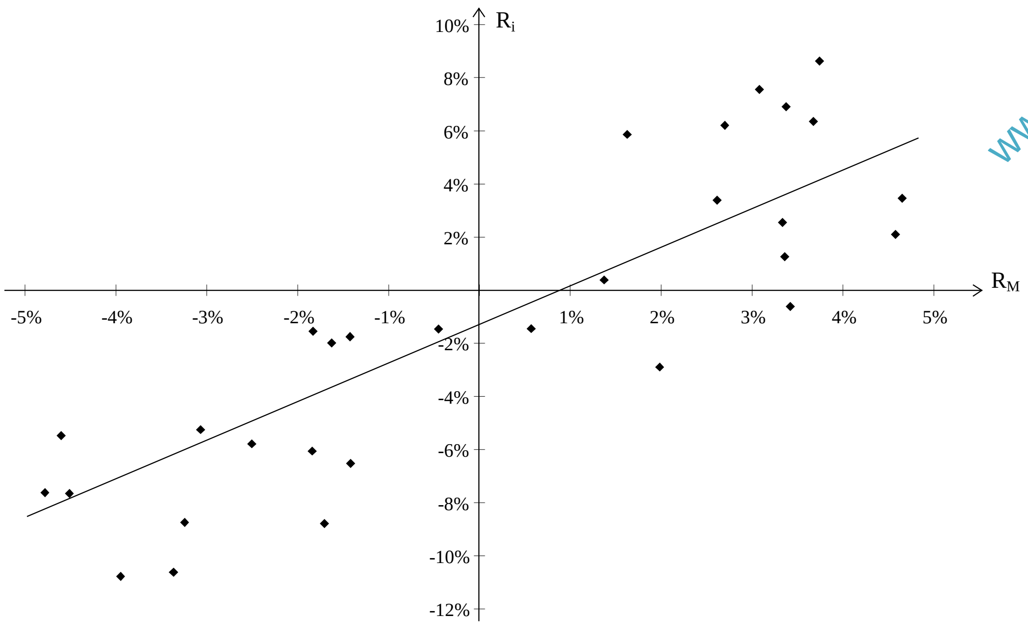


Figure 4-4: Estimating α and β with a regression

If all points were located on a line ($R^2 \cong 1$), an OLS regression would unambiguously identify α and β. However if it is not the case, the α and β will be imprecisely estimated. **It is therefore important not only to focus on the joint estimate of these parameters, but also to pay attention to the degree of precision with which they are estimated**. A measure of this precision is given by

$$\sigma(\beta) = \sigma_{\beta_i} = \sqrt{\frac{\sigma^2_{\varepsilon_i}}{\sum_{t=1}^{30}\left(R_{Mt} - E(R_M)\right)^2}} \cong \frac{\sigma_{\varepsilon_i}}{\sigma_M \cdot \sqrt{n}}$$

In our example, we have an R^2 of about 74.7%, and the precision of our beta is given by

$$\sigma(\beta) = \sigma_{\beta_i} = \frac{\sqrt{\frac{1}{28} \cdot \sum_{t=1}^{30}\left(R_{it} - (\alpha_i + \beta_i \cdot R_{Mt})\right)^2}}{\sqrt{\sum_{t=1}^{30}\left(R_{Mt} - E(R_M)\right)^2}} \cong \frac{\sigma_{\varepsilon_i}}{\sigma_M \cdot \sqrt{n}} \approx 0.18$$

As the beta estimate is 1.60, we can conclude that it is significantly larger than 1. To come to this conclusion we use the t-statistics computed as follows

$$t - \text{stat}(\beta): \quad \frac{\hat{\beta}_i - \beta_o}{\sigma_{\beta_i}} = \frac{1.6 - 1}{0.18} \cong 3.34$$

The t-statistics is above the 95% confidence interval critical value of 1.96, and we can, therefore, assess that it is statistically different from 1.

4.4.5 Predicting future betas

In an investment perspective, using betas derived directly from past data to estimate future betas implicitly assumes their stability over time. Unfortunately, empirical investigations show that portfolio betas remain stable over time, but betas of individual securities are unstable. For this reason, more sophisticated methods have to be applied to forecast future betas.

One simple approach would be to regress current betas against a large sample of past betas to estimate correction factors, A and B, as follows:

$$\text{Current beta} = A + B \cdot (\text{Past beta})$$

Then, using the estimates of A and B, we may write

$$\text{Forecast beta} = A + B \cdot (\text{Current beta})$$

But this methodology does not provide significant improvements. In fact, there is no reason to assume that there exists a linear relationship between current betas and past ones. Even if such a relationship did exist, it may not be stable over time.

Moreover, other financial variables can have some predictive power in forecasting betas, such as the variance of earnings, of cash flows, growth in earnings per share, market capitalisation (firm size), dividend yield, debt ratio, etc. An example of a regression model using such variables could be:

$$\text{Current beta} = A + B_1(\text{past beta}) + B_2(\text{variance of earnings}) + B_3(\text{dividend yield})$$

Empirical studies have suggested that betas tend to move toward 1 over time (as the firm grows and diversifies). Hence, a forecast of the future beta coefficient should take this into consideration, and use an adjusted beta. For example, Blume suggests for the US market

$$\beta_a = (0.66 \cdot \beta_h) + (0.34 \cdot 1.0)$$

where β_a is the adjusted beta, and β_h a historical beta.

4.5 Two applications of the market model

4.5.1 Computing the efficient frontier

The link between past and future returns provided by the market model is very useful in the computation of the efficient frontier. As a matter of fact, this was the main motivation behind the first exploration of the market model.

We have seen that using the market model,

$$R_i = \alpha_i + \beta_i R_M + \varepsilon_i$$
$$\sigma_{\varepsilon_i \varepsilon_j} = 0 \qquad \forall i \neq j$$

This implies

$$\sigma_i^2 = \beta_i^2 \cdot \sigma_M^2 + \sigma_{\varepsilon_i}^2$$

that is,

$$\sigma_{ij} = \beta_i \cdot \beta_j \cdot \sigma_M^2$$

The use of the market model to derive the inputs needed for the MPT significantly reduces the volume of information needed to compute the efficient frontier and simplifies computational difficulties.

- using the Markowitz procedure to calculate the efficient frontier for a set of N stocks would require N estimates of expected returns, N estimates of variances, and $(N^2 - N) / 2$ estimates of covariances.
- using the market model, we only need N estimates of expected returns, N estimates of the firm-specific variances, and (N + 1) terms (N estimates of the sensitivity coefficients β_i, and one estimate for the variance of the market); this would enable us to determine all the σ_{ij}.

The following table shows the gain for various values of N (number of stocks).

N	Markowitz	Market model
1	2	4
2	5	7
3	9	10
4	14	13
5	20	16
10	65	31
50	1'325	151
100	5'150	301
1'000	501'500	3'001
2'000	2'003'000	6'001
5'000	12'507'500	15'001

Table 4-2: Required data to compute the market model

This explains why the market model has been a considerable improvement over the original Markowitz model!

4.5.2 Components of market risk

One can also use the market model to decompose market risk in three components:

- the **world market risk** considers changes in the returns of all the stock markets of the world (such as the 1987 crash).
- the **national market risks** are changes in the returns of all the stocks of a specific country.
- the **industry risks** are risks affecting particularly all the firms of a specific sector of the national economy (banks, chemicals, etc.).

For this reason, we use specific indices for the world stock markets, the national stock market, and the specific industry. Note that these indices are not independent, since the latter is included in the former.

There are three steps for the evaluation:

1) Using the following regression

$$R_i = \alpha_i + \beta_i \cdot R_w + \varepsilon_i$$

calculate the percentage of the total variance explained by the world index return, denoted R^2 (W), where W stands for the world index. It is given by

$$R^2(W) = \frac{\beta_i^2 \cdot \sigma^2(R_w)}{\sigma^2(R_i)} = 1 - \frac{\sigma^2(\varepsilon_i)}{\sigma^2(R_i)}$$

2) Re-estimate the regression equation by adding the national market index:

$$R_i = \alpha_i + \beta_{1i} \cdot R_w + \beta_{2i} \cdot R_N + \varepsilon_i$$

This step will yield R^2(N), which is the total percentage of the variance of returns explained by the two indices (world and national). In order to get the pure national impact, subtract the variance already explained by the world index

$$\text{National component} = R^2(N) - R^2(W)$$

3) The last step is similar to the second one. Use the additional explanatory variable representing the industrial sector and estimate the regression:

$$R_i = \alpha_i + \beta_{1i} \cdot R_w + \beta_{2i} \cdot R_N + \beta_{3i} \cdot R_I + \varepsilon_i$$

Again, R^2(I) gives us the total percentage of the variance of returns explained by the three indices. In order to get the pure industrial impact, we have to subtract the variance already explained by the world and the national indices

$$\text{Industrial sector component} = R^2(I) - R^2(N)$$

The remaining variance is the variance specific to the asset and is therefore diversifiable. It is given by

$$\text{Firm specific component} = 1 - R^2(I)$$

This last variance should tend towards zero in a well diversified portfolio.

4.6 Multi-index models

4.6.1 Multi-index models

The assumption underlying any single-index model is that stock prices move together only because of their common movement with the single factor (generally: the market index). But there can be influences besides this factor that can cause stocks to move together. To consider other sources of covariance between securities, one has to use a **multi-index model**.

Multi-index models attempt to capture some of the non-market influences that cause securities to move together by introducing additional terms in the general return equation.

$$R_i = \alpha_i^* + \beta_{i1}^* \cdot I_1^* + \beta_{i2}^* \cdot I_2^* + \ldots + \beta_{in}^* \cdot I_n^* + \varepsilon_i$$

Such a model has very convenient mathematical properties if the indices are uncorrelated. As it is always possible to convert any set of correlated indices into a set of uncorrelated indices, we will assume that indices are uncorrelated, change the notation, and specify the model as:

$$R_i = \alpha_i + \beta_{i1} \cdot I_1 + \beta_{i2} \cdot I_2 + \ldots + \beta_{in} \cdot I_n + \varepsilon_i$$

where, by construction, and as we have assumed for the single index model,

- the expected mean of the residual (error terms) is zero $E(\varepsilon_i) = 0$ for all stocks
- the covariance between any two different indices equals zero:

$$E\left[\left(I_j - \bar{I}_j\right) \cdot \left(I_k - \bar{I}_k\right)\right] = 0$$

- the covariance between the residual returns for any stock and the returns on each index is zero:

$$E\left[\varepsilon_i \cdot \left(I_j - \bar{I}_j\right)\right] = 0$$

- and by assumption, the covariance between the residuals for any two different stocks is zero:

$$E\left[\varepsilon_i \cdot \varepsilon_j\right] = 0$$

The last assumption implies that there are no factors beyond the selected indices that account for co-movements between any two securities. There is nothing in the estimation model that forces this to be true, but if it were not the case, it would imply that there exists another factor (not considered in the model) that explains some of the co-movement between securities.

The simplest form of a multi-index model is a two-index model; for example, assuming that the two indices are the market return (R_M) and the unanticipated inflation (I), we could have:

$$R_i = \alpha_i + \beta_{iM} \cdot R_M + \beta_{iI} \cdot I + \varepsilon_i$$

The two betas respectively give us the sensitivity of the returns of the asset to changes in the market (the traditional beta) and to the unanticipated changes in inflation[24]. Just as in the context of the single-index model, the betas can be estimated by relating the stock's returns to the unexpected inflation.

The regression is performed the same way as for the single-index model, but instead of having the line of best fit, we will obtain the plane of best fit or even a hyperplane for more than two indices. Nevertheless, the principle remains the same: it is the locus of points that minimises the squared deviations from it relative to all the observed states of nature.

24 Which can be estimated by comparing the effective inflation with the forecasts of the leading forecasting groups of economists. Note however, that this type of data often causes problems, since it is not very accurate, it is only available on a monthly basis, etc.

4.6.2 The portfolio variance under a multi-index model

Using our previous model (with uncorrelated indices), we can write the variance of a portfolio of N stocks as:

$$\underbrace{\sigma_p^2}_{\text{total variance}} = \underbrace{\beta_{P,M}^2 \cdot \sigma_M^2}_{\text{syst. risk of market}} + \underbrace{\beta_{P,I}^2 \cdot \sigma_I^2}_{\text{syst. risk of inflation}} + \underbrace{\sigma_{\varepsilon P}^2}_{\text{residual variance}}$$

If these factors were correlated, the formulas would become more complex since the covariance terms would have to be introduced, but it would not affect the quality of the model. As in the case of the single-factor model, once all the parameters have been determined, the Markowitz's approach can be used.

Example:
The returns on a security "i" are generated by the following three factor model

$$R_i = 5\% + 0.2 \cdot F_1 + 1.1 \cdot F_2 + 0.9 \cdot F_3 + \varepsilon_i$$

where F_1, F_2, and F_3 are uncorrelated factors. If $E(F_1) = 4\%$, $E(F_2) = 3\%$, $E(F_3) = 2\%$, $\sigma_{F1} = 10\%$, $\sigma_{F2} = 11\%$, $\sigma_{F1} = 8\%$, $\sigma_{\varepsilon i} = 10\%$, we want to know what is the expected return on security i, as well as its standard deviation.
The expected return of security i is given by

$$\begin{aligned} E(R_i) &= 5\% + 0.2 \cdot E(F_1) + 1.1 \cdot E(F_2) + 0.9 \cdot E(F_3) \\ &= 5\% + 0.2 \cdot 4\% + 1.1 \cdot 3\% + 0.9 \cdot 2\% \\ &= 10.9\% \end{aligned}$$

and its standard deviation by:

$$\begin{aligned} \sigma_i &= \sqrt{0.2^2 \cdot \sigma_{F1}^2 + 1.1^2 \cdot \sigma_{F2}^2 + 0.9^2 \cdot \sigma_{F3}^2 + \sigma_{e_i}^2} \\ &= \sqrt{0.2^2 \cdot 10^2 + 1.1^2 \cdot 11^2 + 0.9^2 \cdot 8^2 + 10^2} \\ &\approx 17.39\% \end{aligned}$$

As in the single-factor model, the sensitivity of a portfolio to a particular factor in a multiple-factor model is a weighted average of the sensitivities of the securities, where the weights are equal to the proportion invested in each security. This can be seen by noting that the return on a portfolio is a weighted average of the returns of its component securities

$$\begin{aligned} R_p &= \sum_{i=1}^{N} x_i \cdot R_i \\ &= \sum_{i=1}^{N} x_i \cdot \left(\alpha_i + \beta_{iM} \cdot R_M + \beta_{iI} \cdot I + \varepsilon_i\right) \\ &= \left(\sum_{i=1}^{N} x_i \alpha_i\right) + \left(\sum_{i=1}^{N} x_i \cdot \beta_{iM} \cdot R_M\right) + \left(\sum_{i=1}^{N} x_i \cdot \beta_{iI} \cdot I\right) + \left(\sum_{i=1}^{N} x_i \cdot \varepsilon_i\right) \\ &= \alpha_p + \beta_{pM} \cdot R_M + \beta_{pI} \cdot I + \varepsilon_p \end{aligned}$$

where $\alpha_p = \sum_{i=1}^{N} x_i \cdot \alpha_i$, $\beta_{pM} = \sum_{i=1}^{N} x_i \cdot \beta_{iM}$, $\beta_{pI} = \sum_{i=1}^{N} x_i \cdot \beta_{iI}$, and $\varepsilon_p = \sum_{i=1}^{N} x_i \varepsilon_i$

If we assume that the residuals are uncorrelated, we can write

$$\sigma_{\varepsilon_p}^2 = \sum_{i=1}^{N} x_i^2 \cdot \sigma_{\varepsilon_i}^2$$

This equation for the residual variance should hold in a multi-index model. If the errors were correlated, this would mean that either we did not choose the right indices (for example, the unanticipated changes in inflation have no impact on asset returns) or that there is a need for an additional explanatory variable. As we will see in the next chapter, one may need to use at least four or five factors (indices) to specify an adequate model.

4.6.3 An example of a multi-index model

Salomon Brothers use a multi-index model with six variables to explain the returns on securities[25]. They consider:

- the economic growth (year-to-year changes in total industrial production), as a gauge of general economic well-being.
- the spread between the yields on the US Treasuries and investment grade corporate bonds, as a proxy for the default risk.
- the long-term interest rates (the yield change in 10 years US Treasuries) as an indicator of the attractiveness of default-free bonds.
- the short-term interest rates (the yield change in 1 month US T-bills) as an indicator of the attractiveness of short-term maturities versus longer-term instruments.
- the inflation shock which is measured by the difference between the realised inflation (Consumer Price Index CPI) and the expected inflation (derived from T-bills rate using an econometric method).
- the USD fluctuations against a trade-weighted basket of 15 currencies.

Salomon Brothers report that using monthly data, this model explains on average 41% of the fluctuations in returns for a sample of 1'000 stocks.

4.7 Conclusion

The market model has the advantage of being relatively simple. To apply the model, it is only necessary to know the covariance of each asset with the market. Thus, it drastically reduces the number of inputs needed for the determination of the efficient frontier.

The beta coefficient is the contribution of a single asset to the risk of the market portfolio. As such, it can only be used to determine the risk of a portfolio if the portfolio is efficient. Theoretically, all investors are supposed to hold an efficient portfolio. In reality, this is hardly ever the case. From a practical standpoint, the market model is not used for stock-picking, but for the analysis of the portfolio composition.

Nevertheless, the simplifications of a one-factor model also bring with them some limitations. We implicitly assume that all security returns can be explained by the market and the specific return. This makes it impossible for the model to account for shifts in some industries. These shifts might not overly affect the market as a whole, and hence, may not be reflected in the market returns. For this reason multi-factor models might be more suitable in explaining the returns on risky securities.

25 See SORENSEN E., MEZRICH J. and THUM C., 1989, “The Salomon Brothers U.S. Stock Risk Attributes Model", Salomon Brothers