Table of Contents

* final level

1. Modern Portfolio Theory

1.1 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 from 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.

1.1.1 Return and measures of return

1.1.1.1 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 CHF 350 and sells it at time t=1 at CHF 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 CHF 100; at time t=1 a dividend of CHF 10 is paid; at the same time, the stock is priced at CHF 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 CHF 100; at time t=0.5 a dividend of CHF 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 CHF 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 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.

1.1.1.2 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	CHF 100	CHF 110	10%	CHF 121	10%
B	CHF 100	CHF 150	50%	CHF 121	-19.3%
C	CHF 100	CHF 200	100%	CHF 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 $\frac{1}{2}\cdot\left(\frac{P_1 - P_0}{P_0} + \frac{P_0 - P_1}{P_1}\right)$ 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:

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

$$\text{for B}: \overline{R}^{(g)}_{0,2} = \sqrt{\left(\frac{P_1}{P_0}\right)\cdot\left(\frac{P_2}{P_1}\right)} - 1 = \sqrt{\left(\frac{150}{100}\right)\cdot\left(\frac{121}{150}\right)} - 1 = \sqrt{1.50\cdot 0.81} - 1 = 10\%$$

$$\text{for C}: \overline{R}^{(g)}_{0,2} = \sqrt{\left(\frac{P_1}{P_0}\right)\cdot\left(\frac{P_2}{P_1}\right)} - 1 = \sqrt{\left(\frac{200}{100}\right)\cdot\left(\frac{121}{200}\right)} - 1 = \sqrt{2.00\cdot 0.605} - 1 = 10\%$$

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

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

1.1.1.3 Time value of money: compounding and discounting

Compounding period equal to the holding period

As we have seen receiving CHF 1 today is worth more than receiving CHF 1 tomorrow. Both CHF 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.

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:

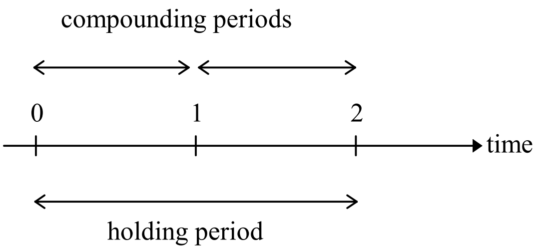


Figure 1-1: Holding period longer than compounding period

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 CHF 100 on his bank account at a 10% annual rate, he will receive CHF 10 at the end of the first year. At the end of the second year, he will receive again CHF 10, corresponding to the interest paid on the principal amount, plus an extra amount of CHF 1 corresponding to the second year interests on the first year interest (10% on CHF 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 CHF 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 $100 \cdot 1.07 \cdot 1.09 \cdot 1.10 =$ CHF 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 CHF 100 invested at a simple versus a compound rate, for various rates and various holding periods.

Time (Years)	Simple Rate	Compoun d Rate	Simple Rate	Compoun d Rate	Simple Rate	Compoun d 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.

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:

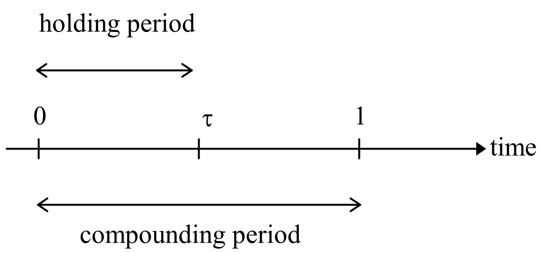


Figure 1-2 Holding period shorter than compounding period

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:

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

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

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

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

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.

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

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

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

and that

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

one has:

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

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.

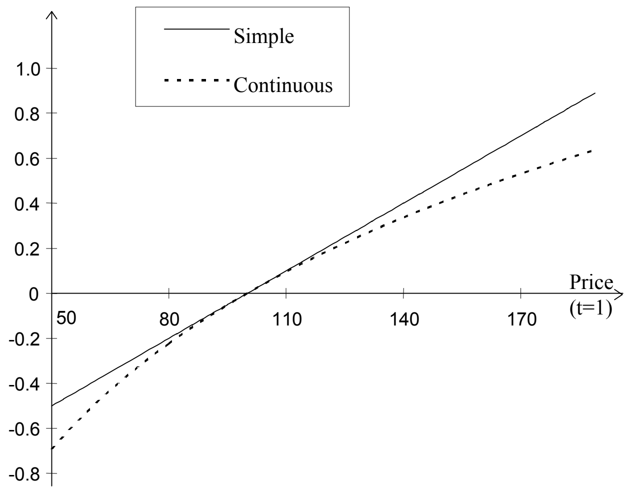


Figure 1-3: Approximating simple returns by continuously compounded returns

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

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

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.

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.

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

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	CHF 100	CHF 110	9.53%	CHF 121	9.53%
B	CHF 100	CHF 150	40.54%	CHF 121	–21.48%
C	CHF 100	CHF 200	69.31%	CHF 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:

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

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:

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

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

1.1.1.4 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.

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:

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

Example:
Let there be a stock worth CHF 100 on 31st of December and CHF 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.

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 CHF 100 on 31st of December and CHF 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.

1.1.2 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.

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

---

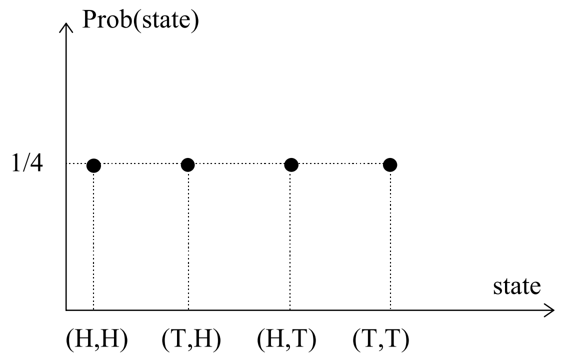
2 Otherwise every investor would simply invest his/her entire wealth in the one asset paying the highest return.

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).

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.

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:

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 modeling 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).

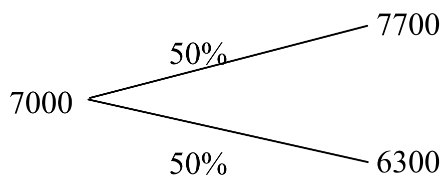


Figure 1-4: A simple binomial tree

This evaluation is relatively rough, since there are only two possible outcomes after one year. Let us refine the estimates by assuming that our investor believes that the first half of the year will be stronger than the second. For instance he might place 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:

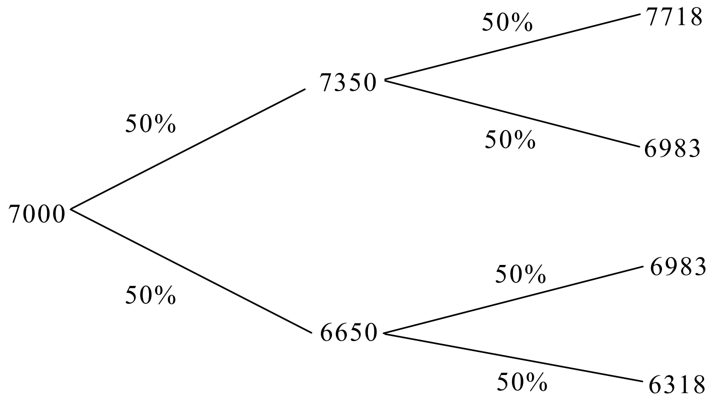


Figure 1-5: A two-period binomial tree

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 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.

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 Swiss stock market, and who is interested in predicting the return on the Swiss market over the next year. The only available information is the list of the last 50 years of annual returns on a market index.

1950	9.68%	1960	44.46%	1970	–10.65%	1980	6.07%	1990	–19.31%
1951	19.53%	1961	49.39%	1971	15.50%	1981	–11.91%	1991	17.67%
1952	8.37%	1962	–17.71%	1972	20.73%	1982	13.26%	1992	17.64%
1953	10.48%	1963	–0.16%	1973	–20.00%	1983	27.29%	1993	50.81%
1954	26.14%	1964	–6.93%	1974	–33.14%	1984	4.52%	1994	-12.26%
1955	5.99%	1965	–7.00%	1975	46.76%	1985	61.36%	1995	25.45%
1956	2.12%	1966	–12.09%	1976	7.89%	1986	9.71%	1996	19.54%
1957	–10.25%	1967	47.19%	1977	8.09%	1987	–27.48%	1997	58.93%
1958	22.76%	1968	39.49%	1978	–0.51%	1988	23.61%	1998	14.29%
1959	29.20%	1969	4.48%	1979	10.93%	1989	22.59%	1999	5.72%

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 36/50=0.72 chances out of one to have a return greater than zero, and 14/50=0.28 chances out of one to have a return lower than zero. This defines a (very simple) **discrete distribution** that we can write as:

$$\begin{cases} \text{Prob}(R \le 0\%) = 0.28 \\ \text{Prob}(R > 0\%) = 0.72 \end{cases}$$

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 $R \le 0\%$ will have an area of 0.28 (or 28% of all boxes total surface), and the box associated with $R > 0\%$ will have an area of 0.72 (or 72% of all boxes total surface).

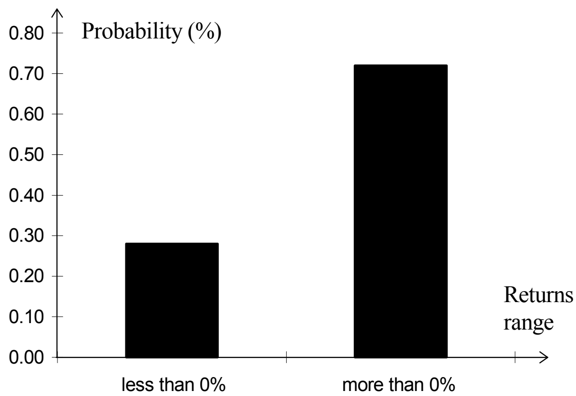


Figure 1-6: A first probability distribution

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%.

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.

Applying the same methodology would yield the following distribution:

$$\begin{cases} \text{Prob}(R \le -15\%) = 0.10 \\ \text{Prob}(-15\% < R \le 0\%) = 0.18 \\ \text{Prob}(0\% < R \le 15\%) = 0.30 \\ \text{Prob}(15\% < R \le 30\%) = 0.26 \\ \text{Prob}(R > 30\%) = 0.16 \end{cases}$$

this can also be represented as a histogram:

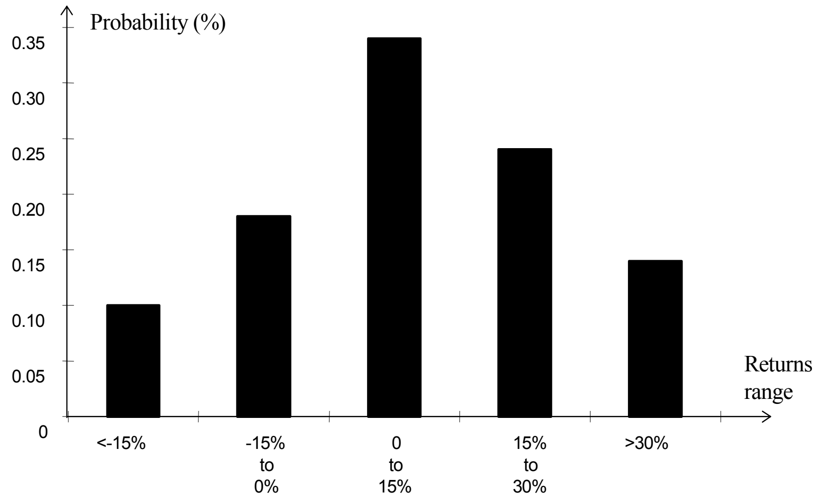


Figure 1-7: A second probability distribution

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.

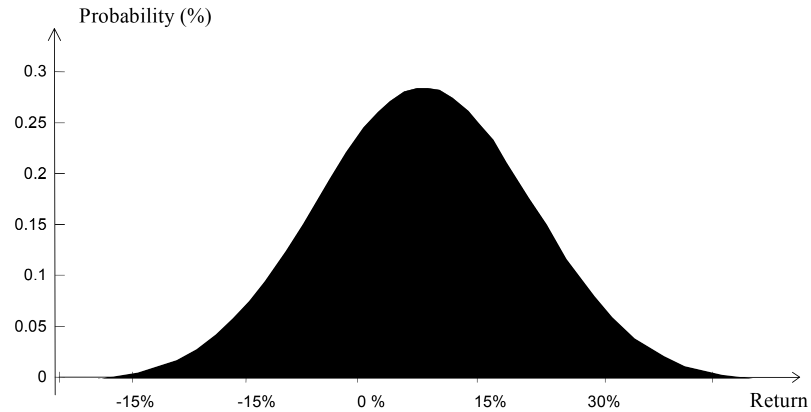


Figure 1-8: Limiting case: a continuous distribution

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).

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.

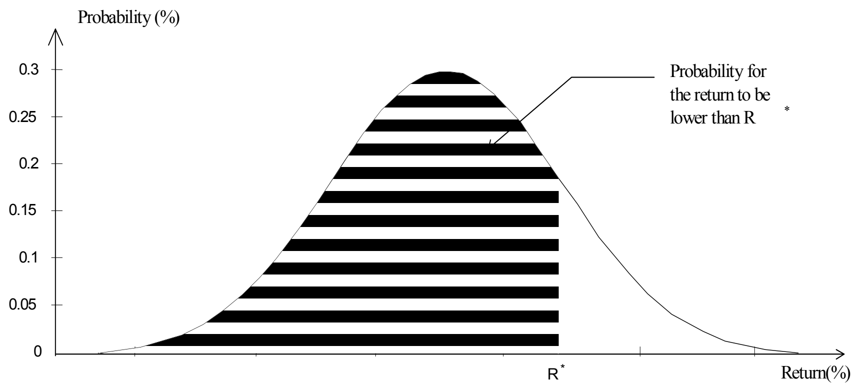


Figure 1-9: Probability distribution and upper bound

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

$$\text{Prob}[R > R^*] = 1 - \text{Prob}[R \leq R^*]$$

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.

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.

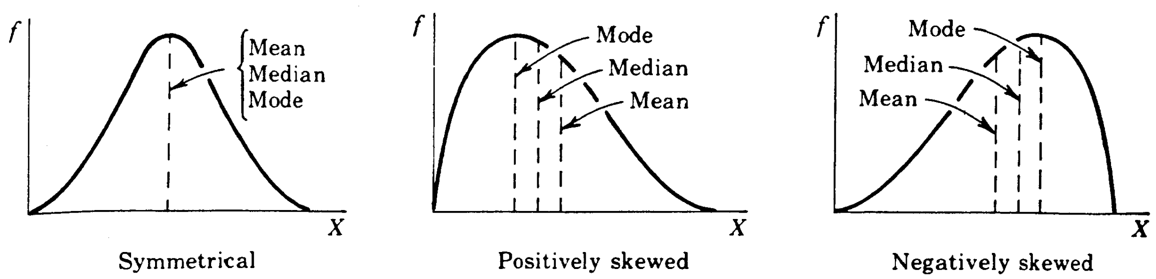


Figure 1-10: Mean, median, and mode of various continuous distributions

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[7]. 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**:

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

7 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.

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:

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

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

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	
Investment 1	CHF 100	CHF 95	CHF 115
Investment 2	CHF 100	CHF 90	CHF 120

The expected value in t=1 for both investments is equal to CHF 105, but the standard deviation is different:

$$\sigma_1 = \sqrt{0.5 \cdot (115-105)^2 + 0.5 \cdot (95-105)^2} = 10$$
$$\sigma_2 = \sqrt{0.5 \cdot (120-105)^2 + 0.5 \cdot (90-105)^2} = 15$$

The second investment opportunity is clearly 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.

In order to compare assets with very different price levels we need a relative measure of the dispersion of the returns. A way to do this is to express the dispersion per unit of expected price; this is the **coefficient of variation (CV)**:

$$CV = \frac{\sigma_i}{E(P)}$$

In the example above we used two investments with the same price level. We will now see what happens if the price levels are different.

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	
Investment 1	CHF 1000	CHF 950	CHF 1150
Investment 2	CHF 100	CHF 90	CHF 120

In t=1 we have following standard deviation:

$$\sigma_1 = \sqrt{0.5 \cdot (1150-1050)^2 + 0.5 \cdot (950-1050)^2} = 100$$
$$\sigma_2 = \sqrt{0.5 \cdot (120-105)^2 + 0.5 \cdot (90-105)^2} = 15$$

The first investment opportunity looks now clearly riskier than the first one. But we know that this is not true. We need to adjust the units. If we use the coefficient of variation we have

$$CV_1 = \frac{100}{1050} = 0.095238$$
$$CV_2 = \frac{15}{105} = 0.142857$$

Now we have the same result as in the previous example.

The coefficient is often used as a measure of relative risk, in which case we express the risk per unit of expected return. The formula then becomes:

$$CV = \frac{\sigma_i}{E(R)}$$

Let us illustrate this with a further example:

Example:
Let there be two investments with following expected return and standard deviation.

	E(R)	σ
Investment 1	0.08	0.05
Investment 2	0.12	0.07

Based on absolute values, investment 2 looks riskier with a standard deviation of 7 percent than investment 1 which has a standard deviation of only 5 percent. However if we take the expected return into account we have:

$$CV_1 = \frac{0.05}{0.08} = 0.6250$$
$$CV_2 = \frac{.07}{0.12} = 0.5833$$

Now we see that investment 1 actually shows a higher risk per unit of expected return than investment 2.

The normal distribution
A common continuous probability distribution used in finance is the **normal distribution**. Its probability density is given by the following function[8]:

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

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 CHF 1 and each time a head comes up, the player must pay CHF 1. If the coin is well balanced (no cheating is possible), the player has equal chances of paying or receiving CHF 1. After the first round, he will either have won or lost CHF 1. If he won the first round, after the second round, he will

8 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.

either have CHF 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 CHF 2. Hence after two periods, he has out of 4 possible outcomes either lost or won CHF 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 $3\sqrt{n}$ CHF 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.
- 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 $3 \cdot \sigma$.

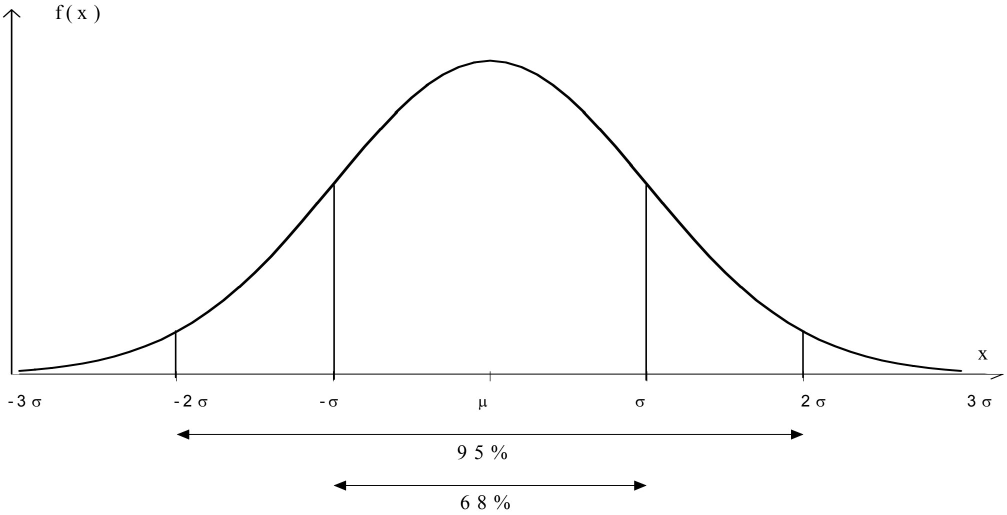


Figure 1-11: The normal distribution

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

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.

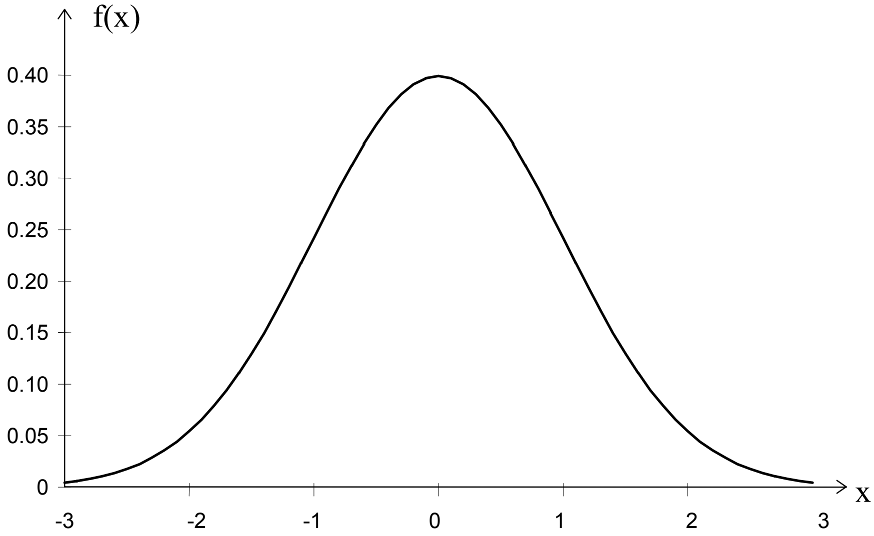


Figure 1-12: The standard normal distribution

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%?

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

As $\frac{(r-11.5\%)}{31\%}$ 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

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

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:

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

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

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

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

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.

1.1.2.2 Computing and annualising volatility and practice

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,

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

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[9]. 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[10]. But we have seen that returns are multiplicative over time[11].

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%.

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

10 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.

11 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 $(R_{0,1}+R_{1,2})$.

Annualising volatility

Another practical problem often encountered in practice is that the sample from which the volatility is computed has a time length that differs from the desired one. For instance, we have estimated a volatility of 6% using six month of data, but we are interested in the annual volatility.

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 volatility of the ABC stock is 3% for one month. 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.

1.1.2.3 Statistical concepts

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

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

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

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

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.

1.2 Efficient markets

To introduce the idea of market efficiency, a relatively abstract concept, it is easiest to start with an example. Consider the following stock price time series of a drug company called Firm X.

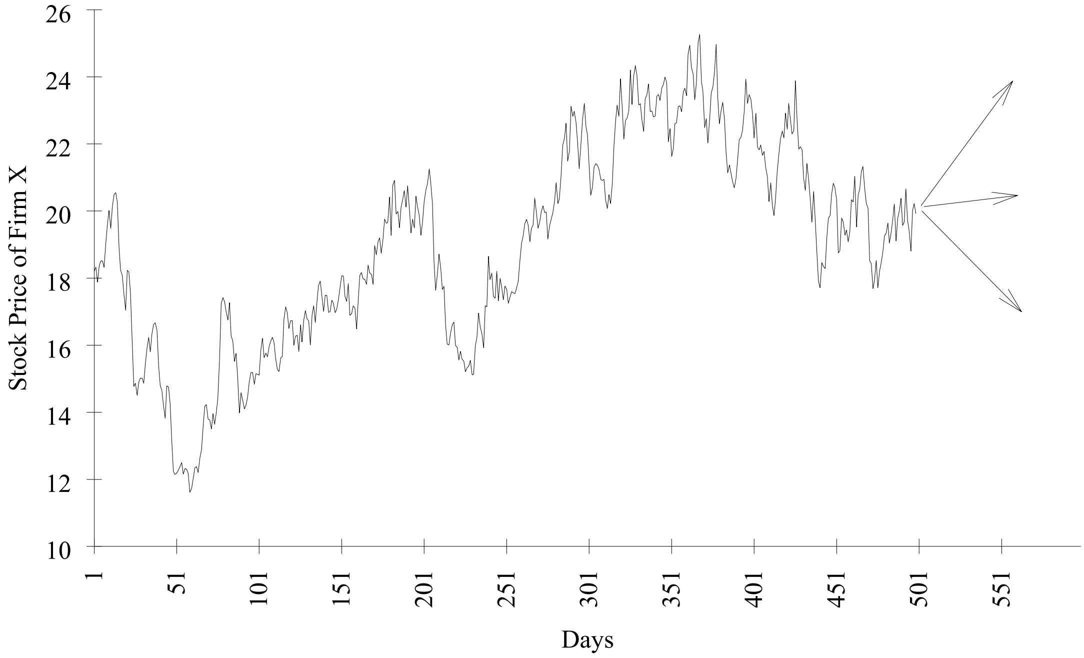


Figure 1-13: Stock price of firm X: value over time

Assume that on the 501st day the price of the stock of the firm X is CHF 20. What will be the price of the stock of the Firm X on the 502nd day? Also, assume that the news that Firm X is about to undertake a project with very poor prospects becomes known to public. Hence, some of the investors who own stock X will try to sell it as soon as possible. Others will be reluctant to buy the stock at the prevailing price. Because of these reactions, the stock price will quickly fall to a lower equilibrium level on that day.

Alternatively, what would happen if Firm X announces a breakthrough in the research for a new drug? If an assumption is made that will take another two years for the drug to become available to patients, what will rational investors do in this situation? They will try to buy the stock as soon as possible at the lowest possible price. Of course, many other investors in the market will simultaneously try to do the same. Nobody will wait two years until the higher cash flows materialize. In this case, the stock price will increase as soon as the new information becomes available.

To illustrate the process of settling down, consider the effect of bad news on the price of a five-year zero coupon bond. The bond has a face value of CHF 1'000. The relevant opportunity rate for a bond with the same maturity and the same risk is 7% (assume a flat term structure of interest rates).

Given this discount rate, today's bond price can be calculated as follows:

$$P_0 = \frac{1000}{1.07^5} = 712.99$$

We therefore discount the expected future payout of the bond with the appropriate opportunity rate of return. The discount rate of 7% represents the required rate of return any investor will get on average in the capital market for an investment with the same risk characteristics.

If we assume that the relevant time horizon of the investor is only one year, what's the expected rate of return for an investor who holds the bond from year 0 to year 1? Assume interest rates do not change.

$$\text{Expected rate of return} = \frac{\text{price}_1 - \text{price}_0}{\text{price}_0} = \frac{\frac{1000}{1.07^4} - 712.99}{712.99} = 7\%$$

What happens if investors expect an increase in the opportunity rate of return, i.e. if the expected rate of return rises to 10% p.a.? Today's bond price will drop. What will the new, lower equilibrium price of the bond be? The bond price has to fall until the expected rate of return of the bond equals 10%. So how can we determine the new bond price?

$$P_0 = \frac{1000}{1.10^5} = 620.92\,\text{CHF}$$

The new price of this bond will be CHF 620.92. This price implies an expected rate of return of 10%.

$$\text{Expected rate of return} = \frac{\frac{1000}{1.10^4} - 620.92}{620.92} = 10\%$$

Now what will happen, if the new price is higher (or lower) than CHF 620.92? We will answer this question for the case in which the price is higher, say CHF 649.93. Assume any investor can buy the bond for this price. What return, y, does he expect to receive on average for five years?

$$P_0 = \frac{1000}{(1+y)^5} = 649.93\,\text{CHF}$$

$$y = \left(\frac{1000}{649.93}\right)^{\frac{1}{5}} - 1 = 9\%$$

What will investors do if they can choose between two investments, both with the same risk characteristics, one offering a return of 9% (a bond) and the other offering 10% (an alternative investment in the capital market)? They will invest in the capital market. Will any rational investor put his wealth in the bond? Ignoring risk considerations, the answer is no, since everybody prefers a rate of return of 10% to a rate of return of 9%. What will happen to the price of the bond under these circumstances? Since no one wants to hold the bond, its price will fall. It will fall, in fact, until investors are indifferent about investing either in the bond or in the capital market. This will take place as soon as the bond offers a rate of return of 10%. Therefore, the price has to fall to exactly CHF 620.92.

We can think also about *equilibrium* expected price changes (opportunity rates of return). If we ignore cash payouts, the price of a stock increases at the same rate on average as the opportunity rate of return of an asset with the same risk. For example, assume that today's stock price of Firm X is CHF 20, and the market requires a rate of return of 20% per year for investment opportunities with a risk similar to that of Firm X. In such a case, we would expect the stock price of Firm X to increase at a rate of 20% per year. And in equilibrium, we expect the stock price to be CHF 20·1.2 = CHF 24 in one year.

What can we learn from this discussion? Any investment strategy has to cover its costs. In other words, the invested capital has to earn a return that is appropriate for the risk of the investment, the time value of money and the transaction costs of the strategy. In a competitive capital market, most of the long-term investment strategies are just covering these costs. But just as capital markets are competitive, so, too, are markets for goods and services. What income do you expect, for instance, from opening a new barber shop? Can you expect to make large profits? The answer, in general, is no, as long as you have no comparative advantage as a barber. If it were possible, in fact, to make large profits, what would happen? Other people would recognise this, open new barbershops and try to imitate the successful business strategies. Of course, in the long run this will reduce everybody's profits to a level that covers the costs of running a barbershop.

Capital markets are competitive. They produce and process information. This process continuously repeats itself for all the securities traded in the market; as in any other industry, there is entry where there are profits, and there is exit where are losses.

News that is cheap to process is quickly reflected in prices. This includes easily-observed regularities or trends that are easy to spot. Additionally, the market cannot overreact systematically to dividend announcements. Any such patterns are self-destroying in a competitive market.

News that is more expensive to process may take longer to be reflected in prices. This is a source of potential profits. In the long run, however, and unless there are superior talents, competition will drive these profits down. This does not mean necessarily that competition makes all information production technologies obsolete. If that were the case, new information production opportunities would be created. All it means is that these technologies simply cover their costs (opportunity costs of time etc.). There are many industries that survive simply by covering production costs-the real world abounds with examples. Of course, in the short run, many investment strategies can be profitable (and many can be catastrophic).

A market in which security prices adjust rapidly to the publication of new information, and therefore in which the current prices of the securities fully reflect all information about that security is commonly said to be an **efficient market**, or more precisely an **informationally efficient market**.

This precision is noteworthy because the reason of being of a market is the allocation of capital to the most promising investment opportunities in the market. This is what we call an **allocationally efficient market**. In order for a market to be allocationally efficient it must be both externally and internally efficient. An **internally efficient market** is a market in which transaction prices are low and execution speed is high thanks to fierce competition among brokers and dealers. An **externally efficient market** is what we defined as an informationally efficient market.

Note that from now on, we will use the term efficient market in the sense of external market efficiency.

1.2.1 Information efficient markets

Prices of stocks, bonds and other financial assets are determined in the capital markets. For example, the stock price P_0 of Firm X can be determined as follows:

$$P_0 = \sum_{t=1}^{\infty} \frac{E(Div_t)}{(1+R_t)^t}$$

where:

E	Expectation operator
Div_t	Dividend payment in t
R_t	Opportunity rate of return

Therefore, to price assets, market participants form their expectations of future interest rates, future risk characteristics of the firm, and future cash distributions, from various types of information. In the case of expected future distributions of a stock, this information can include the following characteristics of a firm:

- Product quality of Firm X
- Capital budgeting policy of Firm X
- Financial policy of Firm X
- Experience and abilities of Firm X's management
- Future macroeconomic perspectives
- Growth opportunities of Firm X's industry
- Main competitors of Firm X

1.2.1.1 Assumptions

An information efficient market requires a *large number of competing market participants, each of which independently analyses and values securities in order to optimise their profit.*

A second assumption is that *the competing investors attempt to adjust the price of securities immediately to all available and relevant information in order to reflect the effect of it.*

If prices are bid immediately to fair levels, it must be that they increase or decrease only in response to new information. Thus, the third assumption is that *the price changes at any moment are based solely on the random arrival of new information*. This is the essence of the argument that stock prices should follow a random walk, that is, that price changes should be random and unpredictable. Any information that could be used to predict stock performance must already be reflected in stock prices.

We have just seen that competitive pressure forces the security's price to the new level instantly. Moreover, as long as everyone is attempting to draw reasonable judgements regarding the implications of the information for the security's price, the new price level will unbiasedly represent the market's summary judgement of the value of the security. Neither systematic underreaction nor systematic overreactions are possible in an information efficient market.

Although it may not literally be true that 'all' relevant information will be uncovered, it is virtually certain that there are many investigators that may improve investment performance. Competition among these analysts ensure that, as a general rule, stock prices ought to reflect all available information.

1.2.1.2 Characteristics of perfect information efficient markets

There are some noteworthy observations about **perfect** information efficient markets.

1. *Investors should expect to make a return on their investment that just covers its costs (fair return)*

   This means that using fundamental analysis or technical analysis in order to find mispriced securities will not generate above average returns.

   Technical trading refers to the attempt to predict future prices from the pattern of past price movements. It is essentially the search for recurrent and predictable patterns in stock prices. This approach is diametrically opposed to the notion of an efficient market.

   To demonstrate this, let us examine the stock price pattern of Firm X again. In Figure 1-14, we've added three straight lines, indicating possible stock price tendencies over periods of 150 days. The example shows one possible pattern. Of course, there are many other patterns that could be detected. Note that it is always possible to identify patterns in a series of historical prices ex post.

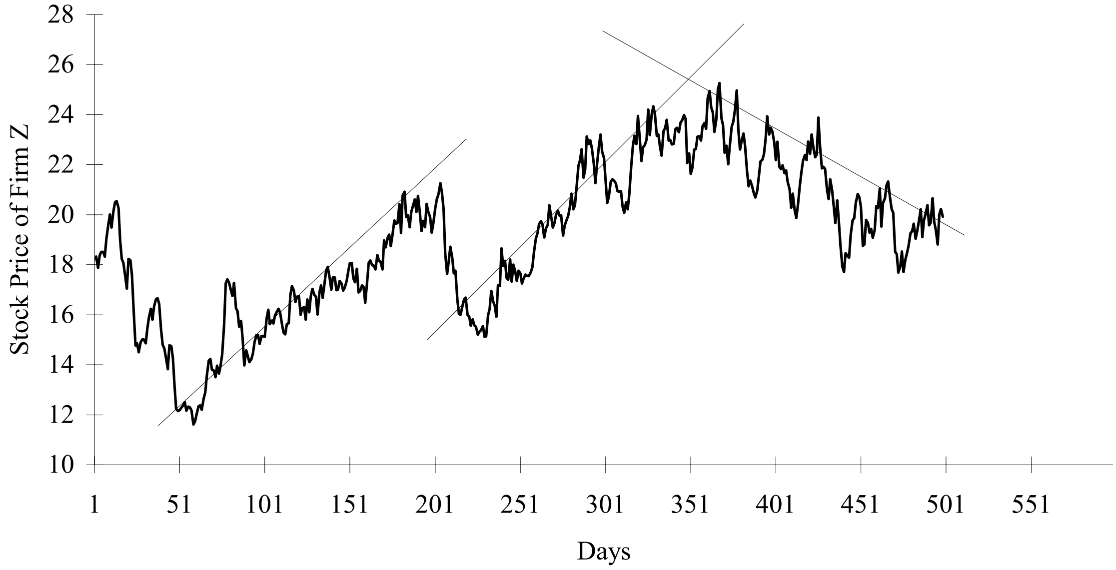


Figure 1-14 Stock price of firm X

   Alternatively, in Figure 1-15 we take the same stock price data and calculated a 20-day moving average from day 1 to 100. This is a very common approach to generate buy and sell signals. One possible rule to follow would be to buy the asset as soon as the actual stock price becomes higher than the 20-day moving average, and vice versa.

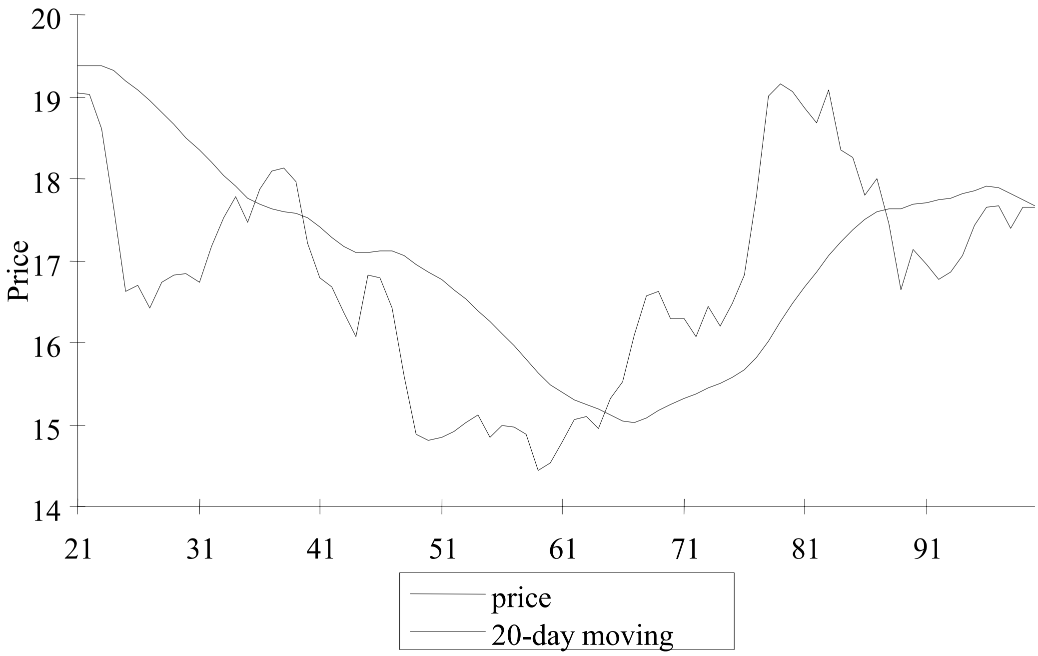


Figure 1-15: Price observations and moving average

If capital markets are information efficient, it is not possible to generate positive abnormal returns with this investment strategy. The calculation of the 20-day moving average relies exclusively on historical stock data. But since, in an efficient market, all available information is already reflected in the actual stock price, historical data have no power to predict future stock prices.

What neither Figure 1-14 nor Figure 1-15 show is that the data are not the actual prices of any real market; they were artificially computed with a random number generator in a spreadsheet program. The point is, that it is always possible, on hindsight, to discover 'patterns' in any time series. But often these patterns are merely a fiction of our imagination.

Some chartists also work with *filter rules,* mathematical rules that can be applied to produce buy and sell signals. For example, a filter rule may dictate that a stock should be purchased when it moves up in price by z percent or sold (or shorted) when it falls from its previous high by z percent. Fama and Blume[12] compared the rates of return earned from applying such a trading rule with the rates of return earned from a policy of buying and holding the common stock for each of the 30 Dow-Jones Industrial Securities. The following table shows the average rates of return of the two strategies at different filter levels.

12 FAMA Eugene F. and BLUME M., 1966, “Filter Rules and Stock Market Trading Profits”, Journal of Business (Supplement)

Filter	Filter Rule Return (average)	Buy and Hold Return (average)
0.005	0.115	0.104
0.010	0.055	0.103
0.020	0.002	0.103
0.050	–0.019	0.100
0.100	0.030	0.093
0.200	0.043	0.098
0.300	–0.005	0.064

Table 1-5: Average annual rates of return applied to 30 common stocks (1956-1962)[13]

These figures show that only the 0.5% filter, with its 11.5% return is higher than the 10.4% return of the buy and hold strategy. Telling whether this difference is statistically significant requires more information than is available. In particular, these calculations ignore transaction costs. Over the period concerned, 12'514 transactions would have been necessary to follow the 0.5% filter rule. Taking the commissions of the NYSE into consideration, the average annual return from this filter strategy turns out to be negative (–103.5%).

2. *Future performance cannot be deduced from past performance*
   Investment strategies that where successful in the past are no more likely to perform better than strategies that where not successful in the past.

3. *Markets can only be efficient if enough person believe the market is not efficient*
   This statement is only in appearance a paradox. Indeed, what makes the price of securities reflecting the true value is precisely the independent analysis of securities done by of numerous investors.

   If everybody believed that the markets are perfectly efficient and that it is therefore not possible to generate above average profit by searching for undervalued securities nobody would bother to analyse securities. Consequently the price of the securities could not adjust to new information and thus the security might get mispriced.

4. *Capital markets react quickly and completely to new information*
   As soon as new information is available to capital markets, price reactions are very quick. And to profit to the full extent from this new information, investors trade stocks quickly according to their changing expectations. So, competitive pressure forces the security's price to the new level instantly. Moreover, as long as everyone is attempting to draw reasonable judgements regarding the implications of the information for the security's price, the new price level will unbiasedly represent the market's summary judgement of the value of the security. Neither systematic underreaction nor systematic overreaction are possible in an information efficient market.

   Some empirical studies have examined the stock price reaction of a sample of firms that became targets in a merger before and after the merger announcement. In most take-overs, stockholders of the acquired firms (targets) sell their shares to the acquirer at substantial premiums over market value. The announcement of a take-over attempt is good news for shareholders of the target firm and should therefore cause stock prices to jump.

13 FAMA Eugene F. and BLUME M., 1966, “Filter Rules and Stock Market Trading Profits”, Journal of Business (Supplement)

Figure 1-16 shows the typical price pattern around take-over announcements.

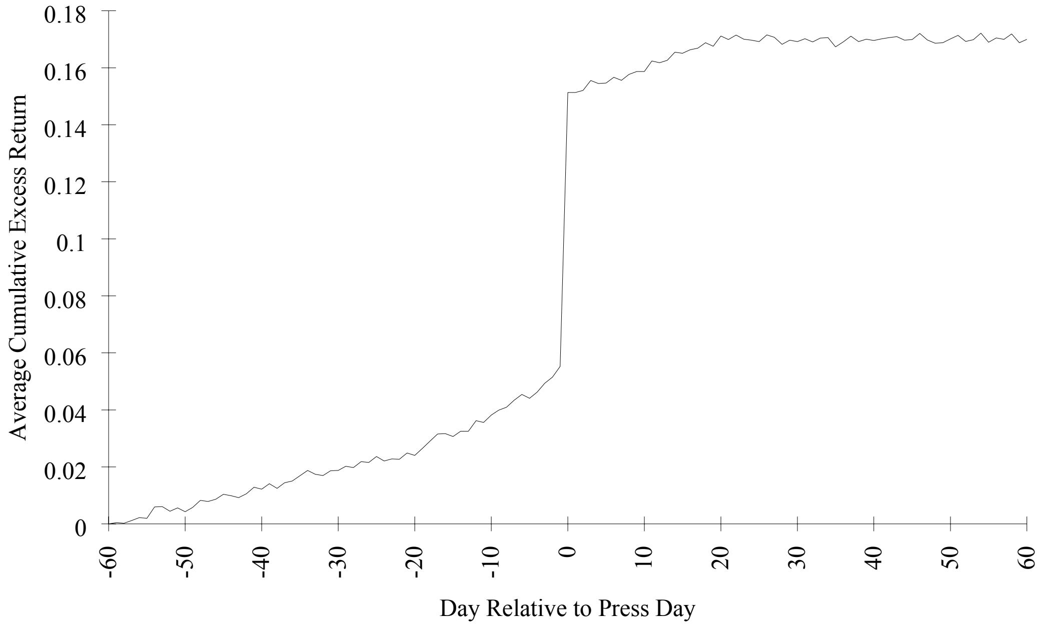


Figure 1-16: Price pattern of selling firms around take-over announcements

On the announcement day (day 0) one can observe a large, positive abnormal return (defined below). Immediately after the announcement, the stock price no longer increases or decreases abnormally. This price reaction pattern is perfectly consistent with the notion of information efficient capital markets. The price reaction occurs quickly (within one day) and completely on the day the new information becomes public (announcement day). On average, no new information becomes public on this particular sample of firms after the announcement day. Consequently, there are no apparent systematic patterns in the cumulative abnormal returns after day zero.

5. *On average, capital markets participants ignore irrelevant information*

If we ignore tax considerations and information signalling, stock splits per se should be irrelevant events. Assume an investor owns a share with a nominal value of CHF 100 and a market value of CHF 500. This firm wants to split the shares 2:1. This means that every investor gets two new shares in exchange for one old share. The investor now owns two shares with a face value of CHF 50 and a market value of CHF 250. In this example, one would expect the market not to react with an abnormal return on the split announcement date. This problem has been examined extensively in the empirical literature.

Figure 1-17 shows the typical reaction pattern of stock prices before and after a stock split.

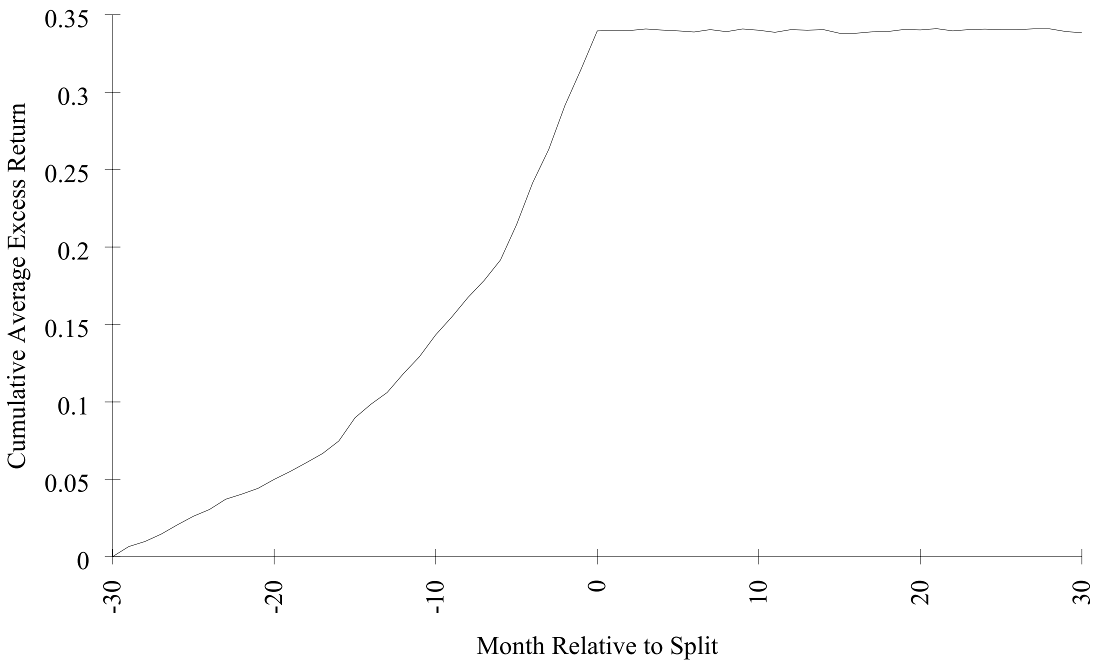


Figure 1-17: Stock price pattern surrounding stock splits

Positive abnormal returns are observed before the stock split, because firms tend to split in times when prices have gone up. In the month of the split and thereafter, however, there is no tendency for the cumulative abnormal return to increase. This is consistent with efficient capital markets. Market participants only react to new (unexpected) information, announcement of an unexpectedly high dividend, unexpectedly high earnings, etc.

1.2.1.3 An example of market efficiency (the Swiss evidence)

It is commonly argued that everything is different in Switzerland. Some people claim that although market efficiency is an important theoretical concept, the Swiss stock markets are not efficient. The point is that these people usually mix up efficiency with liquidity. There is no doubt that the market for some, possibly most Swiss stocks is illiquid. But this does not mean that the market does not process information quickly and to the full extent, i.e. that the market is information inefficient. Probably one of the best demonstrations that Swiss stock markets are, in fact, efficient is the case of Nestlé (by all measures, a very liquid stock).

On November 17, 1988 the Board of Directors of Nestlé AG decided to allow foreign investors to hold Nestlé's registered stock, reversing a long-standing practice. Until 1988, only domestic investors acceptable to management could hold Nestlé’s registered stock. Foreign investors could trade freely in the other two classes of stock, voting and non-voting bearer shares. On November 15, Nestlé's capital structure was as given in Table 1-6.

Shares	Voting bearer	Nonvoting bearer	Registered
Number of shares	1'073'000	1'150'000	2'227'000
Par value per share	CHF 100	CHF 20	CHF 100
Price per share	CHF 8'790	CHF 1'280	CHF 4'310

Table 1-6: Nestlé's capital structure on the 15th of November 1988[14]

As Table 1-7 and Figure 1-18 show, Nestlé's decisions had a dramatic impact on its share prices. From the closing on the 15th of November 1988 to the closing on the 22nd of November 1988, the price of the voting bearer stock dropped by 24%, the non voting bearer stock price fell by 15%, and the price of the registered stock surged by 33%.

Shares	Voting bearer	Nonvoting bearer	Registered
Price on the 15th of Nov.	CHF 8’790	CHF 1’280	CHF 4’310
Price on the 22nd of Nov.	CHF 6’650	CHF 1’090	CHF 5’740
Change	–24.35%	–14.84%	33.18%
Aggregate change	CHF –2'296 Mio.	CHF –219 Mio.	CHF 3'185 Mio.

Table 1-7: Announcement effect of Nestlé's policy change[15]

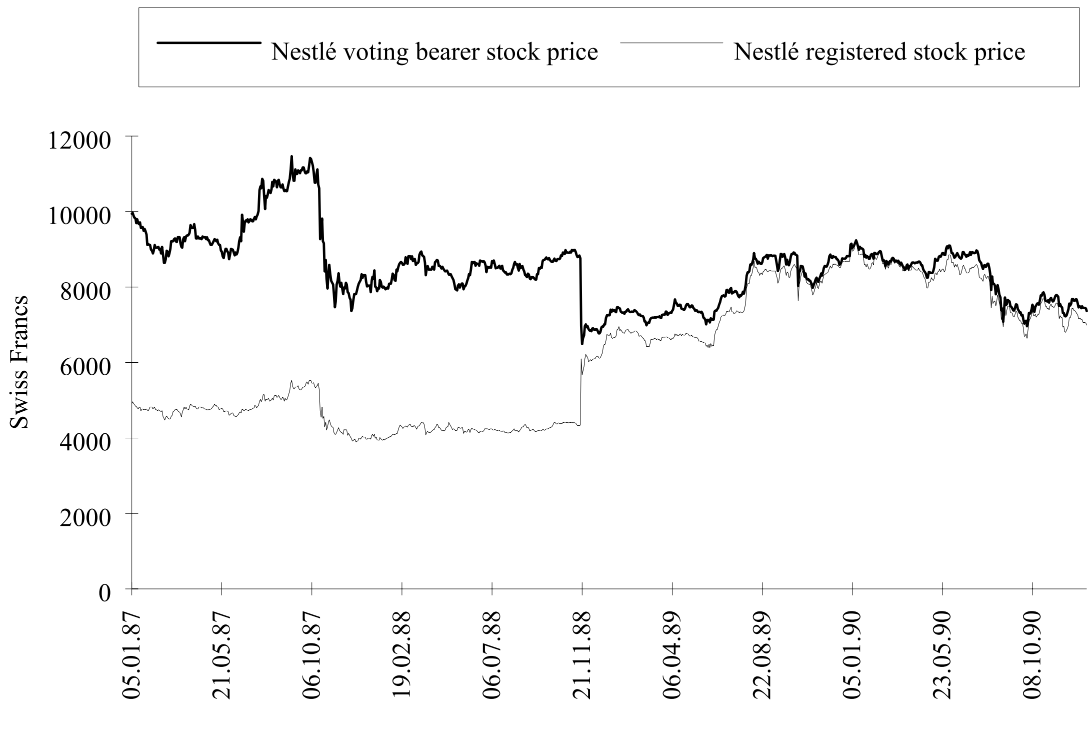


Figure 1-18: Daily prices of Nestlé voting bearer stock and registered stock[16]

Additional analysis shows that Nestlé's stock prices did not change abnormally during the following 50 days. This example demonstrates clearly that Swiss stock markets are efficient. They react quickly and to the full extent to new publicly available information.[17]

14 LODERER Claudio. F and JACOBS Andreas, 1995, “The Nestlé Crash”, Journal of Financial Economics, Vol. 37, pp. 315 - 339

15 Same as 14

16 Same as 14

1.2.2 Efficient market hypothesis

The notion that stocks already reflect some type of information is referred to as the **Efficient Market Hypothesis** (EMH).

EMH implies that market price always reflects the *true* value of the asset. If markets are efficient, then purchase or sale of any asset at the prevailing market price is never a positive net present value (NPV) transaction. In other words, on average, you always receive a fair compensation for the risk effectively taken. The invested capital has to earn a return that is appropriate for the risk of the investment, the time value of money and the transaction costs of the strategy. In a competitive capital market, most of the long-term investment strategies are just covering these costs; there is no 'free lunch'.

More formally Fama described in his article[18] how investors generate price expectation notionally as follows:

$$E\left(\tilde{P}_{j,t+1}\middle|\Phi_t\right) = \left(1 + E\left(\tilde{r}_{j,t+1}\middle|\Phi_t\right)\right)P_{jt}$$

where:

E	expected value operator
$P_{j,t}$	price of security j at time t
$\tilde{P}_{j,t+1}$	price of security j at time t+1
$\tilde{r}_{j,t+1}$	the one period percent rate of return for security *j* during period t+1
Φ_t	the set of information available at time t

The equation above denotes that expected price of the security j, given the information available at time t, is equal to the current price multiplied by 1 plus the expected return on security j, given the set of available information. The question is, what does the available information set Φ_t consist of? The answer depends on the particular form of the market hypothesis being considered.

1.2.2.1 Forms of market efficiency

In his original article Fama divided the general efficient market hypothesis into different sub-hypothesis, according to the kind of information that is already reflected in the asset's price. We traditionally distinguish three forms of market efficiency.

Weak Form

If the market is efficient with regard to *past information*, in other words if all historical information is already discounted in prices, the market is said to be **weak-form efficient**.

This hypothesis implies that there should be no gain from any trading rule that decides whether to buy or to sell a security based on past rate of returns or any other past market data.

---

17 Same as 14

18 Eugene F. Fama, 'Efficient Capital Markets: A Review of Theory and Empirical Work', Journal of Finance 25, no. 2 (May 1970): 383-417

Semi-strong form

If market prices incorporate all the *publicly available information*, it is said to be **semistrong-form efficient**. The semistrong-form hypothesis includes the weak-form hypothesis because all the past market data considered in the weak-form hypothesis is public. Public information does also include non-market information such as political news, news about the economy, earnings and divided announcements, publication of analyst reports, ratios and so on.

Therefore there should be no gain from decision based on new information after it has been made available to the public because the security price already reflects all such new public information.

Strong Form

A market in which all information, including *privately held information*, is reflected in prices is said to be **strong form efficient**. The strong-form efficiency implies the semistrong-form, which in turn implies weak-form efficiency. This means that no group of investors in the market has exclusive access to relevant information, thus no group is capable of generating consistent excess return.

1.2.2.2 Testing market efficiency

Now that we have defined what a perfect efficient market is and that we have seen the three forms of the EMH, we can ask ourselves how to determine the form of efficiency that can be attributed to a market under investigation.

The methodology used to test the validity of market efficiency depends on the efficiency form under investigation.

1. Weak-form hypothesis

To test the validity of the weak-form hypothesis *statistical test of independence* (autocorrelation and runs tests) and *trading rule test*, whereby the risk-return results derived from trading simulation are compared to simple buy-and-hold policy, are used.

2. Semistrong-form hypothesis

The studies used to test this form of market hypothesis can be divided into two sets:

1. *Event studies* are used to test how fast the stock prices reflect the arrival of new information. Defenders of the EMH would expect the price reacts so quickly that it is not possible for investors to earn excess risk-adjusted return by investing after the public release of new information.

2. *Time series analysis* of returns, *cross-section distribution* of returns or other individual stock ratios are used to predict future rate of returns using available public information beyond the market information considered in the weak-form hypothesis.

Any of theses tests examines whether a given investment strategy yields *abnormal* or *excess returns*. To perform these tests, we need to define what normal return is. In other words, we need a benchmark, a model that tells us what the required (or equilibrium, or opportunity) rate of return is for a given risk and a given investment horizon. One possible benchmark is the Capital Asset Pricing Model (CAPM) that we will discover in later. Note that all market efficiency tests are conditional on a given model of equilibrium returns; they may therefore not yield the same conclusions.

Consequently these studies on market efficiency are dual tests of the EMH and the CAPM. Abnormal returns may occur because the markets are not efficient or because the CAPM does not provide the correct estimates of the expected returns.

3. Strong-form hypothesis

Typical tests for the strong-form hypothesis do compare the returns over time of different identifiable investment groups in order to determine if one constantly earns above risk-adjusted return.

*1.2.2.3 Market anomalies**

Like many other hypothesis formulated in finance and economics, the evidence of the EMH is mixed. A lot of studies have been published about the subject, some of which do support the EMH and thus indicate that capital markets are efficient. Other studies however have shown results that are not consistent with the hypothesis and have revealed some anomalies, raising questions about the support for them (Note that the test and the results depend on the form of the EMH tested). The following anomalies are observable, thus public. This implies that they are challenging the EMH in its semi-strong form.

Size Effect (i.e. Small Firm Effect)

Several studies[19] have found that the risk-adjusted return for extended periods (10-15 years) indicated that the small firms (expressed in market value) consistently experienced significantly larger risk-adjusted returns than the larger firms.

Because of the use of the CAPM as model to predict normal returns, some authors argued that the observed difference in risk-adjusted return between small and large firms might actually be much smaller[20] than initially assumed, the CAPM being not the right model to predict the returns of small firms.

Other authors confirmed the first observations, but also found a strong positive relationship between average price per share and the market value, e.g. firms with small market value have low stock prices. Because transaction costs vary inversely with the price per share, they must be considered when analysing the small firm effect.

19 R.W. Banz, 'The Relationship Between Market Return and Market Value of Common Stocks', Journal of Financial Economics 9, no.1 (March 1981): 3-18; and Marc R. Reinganum, 'Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings Yield and Market Values', Journal of Financial Economics 9, no. 1 (March 1981): 19-46

20 Marc M. Reinganum, 'Abnormal Returns in Small Firm Portfolios', Financial Analyst Journal 37, no. 2 (March-April 1981): 52-57; and Richard Roll, ' A Possible Explanation of the Small Firm Effect', Journal of Finance 36, no. 4 (September 1981): 879-888.

Book Value/Market Value
Significant positive relationship between book value (BV) and market value (MV) and future stock returns have been found by Rosenberg, Reid and Lanstein[21]. They contended that this relationship was evidence against the EMH.

However Fama and French provided the strongest support for the importance of this ratio by evaluating the joint effects of market beta, size, E/P ratio, leverage, and the BV/MV ratio on the cross-section average returns on the NYSE, AMEX and NASDAQ stocks. Like Rosenberg, Reid and Lanstein, they found a significant positive relationship between the BV/MV ratio and future stock returns that persisted when other variables were included. They also found that the BV/MV ratio in combination with size dominated other ratios.

Several studies followed the publication of Fama and French, the majority of which reached the same conclusion.

In summary, the tests of publicly available ratios that are used to predict future stock returns have provided substantial evidence against the semistrong-form EMH.

High P/E Ratio Effect
Some practitioners have suggested that low P/E stocks are likely to outperform high P/E stocks. This is because the price of growth companies tends to be overvalued (Overestimation of the growth), whereas low-growth firms trends to be undervalued.

A relationship between historical P/E rations and risk-adjusted return would constitute evidence against the semistrong-form EMH.

A study conducted by Basu concluded that low P/E ratio stocks experienced significantly higher rates of return relative to the market, whereas the opposite was true for high P/E ratio stocks[22].

Other studies reached the same conclusion.

Year-End or January Effect
It can be observed that returns in November and December tend to be below average and on the opposite the returns in the first two weeks of January tend to be above average.

Obviously those who believe in efficient markets would not expect such seasonal pattern to persist. Typically arbitrageurs should eliminate this anomaly.

The January anomaly is intriguing in that it is so persuasive and despite numerous studies, the January anomaly poses as many questions as it answers.

---

21 Barr Rosenberg, Kenneth Reid, Ronald Lanstein, 'Persuasive Evidence of Market Inefficiency', Journal of Portfolio Management 11, no. 3 (Spring 1985): 9-17.

22 S. Basu, 'Investment Performance of Common Stocks in Relation to Their Price-Earning Ratios: A Test of the Efficient Market Hypothesis', Journal of Finance 32, no. 3 (June 1977): 622-682; and S. Basu, 'The Information Content of the Price-Earnings Ratios', Financial Management 4, no. 2, (Summer 1975): 53-64.

Day of the Week Effect
Besides the January effect, there are several other *calendar* effects, one of which is the day of week effect or weekend effect. French has observed that the mean return for Monday was significantly negative, whereas the mean return of the four remaining days was positive[23]. Other authors like Gibbons and Hess found similar results[24].

The following tests do challenge the strong-form of the EMH.

The Value Line Enigma
Several Studies have been conducted in order to establish whether it is possible to identify groups of analysts that have the ability to consistently select undervalued stocks. The idea behind this is that theses investments professional have to obvious advantage to other types of investors except for their training and experience. If anybody is able to consistently select undervalued stock, then it should be these professionals. One of these tests is the Value Line enigma

The value line is a well-known advisory service that publishes several reports. One of theses reports indicates the Value Line's expectations of stock performances over the coming 12-month. Several factors are used to define the ranking that goes from 1 to 5, from the best to the worst.

The Value Line indicated several years after the initial publication that the risk-adjusted return between the different rankings varied significantly, the stocks ranked as 1 significantly (20%) outperformed the one ranked as 2. Subsequently several studies have been conducted in order to analyse this phenomenon. The older studies[25] tend to support the Value Line enigma and thus to challenge the EMH in its strong form. However, more recent studies[26] do indicate the opposite. Once again, there is no clear-cut conclusion possible from these findings.

Note that the above-described anomalies are amongst the most interesting ones, but that there are many other ones that simply would be too long to describe in this document.

1.2.3 Are markets efficient?

We have just seen that in the academic community there is no unanimity concerning the relevance of the EMH. In summary we can say that according to most of the early empirical studies, it appeared effectively that markets were at least semistrong-form efficient. But there has recently been a renewal in the 'not-so-efficient' markets hypothesis. The predictive power of some economic and financial indicators as well as some regularities in historical data have effectively been uncovered on most of stock markets. Determining if these findings are the fruits of statistical illusions or real opportunities is, however, tricky.

Among the community of professional portfolio managers the EMH is not a very popular theory. Consequently, the EMH has never been widely accepted on Wall Street, and debate

---

23 Kenneth R. French, 'Stock Returns and the Weekend Effect', Journal of Financial Economics 8, no.1 (March 1980): 55-70

24 Michal R. Gibbons and Patrick Hess, ' Day of the Week Effect and Asset Returns', Journal of Business 54, no. 4, (October 1981): 579-596

25 Fischer Black, "Yes, Virginia, There is Hope: Tests of the Value Line Ranking System", Financial Analysts Journal 29, no.5 (September-October 73): 10-14.

26 Mark Hulbert, "Proof of Pudding", Forbs, (December 10, 1990): 316.

continues today on the degree to which security analysis can improve investment performance.

Ultimately, however, the issue of market efficiency boils down to whether skilled investors can make consistent abnormal trading profits. Casual evidence does not support claims that professionally managed portfolios can consistently beat the market.

1.2.4 Market efficiency and investment policy

In a weak-form efficient market, the use of **technical analysis** and technical rules, i.e. the search for recurrent and predictable patterns in stock prices, does not lead to superior performance since the information used in building such strategies (the past prices) is already reflected in actual prices. In order technical analysis to be successful, stock prices must respond slowly to fundamental supply-and-demand factors. This prerequisite, of course, is diametrically opposed to the notion of an efficient market.

Fundamental analysis uses financial statements, earnings and dividend prospects of the firm to determine proper stock prices. Fundamental analysis is difficult because it is not enough to do a good analysis of a firm; you can make money only if your analysis is better than that of your competitors. The goal of fundamental analysis is to attain insight into future performance of the firm that is not yet recognized by the rest of the market. If the market is semistrong-form efficient, fundamental analysis does not permit to achieve superior performances since all publicly available information is already reflected in prices.

At this stage, one could cast some doubt on the utility of fundamental analysis. The situation may indeed seem to be quite paradoxical because if markets are efficient, there's no incentive for anyone to do some analysis, but if nobody does, information won't be reflected in prices and as a result, the market won't be efficient anymore, thus creating an incentive for analysis. This paradox is resolved once we are aware that information is costly. In this case the equilibrium is reached when, at the margin, the cost of acquiring information is equal to the benefits this information provides to his acquirer.

The strength of your belief in the level of the market's efficiency determines the kind of management you will adopt. Defenders of the efficient markets hypothesis will adopt a **passive strategy**, i.e. they won't try to beat the market. They will simply adopt a *buy-and-hold* strategy, holding the market (or a mix of the market and the riskless security depending on their risk tolerance) since the market is supposed to be fairly evaluated.

On the other side, those who think markets leave room for opportunities will engage in **active strategies**, trying to buy (sell) undervalued (overvalued) stocks and/or trying to forecast the evolution of the market as a whole. These two strategies are respectively known as *stock picking* and *market timing*.

There is, however, a role for rational portfolio management, even in perfectly efficient markets. Investors' optimal positions will vary according to factors such as age, tax bracket, risk aversion, and employment. The role of the portfolio manager in an efficient market is to tailor the portfolio to these needs, rather than to beat the market.

1.2.5 Lessons from market efficiency

1.2.5.1 For portfolio managers

Stay realistic. If you try to realise a systematically higher return than the market offers for the same risk, you don't have to beat only one or two investors; your competitors are thousands of well-educated, intelligent traders and investors who are, on average, as smart as you are. You have to have some significant comparative advantages.

Always remember hamburgers, it is difficult to enter the burger market and make money. It is also difficult to be better than your competitors. This doesn't mean that selling hamburgers cannot be profitable. It simply means that it is extremely difficult to do better than covering costs in the medium to long run. You have to ask yourself under what conditions it is profitable to be in such a competitive market. If you just replicate what your competitors do, unless they are currently running profitable operations, entry will not make you any richer.

Diversify your risk. If you don't have a comparative advantage in the capital market (i.e. if you're unable to beat the market) try to diversify your risk. And don't forget taxes and transaction costs. For instance, if you engage in stock picking instead of simply following a buy-and-hold strategy, every time you trade in a security, you have to bear transaction costs. Your strategy must cover these costs.

1.2.5.2 For financial managers

Financial assets are traded in competitive financial markets. This doesn't mean that banks engage in irrelevant activities. Banks offer intermediation.

1.2.5.3 For others

For non-banks, financial decisions are not likely to create much value. If a firm wants to raise a specific amount of cash in the capital market, it has to promise future cash flows with a present value that is equal to the amount it desires today. It doesn't have to promise more, and it cannot promise less. But this means zero value creation on average. Taxes, bankruptcy and their transaction costs, information asymmetries and agency costs are the factors that could change this conclusion.[27]

27 The interested reader should consult WELCH Ivo, 1995, “A Primer on Capital Structure”, Finanzmarkt und Portfolio Management, pp. 232-249, or BREALEY Richard A. and MYERS Stewart C., 1996, “Principles of Corporate Finance”, 5th edition, McGraw-Hill, New York

1.3 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[28], showed that the variance of the rate of return could be used as a measure of the risk of a portfolio.

1.3.1 Diversification and portfolio risk

1.3.1.1 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 CHF 1'000 invested, CHF 500 will be invested in asset 1, CHF 600 in asset 2 with the difference between the initial capital (CHF 1'000) and the two long positions (CHF 1'100) being the proceeds from selling short asset 3 for a value of CHF 100.

1.3.1.2 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
N	number of assets available.

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

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?

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

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?

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

1.3.1.3 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 average standard deviation 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 average standard deviation 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

$$\begin{aligned}\sigma_P &= \sqrt{\frac{1}{3^2} \cdot \left(0.15^2 + 0.18^2 + 0.25^2 + 2 \cdot 0.50 \cdot 0.15 \cdot 0.18 + 2 \cdot 0.70 \cdot 0.15 \cdot 0.25 + 2 \cdot 0.55 \cdot 0.18 \cdot 0.25\right)} \\ &= 16.54\%\end{aligned}$$

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

$$\begin{aligned}\sigma_P &= \sqrt{\frac{1}{3^2} \cdot \left(0.15^2 + 0.18^2 + 0.25^2 + 2 \cdot 0.00 \cdot 0.15 \cdot 0.18 + 2 \cdot 0.10 \cdot 0.15 \cdot 0.25 + 2 \cdot 0.80 \cdot 0.18 \cdot 0.25\right)} \\ &= 14.78\%\end{aligned}$$

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.

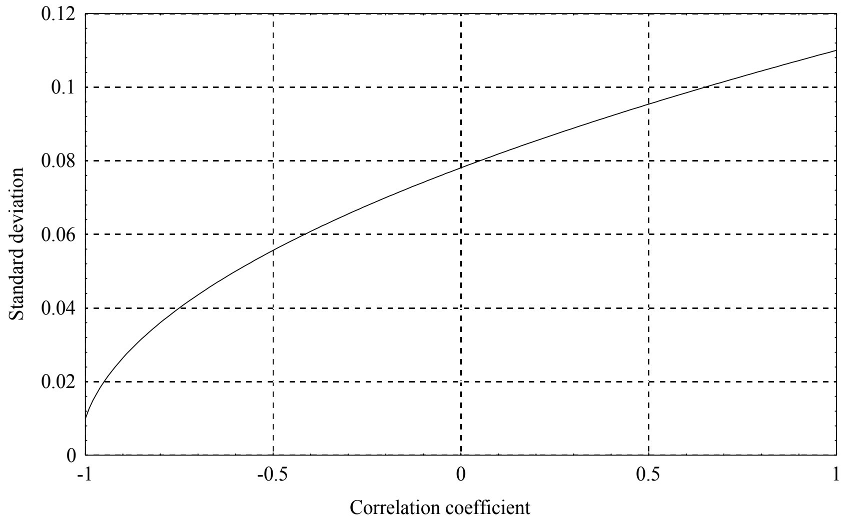


Figure 2-1 Effect of correlation on the standard deviation of a portfolio of two assets

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 benefit of diversification will be the larger, the lower the correlation between the individual assets. 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 on this later.

1.3.1.4 Risk aversion and risk premiums

One of the basic assumptions of the portfolio model is that that investors are basically **risk adverse** 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).

1.3.2 *Markowitz model and efficient frontier*

1.3.2.1 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?

1.3.2.2 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	CHF 100	CHF 95	CHF 115
Investment B	CHF 100	CHF 90	CHF 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	CHF 100	CHF 95	CHF 115
Investment B	CHF 100	CHF 90	CHF 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!

1.3.2.3 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).

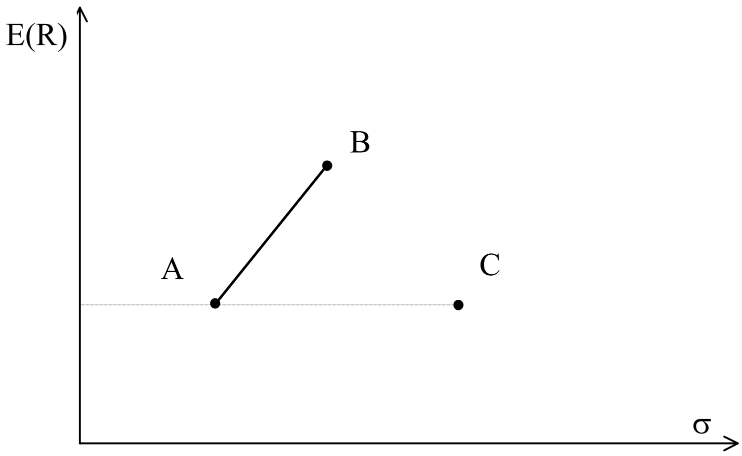


Figure 1-19: A basic example of indifference curves

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.

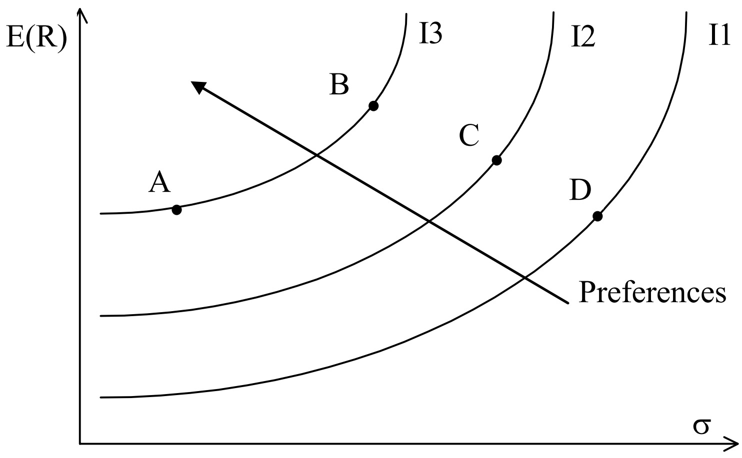


Figure 1-20: Indifference curves and utility levels

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.

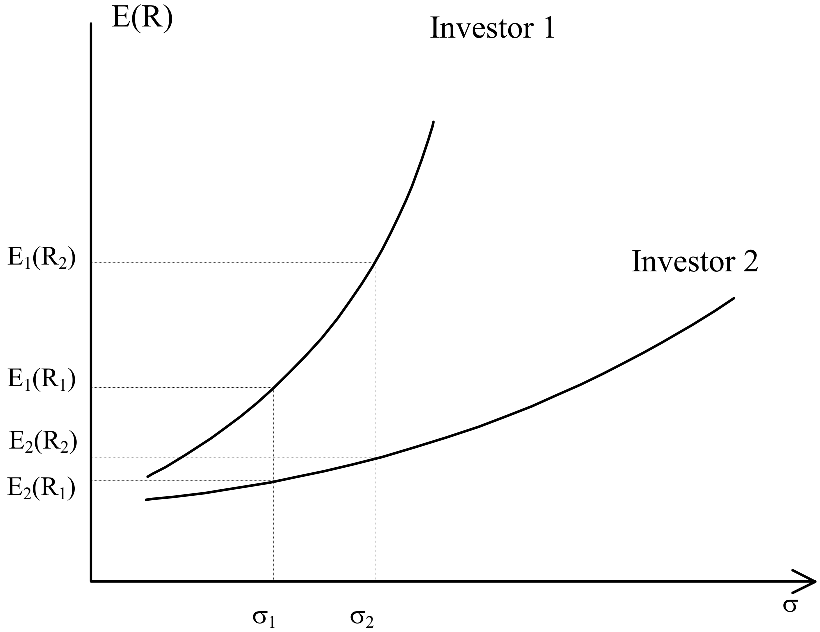


Figure 1-21: Indifference curves and risk aversion

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: [29]

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

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

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

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%.

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 1-8: Risk return combination with equal risk aversion (λ=2)

1.3.2.4 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.

It can be proven (and we will do it later) that in such a case, the **feasible investment set** takes the form of an "umbrella" (see figure below). 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.

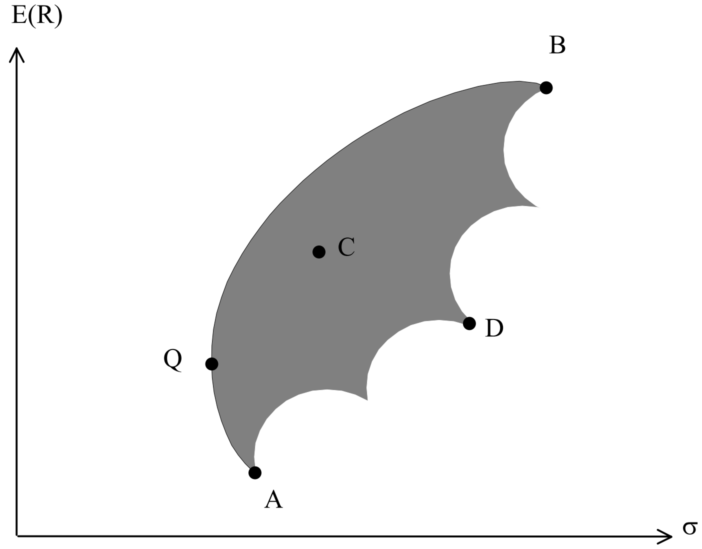


Figure 1-22: The feasible investment set

However, not all these portfolios are relevant. The set of candidate ‘best’ portfolios is a 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 C located inside the grey area, it is possible to find a portfolio C’ on the boundary with a higher expected return for the same risk.

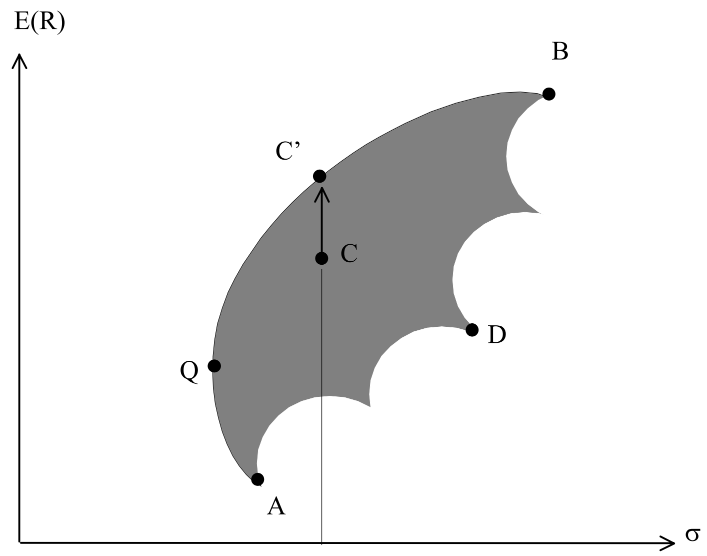


Figure 1-23: A higher return for the same risk

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

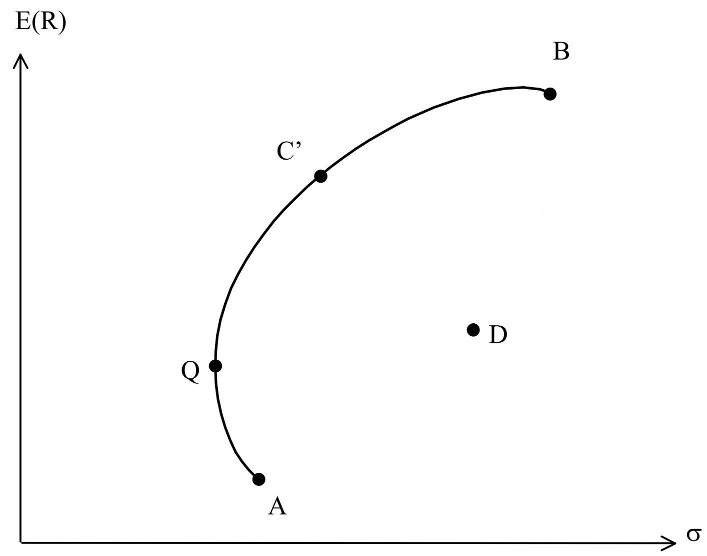


Figure 1-24: The minimum variance frontier

But actually, for any portfolio located in the lower part of the feasible set boundary, the same procedure can be applied. 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 umbrella is called the **efficient set** or **efficient frontier**.

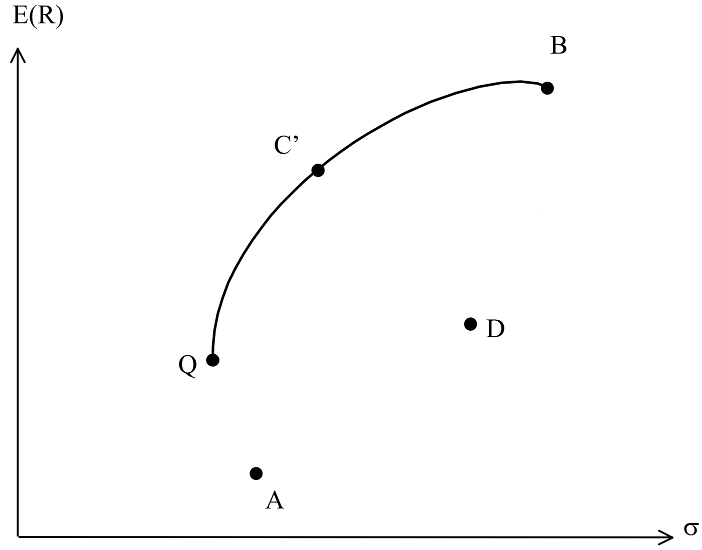


Figure 1-25: The efficient frontier

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 'umbrella' to this curve. But how can we select the optimal portfolio for an investor in this curve?

1.3.2.5 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.

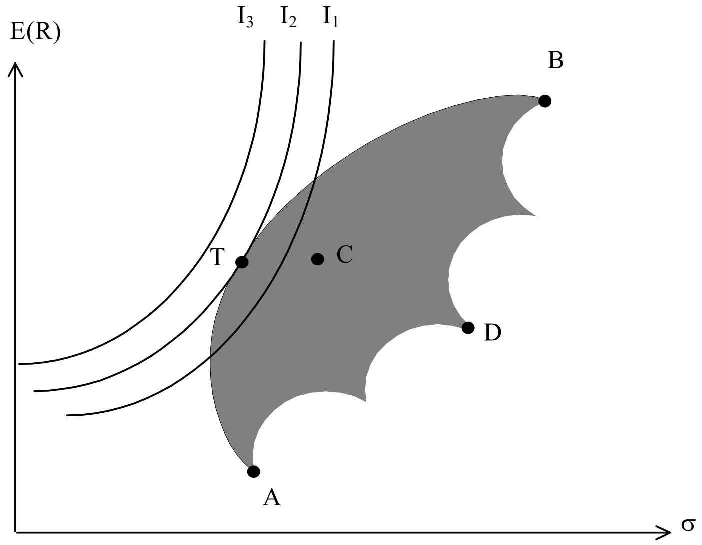


Figure 1-26: The efficient frontier

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 infinity 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.

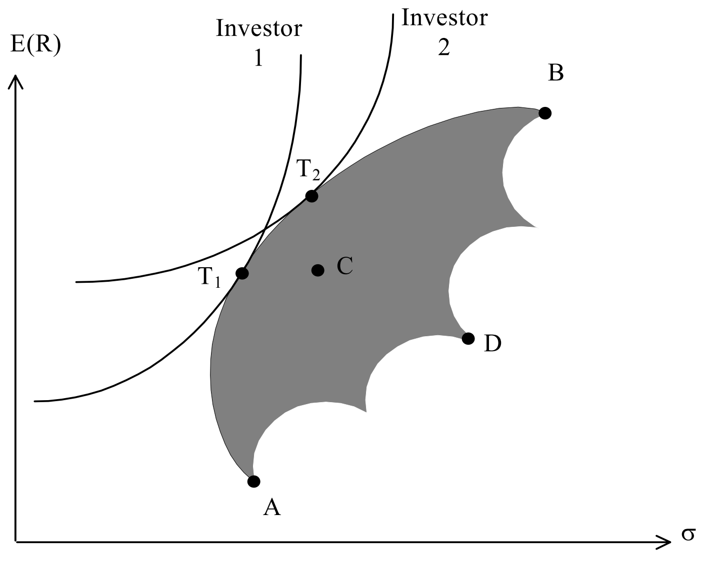


Figure 1-27: Optimal portfolios for two investors

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

1.3.2.6 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.

1.3.2.7 Two risky assets

Let us consider an investor who has the choice to invest his wealth between two risky assets (characterised by ($E(R_1)$, σ_1) and ($E(R_2)$, σ_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:

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

Collecting terms in σ_P and $(\sigma_P)^2$ gives:

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

This is the equation of a *parabola* in a (E(R), σ^2) plane[30]. 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.

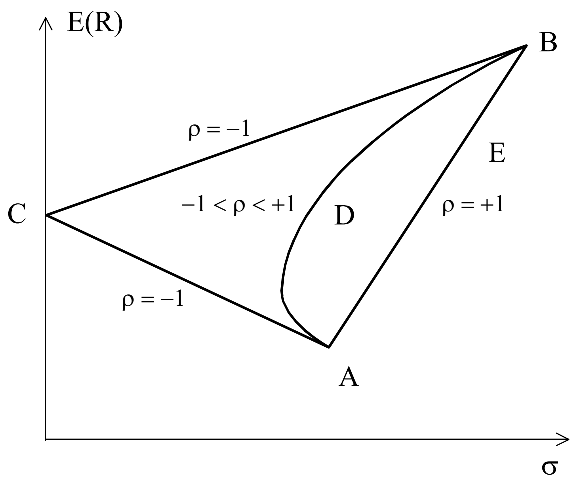


Figure 1-28: The efficient frontier (two risky assets)

---

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

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:

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

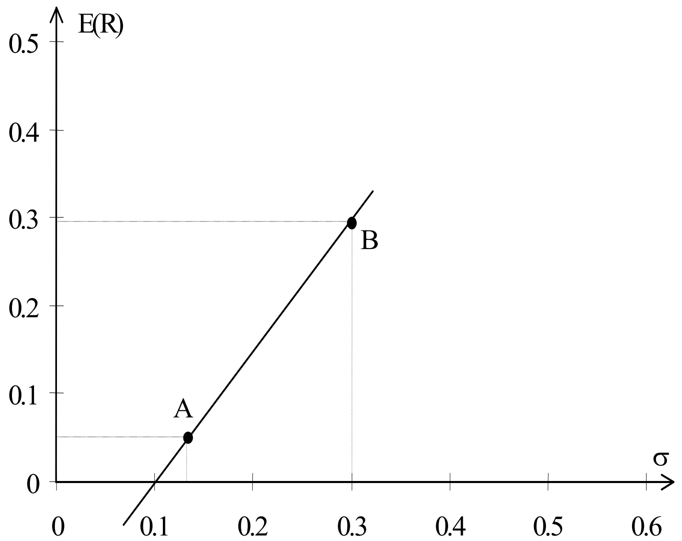
Note that, in this case, the standard deviation of a portfolio made of A and B is the weighted average of σ_A and σ_B. (This is the only exception to the statement made earlier about the standard deviation of a portfolio).

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:

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

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%.

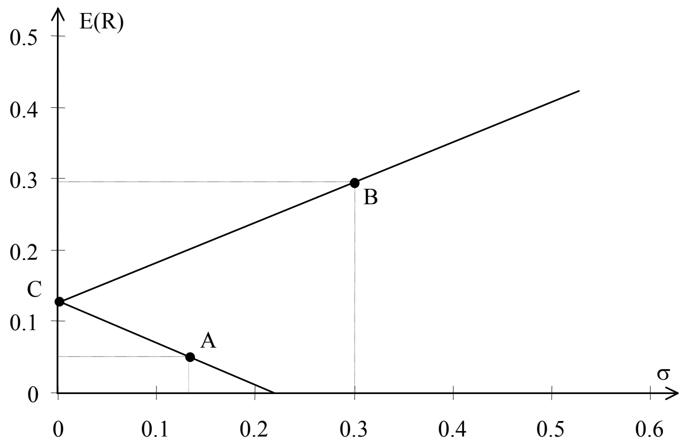
- If the correlation is perfectly negative (ρ=–1), the benefit of diversification is the largest; it is in fact possible to create a risk-free portfolio having a positive return (point C). ACB is the minimum variance set. Only BC belongs to the efficient set.

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:

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

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):

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

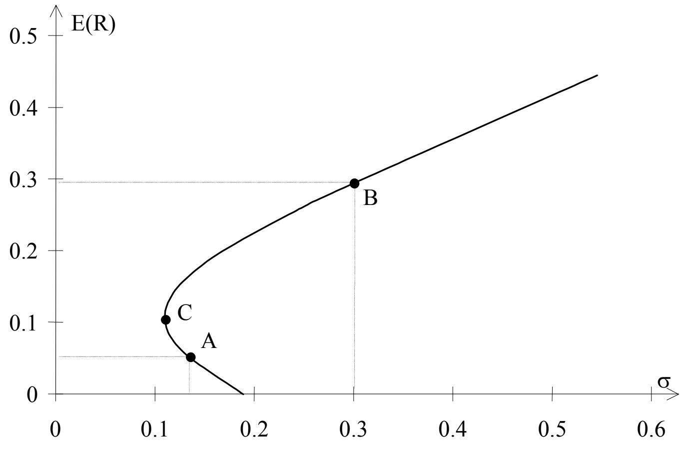
where x_A is the proportion invested in A, $1 - x_A$ in B. This is the general case. It justifies our using the 'umbrella' as the standard shape for the efficient frontier in the previous sections.

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%.

1.3.2.8 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.

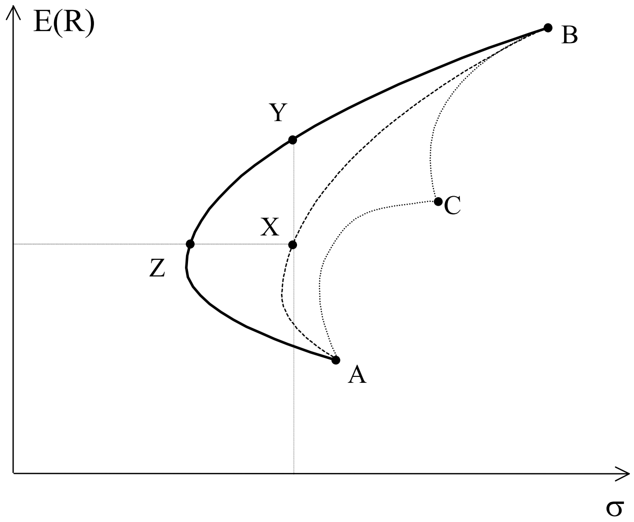


Figure 1-29: The increase in diversification when adding an asset

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.**

1.3.2.9 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 $N \to \infty$.

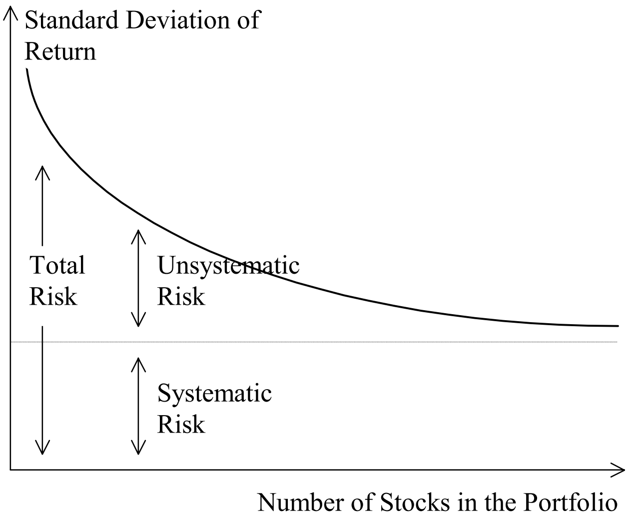


Figure 1-30 Diversification effect

1.3.2.10 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[31] 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 · (N–1) correlations[32] to compute the efficient frontier.

[31] For investors with the same time horizon.

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

1.4 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.

1.4.1 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. 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.

1.4.1.1 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[33]. 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.

1.4.1.2 One risky and one risk-free asset

Let us consider an investor who has the choice to invest his wealth between a risky asset[34] (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:

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

33 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.

34 which might also be interpreted as a portfolio.

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):

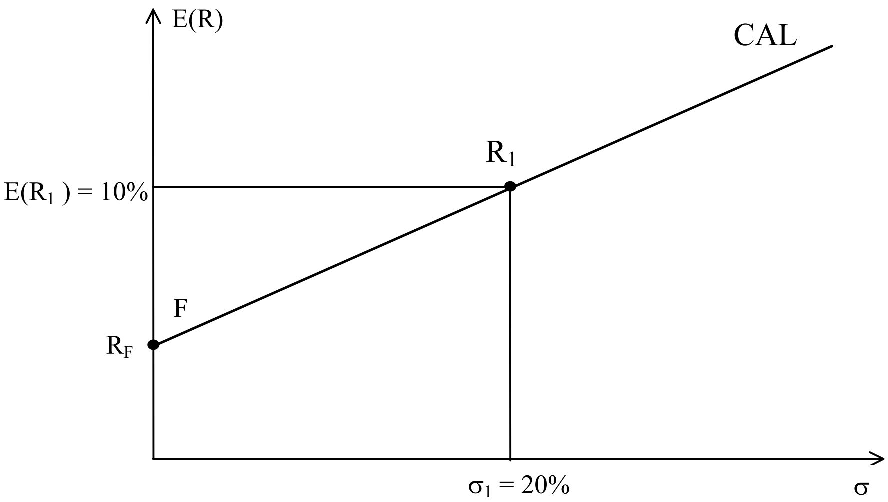


Figure 1-31: The Capital Allocation Line (CAL)

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.

1.4.1.3 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 frontier of the umbrella. This implies that the efficient frontier in this more

general case is the straight line from R_F which is tangent to the umbrella representing the efficient frontier when there are only risky assets.

The efficient frontier is represented below:

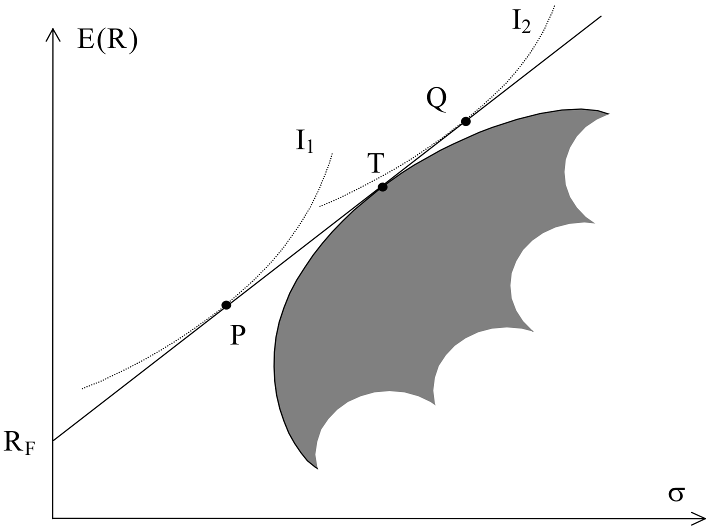


Figure 1-32: N risky assets+one risk-free asset

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.

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.

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 the one known as market imperfections.

1.4.1.4 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:

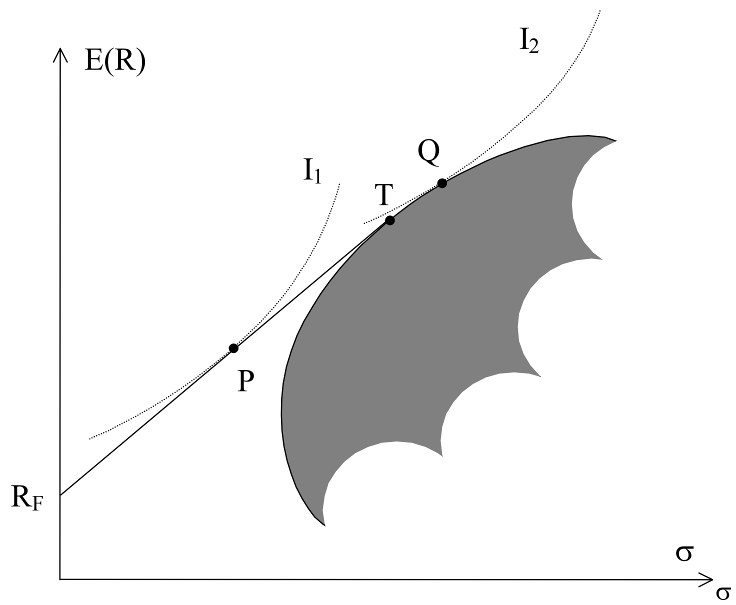


Figure 1-33: Short selling restrictions

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:

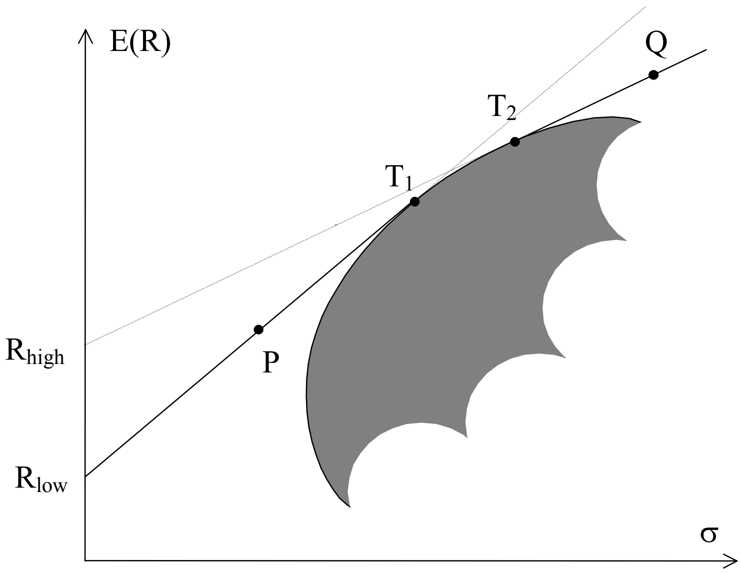


Figure 1-34: Different borrowing and lending rates

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.

1.4.1.5 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.

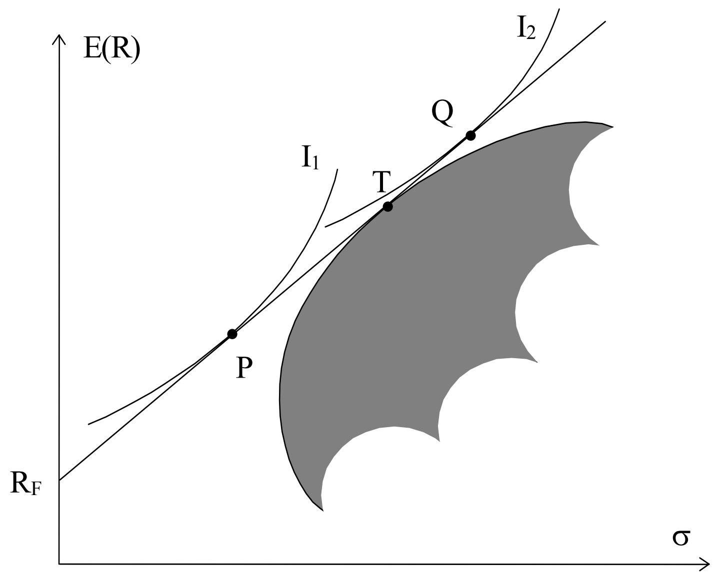


Figure 1-35: The separation theorem

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.

1.4.1.6 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).

1.4.2 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*.

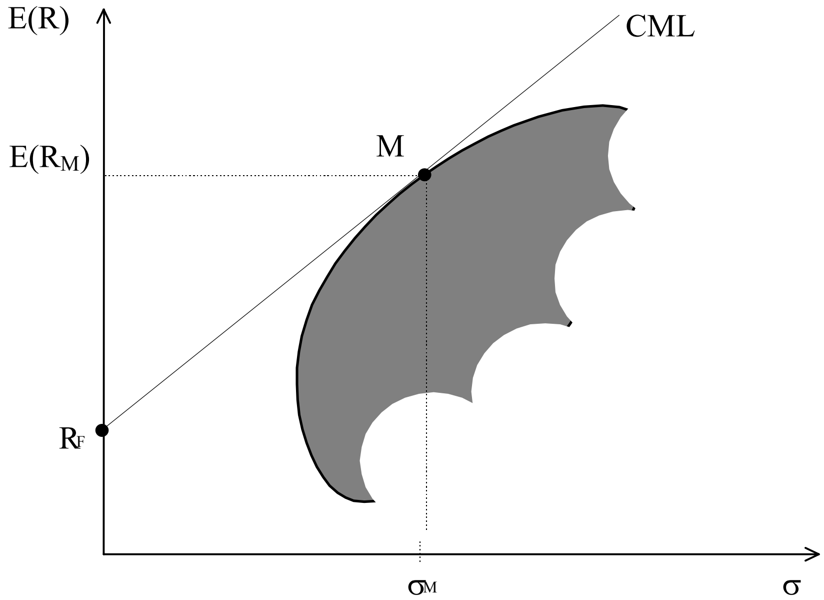


Figure 1-36: The capital market line (CML)

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).

1.4.3 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 investigate individual asset risks and their impact on the total risk of the market portfolio, we have to decompose its standard deviation:

$$\sigma_M = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} X_{iM} X_{jM} \sigma_{ij}\right]^{1/2}$$

$$= \left[X_{1M}\sum_{j=1}^{N} X_{jM}\sigma_{1j} + X_{2M}\sum_{j=1}^{N} X_{jM}\sigma_{2j} + \cdots + X_{NM}\sum_{J=1}^{N} X_{jM}\sigma_{Nj}\right]^{1/2}$$

where x_{iM} and x_{jM} are the proportions of the market portfolio invested in assets i and j .

Now, the covariance of the security i with the market portfolio is given by:

$$\sum_{j=1}^{N} X_{jM}\sigma_{ij} = \sigma_{iM}$$

which can be rewritten as:

$$\sigma_M = \left[X_{1M}\sigma_{1M} + X_{2M}\sigma_{2M} + \cdots + X_{NM}\sigma_{NM}\right]^{1/2}$$

This, together with the above reasoning, implies that the relevant measure of risk for an individual asset is its covariance with the market portfolio, and it will appear in the form

$$\frac{\sigma_{iM}}{\sigma_M^2}$$

Thus, the central equation of the CAPM is the following expected return relationship for **individual securities** or **non-efficient portfolios**:

$$E(R_i) = R_F + \left[\frac{E(R_M) - R_F}{\sigma_M}\right] \cdot \frac{\sigma_{iM}}{\sigma_M}$$

where the unit price of risk is given by

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

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

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

where β_i (beta) is defined as follows:

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

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. β=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:

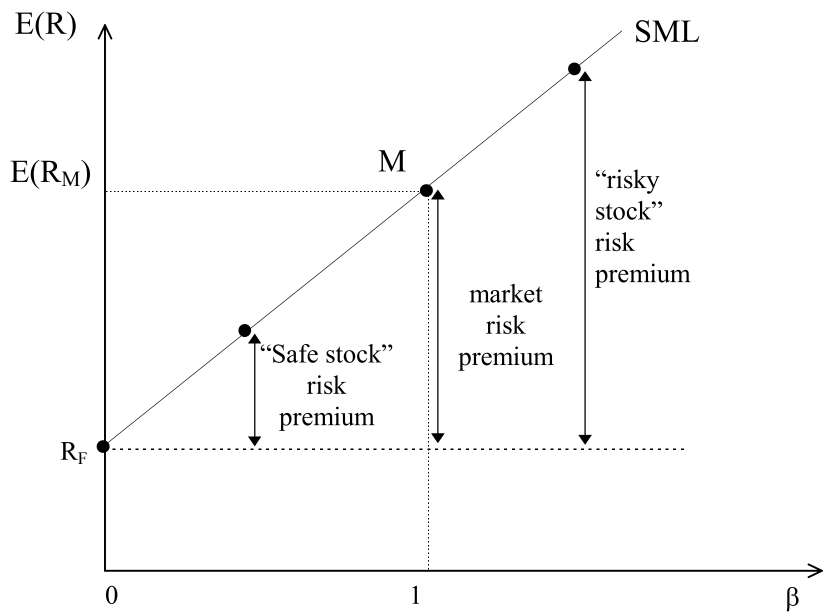


Figure 1-37: The security market line (SML)

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.

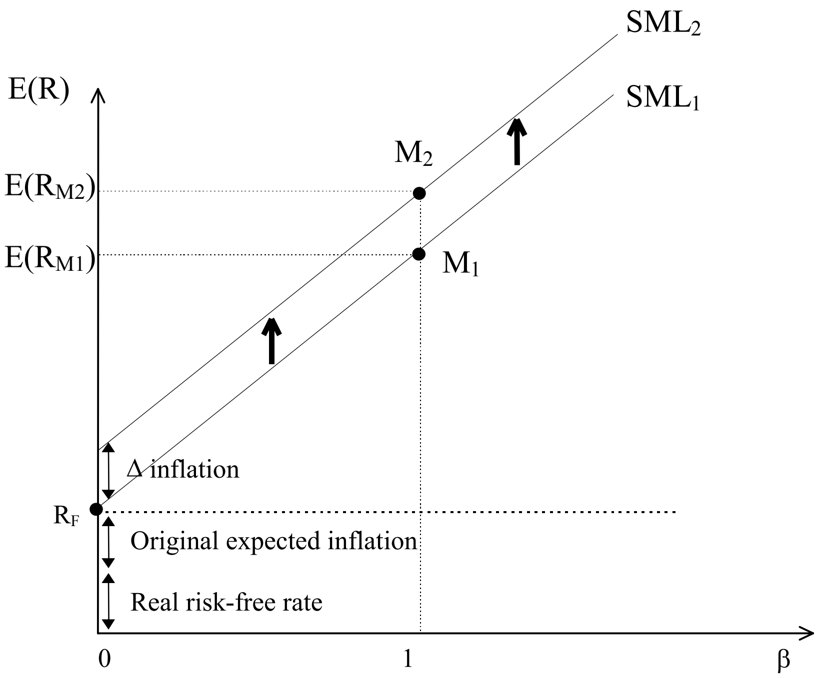


Figure 1-38: The SML and an increase in anticipated inflation

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.

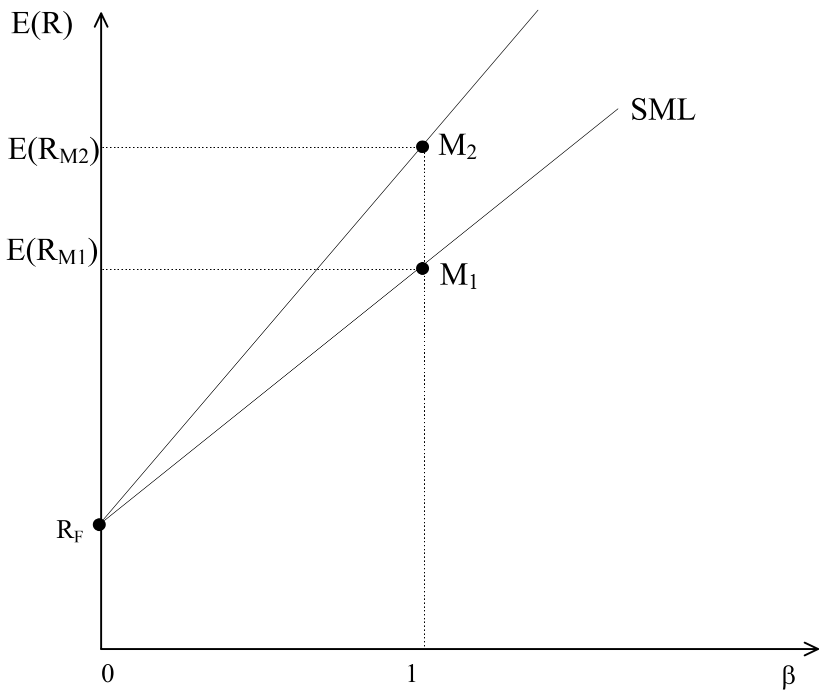


Figure 1-39: The SML and an increase in risk aversion

1.4.3.1 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 to be efficient and thus need not to 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.

1.4.3.2 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.

1.4.3.3 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.

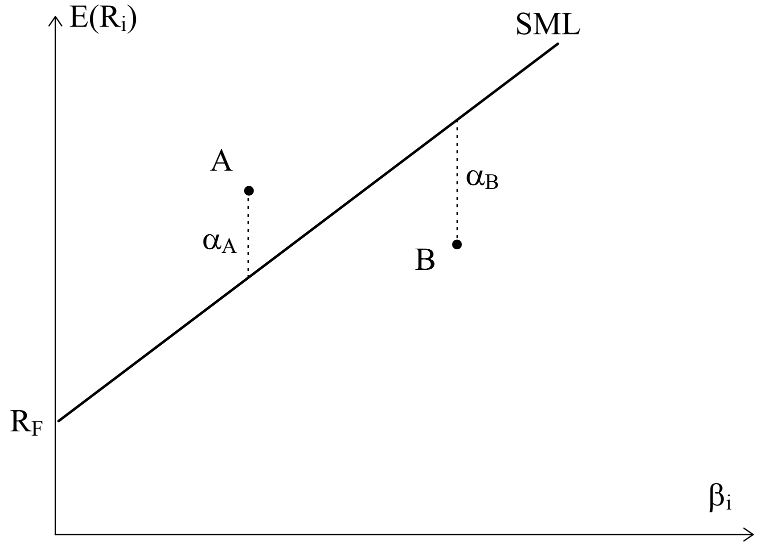


Figure 1-40: Overvalued and undervalued securities

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:

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

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 1-39, 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 CHF 460. In one year, you expect its price to be CHF 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.

1.4.4 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[35] 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).

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

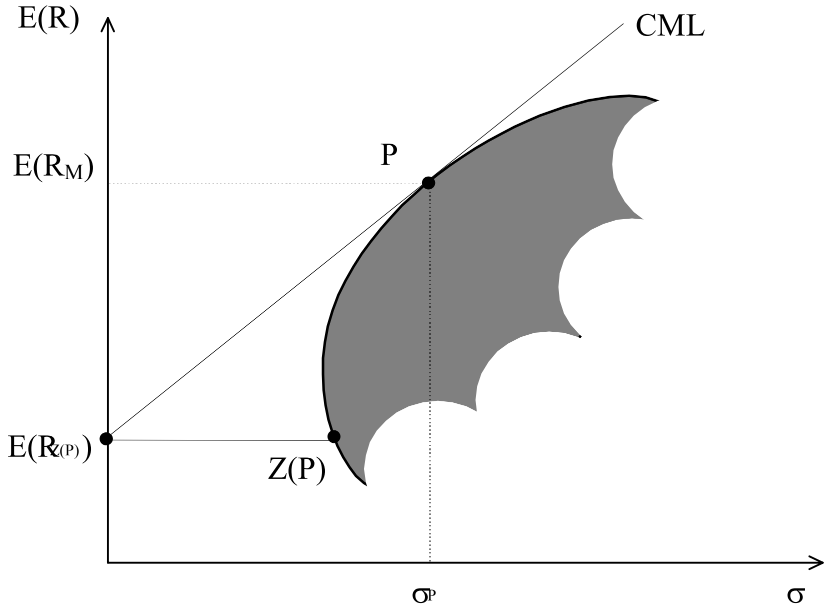


Figure 1-41: Example of a zero-beta portfolio

- 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 to have a SML for cases where there is no risk-free assets, 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.

1.4.5 *The international CAPM**

To present an international version of the CAPM in which PPP is assumed not to hold, we shall first present another way to look at the traditional CAPM. Then we shall derive a 2-country CAPM, relying heavily on the presentation of P. Sercu[36], and finally present a general international CAPM (with an arbitrary number of countries)[37].

Let us go back to the traditional CAPM. By replacing β_i by its value as given by the market model and after rearranging terms we get:

$$\begin{aligned} E[r_i] &= r_F + \left(E[r_M] - r_F\right) \cdot \beta_i \\ &= r_F + \frac{E[r_M] - r_F}{Var[r_M]} \cdot Cov[r_i, r_M] \\ &= r_F + \lambda \cdot Cov[r_i, r_M] \end{aligned}$$

λ is called the **unit market price of risk**. This concept needs some comments:

- "unit" refers to the fact that λ is a normalised value because of the division by $Var[r_M]$;
- this unit price is common to all stocks; the risk premium for a particular stock is then equal to the unit price of risk times the quantity of risk of this stock, which is measured by $Cov[r_i, r_M]$. The expected return is then equal to the risk-free rate plus the risk premium, as stated above. λ simply measures the trade-off between contributions to the market portfolio's variance and expected excess return (remember that this term is equal among all assets as a result of no-arbitrage arguments);
- the unit price of risk depends on the **average degree of risk aversion** of the market. Indeed, the composition of the market portfolio is function of the aggregation of investors' demand, which, in turn, depends on their risk aversion.

But consider now the case of a Swiss and a French investor investing in both the Swiss and French market; this imposes two important modifications to the standard CAPM (let us assume no inflation to keep things simple):

- first the market portfolio for each investor may no more be proxied by its national stock market; each investor now has access to both markets and as a result, the market portfolio he faces is the aggregate of the Swiss and French market;
- second, the crucial assumption of homogenous expectations among agents we had to made to derive the standard CAPM is now violated. Since agents do face real exchange risk, the distribution of real returns will depend upon investors' nationality. For example under real CHF/EUR exchange rate risk, the distribution of the real return of Nestlé in Zurich will not be the same for the Swiss investor (who reasons in terms of real CHF) and for the French

---

36 SERCU P. and UPPAL R., 1995, "International Financial Markets and the Firm", South-Western, Chapman & Hall, pp. 598-607 and Appendix 22A, pp. 617-619

37 More sophisticated ICAPM can be found in SOLNIK B., 1974, "An Equilibrium Model of the International Capital Market", Journal of Economic Theory, Vol. 8, pp.500-524, SOLNIK B., 1983, "International Arbitrage Pricing Theory", Journal of Finance, Vol. 38, pp.449-457, or ADLER M. and DUMAS B., 1983, "International Portfolio Choice and Corporation Finance: A Synthesis", Journal of Finance, Vol. 38, pp.925-984.

investor (who reasons in terms of real EUR). Thus we say that investors' perceptions are segmented across countries.

Each asset contributes to the total variance and to the excess expected return of the market portfolio. The Swiss investor is concerned with the latter contributions in terms of CHF, whereas the French investor measures these in EUR. To make aggregation possible we have to choose a common numeraire, say the CHF. The French investor is now concerned with the joint distribution of the nominal returns on the stock and of that of the CHF/EUR exchange rate. Therefore investors are concerned with **two sources of risk**: the covariance risk of each asset with the newly defined market portfolio and the exchange rate covariance risk (also called **exchange rate exposure**). This can be seen as a simple extension of the standard CAPM since exactly the same approach has been used: **a risk is faced and each asset "contributes to this risk", so there is a risk premium attached to it.**
With no surprise the 2-country CAPM equation is:

$$E[r_i] = r_F + \lambda \cdot Cov[r_i, r_M] + \eta \cdot Cov[r_i, s_{CHF/FRF}]$$

where η is the risk premium attached to the exchange rate covariance risk of the asset.
The next step is to find the expressions for λ and η. Again, this is done as in the traditional setting (we used the market model) but the estimation equation involves one more term, the "returns" on the exchange rate, since we have two sources of risk:

$$r_i = \alpha_i + \beta_i \cdot r_M + \gamma_i \cdot s_{CHF/FRF} + \varepsilon_i$$

First we derive the regression coefficients through computation of the covariance of asset i's returns with the market portfolio returns and with the exchange rate. The covariance with the market portfolio is:

$$\begin{aligned} Cov[r_i, r_M] &= Cov[\alpha_i + \beta_i \cdot r_M + \gamma_i \cdot s_{CHF/FRF} + \varepsilon_i, r_M] \\ &= \beta_i \cdot Cov[r_M, r_M] + \gamma_i \cdot Cov[s_{CHF/FRF}, r_M] \\ &= \beta_i \cdot Var[r_M] + \gamma_i \cdot Cov[r_M, s_{CHF/FRF}] \end{aligned}$$

And similarly for the exchange rate:

$$Cov[r_i, s_{CHF/FRF}] = \beta_i \cdot Cov[r_M, s_{CHF/FRF}] + \gamma_i \cdot Var[s_{CHF/FRF}]$$

This can be written in matrix form and solved to find the coefficients:

$$\begin{bmatrix} Cov[r_i, r_M] \\ Cov[r_i, s_{CHF/FRF}] \end{bmatrix} = \begin{bmatrix} Var[r_M] & Cov[r_M, s_{CHF/FRF}] \\ Cov[r_M, s_{CHF/FRF}] & Var[s_{CHF/FRF}] \end{bmatrix} \cdot \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix}$$

$$\begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} = \begin{bmatrix} Var[r_M] & Cov[r_M, s_{CHF/FRF}] \\ Cov[r_M, s_{CHF/FRF}] & Var[s_{CHF/FRF}] \end{bmatrix}^{-1} \cdot \begin{bmatrix} Cov[r_i, r_M] \\ Cov[r_i, s_{CHF/FRF}] \end{bmatrix}$$

Then we use the 2-country CAPM equation to derive the risk premia λ and η. To identify these unknown premia we need two benchmark portfolios; since the two premia come from

the market and exchange rate risk, the natural choice is the market portfolio and a risk-free investment in the foreign country. For these investments the 2-country CAPM equation is:

$$E\left[r_M\right]-r_F^{CHF}=\lambda\cdot Var\left[r_M\right]+\eta\cdot Cov\left[r_M,s_{CHF/FRF}\right]$$
$$E\left[s_{CHF/FRF}\right]+r_F^{FRF}-r_F^{CHF}=\lambda\cdot Cov\left[r_M,s_{CHF/FRF}\right]+\eta\cdot Var\left[s_{CHF/FRF}\right]$$

with obvious notations; just notice that the return on the foreign risk-free investment is equal to the foreign risk-free rate plus the currency appreciation over the period.
Using the same matrix technique as above we obtain:

$$\begin{bmatrix} E\left[r_M\right]-r_F^{CHF} \\ E\left[s_{CHF/FRF}\right]+R_F^{FRF}-R_F^{CHF} \end{bmatrix}=\begin{bmatrix} Var\left[r_M\right] & Cov\left[r_M,s_{CHF/FRF}\right] \\ Cov\left[r_M,s_{CHF/FRF}\right] & Var\left[s_{CHF/FRF}\right] \end{bmatrix}\cdot\begin{bmatrix}\lambda \\ \eta\end{bmatrix}$$
$$\begin{bmatrix}\lambda \\ \eta\end{bmatrix}=\begin{bmatrix} Var\left[r_M\right] & Cov\left[r_M,s_{CHF/FRF}\right] \\ Cov\left[r_M,s_{CHF/FRF}\right] & Var\left[s_{CHF/FRF}\right] \end{bmatrix}^{-1}\cdot\begin{bmatrix} E\left[r_M\right]-r_F^{CHF} \\ E\left[s_{CHF/FRF}\right]+r_F^{FRF}-r_F^{CHF} \end{bmatrix}$$

The last step is to link the regression coefficients (β_i and γ_i) with the risk premia (λ and η) and the covariances. This is done by substituting the premium equations into the 2-country CAPM equation (in matrix notation):

$$E\left[r_i\right]-r_F^{CHF}=\left[Cov\left[r_i,r_M\right]\quad Cov\left[r_i,s_{CHF/FRF}\right]\right]\cdot\begin{bmatrix}\lambda \\ \eta\end{bmatrix}$$
$$=\left[Cov\left[r_i,r_M\right]\quad Cov\left[r_i,S_{CHF/FRF}\right]\right]\cdot\begin{bmatrix} Var\left[r_M\right] & Cov\left[r_M,s_{CHF/FRF}\right] \\ Cov\left[r_M,s_{CHF/FRF}\right] & Var\left[s_{CHF/FRF}\right] \end{bmatrix}^{-1}\cdot\begin{bmatrix} E\left[r_M\right]-r_F^{CHF} \\ E\left[s_{CHF/FRF}\right]+r_F^{FRF}-r_F^{CHF} \end{bmatrix}$$

The fist two matrices above are just the transpose of the system we derived for the regression coefficients, and they can be replaced; this yields the final 2-country CAPM equation:

$$E\left[r_i\right]-r_F^{CHF}=\left[\beta_i\quad \gamma_i\right]\cdot\begin{bmatrix} E\left[r_M\right]-r_F^{CHF} \\ E\left[s_{CHF/FRF}\right]+r_F^{FRF}-r_F^{CHF} \end{bmatrix}$$
$$E\left[r_i\right]-r_F^{CHF}=\beta_i\cdot\left(E\left[r_M\right]-r_F^{CHF}\right)+\gamma_i\cdot\left(E\left[s_{CHF/FRF}\right]+r_F^{FRF}-r_F^{CHF}\right)$$

Clearly, this last result looks very much like the standard CAPM with one more risk premium. This derivation in matrix form is straightforwardly extended to the multi-country framework, so only the final equation is given here:

$$E\left[r_i\right]-r_F=\beta_i\cdot\left(E\left[r_M\right]-r_F\right)+\sum_{k=1}^{K-1}\gamma_{i,k}\cdot\left(E\left[s_k\right]+r_F^k-r_F\right)$$

where:

r_F	Continuous compounded risk-free rate in the domestic country
s_k	Exchange rate of country k
r_F^k	Continuous compounded risk-free rate in country k
K	Number of countries considered
β_i and γ_I	As usual

In words, the excess return on assets is modelled as a linear function of the excess return on the world market portfolio (including all assets of the K countries) and of K exchange rate risk premia.
The first empirical work in international asset pricing models was conducted by Solnik[38]; he tested a simplified version of the ICAPM and found that the data were consistent with his ICAPM. However, he did not compare it to the domestic CAPM. Stehle[39] tested another simplified ICAPM on the US market (all the exchange rate covariance risk premium are equal to zero as a result of selecting a particular form for the utility functions; the only difference between the domestic CAPM and this ICAPM is then the definition of the market portfolio); he found no evidence that one model performed better than the other. More recently, Dumas and Solnik[40] test an unrestricted version of the ICAPM for major OECD countries and they reject the domestic CAPM. This gives support to the general view that the ICAPM should do better than its domestic counterpart for countries where capital flows restrictions are low.

38 SOLNIK B., 1974, "The International Pricing of Risk: An Empirical Investigation of the World Capital Market Structure", Journal of Finance, Vol. 29, pp.365-378

39 STEHLE R., 1977, "An Empirical Test of the Alternative Hypothsese of National and International Pricing of Risky Assets", Journal of Finance, Vol. 32, pp.493-502

40 DUMAS B. and SOLNIK B., 1995, "The World Price of Foreign Exchange Risk", Journal of Finance, Vol. 50, pp.445-480

1.5 Index and market models*

*1.5.1 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.

*1.5.2 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,

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

- 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:

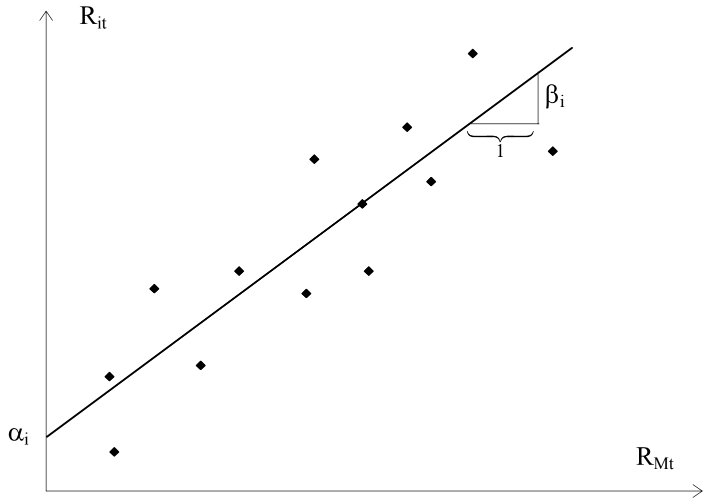


Figure 1-42: Single index model regression estimates

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.

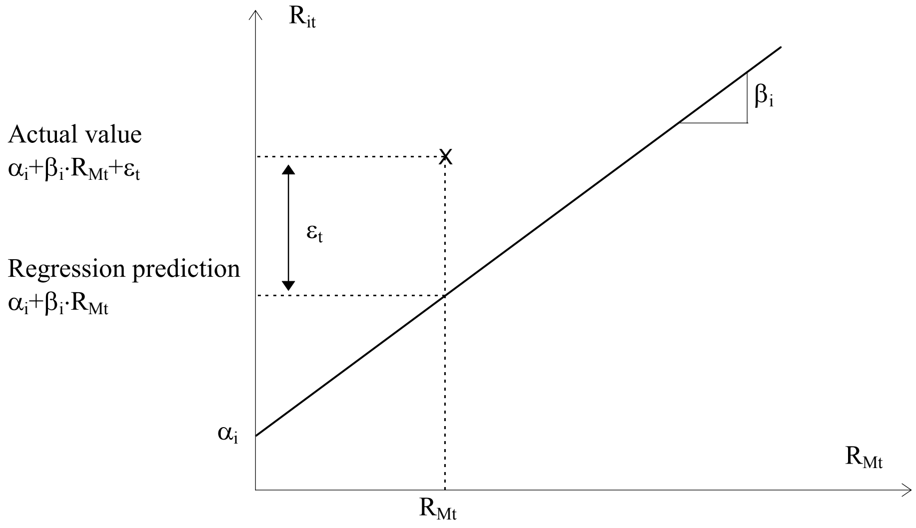


Figure 1-43: Simple regression estimates and residuals

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.

1.5.3 Decomposing variance into systematic and diversifiable risk*

1.5.3.1 In the case of a single security*

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

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

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:

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

Squaring terms in the parenthesis and taking expectations gives

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

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

$$\sigma_i^2 = \beta_i^2 \cdot E\Big(R_M - E\big(R_M\big)\Big)^2 + E(\varepsilon_i^2)$$

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 Swiss market:

Asset i	β_i	σ_i
Swiss Market Index	1.00	0.1346
Novartis	1.28	0.1729
Credit Suisse	1.33	0.2546

From there, we want to find the idiosyncratic risk of Novartis and Credit Suisse returns.

For Novartis, we have

$$\sigma_{\varepsilon_1} = \sqrt{\sigma_1^2 - \beta_1^2 \cdot \sigma_M^2} = \sqrt{0.1729^2 - 1.28^2 \cdot 0.1346^2} \approx 0.0145$$

And for Credit Suisse

$$\sigma_{\varepsilon_2} = \sqrt{\sigma_2^2 - \beta_2^2 \cdot \sigma_M^2} = \sqrt{0.2546^2 - 1.33^2 \cdot 0.1346^2} \approx 0.1810$$

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:

$$\sigma_{ij} = \text{Cov}(R_i, R_j) = E\Big[\big(R_i - E(R_i)\big)\big(R_j - E(R_j)\big)\Big]$$

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

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

Multiplying the terms, we get:

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

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:

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

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 a Swiss stocks research paper:

Asset i	β_I	σ_i
Swiss Market Index	1.00	0.1346
Novartis	1.28	0.1729
Credit Suisse	1.33	0.2546

From there, what is the covariance and the correlation coefficient of Novartis and Credit Suisse returns?

The covariance is given by:

$$\sigma_{12} = \beta_1 \cdot \beta_2 \cdot \sigma_M^2 = 1.28 \cdot 1.33 \cdot 0.1346^2 \approx 0.031$$

From there, we can easily infer the correlation coefficient:

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2} = \frac{0.031}{0.1729 \cdot 0.2546} \approx 0.70$$

*1.5.3.2 In the case of a portfolio: implications for diversification**

The equation

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

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

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

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, $\sigma_{\varepsilon_p}^2$, 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.

*1.5.3.3 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

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

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.

*1.5.4 The link with the CAPM**

*1.5.4.1 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

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

Which can be rewritten in the following form

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

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 relationships 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.

*1.5.4.2 Estimating the alphas (α)[41]**

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

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

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.

41 Not the same α as in the previous chapter

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[42].

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.

*1.5.4.3 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.

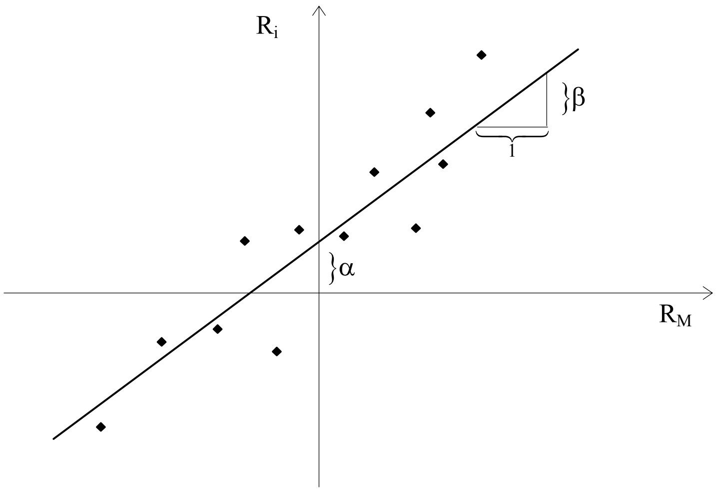


Figure 1-44: Estimating the βs with a regression

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,

42 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.

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.

*1.5.4.4 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}	$R_{Mt}-E(R_M)$	$R_{it}-E(R_i)$	$[R_{Mt}-E(R_M)]\cdot[R_{it}-E(R_i)]$	$[R_{Mt}-E(R_M)]^2$
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 1-9: 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

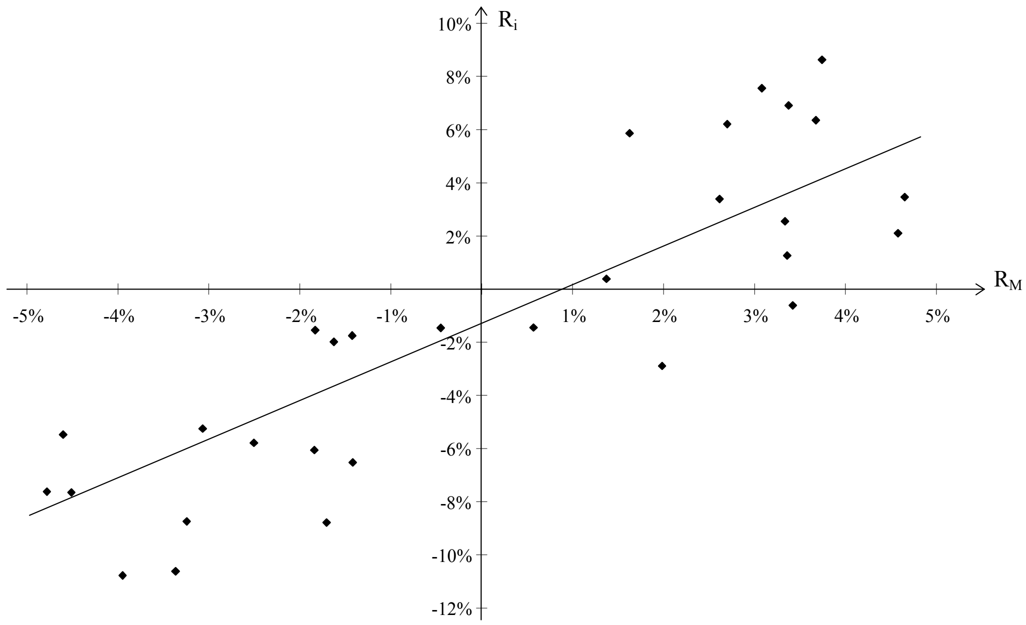


Figure 1-45: Estimating α and β with a regression

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

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

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.

*1.5.4.5 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.

*1.5.5 Two applications of the market model**

*1.5.5.1 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,

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

This implies

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

that is,

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

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 1-10: Required data to compute the market model

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

*1.5.5.2 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

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

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.

1.5.6 *Multi-index models**

*1.5.6.1 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[43]. 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.

*1.5.6.2 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

[43] 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.

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

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

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

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

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

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.

*1.5.6.3 An example of a multi-index model**

Salomon Brothers use a multi-index model with six variables to explain the returns on securities[44]. 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.

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

1.5.7 *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.

1.6 Arbitrage Pricing Theory*

1.6.1 Assumptions Underlying the APT*

The Arbitrage Pricing Theory[45] (APT) is an alternative approach to the equilibrium determination of asset prices. The model describes expected returns and is built on two key assumptions:

- Asset returns are generated by a multi-index model.
- There is an absence of arbitrage opportunities.

*1.6.1.1 Return Generating Process**

Asset returns are all generated by the same linear model, which has been introduced in the previous chapter, the multifactor (or multi-index) model. The return of an asset i is:

$$R_i = a_i + b_{i1} \cdot F_1 + b_{i2} \cdot F_2 + \ldots\ldots + b_{ik} \cdot F_k + \varepsilon_i$$

Where R_i is the return on asset i, a_i is the return on asset i if the realization of all factors equals zero, b_{ij} are the sensitivities (or systematic risks) of asset i with respect to factor j, F_j is the realized return on factor j and ε_i is the residual component of asset i (i.e. the fraction of return not determined by the factors and which is specific to the company).

This model is based on several assumptions:

- The residual component of asset i is on average equal to zero, $E(\varepsilon_i) = 0$
- The residual component of assets i and j are not correlated, $Cov(\varepsilon_i, \varepsilon_j) = 0$, meaning that all common variation in the two returns are captured by the factors.
- The residual components of asset i is not correlated with the factors, $Cov(\varepsilon_i, F_k) = 0$.

Note that:

- It is only assumed that a k-factor model generates asset returns, but there is no indication about the number and nature of these factors.
- The risk factors may be correlated.
- Every asset i has its own set of sensitivities to the different factors b_{i1}, b_{i2}, …, b_{ik}. These sensitivities can take negative and positive values.

Example:
Assume that a 3-factor model determines the returns of Swiss stocks where the factors are: long-term interest rate (factor 1), industrial production (factor 2), variation of the exchange rate (factor 3). The sensitivities are determined through regression analysis, and the following results have been obtained[46]:

[45] See ROSS Stephen A., 1976, "The Arbitrage Theory of Capital Asset Pricing", Journal of Economic Theory, Vol. 13, pp. 341-360

[46] Adapted from the results of CUENOT Elisabeth and REYES Cecilia, 1992, "Multi-Factor APT Model for the Swiss Equity Market", CS Investment Research Basic Report

Stock	b_{i1}	b_{i2}	b_{i3}
Nestlé	-0.20	2.35	-0.98
Roche	-0.45	2.21	-1.21
EG Laufenburg	0.13	-1.08	-0.42
SMH	-0.28	0.56	-0.02

The multifactor model is often described with an alternative but equivalent expression in terms of factor deviation to their mean. In order to get this expression, take the expectation of the previous equation

$$E(R_i) = a_i + b_{i1} \cdot E(F_1) + b_{i2} \cdot E(F_2) + \ldots\ldots + b_{ik} \cdot E(F_k) + E(\varepsilon_i)$$

The last term of the right-hand side expression is equal to zero, as it is the average of the residual component. This expression is then subtracted from the original multifactor model to get

$$R_i = E(R_i) + b_{i1} \cdot (F_1 - E(F_1)) + b_{i2} \cdot (F_2 - E(F_2)) + \ldots\ldots + b_{ik} \cdot (F_k - E(F_k)) + \varepsilon_i$$

This expression is equivalent to the original multifactor model. Note that, by construction, the variables $(F_i - E(F_i))$ have an expected value of zero.

1.6.1.2 Absence of Arbitrage Opportunities*

This assumption is one of the pillars of modern financial theory. It is also widely used in other fields of finance (e.g. option pricing theory). This assumption is derived from the very general economic principle known as the law of one price, which states that two identical items cannot sell at different prices. Translated in financial terms, this principle states that two securities with identical risks cannot have different expected returns. If such a situation does exist, arbitrageurs would buy the high return security and sell short the low return security. They would have a portfolio without an initial investment and risk, which would offer a positive profit known as arbitrage (or riskless) profits. This process, known as arbitrage, would last until the equilibrium is reached where securities with the same risk will provide the same expected return. Stated differently, this assumption states that it is impossible for an investor to earn a positive expected return from a portfolio without assuming some risk and making some net investment of funds.

1.6.1.3 Other Assumptions and a Definition*

The APT also requires the following assumptions:

- Investors prefer more to less but there is no specific assumption about their risk-aversion.
- The number of available securities in financial markets is much larger than the number of factors generating the asset returns.
- Short sales are allowed and assets are infinitely divisible.

Before turning to the description of the APT, let us define a well-diversified portfolio. Such a portfolio is supposed to be diversified over a large number of securities and to have very small proportions invested in each security. This implies that the residual component of this portfolio ε_p is equal to zero (by the law of large numbers and assuming $E(\varepsilon_p) = 0$) and that its specific risk $\sigma_{\varepsilon p}$ is negligible (close to or equal to zero).

1.6.2 *The APT and its Derivation**

*1.6.2.1 Development of the APT**

Under the previous set of assumptions, the APT states that the equilibrium expected return on every available asset on the market is

$$E(R_i) = \lambda_0 + b_{i1} \cdot \lambda_1 + b_{i2} \cdot \lambda_2 + \ldots\ldots + b_{ik} \cdot \lambda_k$$

Where b_{ij} is the sensitivity of the security i to the factor j and λ_j is the risk premium on the factor j.
Moreover, it can be shown that

$$\lambda_0 = R_f$$

Where R_f is the risk-free rate, and also that

$$\lambda_j = E(PF_j) - R_f$$

Where PF_j is the return on a pure factor j portfolio. A pure factor j portfolio is defined as a portfolio which has a sensitivity b_j to factor j equal to one and sensitivities b_m ($m \neq j$) to other factors all equal to zero. In the special case when the factors of the multifactor model are independent (not correlated), the pure factor portfolios are equivalent to factors themselves.
Under the above assumptions we have

$$\begin{aligned} E(PF_j) &= R_f + 0 \cdot \lambda_1 + 0 \cdot \lambda_2 + \cdots + 1 \cdot \lambda_j + 0 \cdot \lambda_k \\ &= R_f + \lambda_j \end{aligned}$$

At this point, note that:

- The APT is *different* from the original multifactor model. APT determines equilibrium expected returns and is an exact relationship. The multifactor model determines realized returns and includes a residual undetermined component. However, the assumption that realized returns are generated by the multifactor model is necessary to obtain the APT.
- The APT does not specify the number and the nature of the factors. It only presents the structure of expected returns.
- The equilibrium relationship of the APT has a structure similar to the CAPM, except that it allows multiple sources of risk and that the assumptions underlying both models are different.

Example

Using the data of the previous example, we assume that the 3 factors are uncorrelated and that their expected return are respectively $E(F_1)$=6%, $E(F_2)$=4%, $E(F_3)$=3% and that the risk-free rate is equal to 2%. The risk premia associated to the factors are equal to: λ_0=2%, λ_1=4%, λ_2=2%, λ_3=1%.

The APT equilibrium expected returns on the securities are:

Nestlé	E(R)=2%-0.20 (4%)+2.35 (2%)-0.98 (1%)	=	4.92%
Roche	E(R)=2%-0.45 (4%)+2.21 (2%)-1.21 (1%)	=	3.41%
EG Laufenburg	E(R)=2%+0.13 (4%)-1.08 (2%)-0.42 (1%)	=	-0.06%
SMH	E(R)=2%-0.28 (4%)+0.56 (2%)-0.02 (1%)	=	1.98%

1.6.2.2 Formal Derivation of the APT*

We derive the equilibrium relationship of the APT by assuming a two-index return generating process. This is enough to allow generalization to any arbitrary number of factors. The two-factor model is

$$R_i = a_i + b_{i1} \cdot F_1 + b_{i2} \cdot F_2 + \varepsilon_i$$

A sufficient condition to prove the existence of the APT is that a well-diversified portfolio with the following characteristics exists:

- Its net investment is zero.
- It has no systematic risk.

The fact that this portfolio is well diversified implies that it has no specific risk. By the absence of arbitrage opportunities condition, the expected return of this portfolio has to be equal to zero.

Denoting x_i as the fraction of the portfolio invested in asset i and n as the total number of available securities, we can rewrite the condition on this portfolio more formally. Since the portfolio has zero investment, the fractions of investments in different assets must be such that

$$\sum_{i=1}^{n} x_i = 0$$

Recall that the systematic risk of a portfolio is the weighted sum of the risk of its components. A portfolio without risk due to factor 1 and 2 is such as

$$\sum_{i=1}^{n} x_i \cdot b_{i1} = 0 \text{ and } \sum_{i=1}^{n} x_i \cdot b_{i2} = 0$$

As this portfolio is riskless and without investment, by the absence of arbitrage opportunities condition, it has to yield a zero expected return, that is

$$\sum_{i=1}^{n} x_i \cdot E(R_i) = 0$$

Assuming that we know the weights x_i, this leaves us with the question of what should be the value of the expected returns of individual stocks. As we have n securities on the market, we have 4 equations and n unknown variables. A solution to this system is when the expected returns are a linear combinations of their respective risk sensitivities, such as

$$E(R_i) = \lambda_0 + \lambda_1 \cdot b_{i1} + \lambda_2 \cdot b_{i2}$$

This is the equilibrium equation of the APT. In order to prove this result, let us take an example with (only!) 5 securities in the economy, but which can be generalized to any number of assets. We have to solve the following system:

x_1	$+x_2$	$+x_3$	$+x_4$	$+x_5$	$=0$	(1)
x_1b_{11}	$+x_2b_{21}$	$+x_3b_{31}$	$+x_4b_{41}$	$+x_5b_{51}$	$=0$	(2)
x_1b_{12}	$+x_2b_{22}$	$+x_3b_{32}$	$+x_4b_{42}$	$+x_5b_{52}$	$=0$	(3)
$x_1E(R_1)$	$+x_2E(R_2)$	$+x_3E(R_3)$	$+x_4E(R_4)$	$+x_5E(R_5)$	$=0$	(4)

Assuming that we know the weights x_i, we have 4 equations and 5 unknown variables, the $E(R_i)$. If we replace $E(R_i)$ by the APT equation in equation (4) we obtain the following equation:

$$x_1 \cdot (\lambda_0 + \lambda_1 \cdot b_{11} + \lambda_2 \cdot b_{12}) + x_2 \cdot (\lambda_0 + \lambda_1 \cdot b_{21} + \lambda_2 \cdot b_{22}) + x_3 \cdot (\lambda_0 + \lambda_1 \cdot b_{31} + \lambda_2 \cdot b_{32})$$
$$+ x_4 \cdot (\lambda_0 + \lambda_1 \cdot b_{41} + \lambda_2 \cdot b_{42}) + x_5 \cdot (\lambda_0 + \lambda_1 \cdot b_{51} + \lambda_2 \cdot b_{52})$$

We group terms in λ_0, λ_1, and λ_2 and obtain

$$\lambda_0 \cdot (x_1 + x_2 + x_3 + x_4 + x_5) + \lambda_1 \cdot (x_1 \cdot b_{11} + x_2 \cdot b_{21} + x_3 \cdot b_{31} + x_4 \cdot b_{41} + x_5 \cdot b_{51})$$
$$+ \lambda_2 \cdot (x_1 \cdot b_{12} + x_2 \cdot b_{22} + x_3 \cdot b_{32} + x_4 \cdot b_{42} + x_5 \cdot b_{52})$$

The coefficient of λ_0 is the left-hand side of equation (1), the coefficient of λ_1 is the left-hand side of equation (2) and the coefficient of λ_2 is the left-hand side of equation (3). These 3 coefficients are all equal to zero as indicated by equation (1), (2) and (3). This means that $E(R_i) = \lambda_0 + \lambda_1 \cdot b_{i1} + \lambda_2 \cdot b_{i2}$ is a solution to equation (4) and solves the system. We have shown that without arbitrage opportunities and if asset returns are driven by a linear multifactor model then expected returns on assets must be linearly related to their risks.
The only question that remains unanswered is the magnitude of the coefficients λ_0, λ_1, and λ_2 λ_2. The equilibrium model produced by the APT when a two-factor model generates the returns is

$$E(R_i) = \lambda_0 + \lambda_1 \cdot b_{i1} + \lambda_2 \cdot b_{i2}$$

Intuitively λ_1 and λ_2 λ_2 are the returns for bearing risks associated with factors 1 and 2. More insight can be gained by examining certain type of portfolios. Assume a portfolio that is insensitive to factors 1 and 2. Such portfolio has therefore $b_{i1=}0$ and $b_{i2=}0$ and is riskless, its expected return has to yield the riskless rate R_f. Therefore λ_0 should be equal to R_f. Let us now examine a portfolio which is only sensitive to factor 1 with a unit sensitivity. This portfolio is such that $b_{i1}=1$ and $b_{i2}=0$. From the equilibrium equation, the expected return of this portfolio is equal to $\lambda_0 + \lambda_1$. We know that and can therefore claim that λ_1 is the expected return of a portfolio only subject to risk of factor 1, having a unit measure of this risk (a pure factor 1 portfolio) minus the risk-free rate. The same type of analysis applies to λ_2.
All the analyses of this section can be generalized to show that if the securities are generated by a multifactor model with k factors, then the asset expected returns are described by a k-dimensional hyperplane

$$E(R_i) = \lambda_0 + \lambda_1 \cdot b_{i1} + \lambda_2 \cdot b_{i2} + \ldots\ldots + \lambda_k \cdot b_{ik}$$

where

$\lambda_0 = R_f$
$\lambda_j = E(PF_j) - R_f$

1.6.2.3 An Illustration of the APT*

Assume that the returns are generated by a two-factor model:

$$R_i = a_i + b_{i1} \cdot F_1 + b_{i2} \cdot F_2 + \varepsilon_i$$

Let us assume that we have 3 stocks that are priced according to the APT. These securities have the following characteristics:

Security	$\mathbf{E(R_i)}$	$\mathbf{b_{i1}}$	$\mathbf{b_{i2}}$
UBS	5%	0.4	-0.3
Novartis	8%	0.6	0.3
ABB	9%	0.4	0.7

We obtain the following equations:

UBS	$5=\lambda_0+0.4\cdot\lambda_1-0.3\cdot\lambda_2$
Novartis	$8=\lambda_0+0.6\cdot\lambda_1+0.3\cdot\lambda_2$
ABB	$9=\lambda_0+0.4\cdot\lambda_1+0.7\cdot\lambda_2$

This results in the APT equilibrium equation:

$$E(R_i) = 5 + 3\cdot b_{i1} + 4\cdot b_{i2}$$

where:

$\lambda_0=5$, $\lambda_1=3$ and λ_2 $\lambda_2=4$

This is also the equation of a plane in the $E(R_i)$, b_{i1}, b_{i2} space. Any linear combination of these securities must be on the same plane.

Let us now consider what happens to a security that is not priced according to the APT (i.e. that is not on that plane). Assume the SMH stock has an expected return of 3.75% a b_{i1} of 0.45 and a b_{i2} equal to 0.35. We compare the SMH stock with a portfolio p by placing 25% of the funds in the UBS stock, 25% of the funds in the Novartis stock and 50% in the ABB stock. The sensitivities of this portfolio p to factor 1 and 2 are:

$$b_{p1} = 0.25\cdot(0.40) + 0.25\cdot(0.60) + 0.50\cdot(0.40) = 0.45$$
$$b_{p2} = 0.25\cdot(-0.30) + 0.25\cdot(0.30) + 0.50\cdot(0.70) = 0.35$$

The risks of portfolio p are therefore identical to the risks of the SMH stock. The expected return on portfolio p is:

$$E(R_p) = 0.25\cdot(5\%) + 0.25\cdot(8\%) + 0.50\cdot(9\%) = 7.75\%$$

Alternatively, since portfolio p must lie on the plane described above (the APT equation) we could have obtained its expected return from the equation of the plane:

$$E(R_p) = 5 + 3\cdot(0.45) + 4\cdot(0.35) = 7.75\%$$

By the absence of arbitrage opportunities assumption, two portfolios with the same risk cannot have different expected returns. In fact, if such a situation existed it would quickly disappear as arbitrageurs would step in and would buy portfolio p and sell short the SMH stock, thereby obtaining an arbitrage profit of 4%. Let us illustrate this by assuming that an investor buys CHF 1000 of portfolio p and sells short CHF 1000 of SMH stock.

Security	**Initial Cash Flow**	**End of Period Cash Flow**	b_{p1}	b_{p2}
Portfolio p	-1000	1077.50	0.45	0.35
SMH stock	1000	-1037.50	-0.45	-0.35
Arbitrage portfolio	0	40.00	0.00	0.00

The arbitrage portfolio involves zero investment, no systematic risk (b_{p1}, b_{p2}) and earns CHF 40. Arbitrage would continue until the stocks SMH lies on the same plane as UBS, Novartis and ABB stocks.

1.6.3 *The Link between the APT and the CAPM**

The CAPM and the APT should not be considered as mutually exclusive models but as models having different approaches to the same reality. This section describes the link between the two models and also their major differences.

Let us first review the link between the two models. The simplest case is where returns are generated by a one-factor model and that the single factor is the market portfolio. In this case, the CAPM and the APT have the same equilibrium pricing equation, despite the fact that they are based on different assumptions. This case is trivial and uninteresting since there is only one source of risk.

A more interesting case is where the security returns are generated by a k-factor model. It is then possible to relate the beta of a security to its sensitivities to the k factors. In a two-factor model and given the properties of covariance, the covariance of the return on the security i with the return of the market portfolio M is:

$$\begin{aligned}\mathrm{Cov}(R_i,R_M) &= \mathrm{Cov}((a_i + b_{i1}\cdot F_1 + b_{i2}\cdot F_2 + \varepsilon_i),\, R_M)\\ &= \left[\mathrm{Cov}(F_1,R_M)\cdot b_{i1}\right] + \left[\mathrm{Cov}(F_2,R_M)\cdot b_{i2}\right] + \mathrm{Cov}(\varepsilon_i,R_M)\end{aligned}$$

Dividing both sides of the equation by σ_M^2 and knowing that, by the CAPM definition, $\beta_i = \mathrm{Cov}(R_i, R_M)/\sigma_M^2$ we get:

$$\beta_i = \left[\frac{\mathrm{Cov}(F_1,\, R_M)}{\sigma_M^2}\cdot b_{i1}\right] + \left[\frac{\mathrm{Cov}(F_2,\, R_M)}{\sigma_M^2}\cdot b_{i2}\right] + \frac{\mathrm{Cov}(\varepsilon_i,\, R_M)}{\sigma_M^2}$$

Assuming that the error terms are not correlated with the market returns, we obtain:

$$\beta_i = \beta_{F1}\cdot b_{i1} + \beta_{F2}\cdot b_{i2}$$

Where β_{F1} and β_{F2} are the respective betas of factors 1 and 2 (with respect to the market). Since β_{F1} and β_{F2} are constant and independent of the share (they only depend on correlations between the factors and the market), the beta of a security is a function of b_{i1} and b_{i2}. Hence, if two securities have different betas, it is due to different sensitivities to the factors.

As a result of the CAPM assumptions, the expected returns of security i is related to the beta:

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

Hence, in a two-factor model, the returns are linked to the two factors through the following relationship:

$$\begin{aligned}E(R_i) &= R_f + (E(R_M) - R_f)\cdot(\beta_{F1}\cdot b_{i1} + \beta_{F2}\cdot b_{i2})\\ &= R_f + \left[(E(R_M) - R_f)\cdot\beta_{F1}\right]\cdot b_{i1} + \left[(E(R_M) - R_f)\cdot\beta_{F2}\right]\cdot b_{i2}\\ &= R_f + \lambda_1\cdot b_{i1} + \lambda_2\cdot b_{i2}\end{aligned}$$

where:

$$\lambda_1 = (E(R_M) - R_f)\cdot\beta_{F1}$$
$$\lambda_2 = (E(R_M) - R_f)\cdot\beta_{F2}$$

The CAPM gives another meaning to the risk premiums λ_1 and λ_2, which were defined as expected excess returns on pure factor portfolios. Assuming that factor 1 is positively correlated with market return means that $\mathrm{Cov}(F_1,R_M)$ is positive. Consequently, β_{F1} is positive

and, since $E(R_M)–R_f > 0$, λ_1 is also positive. Thus, the higher b_{i1}, the higher expected return of the security. The reverse will be true for negative betas.

This result means that the two theories are consistent. Nevertheless, the CAPM should not be considered to be a particular case of the APT since their assumptions are different.

- The CAPM makes strong assumptions on investors' behaviour, on the role of the market portfolio and on the way equilibrium is achieved.
- The APT makes less stringent assumptions on investors' behaviour, does not assign a particular role to the market portfolio and is more realistic by assuming multiple sources of risk.

The apparent superiority (or generality) of the APT from a theoretical point of view is however lost from an empirical point of view. The CAPM makes very precise predictions on the way expected returns are determined. On the other hand, the APT remains silent on the number and nature of factors generating asset returns. As we will see in the next section, despite empirical tests of the APT use sophisticated statistical techniques, no consensus and definitive conclusion on the nature and number of APT factors has been obtained so far in the academic literature.

*1.6.4 Empirical Tests of the APT**

*1.6.4.1 Identifying factors**

Since the APT is based on a factor model of security returns, any test of its predictions must incorporate such a factor model. The test will be, in fact, a joint test of an equilibrium theory and of the appropriateness of the selected factor model. This means that it may be very difficult to interpret the results of the tests since we do not know if the APT does not work or if the model was only badly specified (a set of wrong factors or the wrong number of factors was used).

Another problem is that it is very difficult to determine which the relevant factors to use in the model are. Hence, there are two alternative approaches that can be used to estimate an APT model:

- The risk factors and the sensitivities of the assets to those factors can be simultaneously computed using statistical techniques, such as factor analysis or principal components analysis[47]. But these methods have drawbacks. First, the factors are not identified economic variables. It is necessary to compare the factors with existing variables to identify them. Moreover, the economic interpretation of factors may change over time, so that, for example, factor 2 for a sample period will certainly differ from factor 2 in another sample period. Second, the number of relevant factors found appears to vary according to the number of stocks used in the analysis, while it should be independent of the sample used.

47 Both techniques extract from the data a set of indices that best explain the variance of the data, so that the covariance of residual returns (returns after the influence of the indices have been removed) is as small as possible.

- Economic theory and knowledge of financial markets can be used to hypothesise and pre-specify an intuitively appealing set of factors that can be measured from available macroeconomic and financial data. But the right selection of an appropriate set of factors involves as much art as it does science.

*1.6.5 Pre-Specifying Factors**

Several studies have used the second approach. For instance, Chen, Roll and Ross (1986)[48] have determined a large fraction of the covariance between the securities using the following macroeconomic factors:

- The spread in yield between a long-and a short-term treasury bond, as differences affect the value of payments far in the future relative to near-term payments.
- The rate of inflation, which impacts both the level of the discount rate and the size of the future cash flows.
- The spread between low-grade bonds and treasury bonds, which is a measure of market reaction to risk.
- The growth rate of industrial production, as changes in industrial production affects the opportunities facing investors and the real value of cash flows.

The problem with this method is that i) it is not necessary that there are only four factors and ii) that the above factors are the proper ones. Hence, even if these factors explain most of the variance, it does not necessarily validate the APT. The sample used in estimating the factors may not be large enough or may not be representative. It is possible that a different sample may yield entirely different factors from those found using the first sample.

Nevertheless, it was possible to find evidence that there are risk premia associated with these factors, but it must be noted that the market index often proved to be one of the (if not the only one) relevant factors. This means that it is very hard to determine which factors really drive stock returns. Chen, Roll and Ross end up with the market as not being a significant factor for portfolio returns in their study.

Another example of a study using pre-specified factors is the work of Fama and French (1993). They found that the following factors significantly affect securities returns:

- The difference in return on a portfolio of small stocks and a portfolio of large stocks.
- The difference in return on a portfolio of high book-to-market value stocks and a portfolio of low book-to-market value stocks.
- The difference between the monthly long-term government bond return and the one-month Treasury bill return.
- The difference in the monthly return on a portfolio of long-term corporate bonds and a portfolio of long term government bonds.

As can be seen from these factors, microeconomic variables reflecting the difference in firm size and book-to-market value ratios are also relevant in explaining the returns. This shows that the search for relevant factors is still going on, and their identification is the crucial step before the APT can really become a recognized model from an empirical point of view.

---

48 CHEN Rai-Fu, ROLL Richard and ROSS Stephen A., 1983, “Economic Forces and the Stock Market”, Journal of Business, Vol. 59, pp. 386-403

*1.6.6 Some Applications of the APT**

In practice, the APT can be used for different purposes. This section describes two applications of the model: tracking an index and active portfolio management.

The first use of the APT is linked to the use of multifactor models in the creation of a portfolio of stocks that closely tracks an index. An obvious way to construct an index fund is to hold stocks constituting the portfolio in the same proportions as the stocks are represented in the index. This way of replicating the index can be very costly when the index is composed of a large number of securities. For instance, an investor who would like to hold the Swiss Performance Index (SPI) with this method will have to hold more than 250 securities and frequently readjust his positions to replicate the exact composition of the index. This would incur a lot of transaction and monitoring costs. Another way to achieve the same goal is to use a multifactor model to determine the sensitivities of the index to the underlying factors. The investor could then hold a reduced set of securities in a portfolio that would have exactly the same sensitivities to the factors as the index. The index would then be perfectly tracked with a reduced number of stocks. Nowadays, several funds avoid certain stocks for ethical or other reasons (e.g. tobacco companies). With the help of a multifactor model, one could still form a portfolio avoiding those stocks but still matching the sensitivities of the index to track. The problem with this application is that the exact identities of factors generating the returns are unknown. This is where APT can help us, with the statistical tools used in its empirical tests. These methods (factor analysis and principal components analysis) allow us to determine statistically factors and sensitivities of the securities and the index, without knowing the identity of the factors. The sensitivities of the stocks are used to form a portfolio replicating the sensitivities of the index.

Example:
Assume that we have the following sensitivities to 4 factors determined with factor analysis.

Stock	b_{i1}	b_{i2}	b_{i3}	b_{i4}
Nestlé	-1.60	0.42	1.02	-4.66
Roche	0.50	0.12	0.98	3.88
SPI Index	-0.55	0.27	1.00	-0.39

It is possible to form a portfolio with 50% invested in Nestlé and the other 50% in Roche and obtain exactly the same sensitivities to the 4 factors as the SPI index.

$$b_{p1} = 0.5\cdot(-1.60) + 0.5\cdot 0.50 = -0.55$$
$$b_{p2} = 0.5\cdot 0.42 + 0.5\cdot 0.12 = 0.27$$
$$b_{p3} = 0.5\cdot 1.02 + 0.5\cdot 0.98 = 1.00$$
$$b_{p4} = 0.5\cdot(-4.66) + 0.5\cdot 3.88 = -0.39$$

Active portfolio management involves making bets about securities or group of securities, i.e., designing a portfolio based on the belief that one or more securities are mispriced. APT can be used to determine the equilibrium (or normal) expected return on a security and then compare it with some predictions about future return on the security. The first kind of expected return is the global return obtained through APT. If an analyst predicts a higher return than that obtained through the APT for a stock, then the investor would buy that stock. On the other hand, if the analyst predicts that a stock will have a lower return than its expected APT return, the investor should sell that security.

Another kind of active portfolio management could be designed on the basis of predictions on realization of a factor. If an analyst predicts a higher return on a pure factor portfolio than that

used in the APT (λ's), then it is worthwhile increasing the investor's exposure to that factor, by investing in securities with high positive sensitivities to that factor.

Example:
We have 3 factors which are uncorrelated and which expected return are respectively $E(F_1)=6\%$, $E(F_2)=4\%$, $E(F_3)=3$. If an analyst predicts that the expected return on the first factor will be 10% instead of the equilibrium 6%, then it would be profitable to invest in securities that have a high positive sensitivity to factor 1.