2. Modern Portfolio Theory

2.1. Overview

Modern Portfolio Theory (MPT) was pioneered by Harry Markowitz in a paper published in 1952. This was considered as a breakthrough in using quantitative methods within the investment process. A key idea of this theory is the emphasis on the tradeoff between *risk* and *return*. Basically, risk is measured via the variance of returns, and the end result of the portfolio management process is the generation of a **mean-variance efficient portfolio.** Portfolio selection is achieved through an optimization process, usually in the presence of various constraints. We can state the investment problem as follows:

- Given the characteristics of the assets (e.g. expected mean and variance of return), we want to find a portfolio that can either:
    - Maximize return for a given level of risk
    - Minimize risk for a given level of return

The solution to the above problem should tell us the optimum portfolio allocation, i.e. the exact proportion (or weights) of the various assets.

2.2. The concept of "risk"

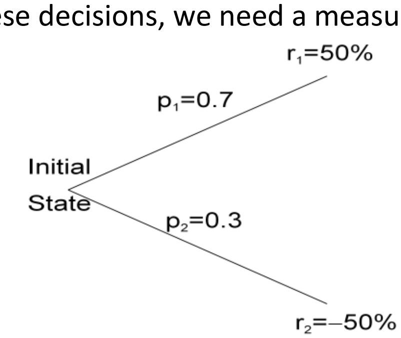
2.2.1. Measures of risk

We define the presence of risk as the situation when more than one outcome is possible and there is a chance that one can suffer a loss. Investors often need to choose between various options. Consider the following example: with an initial investment of $1000, there are two alternatives:

- Buy a 1-year government bill, yielding 2%, i.e. the final amount received = $1020.
- Invest in stocks, which you think there is a probability of 0.7 that the stock would worth $1500 and a probability of 0.3 that it would become $500.

Obviously there is a difference between these two alternatives. In the first option, the outcome is (almost) certain because (it is assumed) there is very little chance that the government will go bankrupt. The investor can receive the initial investment plus interest at the end of one year. This is known as a risk free (or "riskless") investment. In the second case, while there is a substantial chance of a big return, the chance of losing money is not insignificant either. This option is risky because one can possibly lose money at the end.

In order to quantify these decisions, we need a measure of risk.



In the diagram above, p_i is the probability of each outcome; r_i is the possible return. The expected return is given by:

$$E(r) = \bar{r} = \sum_{i=1}^{2} p_i r_i = 0.7 \times 0.5 + 0.3 \times (-0.5) = 20\%$$

Standard deviation of return

$$\sigma = \sqrt{\sum_{i=1}^{2} p_i \left(r_i - \bar{r}\right)^2}$$
$$= 45.8\%$$

σ is the key parameter in measuring riskiness. A higher σ denotes higher risk. By definition, a risk free investment would have σ = 0.

Compare to the first choice above which is a risk free investment (return = 2%), the second choice can have an additional return of 18%. This is known as the *risk premium*, which is defined as the extra return over a risk free investment. If there is no risk premium (i.e. expected returns of two options are the same), the basic assumption here is that investors will always choose the option which would have the lower risk, and thus the first option will never become attractive. Thus risk premium is a compensation for the risk of an investment. Of course, the investment decision in real life is more complex. However, there is still a remaining question: would a risk premium of 18% be enough to entice you into the investment opportunity? What if there is an investment with an expected return of 10%, but σ=20%? How can we rank these choices?

In finance theory, investors are classified into three different types.

- **Risk averse**: although the return is high, these investors would "penalize" high risk investments; a higher risk premium is required for investments with higher risk (although expected return may be higher)

- **Risk neutral**: investors who are happy to receive the expected return, ignoring risk factors; the higher the return the better

- **Risk seeking**: investors who like risk, so they are happy to receive a lower than expected return to enter the game (e.g. in a casino)

Many research results found that most people are risk averse when they consider investments.

Quiz: How would the different types of investors pick from these investment choices?

i. E(r) = 2%, σ = 0
ii. E(r) = 20%, σ = 0.46
iii. E(r) = 21%, σ = 0.90
iv. E(r) = 10%, σ = 0.20
v. E(r) = 1%, σ = 1.25; small probability of extra return = 200%

Answer: A risk seeking investor would choose the option with the highest risk, i.e. (v). A risk neutral investor would choose the option with the highest expected return, i.e. (iii). A risk averse investor may choose (i), (ii), or (iv), depending on the individual's risk preference; (iii) is still possible, but it is quite unlikely because the big increase in risk (from σ= 0.46 to σ= 0.90) is only compensated by a small increase in return (from 20% to 21%).

Investment choice	Risk averse	Risk seeking	Risk neutral
1. E(r) = 2%, σ = 0	Possible	X	X
2. E(r) = 20%, σ = 0.46	Possible	X	X
3. E(r) = 21%, σ = 0.90	Unlikely	X	√
4. E(r) = 10%, σ = 0.20	Possible	X	X
5. E(r) = 1%, σ = 1.25 small probability of extra return = 200%	X	√	X

2.2.2. How do we quantify risk aversion?

In order to represent the different levels of risk aversion of the investors, we introduce the concept of *utility function*. Simply speaking, it can be used as a means to rank portfolios with different risk/return characteristics. Many valid forms exist, and one popular choice is a *quadratic function*:

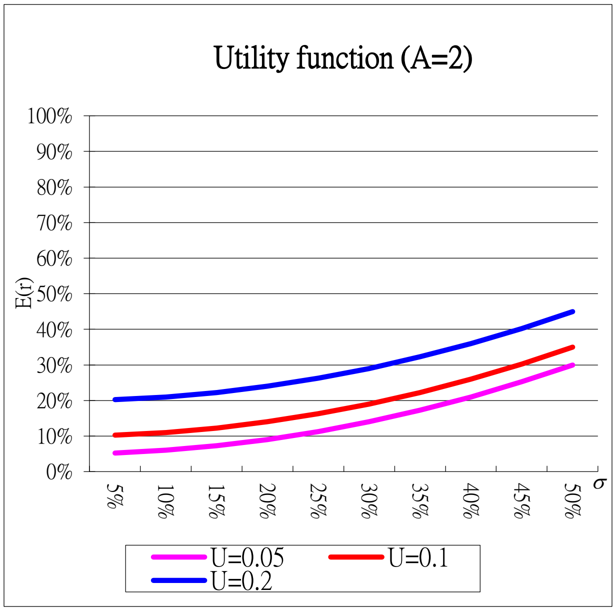
$$U = E(r) - \frac{1}{2}A\sigma^2$$

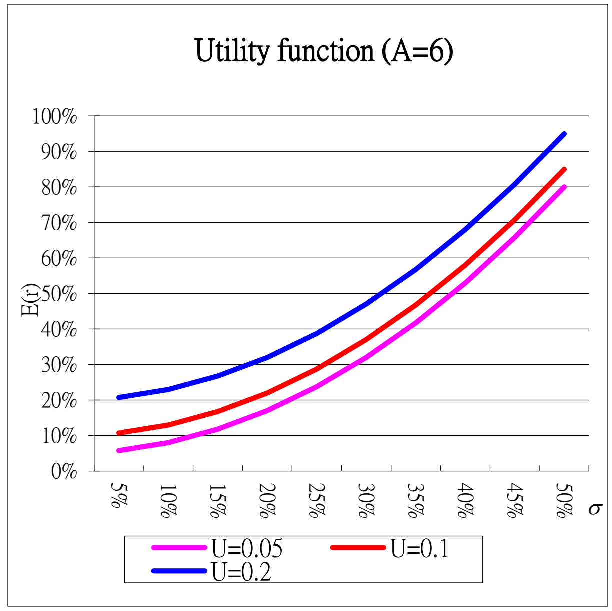
In this formulation, utility only depends on two variables, the expected return and standard deviation of an investment. There is an inherent assumption known as *non-satiation* – more money is preferred to less money, or equivalently, we can assume that the higher the utility the better. This is not always true if we need to take other non-financial factors into account, but it is a simplification underlying most finance theories.

A is a measure of the risk tolerance of the investor. We can distinguish between the different cases of A which correspond to the classification of investors earlier:

- A >0: risk averse
  - In order to get to a higher utility, if an investment has a higher σ, then E(r) has to be even higher to compensate for the risk.
- A =0: risk neutral
  - Changing σ does not affect the utility; E(r) determines U. In other words, these investors seeking the highest utility would only look for the choice with the highest E(r).
- A <0: risk seeking
  - The second term will contribute to the utility, thus the higher σ the better.

We can illustrate these concepts by plotting E(r) against σ, and this plot is known as the *indifference curve*.





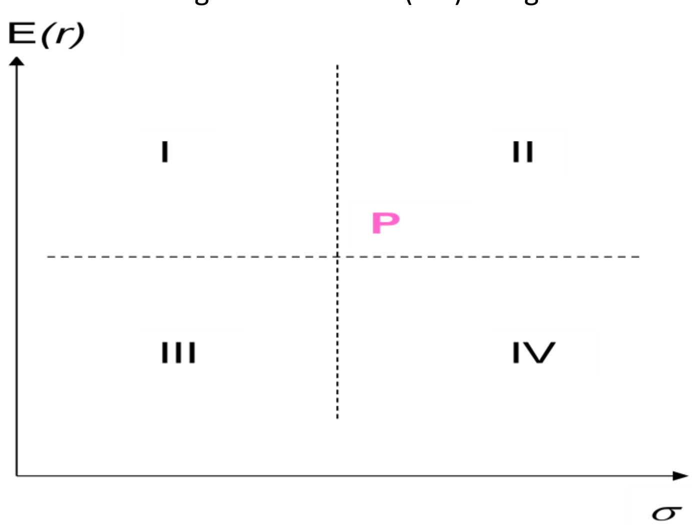
In the left diagram, A = 2, and the lines show the contours when U is kept constant. For example, a point on the blue line represents a particular combination of (σ, E(r)) such that U = 0.2. All points on the same contour line are considered as equivalent choices because the utility remains the same.

Two points are noted from these diagrams. Firstly, with the non-satiation principle, the curves move towards the North-West direction for higher U's and higher U's are preferred. Secondly, when A is larger, the curves are steeper. This is consistent with the understanding that A represents risk aversion, so a higher A means that the investor demands a higher return when σ increases.

2.3. The portfolio selection process

2.3.1. Defining the problem

We can now re-phrase the criterion given in section (2.1) using the **mean-variance criterion**.



The diagram shows a portfolio P with a given expected return and risk. Portfolios in quadrant I are preferable to P, which is preferable to those in quadrant IV, because $\mathrm{E}(r)$ is higher and σ is lower. Formally speaking, if

$$E(r_M) \geq E(r_N) \text{ and } \sigma_M \leq \sigma_N$$

and at least one inequality is strict (i.e. without the equal sign), then we say portfolio M **dominates** portfolio N (or in other words, M is more preferable to N). However, we cannot make a claim to the portfolios in quadrants II and III. *Whether portfolios in quadrants II and III are preferable to P depend on the investor's risk aversion*. For example, if there is a portfolio in quadrant II called Q, it would have a higher expected return than P, but the risk of Q is also higher. We cannot conclude whether the investor would prefer Q to P or not: it depends on the investor's attitude to risk.

The asset allocation problem is a three-step process.

Step 1: select a combination of risky assets in a portfolio
Step 2: combine the risky portfolio with a risk free asset (if the risk free asset is available)
Step 3: select the optimal portfolio according to the individual investor's preference

2.3.2. Step 1: combining risky assets

Assume that we have already decided to select risky assets from all available investments to be put into the portfolio, and know the expected return $\mathrm{E}(r)$ and standard deviation σ of each asset (*this is a big assumption*). The problem is to find out which assets and in what proportions can we achieve an optimal result. We can form a portfolio with a value of P using N risky assets

$$P = \sum_{i=1}^{N} w_i S_i$$

where w_i and S_i are the weight and price of asset *i*. By convention, we could set

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

Mathematically we can write down the characteristics of the portfolio. Firstly, the return of the portfolio is simply the weighted average of the return on the individual assets. Therefore the expected return of the portfolio is the weighted average of the expected return on individual assets:

$$\overline{r_p} = \sum_{i=1}^{n} w_i \overline{r}_i$$

Covariance between two assets:

$$\sigma_{ij} = \sum_{k=1}^{N} \left[\left(r_{ik} - \overline{r_i} \right) \left(r_{jk} - \overline{r_j} \right) \right]$$

Correlation between two assets:

$$\rho_{ij} = \sigma_{ij} \big/ \sigma_i \sigma_j$$

Portfolio variance: $$\sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{j=1}^{N} \sum_{\substack{k=1 \\ k \neq j}}^{N} w_j w_k \sigma_{jk}$$

The above can be expressed more succinctly with a matrix convention. With portfolio weights: $\boldsymbol{w} = (w_1 \ldots w_N)^T$ and expected returns of assets: $\bar{\boldsymbol{r}} = (\overline{r_1}, \ldots, \overline{r_N})^T$, and an NxN covariance matrix Ω, where Ω is symmetric and $\Omega_{ij} = \sigma_{ij}$ = covar(r_i, r_j), the portfolio characteristics are given as:

Portfolio expected return: $\overline{r_p} = \boldsymbol{w}^T \bar{\boldsymbol{r}}$

Portfolio variance: $\sigma_p^2 = \boldsymbol{w}^T \Omega \boldsymbol{w}$

For N=2, we have $r_p = w_1 r_1 + w_2 r_2$

$$\sigma_p^2 = [w_1 \quad w_2] \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= [w_1\sigma_{11} + w_2\sigma_{21} \quad w_1\sigma_{12} + w_2\sigma_{22}] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12}$$

$$= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2$$

Note that we assume Ω to be symmetric and positive definite, so that its inverse Ω^{-1} always exist.

An example of portfolio variance calculation

We have the following data

			Covariance with		
Security	Weight	s.d. (σ)	Cheung Kong	HSBC	MTR
Cheung Kong	1/3	36%		0.0648	0.04968
HSBC	1/2	20%			0.023
MTR	1/6	23%			

The portfolio variance is:

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 + 2w_1w_2\sigma_{12} + 2w_1w_3\sigma_{13} + 2w_2w_3\sigma_{23}$$

$$= \left(\frac{1}{3}\right)^2 \times 0.36^2 + \left(\frac{1}{2}\right)^2 \times 0.20^2 + \left(\frac{1}{6}\right)^2 \times 0.23^2 + 2\times\left(\frac{1}{3}\right)\times\left(\frac{1}{2}\right)\times 0.0648$$

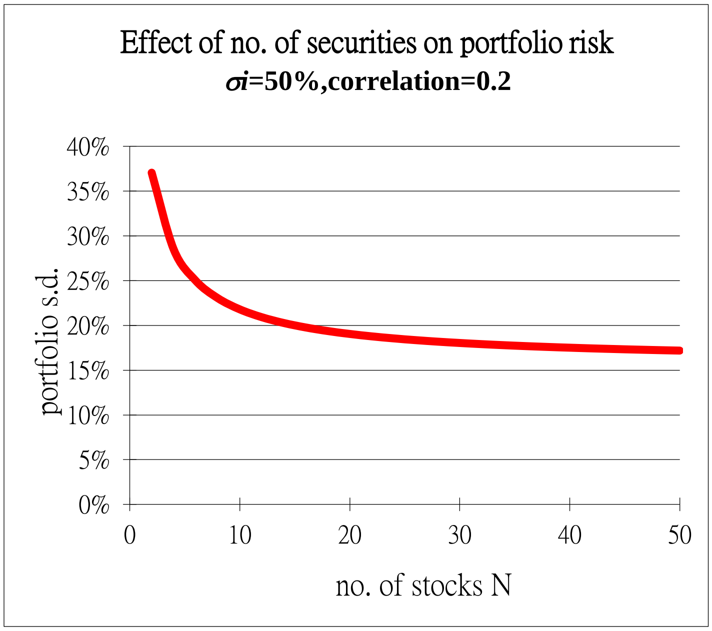
$$+ 2\times\left(\frac{1}{3}\right)\times\left(\frac{1}{6}\right)\times 0.04968 + 2\times\left(\frac{1}{2}\right)\times\left(\frac{1}{6}\right)\times 0.023$$

$$= 0.0568 \quad \text{or} \quad \sigma_P = 23.8\%$$

The effect of diversification

We often heard that we should diversify our investments. In fact this suggestion has sound theoretical basis within the modern portfolio theory. To illustrate the idea, assume that each of N assets has equal weighting, i.e. $w_i = 1/N$. The portfolio variance is given as:

$$\begin{aligned}
\sigma_p^2 &= \sum_{i=1}^{N} (1/N)^2 \sigma_i^2 + \sum_{j=1}^{N} \sum_{\substack{k=1 \\ k \neq j}}^{N} (1/N)^2 \sigma_{jk} \\
&= \frac{1}{N} \sum_{i=1}^{N} \frac{\sigma_i^2}{N} + \frac{N-1}{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ k \neq j}}^{N} \left[\frac{\sigma_{jk}}{N(N-1)} \right] \\
&= \frac{1}{N} \overline{\sigma_i^2} + \frac{N-1}{N} \overline{\sigma_{jk}}
\end{aligned}$$



The first term in the previous equation is known as "**diversifiable risk**", i.e. risk that can be eliminated by diversification (i.e. increasing the number of assets in the portfolio). We can see that there is a limit to the decrease in the portfolio s.d. The second term is sometimes known as "**non-diversifiable risk**" or "**systematic risk**". In the real world, covariance between assets are sometimes attributed to market factors (e.g. country).

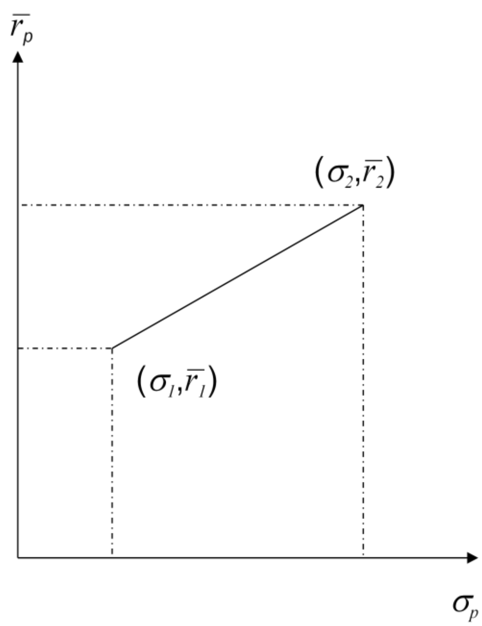
As the number of assets increases, the contribution of the first term becomes smaller. In the limit when $N\text{->}\infty$, portfolio variance would approximate the average of the covariance of the assets in the portfolio. An important conclusion is that in the case of a portfolio with many assets, *covariance between assets is important* whereas Individual asset variance is "irrelevant." We will come back to this point when we look at the Capital Asset Pricing Model (CAPM) in a later chapter.

We are now ready to introduce the mean-variance portfolio optimization process. For illustration, we assume the simplest example: the portfolio just consists of two assets. Assume that these two assets have $\overline{r_1}$ and $\overline{r_2}$ as the expected returns, and σ_1 and σ_2 are the standard deviation of returns. A portfolio can be formed by combining these two assets in different proportions (w_1 and w_2 being the weights assigned to each asset), effectively creating a new "asset" with expected return $\overline{r_p}$ and standard deviation σ_p:

$$
\begin{aligned}
&w_1 + w_2 = 1 \\
&\overline{r_p} = w_1\overline{r_1} + w_2\overline{r_2} = w_1\overline{r_1} + (1-w_1)\overline{r_2} \\
&\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho w_1 w_2 \sigma_1\sigma_2}
\end{aligned}
$$

Depending on the covariance of the two assets, the portfolio characteristics can be represented on an Expected return-Standard deviation diagram.

Two-asset portfolio example: $\rho = 1$



$$
\begin{aligned}
\overline{r_p} &= w_1\overline{r_1} + (1-w_1)\overline{r_2} \\
\sigma_p &= \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2 w_1 w_2 \sigma_1\sigma_2} \\
&= w_1\sigma_1 + (1-w_1)\sigma_2
\end{aligned}
$$

Eliminating w_1, we find

$$
\begin{aligned}
\overline{r_p} &= \frac{\overline{r_2}\sigma_1 - \overline{r_1}\sigma_2}{\sigma_1 - \sigma_2} + \frac{\overline{r_1} - \overline{r_2}}{\sigma_1 - \sigma_2}\sigma_p \\
&= c + \theta\sigma_p
\end{aligned}
$$

This is an equation of a straight line with slope θ and intercept c. However, if $1 > w_1 > 0$, i.e. borrowing is not allowed, the straight line must end at $(\sigma_1, \overline{r_1})$ and $(\sigma_2, \overline{r_2})$, representing the case when 100% of the money is invested in asset 1 or asset 2.

Two-asset portfolio example: $\rho = -1$

The portfolio variance is

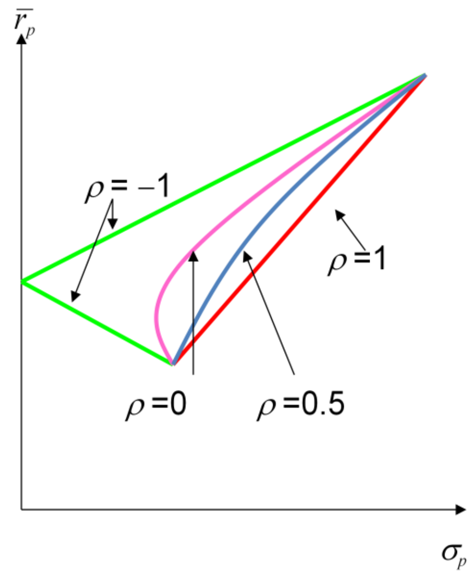
$$
\begin{aligned}
\sigma_p &= \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 - 2 w_1 w_2 \sigma_1\sigma_2} \\
&= \left| w_1\sigma_1 - (1-w_1)\sigma_2 \right|
\end{aligned}
$$

This is equivalent to the following two straight lines

$$
\begin{aligned}
\sigma_p &= w_1\sigma_1 - (1-w_1)\sigma_2, \ w_1 \ge \frac{\sigma_2}{\sigma_1+\sigma_2} \\
&= (1-w_1)\sigma_2 - w_1\sigma_1, \ w_1 < \frac{\sigma_2}{\sigma_1+\sigma_2}
\end{aligned}
$$

Note that when $w_1 = \dfrac{\sigma_2}{\sigma_1 + \sigma_2}$, the portfolio variance becomes 0.

Two-asset portfolio example: other cases



The diagram shows combinations of the portfolio opportunity set for different ρ (only valid when 1> w_i >0, i.e. no borrowing allowed). The different combinations are bound by the lines given in the cases of ρ = 1 and $\rho = -1$. Clearly, the lower the ρ, the more risk reduction through diversification could be achieved (by having a smaller minimum σ_p; for example, in the extreme case of $\rho = -1$, it would be possible to construct a portfolio with σ_p = 0).

We can look at the following numerical example:

weight		$\rho = -1$		$\rho = -0.5$		$\rho = 0.5$		$\rho = 1$	
w_A	$w_B = 1 - w_A$	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
1.0	0.0	10.0	10.00	10.0	10.00	10.0	10.00	10.0	10.00
0.8	0.2	12.0	3.08	12.0	5.04	12.0	8.96	12.0	10.92
0.5	0.5	15.0	0.13	15.0	3.19	15.0	9.31	15.0	12.37
0.2	0.8	18.0	6.08	18.0	8.04	18.0	11.96	18.00	13.92
0.0	1.0	20.0	15.00	20.0	15.00	20.0	15.00	20.0	15.00

The expected returns and risks of two assets A and B are given by: $r_A = 10\%$, $r_B = 20\%$, $\sigma_A = 10\%$, $\sigma_B = 15\%$. The table is generated without the % sign, where the portfolio mean and variance are given by:

$$\overline{r_p} = w_A \times 10 + (1 - w_A) \times 20$$

$$\sigma_p = \sqrt{w_A^2 \times 10 + (1 - w_A)^2 \times 15 + 2\rho w_A (1 - w_A) \times \sqrt{10} \times \sqrt{15}}$$

We can see the effect of correlation: a lower variance is achieved for a given mean when the correlation of the pair of assets' returns becomes more negative. For example, if the mean return is 15%, the lowest variance that can be obtained is 0.13, which is reached when $\rho = -1$.

Two asset-portfolio example: minimum variance portfolio

We can write down an expression for the weights w_1 and w_2 referring to the *minimum variance portfolio*, by performing a simple differentiation:

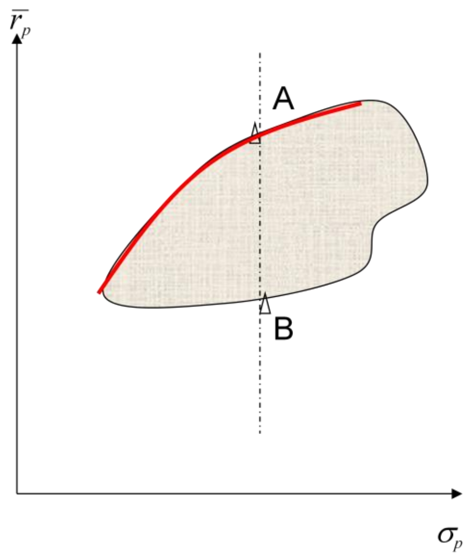
$$\sigma_P^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho w_1 w_2 \sigma_1\sigma_2$$

$$\frac{\partial(\sigma_P^2)}{\partial w_1} = 2w_1\sigma_1^2 - 2(1-w_1)\sigma_2^2 + 2\rho(1-2w_1)\sigma_1\sigma_2 = 0$$

$$\Rightarrow w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

In the special case when $\rho = 0$, this expression is reduced to $w_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.

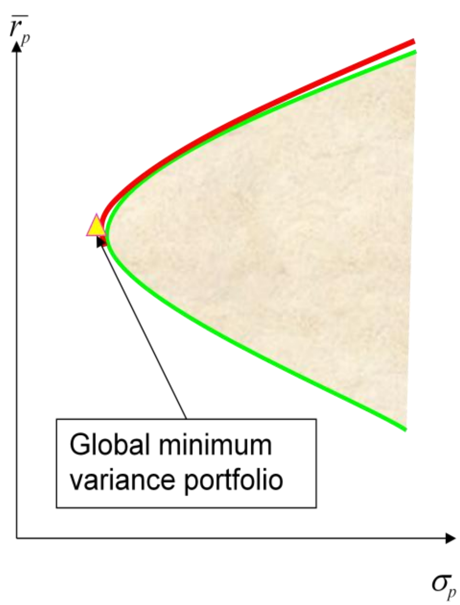
The feasible set for many assets



Using the equations given earlier, we can establish portfolios by combining assets in different proportions. In the diagram on the left, only portfolios within the shaded area are possible; those outside the area are not achievable with the given weight constraints (i.e. $\sum_{i=1}^{N} w_i = 1$).

Within the shaded region, it is obvious that Portfolio A "dominates" portfolio B (having the same risk but higher return), therefore we should always select A. Using the same logic, the portfolios lying on the red line are superior to those within the shaded region. The red line is known as the **efficient frontier** of risky assets.

A similar concept is the **minimum variance frontier**, which is indicated by the green line in the next diagram: at each $\bar{r}_p$, the portfolio on the line is the portfolio with the smallest σ_p. This line is obtained by scanning the portfolios across the shaded region horizontally – the minimum variance portfolio at each given return is the one at the leftmost edge of the region.



The **global minimum variance portfolio** is the portfolio with the smallest σ_p – note that this may not always be the "best" portfolio for an investor. It depends on the risk tolerance of the investor (more about this later).

The main exercise in step 1 is to *find the efficient frontier*. It is an example of a constraint optimization problem. Recall that the portfolio variance is

$$\sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} w_i w_j \sigma_{ij}$$

At each given $\overline{r_p}$, we can minimize σ_p subject to the following constraints

$$\overline{r_p} = \sum_{i=1}^{N} w_i \overline{r_i}, \ \sum_{i=1}^{N} w_i = 1$$

Each point on the efficient frontier represents a particular set of weights. We could solve the above set of equations for w_i, and apply this procedure repeatedly for different $\overline{r_p}$ until the whole efficient frontier is identified.

A useful technique for solving optimization problems is the application of *Lagrange multipliers*.[1] Let's say we want to maximize the function f of N variables x_i subject to two constraints $g(x)$ and $h(x)$, i.e. maximize $f(x_1, x_2, \ldots, x_N)$ where $g(x_1, x_2, \ldots, x_N)=0$ and $h(x_1, x_2, \ldots, x_N)=0$. Introduce the Lagrangian

$$L = f(\boldsymbol{x}) - \lambda_1 g(\boldsymbol{x}) - \lambda_2 h(\boldsymbol{x})$$

We can find the partial derivative of L w.r.t. each x_i and set it to 0 (N equations). Together with the two original constraints, we have $N+2$ equations and $N+2$ unknowns (N weights, λ_1, λ_2) which can be solved as a set of simultaneous equations. λ_1 and λ_2 are known as the Lagrange multipliers.

Formulating the current problem, instead of minimizing σ_P^2, we minimize $\frac{1}{2}\sigma_P^2$ (which would have no impact to the final answer). The Lagrangian is formed by

$$L = \frac{1}{2}\left(\sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} w_i w_j \sigma_{ij} \right) - \lambda_1 \left(\sum_{i=1}^{N} w_i \overline{r_i} - \overline{r_P} \right) - \lambda_2 \left(\sum_{i=1}^{N} w_i - 1 \right)$$

Differentiating L w.r.t to each w_i: $\frac{\partial L}{\partial w_i} = 0, \ i = 1, \cdots, N$

and together with the two original constraints we can set up $N+2$ equations. The Lagrange multipliers and the portfolio weights would satisfy the following:

$$\text{Condition 1}: w_i \sigma_i^2 + \sum_{j=1}^{N} \sigma_{ij} w_j - \lambda_1 \overline{r_i} - \lambda_2 = 0, \ i = 1, \cdots, N$$

$$\text{Condition 2}: \sum_{i=1}^{N} w_i \overline{r_i} = \overline{r_P}$$

$$\text{Condition 3}: \sum_{i=1}^{N} w_i = 1$$

[1] The following sections come from Luenberger(1998), pp.158-165.

An example with 3 assets

Assume the following parameters

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1;\ \overline{r_1} = 1, \overline{r_2} = 2, \overline{r_3} = 3$$

$$\rho_{12} = \rho_{23} = \rho_{13} = 0, \text{or } \sigma_{12} = \sigma_{23} = \sigma_{13} = 0$$

The set of equations is

$$\begin{aligned} w_1 - \lambda_1 - \lambda_2 &= 0 \\ w_2 - 2\lambda_1 - \lambda_2 &= 0 \\ w_3 - 3\lambda_1 - \lambda_2 &= 0 \\ w_1 + 2w_2 + 3w_3 &= \overline{r_P} \\ w_1 + w_2 + w_3 &= 1 \end{aligned}$$

The solution is:

$$\lambda_1 = \left(\overline{r_P}/2\right) - 1,\ \lambda_2 = 2\frac{1}{3} - \overline{r_P}$$

$$w_1 = \frac{4}{3} - \left(\overline{r_P}/2\right),\ w_2 = \frac{1}{3},\ w_3 = \left(\overline{r_P}/2\right) - \frac{2}{3}$$

$$\sigma_P = \sqrt{\frac{7}{3} - 2\overline{r_P} + \frac{\overline{r_P}^2}{2}}$$

In many cases, a short position in an asset is not allowed (i.e. cannot have a weight w_i which is smaller than 0). We can add the following constraint to the optimization problem:

$$w_i \geq 0 \text{ for all } i = 1 \cdots N$$

However, this problem cannot be reduced to a set of linear equations, given that the objective function includes quadratic terms and the constraints are linear equalities and inequalities. Note that when short sale is allowed, most of the w_i will have non-zero values, whereas when short sale is not allowed, many weights are equal to 0.

Using the previous parameters in the example with 3 assets, the solution has the following form:

$1 \leq \overline{r} \leq \frac{4}{3}$	$\frac{4}{3} \leq \overline{r} \leq \frac{8}{3}$	$\frac{8}{3} \leq \overline{r} \leq 3$
$w_1 = 2 - \overline{r}$	$\frac{4}{3} - \frac{\overline{r}}{2}$	0
$w_2 = \overline{r} - 1$	$\frac{1}{3}$	$3 - \overline{r}$
$w_3 = 0$	$\frac{\overline{r}}{2} - \frac{2}{3}$	$\overline{r} - 2$
$\sigma = \sqrt{2\overline{r}^2 - 6\overline{r} + 5}$	$\sqrt{\frac{2}{3} - 2\overline{r}_1 + \frac{\overline{r}^2}{2}}$	$\sqrt{2\overline{r}^2 - 10\overline{r} + 13}$.

Intuitively, the solution is obtained as follows. If short position is not allowed (i.e. each weight is 0 or above), the overall return must lie between 1 (when 100% is invested in asset 1) and 3 (when 100% is invested in asset 3), i.e.:

$$1 \le \overline{r_P} \le 3 \text{ where } \overline{r_P} = w_1 + 2w_2 + 3w_3$$

From the previous solution, $w_3 = \left(\overline{r_P}/2\right) - 2/3$. When $1 \le \overline{r_P} \le 4/3$, w_3 becomes negative; if a short position is not allowed, it would be set to 0. We can then solve w_1 and w_2 by

$$w_1 + 2w_2 = \overline{r_P},\ w_1 + w_2 = 1$$

and obtain $w_1 = 2 - \overline{r_P},\ w_2 = \overline{r_P} - 1$.

A similar argument applies for w_1, which would be set to 0 when $8/3 \le \overline{r_P} \le 3$. When $4/3 \le \overline{r_P} \le 8/3$, we have the same solution as the case without the short sell constraint, as all the weights would be within the range $0 \le w_i \le 1$.

The method described above is the common way of stating the optimization problem, which is based on the risk minimization formulation given a certain level of return. Alternatively, fund managers often find that they are limited to certain risk constraints, and want to find an optimum portfolio to generate the highest return with the given risk. The formulation would then become:

$$\text{Maximize } \overline{r_P} = \sum_{i=1}^{N} w_i \overline{r_i}$$

$$\text{Subject to } \sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \ne j}}^{N} w_i w_j \sigma_{ij}$$

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

In the final part of this section, we look at a surprising application of the portfolio theory. In the computation of the efficient frontier, we are supposed to find the solution from the optimization process by setting different values of the expected return. However, the mean-variance efficient set has an important property, known as the **two-fund theorem**:

> *Two efficient funds (portfolios) can be established so that any efficient portfolio can be duplicated, in terms of mean and variance, as a combination of these two. In other words, all investors seeking efficient portfolios need only invest in combinations of these two funds.* (Luenberger (1998), p.163)

If this is correct, the dramatic implication is that only two (efficient) mutual funds are needed to provide investment service for everyone, and no one needs to buy stocks separately! Of course, the real world does not work this way, because the assumptions used in deriving this result (similar to those in deriving the MPT) may not always hold.

A schematic proof of the result can proceed as follows. Assume that there are two known solutions lying on the minimum variance frontier,

$$w^1 = \left(w_1^1, w_2^1, \cdots, w_n^1\right), \lambda_1^1, \lambda_2^1$$
$$w^2 = \left(w_1^2, w_2^2, \cdots, w_n^2\right), \lambda_1^2, \lambda_2^2$$

which correspond to expected rates of return $\overline{r_P^1}, \overline{r_P^2}$. We form a portfolio by assigning a weight of α to w^1 and $(1-\alpha)$ to w^2 to form a new weight vector, i.e. $\alpha w^1 + (1-\alpha)w^2$. The two-fund theorem is proved if this set of weights give rise to a portfolio which also lies on the minimum variance frontier, i.e. this new portfolio also satisfies the three conditions given on p.11.

Condition 1 is satisfied because if w^1 and w^2 both make the left hand side of that equation equal to zero, it also applies to any linear combination of w^1 and w^2. The new Lagrangian multipliers are given by: $\lambda_1 = \alpha\lambda_1^1 + (1-\alpha)\lambda_1^2 ; \lambda_2 = \alpha\lambda_2^1 + (1-\alpha)\lambda_2^2$.

Condition 2 is satisfied by noting that

$$\sum_{i=1}^{N} \left[\alpha w_i^1 + (1-\alpha) w_i^2\right] \overline{r_i} = \alpha \sum_{i=1}^{N} w_i^1 \overline{r_i} + (1-\alpha) \sum_{i=1}^{N} w_i^2 \overline{r_i}$$
$$= \alpha \overline{r_P^1} + (1-\alpha)\overline{r_P^2}$$

which just indicates that the new portfolio would have an expected return equal to the weighted sum of the returns of the component portfolios.

Condition 3 is satisfied because

$$\sum_{i=1}^{N} \left[\alpha w_i^1 + (1-\alpha) w_i^2\right] = \alpha \sum_{i=1}^{N} w_i^1 + (1-\alpha) \sum_{i=1}^{N} w_i^2 = 1$$

As α varies over $-\infty < \alpha < \infty$, any portfolio on the minimum variance frontier can be obtained.

We now show a numerical example of the two fund theorem (from Luenberger (1998), p.164):

Security	covariance, σ_{ij}					mean, $\overline{r}_i$
1	2.30	0.93	0.62	0.74	−0.23	15.1
2	0.93	1.40	0.22	0.56	0.26	12.5
3	0.62	0.22	1.80	0.78	−0.27	14.7
4	0.74	0.56	0.78	3.40	−0.56	9.02
5	−0.23	0.26	−0.27	−0.56	2.60	17.68

With this set of input parameters, two efficient portfolios can be found by setting:

$$\lambda_1 = 0, \lambda_2 = 1: \quad \sum_{j=1}^{5} \sigma_{ij} v_j^1 = 1$$

$$\lambda_1 = 1, \lambda_2 = 0: \quad \sum_{j=1}^{5} \sigma_{ij} v_j^2 = \overline{r_i}$$

security	v^1	v^2	w_g	w_d
1	0.141	3.652	0.088	0.158
2	0.401	3.583	0.251	0.155
3	0.452	7.284	0.282	0.314
4	0.166	0.874	0.104	0.038
5	0.440	7.706	0.275	0.334
mean			14.413	15.202
variance			0.625	0.659
standard deviation			0.791	0.812

In the solution obtained above, v^1 and v^2 do not have weights that sum to 1, therefore the final solution w_g and w_d are obtained by normalizing the weights to 1

$$w_i^g = v_i^1 / \sum_{j=1}^{5} v_j^1,\ \ w_i^d = v_i^2 / \sum_{j=1}^{5} v_j^2,\ \ i = 1 \text{ to } 5$$

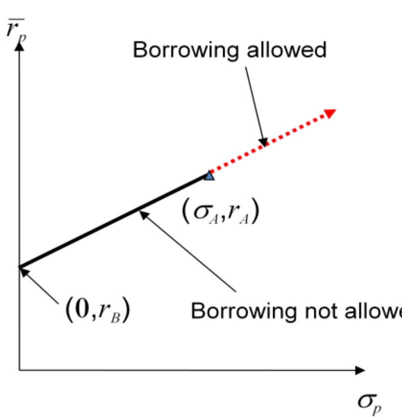
2.3.3. Step 2: Allocation between risky and risk free assets

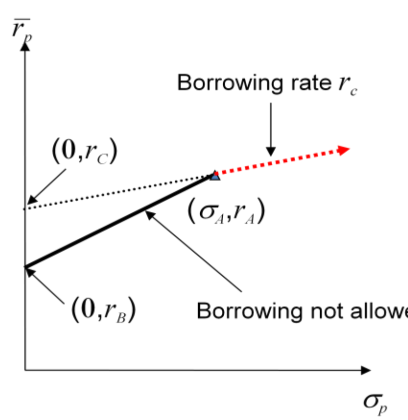
We have already found the efficient frontier of risky assets, but there are many portfolios on the efficient frontier. Which portfolio should we choose to be combined with a risk free asset, in order to find a portfolio with the "best" risk-return characteristics?

We can make use of the characteristics of a portfolio with risk free and risky assets. Say A is the risky portfolio and B is a risk free asset. By definition σ_B=0, so that

$$\sigma_p = w_A\sigma_A + (1-w_A)\sigma_B = w_A\sigma_A$$
$$\overline{r_p} = w_A\overline{r_A} + (1-w_A)\overline{r_B}$$
$$= \overline{r_B} + \frac{(\overline{r_A} - \overline{r_B})}{\sigma_A}\sigma_p$$

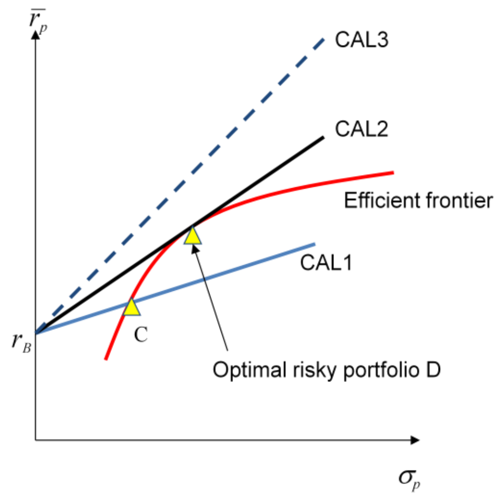
This is a straight line when we plot r_p against σ_p, and it is known as the **Capital Allocation Line (CAL)**. The points on the line represent different weights being assigned to the risky portfolio and the risk free asset. The line passes through the points (0, $\overline{r_B}$) and (σ_A, $\overline{r_A}$), and the slope is called the ***reward-to-variability ratio***: for each unit increase of σ_P, this ratio shows the extra increase in return.



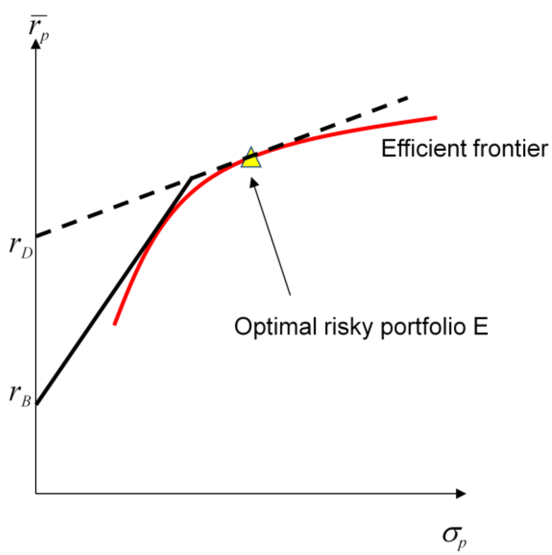


The points $(0, \overline{r}_B)$ and $(\sigma_A, \overline{r}_A)$ represent the case when 100% of the investment is made in either B or A. It may be possible to invest more than 100% in one asset if borrowing is allowed. This is known as a leverage position, part of which is shown in the red dotted line above. In practice, an investor may face different borrowing and lending rates, thus the CAL is no longer a straight line but can have a kinky shape (as shown in the right diagram above).

Which risky portfolio should we choose?



Remember that the CAL represents the combinations of a risky portfolio and a risk free asset. We have also learned from the previous section that all risky portfolios on the efficient frontier are optimal combinations. In the diagram on the left, both portfolios C and D lie on the frontier. However, portfolio D, when combined with the risk free asset B, would produce the **steepest CAL**: CAL2 is a tangent to the efficient frontier. Therefore this is the optimal risky portfolio to be chosen from the efficient frontier. On the other hand, while CAL 3 is even steeper, it is not a valid CAL because no risky portfolio could achieve such a risk/return profile.



The case is slightly more complicated if borrowing is allowed because the CAL is not a straight line if the borrowing rate and lending rate is different. We would probably obtain a different optimal risky portfolio E. Remember this is the optimal portfolio that we should choose when we want to combine a risk free asset with a risky portfolio; if a risk free asset is not needed, other portfolios on the efficient frontier may be chosen.

Finding the optimal CAL is another optimization problem. Recall that the CAL is

$$\overline{r}_p = \overline{r}_B + \frac{(\overline{r}_A - \overline{r}_B)}{\sigma_A}\sigma_p = \overline{r}_B + \theta\sigma_p$$

We just need to optimize the slope of the CAL, i.e. maximize θ

subject to the constraints $\overline{r_A} = \sum_{i=1}^{N} w_i \overline{r_i}, \ \sum_{i=1}^{N} w_i = 1$.

In the basic case when short sell is allowed with riskless borrowing and lending, no further constraint is added to the two conditions stated above. Instead of using Lagrangian multipliers, it can be turned into a simple maximization problem, noting that

$$\overline{r_B} = \sum_{i=1}^{N} w_i \overline{r_B}$$

$$\theta = \frac{\sum_{i=1}^{N} w_i \left(\overline{r_i} - \overline{r_B}\right)}{\left[\sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} w_i w_j \sigma_{ij}\right]^{1/2}}$$

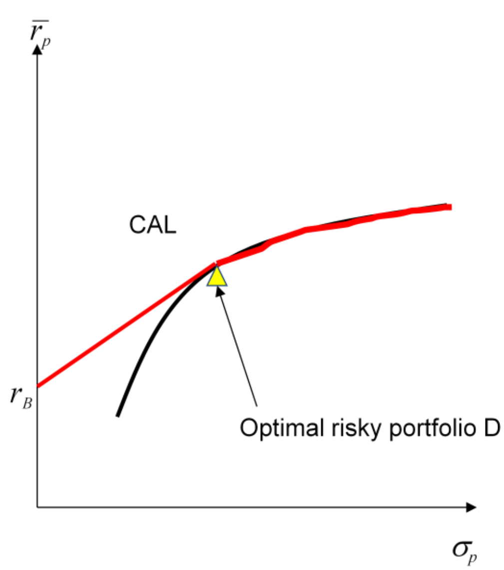
The solution can be found by setting the partial derivative w.r.t. each w_i to be zero, and each of the derivative can be written as

$$\frac{\partial \theta}{\partial w_1} = \frac{\partial \theta}{\partial w_2} = \cdots = \frac{\partial \theta}{\partial w_N} = 0$$

$$\begin{aligned} \frac{\partial \theta}{\partial w_i} &= -(\lambda w_1 \sigma_{1i} + \lambda w_2 \sigma_{2i} + \cdots + \lambda w_i \sigma_i^2 + \cdots \lambda w_N \sigma_{Ni}) + \overline{r_A} - \overline{r_B} \\ &= 0 \end{aligned}$$

$$\lambda = \left(\overline{r_A} - \overline{r_B}\right) / \sigma_P^2$$

where w_i can then be obtained from the solution of a set of simultaneous equations



The case when short sell is allowed with riskless lending only (i.e. no borrowing) is more complex. It means that the maximum proportion that can be invested in the risky portfolio is 100% of the initial wealth. The efficient frontier would be identical to the CAL when $\sigma < \sigma_D$, but will overlap the efficient frontier of the risky portfolio when $\sigma > \sigma_D$

When short sell is not allowed but riskless borrowing and lending is possible, the optimization problem involves one extra constraint compared to the basic case, i.e.

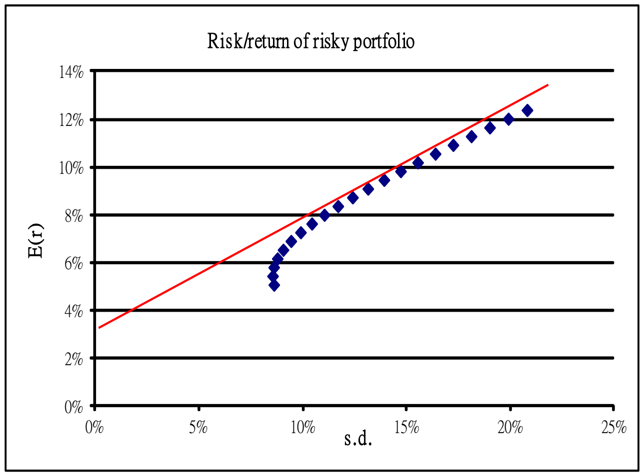
$$w_i \geq 0 \text{ for all } i$$

Again, this becomes a quadratic programming problem as the objective function to be optimization involves quadratic terms, and it cannot be solved by simple differentiation.

Efficient frontier example (with two assets)

sigma 1	20.8%	return 1	12.4%
sigma 2	8.6%	return 2	5.1%
rho	0.30	rate	3.0%

	w1	w2	sigma	E(r)	slope
1	0	1	8.60%	5.10%	0.2442
2	0.05	0.95	8.54%	5.47%	0.2886
3	0.1	0.9	8.60%	5.83%	0.3292
4	0.15	0.85	8.77%	6.20%	0.3644
5	0.2	0.8	9.05%	6.56%	0.3936
6	0.25	0.75	9.42%	6.93%	0.4166
7	0.3	0.7	9.89%	7.29%	0.4340
8	0.35	0.65	10.42%	7.66%	0.4466
9	0.4	0.6	11.03%	8.02%	0.4552
10	0.45	0.55	11.69%	8.39%	0.4608
11	0.5	0.5	12.39%	8.75%	0.4641
12	0.55	0.45	13.13%	9.12%	0.4657
13	0.6	0.4	13.90%	9.48%	0.4660
14	0.65	0.35	14.71%	9.85%	0.4655
15	0.7	0.3	15.53%	10.21%	0.4643
16	0.75	0.25	16.37%	10.58%	0.4626
17	0.8	0.2	17.23%	10.94%	0.4607
18	0.85	0.15	18.11%	11.31%	0.4586
19	0.9	0.1	19.00%	11.67%	0.4564
20	0.95	0.05	19.89%	12.04%	0.4542
21	1	0	20.80%	12.40%	0.4519



Finding the CAL

Use Solver, maximize cell M19 (slope of the CAL) by changing cell I19

w1	w2	sigma	E(r)	slope
0.59042	0.40958	13.75%	9.41%	0.466

Minimum variance portfolio

Use Solver, minimize cell K24 (sigma) by changing cell I24

w1	w2	sigma	E(r)
0.05083	0.94917	8.54%	5.47%

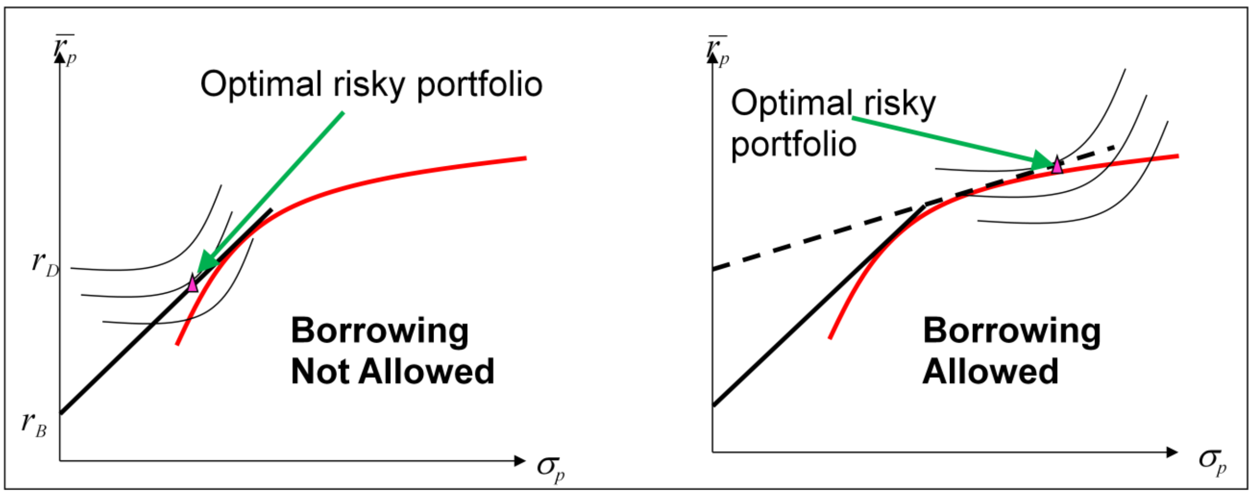
Final remark

Depending on the risky assets available, the optimal portfolio on the efficient frontier can be chosen via the methods described earlier, and the weights of each asset can then be obtained accordingly. On the other hand, one may pursue a *passive* strategy, where there is no need to engage in any security analysis. The portfolio used is a well-diversified portfolio, e.g. a portfolio that mirrors the S&P 500 index – this is sometimes known as the "market portfolio". The CAL that combines a risk-free asset (e.g. US 1-month T-bills) with this passive portfolio is called the **Capital Market Line (CML)**. We will discuss this concept in more detail when we look at the Capital Asset Pricing Model (CAPM) in the next topic.

2.3.4. Step 3: combining with the investor's preference

2.3.4.1. Finding the optimal portfolio

From steps 1 and 2, we have found the optimal CAL, which is the set of the most efficient portfolios that combines a risky portfolio and a risk free asset. An important observation is that this CAL is applicable to any investor ("**the separation property**") – it does not require any input from the investor's own preference in order to identify the CAL. In the final step of the portfolio selection process, we could combine the results with the preference of the investor, by plotting the results above with the set of indifference curves

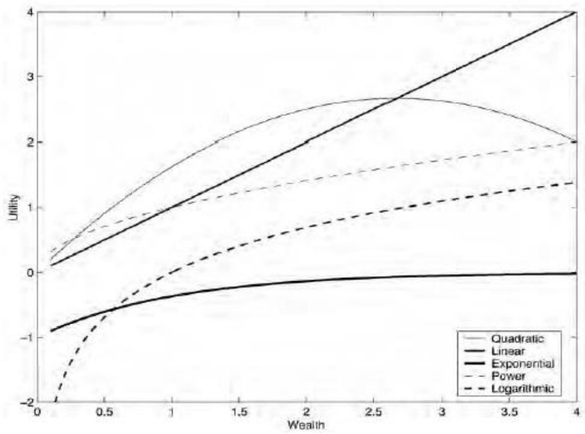


In the left diagram, a set of utility curves are drawn together with the efficient frontier and the CAL. The optimal portfolio for the investor is obtained by identifying the utility curve with the highest utility that is tangential to the CAL. In other words, this portfolio is one which can attain the highest utility for the investor. It is not possible to find a portfolio that lies on the CAL which has a higher utility for this particular investor, according to the utility curves.

If borrowing is allowed (i.e. can invest more than 100% of the initial capital in the risk free asset or the risky portfolio), the optimal portfolio may appear in a different portion of the CAL, as illustrated in the right diagram.

2.3.4.2. Further comments on utility functions

We have introduced the concept of utility function in section 2.2.2. These functions are useful in describing how entities make decisions when faced with a set of choices. The formal theory was first proposed by von Neumann and Morgenstern in 1944 (before the development of the MPT). Simply speaking, we want to assign a (numeric) value to all possible choices faced by the entity. Subsequently we can maximize the utility based on some given constraints.



A few examples of popular utility functions are:

Linear utility $u(x) = a + bx$

Exponential utility $u(x) = 1 - e^{-ax}, x > 0$

Power utility $u(x) = x^{\alpha}, 0 < \alpha < 1$

Logarithmic utility $u(x) = a \ln x + b,\ a > 0$

In Topic 1, we have used the St Petersburg paradox to illustrate the fact that people may not make an investment decision based on expected return only. One way of resolving the problem is to make use of utility functions. It is often noted that investors do not assign the same value per dollar to all payoffs – increasing wealth would become marginally less satisfying, or in other words, the greater their wealth, the less their "appreciation" for each extra dollar. One way to express this "risk aversion" behavior is through a logarithmic utility function:

$$U[R(n)] = ln[R(n)]$$

where $R(n)$ is an expected return or a wealth level, and $U[R(n)]$ is the corresponding utility value. In the St Petersburg paradox, instead of calculating the expected payoff as :

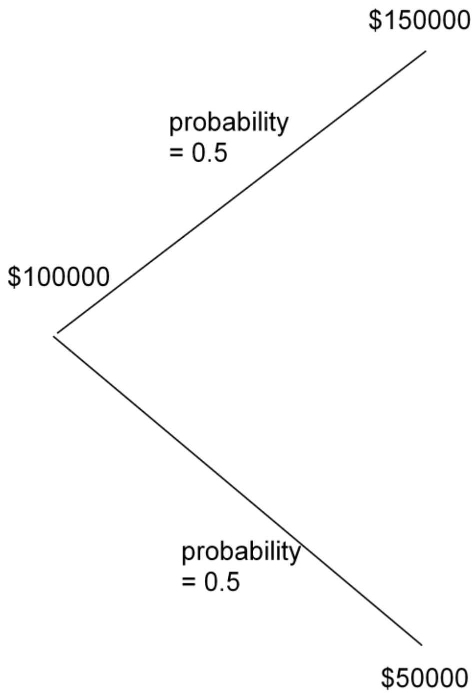
$$\text{Expected payoff } = \sum probability \times payoff = \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \cdots = \infty$$

we calculate the expected utility value instead. If we use a logarithmic utility function:

$$\text{Expected utility} = \sum probability \times \ln(payoff) = \left[\tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16} + \cdots\right] \ln 2 = 0.693$$

The certain wealth level necessary to yield this utility value is \$2, as $\ln(2)=0.693$. This value is known as the certainty equivalent, i.e. the amount of money which can generate the given utility value based on the pre-defined utility function.

We would look at another example:



Expected value

= 0.5x150000+0.5x50000

= 100000

Expected utility

=0.5xln(150000)+0.5xln(50000)

= 11.37

Certainty equivalent:

ln(CE)=11.37=> CE = 86681.87

i.e. this is the amount which has the same utility as if the game is played. In other words, if we have $86681.87, the utility of this wealth amount is equivalent to the utility level which can be obtained by playing the game. However, if this game is rejected, the utility is ln(100000)=11.51. In other words, there is no gain in utility by playing the game.

A quadratic utility function can also be used to explain some of the concepts discussed earlier. It can be defined as:

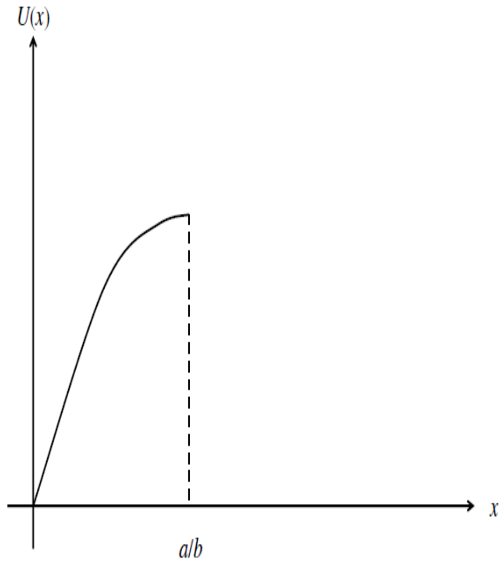
$$u(x) = ax - \frac{b}{2}x^2$$

As we normally assume that utility is an increasing function (i.e. more wealth would lead to higher utility), this function is only applicable to the range $x \le a/b$, as the function is only increasing when this relationship is satisfied. Within this range, the quadratic utility can act as a good approximation to other types of utility functions.

We can observe two properties from this function. Firstly, a risk-averse investor can be represented by setting $b>0$, as the function is strictly concave when the condition is satisfied. Secondly, while the mean-variance analysis formulates the investment problem as the maximization of expected return with a given level of risk, it can be shown that the solution is also given by maximizing expected utility based on the quadratic utility function.

Assume the random wealth value of a portfolio is *y.* Under the expected utility criterion, we calculate

$$\begin{aligned} E[u(y)] &= E\left[ay - \frac{b}{2}y^2\right] \\ &= aE[y] - \frac{b}{2}E[y^2] \\ &= aE[y] - \frac{b}{2}(E[y])^2 - \frac{b}{2}\mathrm{var}(y) \end{aligned}$$



From this simple relation, we can see that the expected utility is dependent only on the mean and variance of y.

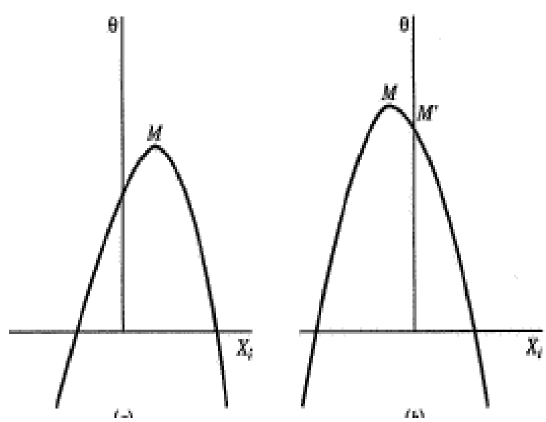
- For a given value of $E[y]$, maximizing $E[u(y)]$ = minimizing $\mathrm{var}(y)$
- For a given $\mathrm{var}(y)$, maximizing $E[u(y)]$ = maximizing $E[y]$

(Note that these results are only applicable in the range $0 \le x \le a/b$)
Therefore, the portfolio optimization problem can be solved if an algorithm is found that can maximize the utility of the portfolio.

2.3.5. Conclusion: what have we achieved so far?

Step 1: Find the efficient frontier of risky assets (e.g. combining stocks and bonds). All portfolios that lie on this frontier are equivalent in the sense that they are optimal combinations which can give rise to the best risk/return characteristics.
Step 2: Find the CAL which has the highest slope tangential to the efficient frontier; hence we can identify ONE optimal risky portfolio (i.e. we know the weightings of the assets of the optimal risky portfolio).
Steps 1 and 2 should be applicable to EVERYONE, i.e. everyone should have the same risky portfolio.
Step 3: Find the highest utility curve that is tangential to the CAL; hence we find the optimal weightings of the risk free asset and the risky portfolio (i.e. identify x such that x% should be invested in the risk free asset and $(1-x)$% should be invested in the risky portfolio).

Appendix: Kuhn-Tucker conditions



In quadratic programming, it is possible that the unconstrained maximum occurs at a point which is limited by some additional constraints, i.e. the solution with constraints may not be the global maximum E.g. in the diagram on the right, the maximum feasible value of θ when there is a constraint $X_i > 0$ occurs at M', not M.

In this case, note that $d\theta / dX_i \le 0$, or we can write $d\theta / dX_i + U_i = 0$.
The optimum point occurs with these conditions

$$X_i > 0, d\theta / dX_i = 0 \Rightarrow U_i = 0$$
$$X_i = 0, d\theta / dX_i < 0 \Rightarrow U_i > 0$$

The four Kuhn-Tucker conditions can be summarized as

$$d\theta / dX_i + U_i = 0$$
$$X_i U_i = 0$$
$$X_i \ge 0$$
$$U_i \ge 0$$

These conditions can then be used as part of the formulation of the optimization problem.

Additional References

Zvi Bodie, Alex Kane and Alan Marcus, *Investments*, 9th edition, McGraw Hill, 2010.
Edwin Elton, Martin Gruber, Stephen Brown and William Goetzmann, *Modern Portfolio Theory and Investment Analysis*, 7th edition, Wiley, 2007.
David Luenberger, *Investment Science*, Oxford University Press, 1998.
Frank Fabozzi, Petter Kolm, Dessislava Pachamanova, Sergio Focardi, *Robust Portfolio Optimization and Management*, Wiley, 2007.

2.4. Implementation Issues
2.4.1. Reducing the complexity of the problem: index models
2.4.1.1. Overview

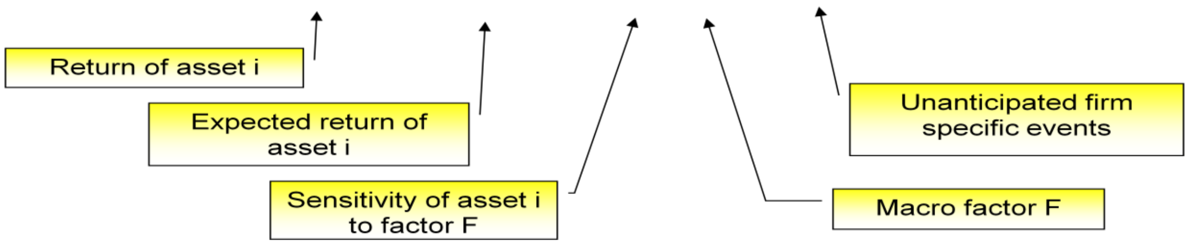
In the previous sections, we have described the basics of the portfolio selection process via a brute force optimization scheme. In practice the method may not become workable. A main issue is the availability of good quality input data. One of the required parameters is the correlation between each pair of assets. If N assets are available for investment, there would be $N(N-1)/2$ terms in the correlation matrix. Typically N can be 1000 or more and it would be impossible to obtain reliable data (with 1000x999/2 = 499500 terms), and it can easily lead to inconsistent results. If the correlation matrix is not positive definite, a negative portfolio variance could be obtained, which is an error by definition (because variance is the square of standard deviation, which must be positive).
For example (taken from Bodie, Kane and Marcus (2005))

			Correlation matrix		
Asset	s.d.	Weight	A	B	C
A	20%	-1	1.00	0.90	0.90
B	20%	1	0.90	1.00	*0.00*
C	20%	1	0.90	*0.00*	1.00

The correlations between (A, B) and (A, C) are both very high (=0.90), but the correlation between (B, C) is given as 0 above. Of course theoretically this cannot happen, but it can be possible due to data error. The portfolio variance thus calculated will be –0.024, which would definitely cause an error in the calculation routines elsewhere.

To simplify the problem, we assume that the co-movements between stocks are due to one or more common influences or indices (known as the *single-* or *multi-index model*). It is often observed that covariance between stocks tend to be positive. Intuitively this could be explained because stocks follow the same economic factors. Assume that we combine all influences to a single factor F, so that

$$r_i = E(r_i) + \beta_i F + e_i$$



We could use a broad based market capitalization weighted stock index, e.g. the S&P 500, as a proxy for the common macro factor. A popular model is to write the return of the stock as:

$$R_i = \alpha_i + \beta_i R_M + e_i$$

where R_M is the return of the market index, α_i is the stock specific expected return, e_i is the stock specific unexpected return, and β_i is the sensitivity factor to the market index

2.4.1.2. Single index model

By construction, we have the following properties:

- Mean of e_i = $E(e_i)$ = 0 {this is a definition of "unexpected return"}
- Assets are unrelated to each other: $E(e_i e_j)$ = 0 {all the dependencies should be captured by the correlation between the asset and the market index}
- Stock specific event unrelated to the market factor:

$$\text{cov}(e_i R_M) = 0 \Rightarrow E\left[(e_i - 0)(R_M - \overline{R_M})\right] = 0$$

- The mean return of the asset can be calculated:

$$\begin{aligned} E(R_i) &= E[\alpha_i + \beta_i R_M + e_i] \\ &= E(\alpha_i) + E(\beta_i R_M) + E(e_i) \\ &= \alpha_i + \beta_i \overline{R_M} \end{aligned}$$

- The variance of an asset's return

$$\begin{aligned} \sigma_i^2 &= E\left(R_i - \overline{R_i}\right)^2 \\ &= E\left[\beta_i (R_M - \overline{R_M}) + e_i\right]^2 \\ &= \beta_i^2 E(R_M - \overline{R_M})^2 + E(e_i)^2 \\ &= \beta_i^2 \sigma_M^2 + \sigma_{e_i}^2 \end{aligned}$$

Thus the variance has two parts: the market-related risk (through $\beta_i\sigma_M$), and the unique risk of the stock (through σ_{ei})

- The covariance between two assets is

$$\begin{aligned}\sigma_{ij} &= E\left[(R_i-\overline{R_i})(R_j-\overline{R_j})\right]\\ &= E\left[\left(\beta_i(R_M-\overline{R_M})+e_i\right)\left(\beta_j(R_M-\overline{R_M})+e_j\right)\right]\\ &= \beta_i\beta_j\sigma_M^2\end{aligned}$$

i.e. **we only need to know the relationship between each asset and the index in order to calculate the covariance between any two assets.** This is the key advantage of the index model, because we no longer need the correlation between each pair of assets. However, the underlying assumption is that assets move as a common response to market movements.

We can also write down the portfolio mean and variance as given by the single index model:

$$\overline{R_P} = \sum_{i=1}^{N} X_i\overline{R_i} = \sum_{i=1}^{N} X_i\alpha_i + \sum_{i=1}^{N} X_i\beta_i\overline{R_M}$$

$$\sigma_P^2 = \sum_{i=1}^{N} X_i^2\beta_i^2\sigma_M^2 + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} X_iX_j\beta_i\beta_j\sigma_M^2 + \sum_{i=1}^{N} X_i^2\sigma_{e_i}^2$$

where X_i is the weight of asset i in the portfolio. The parameters required in their estimation are: α_i, β_i, $(\sigma_{ei})^2$ for each stock, $\overline{R}_M$ and $(\sigma_M)^2$ for the market factor. In particular, we note that there is no need to maintain the covariance matrix for all stocks.

It is illustrative to use the index model to derive some results related to diversification. Assume that we have a portfolio of N stocks, each with a weight = $1/N$. The portfolio return is:

$$\begin{aligned}R_P &= \frac{1}{N}\sum_{i=1}^{N}\left(\alpha_i+\beta_iR_M+e_i\right)\\ &= \frac{1}{N}\sum_{i=1}^{N}\alpha_i + \left(\frac{1}{N}\sum_{i=1}^{N}\beta_i\right)R_M + \frac{1}{N}\sum_{i=1}^{N}e_i\\ &= \alpha_P+\beta_PR_M+e_P\\ \overline{R_P} &= \alpha_P+\beta_P\overline{R_M}\end{aligned}$$

In this convention, α_P and β_P are the portfolio's α and β respectively. The portfolio variance is

$$\begin{aligned}\sigma_P^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} X_iX_j\beta_i\beta_j\sigma_M^2 + \sum_{i=1}^{N} X_i^2\sigma_{e_i}^2\\ &= \beta_P^2\sigma_M^2 + \sum_{i=1}^{N} X_i^2\sigma_{e_i}^2\\ &= \beta_P^2\sigma_M^2 + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sigma_{e_i}^2\end{aligned}$$

When N becomes larger the second term disappears, so that:

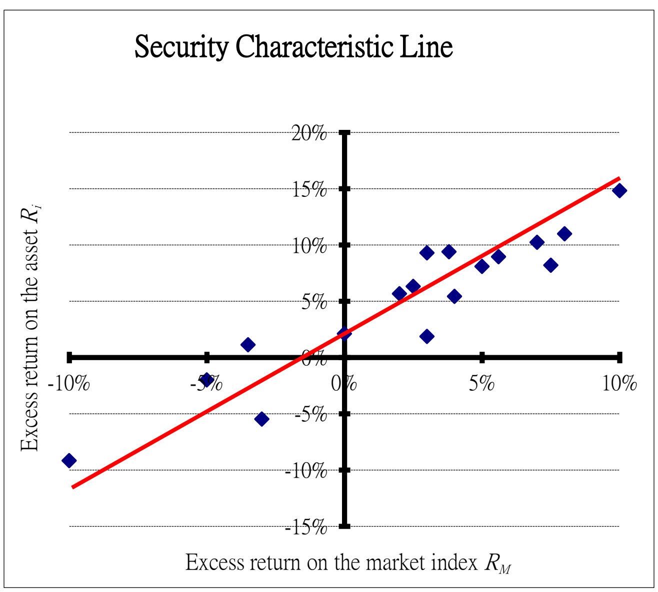
$$\sigma_P \approx \beta_P \sigma_M = \sigma_M \sum_{i=1}^{N} \frac{1}{N} \beta_i$$

As shown earlier, the contribution of an asset to the risk of the portfolio is through β_i, i.e. its correlation with the index, whereas its absolute risk is not relevant.

To summarize the **advantages and disadvantages of this model:**

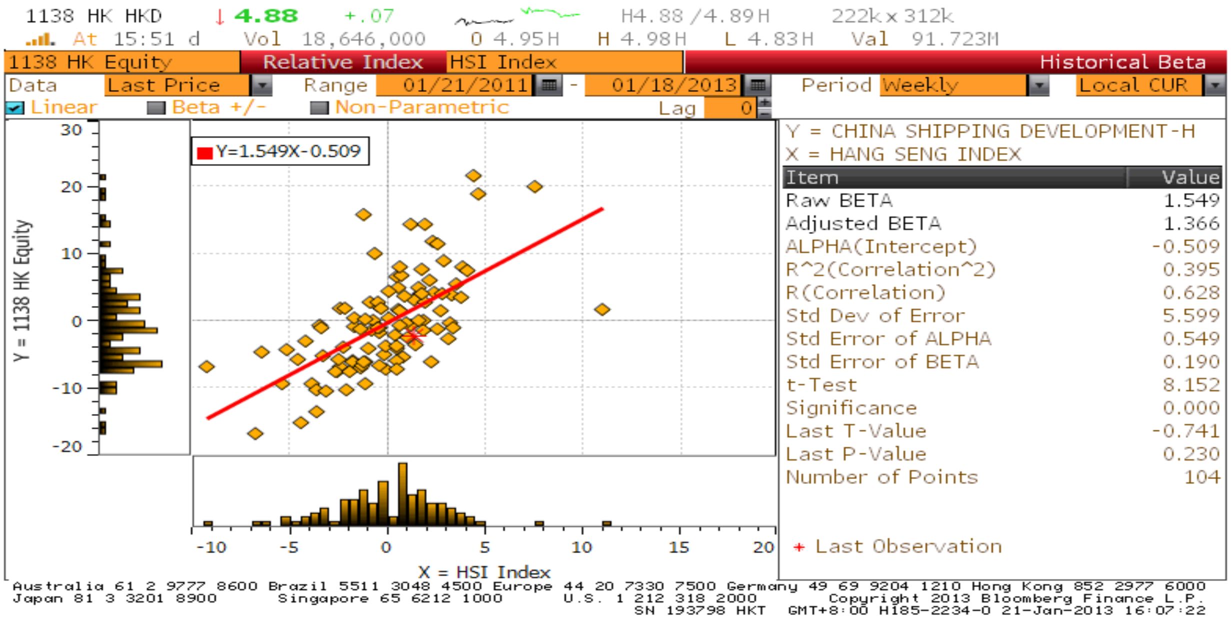
- For large universes of securities, much less data would need to be maintained.
- Adopting the model can affect who is going to perform security analysis. Analysts only need to specialize on one sector and work out a stock's correlation with a market index. For example, the banking analyst only needs to concentrate on analyzing the performance of banks, not the relationship between banks and all other sectors. This would streamline the tasks of the research department in a financial institution.
- Of course, in this much "simplified" world, some kinds of risks may be under-represented. A key omission is something like an industry specific event (not security specific or market specific, therefore not captured by either the macro factor F or the micro factor e). For example, the Chinese government may impose policies to encourage the development of more efficient cars. All car companies may be positively affected, and this kind of event cannot be captured by a single index model because it is neither market specific nor company specific.

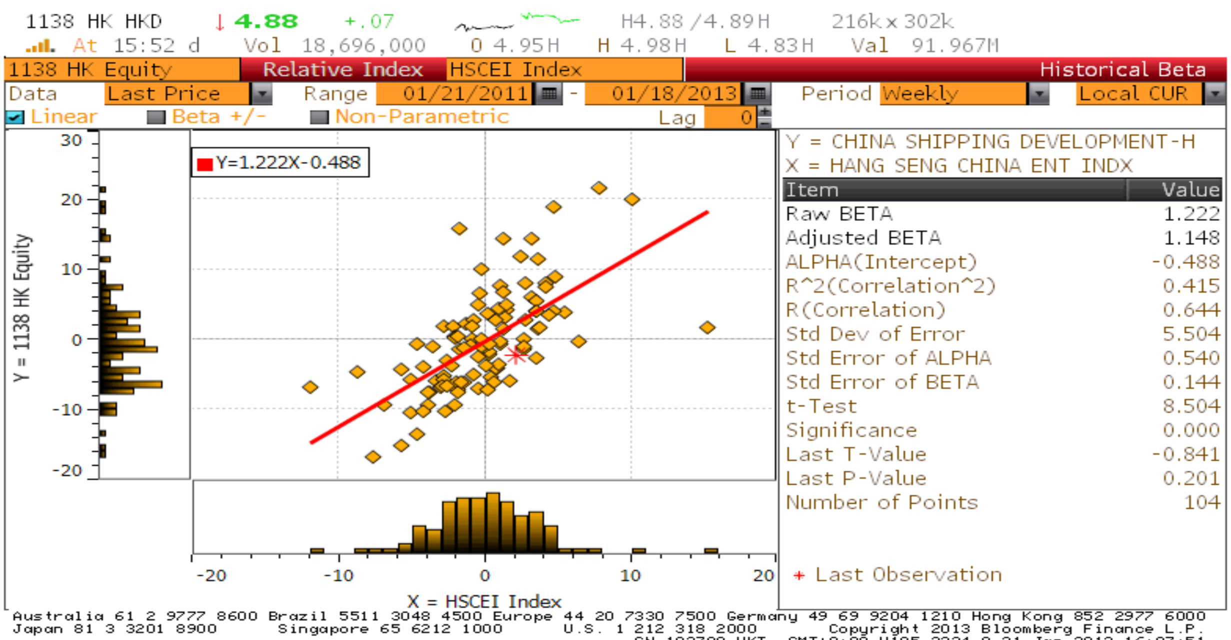
From the above, we can see that the key part of estimating the index model is the ability to obtain a reliable estimate of β. One way of achieving this is to plot a diagram of the return of asset R_i against the return of the market index R_M :



Historical data from different time periods are used and the results are plotted in a scattered diagram. For example, in one period, the return of the asset is 8% when the market index returns 5%. (5%, 8%) is a point in the diagram. After plotting all the data points, the next step is to find the "best-fit" straight line via linear regression. This line is known as the **Security Characteristic Line (SCL)** and the slope of this line is the best estimate for β.

Selecting different market indices can affect the overall results. For example:





Source: Bloomberg

China Shipping Development (stock code 1138) is a Chinese company listed in the Hong Kong stock market. While the most representative stock market index in Hong Kong is the Hang

Seng Index, it may not be appropriate for this company. A possibly better fit may be to use the Hang Seng China Enterprise Index (HSCEI) as the market factor in the index model instead.

2.4.1.3. Multifactor model

Instead of limiting to a single market factor, we could extend the model to include multiple factors. One form of the model is: $R_i = \alpha_i + \beta_{iM} R_{iM} + \beta_{i1} R_{i1} + \beta_{i2} R_{i2} + ... + e_i$
Examples of factors may include economic growth, business cycle, long and short term interest rates, inflation, industrial production, strength of US dollar and so on. For the model to work well, the factors should be independent from each other.

Empirical studies were carried out to test the performance of these models. Given that numerous possibilities exist, we could not draw a definitive conclusion if a model does not perform adequately. Some models give reasonable results, but many multifactor models do not perform as well as a single index model; this illustrates the difficulty in choosing appropriate factors (e.g. factors give rise to high explanatory power in one period may become unsuitable in the future). We will visit this topic again in a later section when we discuss the Arbitrage Pricing Theory.

2.4.2. Considerations in the selection of assets

A key prerequisite in the portfolio selection exercise is the determination of the available opportunity set. Security analysis can be important, in order to select individual stocks/bonds which have the best risk/return characteristics. Furthermore, accurate forecasts are essential because the expected returns and risks can have much impact to the optimization results. However, other selecting criteria may also be imposed. For example, some investors prefer to receive a certain income every year, and thus the stocks with high dividend yields may feature more prominently. Other factors may also affect which assets go into the available set, e.g. ethical consideration is now a popular restriction (some investors may not want to invest in tobacco companies or casinos), whether international stocks or sectors should be included or not, and so on.

The original constraints in Markowitz's formulation often only serve as the starting point. Other complex practical issues include transaction cost modelling, optimization across multiple client accounts, and the need to take tax consequences into considerations. Some of these can be formulated as constraints that can be added to the optimization problem.

2.4.2.1. Linear and quadratic constraints

These constraints take the form of linear and quadratic functions, where the final answers can come from a range of values. A simple example is the *long-only constraint*, where:

$$w_i \geq 0 \text{ for all } i = 1 \cdots N$$

i.e. the weights of all assets have to be positive, which means that short selling is not allowed.

Another kind of constraint relates to *turnover*. High turnover can often result in large transaction costs, and liquidity may also be an issue. To impose a constraint on turnover, it can take the form

$$|x_i| \leq \alpha U_i$$

where U_i is the average daily volume of a stock, x_i is the amount to be traded, and α is a percentage. A typical value of α is 5%, i.e. not more than 5% of the daily volume would be included in the portfolio.

A third kind of linear constraint is based on *holding*. A well-diversified portfolio should not have large concentration in any single asset, industry, sector, or country. Sometimes this constraint is imposed by the regulator. The constraint will take the form

$$L_i \le w_i \le U_i$$

where L_i and U_i are the lower and upper bounds of the holding. The following constraint can also be introduced

$$L_i \le \sum_{j \in I_i} w_j \le U_i$$

which limits the exposure to a particular sector or industry I_i.

A final example is the *benchmark exposure and tracking error constraint*. A fund manager may need to measure the performance against a benchmark, e.g. a stock index. A common constraint is to limit the deviations of the portfolio weights from the benchmark weights (w represents the % weight of an asset, and $w_{benchmark}$ is the corresponding weight in the index):

$$\left\| w - w_{benchmark} \right\| \le M$$

Another constraint is to limit the tracking error against the benchmark, e.g. by limiting the variance of the difference between the return of the portfolio and the benchmark

$$\mathrm{var}(R_p - R_{benchmark}) \le \sigma^2$$

One difficulty for imposing these constraints is that establishing a meaningful benchmark can be arbitrary. For example, one can argue that one does not want to deviate from the benchmark by more than 3% or 5% per stock, but whether these values would be good choices cannot be answered in a concrete manner.

2.4.2.2. Combinatorial constraint

Introduce the notation

$$\delta_i = \begin{cases} 1, \text{ if } w_i \ne 0 \\ 0, \text{ if } w_i = 0 \end{cases}$$

where w_i denotes the portfolio weight of the i^{th} asset. A cardinality constraint is one when a portfolio manager wants to restrict the number of assets in the portfolio, e.g. use a smaller number of assets to replicate a benchmark:

$$\sum_{i=1}^{N} \delta_i = K, \; K << N$$

N is the total number of assets in the benchmark. In order to solve the optimization problem with combinatorial constraints, more sophisticated algorithms are required.

2.4.2.3. Availability of a risk free asset

For the US market, investors usually consider the short dated US Treasury bills as risk free assets. This is only as approximation, but can be a good one as bills are short term instruments, therefore it can safely be assumed to be held to maturity with no price risk. However, other short term money market instruments can also used which may have slightly different characteristics.

2.4.3. Other practical considerations

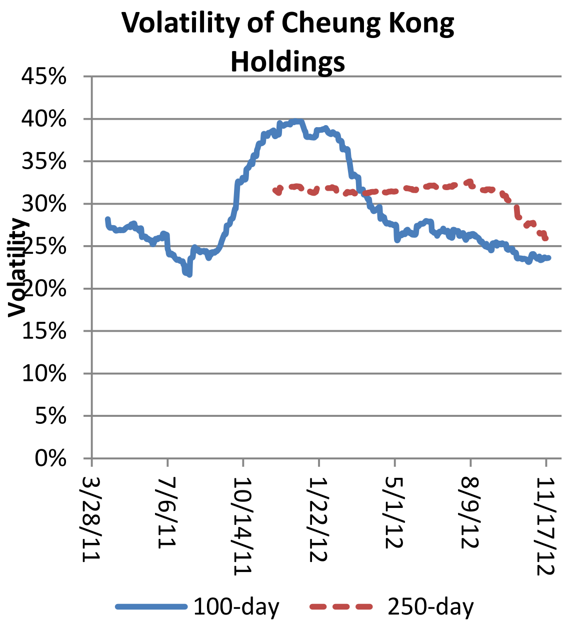
2.4.3.1. Relaxing the modeling assumptions

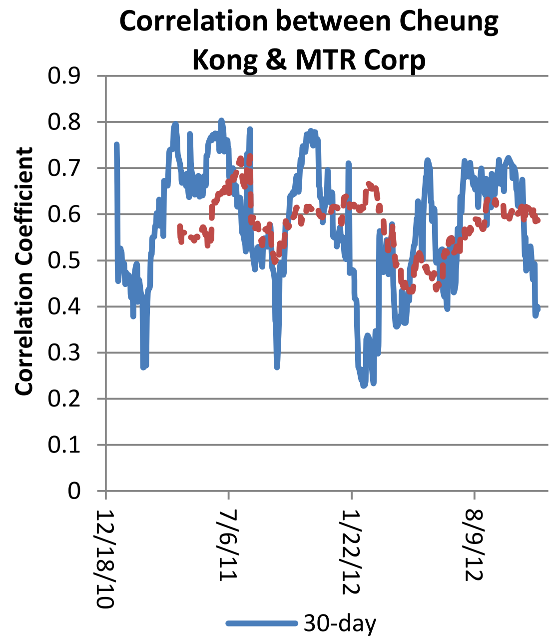
From the descriptions so far, the MPT relies on a large number of assumptions. Some of these assumptions have often been criticized as being unrealistic. Much research has been conducted in the past thirty years in order to assess the validity of the model in the presence of situations which would violate the assumptions. Some of these topics include:

- *The mean-variance criterion*
  While the mean-variance optimization is the cornerstone of MPT, Samuelson (1970) showed that, although the distribution of returns shows higher moments, disregarding these would not affect portfolio choices.
- *Quadratic utility functions*
  It is well known that although a popular form of the utility function is quadratic, other valid and more complex forms exist. In Markowitz's original work in 1959, it was already shown that quadratic utility functions are good local approximates of more complicated forms, hence the results derived from adopting quadratic utility functions should be acceptable.
- *Normal distribution of returns*
  Many studies found that asset returns would not follow a normal distribution, so that their characteristics cannot simply be summarized by expected return and variance. However, even though individual asset returns may not be normal, portfolio returns will resemble a normal distribution (from the Central Limit Theorem).

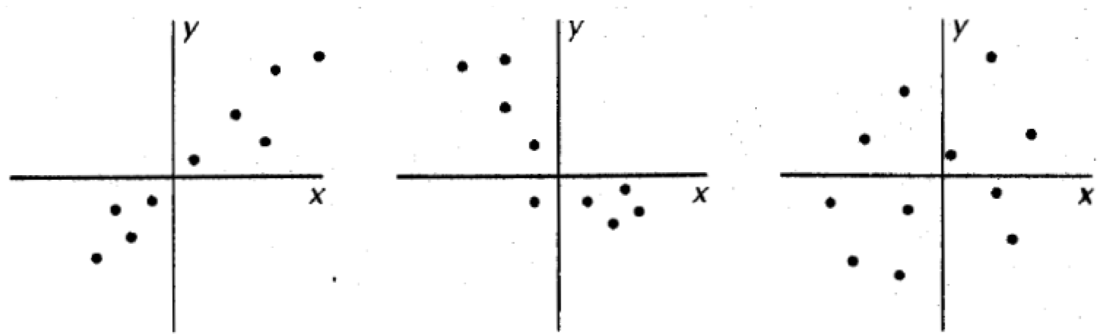
2.4.3.2. Variance and correlation

A more pressing problem is the difficulty in obtaining reliable estimates for variances / covariances. Firstly, financial asset prices are not static: variances often change due to change of circumstances (a phenomenon known as "heteroscedasticity"). If historical data is used in the calculation, it is difficult to decide how many data points should be used. As an example, in the left diagram below, using 100 data points and 250 data points can give very different answers. Secondly, correlation between assets can also change as time progresses. This can be caused by the different economic, social, and other operating conditions. Furthermore, it is observed that short term correlation often increases in turbulent periods – when there is a market downturn, prices of all kinds of assets would go down at the same time. Inherently, the limitation comes from the fact that prediction of future values of variances / covariances is usually based on historical data. More sophisticated econometric models, e.g. GARCH, have been adopted for better predicting the future variances. While the forecast is usually better for variances than for the return, we have to ask: is it accurate enough?



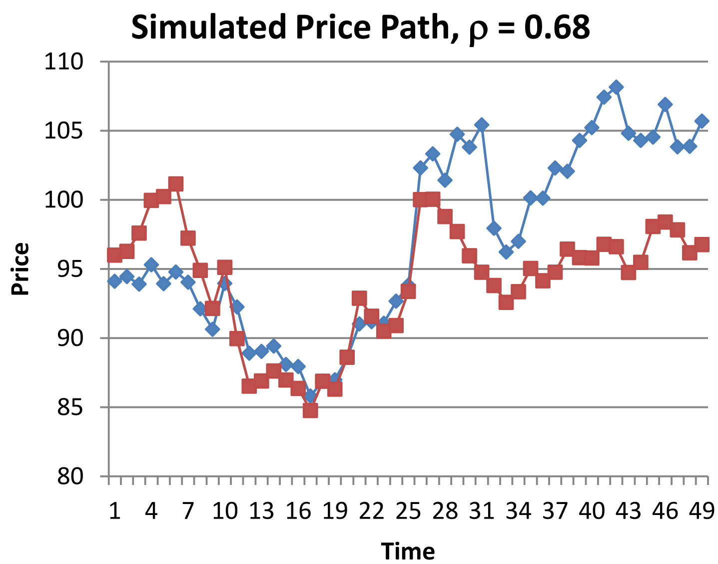


Correlation is a concept that is difficult to understand intuitively. A definition is already given in section 2.3.2. Let's look at the implication in some more detail here.



(*a*) Positively correlated (*b*) Negatively correlated (*c*) Uncorrelated

If two variables are positively correlated, positive deviation from the mean of one variable would give rise to a higher tendency of a positive deviation from mean of the other variable. However, there is no guarantee of the direction of movement. While it is easy to interpret a correlation of +1 or –1, intermediate values are hard to interpret in practice. For example, if two variables A and B have a correlation of 0.8 (which is very high), and price of A increases tomorrow, we can only say that it is more likely for B to be going up as well. If price of B goes down, it is still an acceptable result, although this may go against intuition.



Another result which is often misunderstood is that the impact of correlation can also be affected by the difference in the variance of the assets, which could also be different at different times. In the diagram above, the volatilities of the two variables are 30% and 40% between time 0 and 25, and the volatilities are 60% and 30% between time 26 and 50. Although the correlation is the same in both periods (=0.68), to the naked eye it may appear that the correlation is higher in the earlier period.

2.4.4. Other alternatives

Instead of the standard mean-variance optimizing paradigm, other means of optimization have been suggested. These methods usually make assumptions about some specific aspects which the investors may want to focus on. For example, one can maximize the geometric mean return, which is equivalent to finding the investments with the highest probability of reaching, or exceeding, any given wealth level over a given period of time. While this method corresponds directly with the investor's objective, theoretically speaking it contradicts with the utility maximization theory. Another suggestion is the "Safety first" method, where the optimization is based on minimizing the probability of bad returns (losses). The results thus obtained may not be optimal in terms of achieving the highest possible combination of risk and return. Finally, the method of stochastic dominance makes no assumption about the form of the probability distribution of returns. On the other hand, it requires an assumption about the investor's behavior, e.g. the degree of risk aversion. None of these models are as popular or easy to understand as the MPT and is not widely accepted in practice.

A more recent trend of research aims to correct a key assumption in most finance theories, that investors always behave rationally. Through the Nobel prize winning work of Daniel Kahneman (winner of the Economics prize in 2002, collaboration with Amos Tversky), it is common experience to find investor behaving inconsistently, i.e. they make choices in an irrational (or even random) manner. It is sometimes impossible to quantify their preferences in terms of a utility function. Much work in behavioral finance topics, such as overconfidence, framing, mental accounting, regret avoidance etc. have been published. However, a comprehensive portfolio theory which can take into account of the irrational behavior of investors has not been fully developed yet.

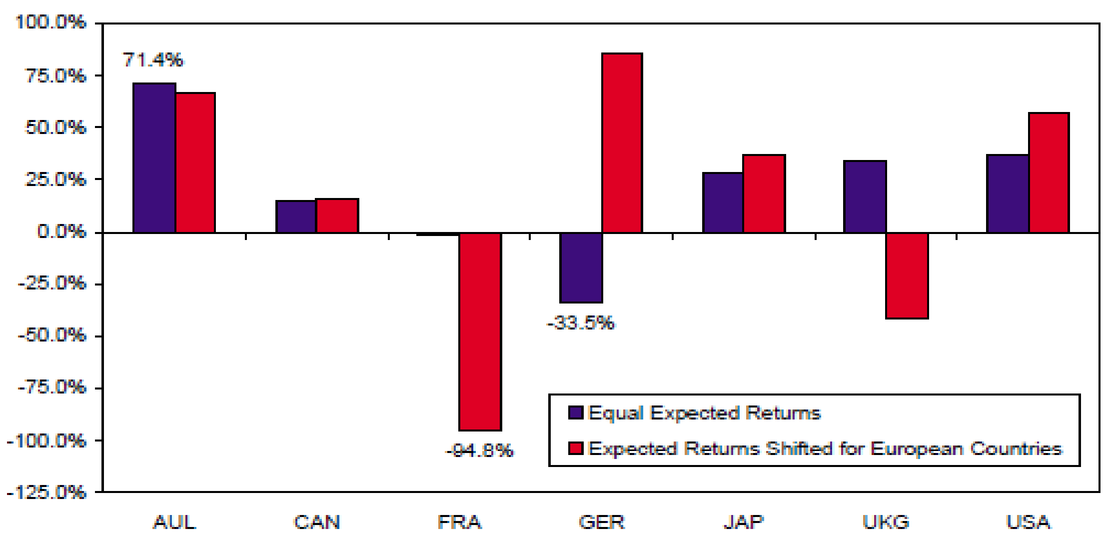
2.4.5. Conclusion: How could we use the MPT?

According to theory, everyone should hold the same risky portfolio (in the same proportion). If indeed this is the case, many people would lose their jobs, because we would not need so many fund managers and security analysts! Because of various reasons, many investors may be reluctant to invest in some asset classes, e.g. Hong Kong investors may find it inconvenient to invest in stocks in other countries. It is often impractical to include all kinds of assets as inputs to the optimization process. However, we could still use Modern Portfolio Theory to optimize within each asset class.

2.5. More advanced use of the MPT

2.5.1. Problems with MPT

While the development of MPT represents a giant step in quantifying the investment process, soon the problems of implementing the model become evident. In the section above we have already mentioned the difficulty in obtaining reliable input parameters. It was noted that while the optimization results are less sensitive to errors in estimating variance, and that the population covariance is more stable over time than estimating returns from historical data, mean-variance optimization using historical data often yields unrealistic results. Firstly, weights of assets in the portfolio are unintuitive – the MPT serves as a kind of "black box" and the allocation does not require any particular view of the attractiveness of the assets in question. Secondly, in many cases, the allocation results are concentrated in a few representative stocks. Furthermore, small changes in the input return parameters lead to drastic changes in portfolio allocations.



An example is given in He and Litterman (1999). In the simulation, it was assumed that initially the expected returns of seven countries are all set to 7%. Then three of the returns are shifted: Germany is shifted up by 2.5%, and both France and England are shifted down by 2.5%. The resulting allocations are shown in the above diagram. Dramatic shifts in weights for all three countries can be seen, while other countries are affected as well

Another noticeable problem in the default MPT algorithm is that if two assets are similar but one has a slighter higher forecast return, the optimization routine would allocate everything to the asset with the higher forecast return. It is a "one-or-the-other" kind of optimization. Of course, these problems can partly be alleviated by imposing various constraints to limit the optimization process so as to obtain more "reasonable" results, e.g. by setting size limits. However, using too many constraints would be equivalent to specifying the portfolio directly, which is not a desirable outcome.

2.5.2. Black-Litterman model

An attractive solution to some of the problems mentioned above has been proposed by Fischer Black and Robert Litterman in the early 1990s (Black and Litterman (1992)). The model allows investors to impose their own views on certain stocks, to be combined with the equilibrium returns, in order to find an optimized portfolio.

The starting point of this model is similar to the MPT: asset returns are assumed to be normally distributed, and variances of the equilibrium returns and the conditional distributions of the true mean are known. In addition, an important underlying concept is the adoption of Bayesian probability. Bayes theorem states that:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$P(A|B)$: Conditional probability of A, given B; also known as the *posterior distribution*
$P(B|A)$: Conditional probability of B, given A; also known as the *conditional distribution*
$P(A)$: Probability of A; also known as the *prior distribution*
$P(B)$: Probability of B; acts as a normalizing constant

Implementing the model requires the following procedure:

1) Calculate the expected excess return through the Capital Asset Pricing Model (known as the prior distribution). (More details of CAPM are discussed in the next topic).
2) Express the investor's views by forming a matrix of expected returns and confidence intervals for some of the assets (known as the conditional distribution)
3) Combine the two distributions to form the expected returns for all assets (known as the posterior distribution).

The results can then be fed into the portfolio optimization process.

2.5.2.1. Computing the equilibrium returns

Starting from the CAPM relationship

$$E(r_i) - r_f = \beta_i \left\lfloor E(r_M) - r_f \right\rfloor$$

$$\beta_i = \frac{\operatorname{cov}(r_i, r_M)}{\sigma_M^2}$$

The return of the market portfolio can be written as $r_M = \sum_{j=1}^{N} w_j r_j$

The expected excess return on asset $i = \Pi_i$

$$\begin{aligned}\Pi_i &= E(r_i) - r_f \\ &= \frac{\operatorname{cov}(r_i, r_M)}{\sigma_M^2}\left[E(r_M) - r_f\right] \\ &= \frac{\left[E(r_M) - r_f\right]}{\sigma_M^2}\sum_{j=1}^{N}\operatorname{cov}(r_i, r_j) w_j\end{aligned}$$

The expression can be written in matrix form:

$$\mathbf{\Pi} = \lambda \mathbf{\Sigma} \mathbf{w}$$

where λ is the market price of risk (risk aversion coefficient), $\mathbf{\Sigma}$ is the covariance matrix, and $\mathbf{w}$ is a matrix of weights in the benchmark portfolio based on market capitalization. This equation can be used to obtain the estimated expected returns of all assets once all the terms in the right hand side are specified (a process known as reverse optimization).

The true expected return $\boldsymbol{\mu}$ follows a distribution

$$\mathbf{\Pi} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{\mathbf{\Pi}},\ \boldsymbol{\varepsilon}_{\mathbf{\Pi}} \sim N(0, \tau\mathbf{\Sigma}),\ \tau << 1$$

where $\tau\mathbf{\Sigma}$ is a confidence level in how well we can estimate the equilibrium returns.

2.5.2.2. Expressing the investor's views

It is very common that an investment manager has views on some of the assets which may differ from the equilibrium expected returns. These K views in the model are expressed as a K-dimensional vector **Q:**

$$\mathbf{Q} = \mathbf{P}\boldsymbol{\mu} + \boldsymbol{\varepsilon}_{\mathbf{q}},\ \boldsymbol{\varepsilon}_{\mathbf{q}} \sim N(0, \mathbf{\Omega})$$

- $\mathbf{Q}$ is a $K \times 1$ vector representing the excess returns for each view
- $\mathbf{P}$ is a $K \times N$ matrix expressing the views. Each row of $\boldsymbol{P}$ sums to 0 if the view is relative, or sums to 1 if the view is absolute.
- $\mathbf{\Omega}$ is a diagonal $K \times K$ matrix expressing the confidence in the views
- $\boldsymbol{\varepsilon}_{\mathbf{q}}$ is a normally-distributed error term vector. Note that this term does not enter into the Black-Litterman formula; only its variance is relevant

As an example, assume that there are 5 stocks, and the investor's views are:
- Stock 1 will have a return of 1.5%
- Stock 3 will outperform stock 2 by 4%

These two views can be expressed as

$$\begin{bmatrix}1.5\% \\ 4\%\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0\end{bmatrix}\begin{bmatrix}\mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5\end{bmatrix} + \begin{bmatrix}\varepsilon_1 \\ \varepsilon_2\end{bmatrix}$$

If the confidence on these relationships is high, the variance matrix would have small values, such as

$$\mathbf{\Omega} = \begin{bmatrix} 1\%^2 & 0 \\ 0 & 1.5\%^2 \end{bmatrix}$$

2.5.2.3. Combining the two distributions

The key result in this model is the formula for the new combined returned vector is:

$$E[\mathbf{R}] = [(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^{\mathbf{T}}\mathbf{\Omega}^{-1}\mathbf{P}]^{-1}[(\tau\mathbf{\Sigma})^{-1}\mathbf{\Pi} + \mathbf{P}^{\mathbf{T}}\mathbf{\Omega}^{-1}\mathbf{Q}]$$

where the variables follow the definitions as given above. This is the expected mean of the posterior distribution.

Also from the above, we can interpret the Black-Litterman model as a complex, weighted average of the Implied Equilibrium Excess Return Vector $\mathbf{\Pi}$ and the View Vector $\mathbf{Q}$ in which the relative weightings are a function of a scalar τ and the uncertainty of the views $\mathbf{\Omega}$ (see the alternative notation below).

The variance of the mean of the posterior distribution about the true mean is:

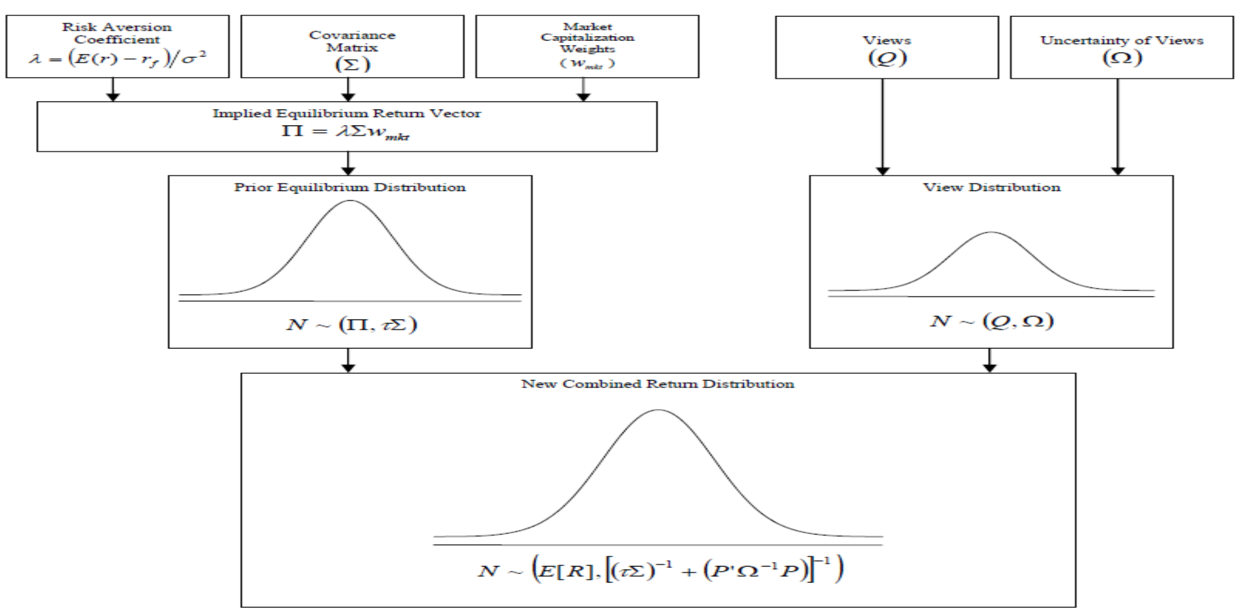
$$M = \left[(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^{\mathbf{T}}\mathbf{\Omega}^{-1}\mathbf{P}\right]^{-1}$$

The variance of the expected mean is:

$$\mathbf{\Sigma}_{\mathbf{BL}} = \mathbf{\Sigma} + \left[(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^{\mathbf{T}}\mathbf{\Omega}^{-1}\mathbf{P}\right]^{-1}$$

This is the variance to be used in the mean-variance optimization procedure, which is considered as a better estimate of the variance of the returns as more information has been used to derive its value.

A schematic diagram of the model is given by Idzorek (2005).



2.5.2.4. Alternative forms of the expected return relationship

We can also write the excess expected return as:

$$E[\mathbf{R}] = \mathbf{w}_{\mathbf{\Pi}}\mathbf{\Pi} + \mathbf{w}_{\mathbf{q}}\boldsymbol{\mu}$$

$$\mathbf{w}_{\mathbf{\Pi}} = [(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^{\mathbf{T}}\mathbf{\Omega}^{-1}\mathbf{P}]^{-1}[(\tau\mathbf{\Sigma})^{-1}]$$

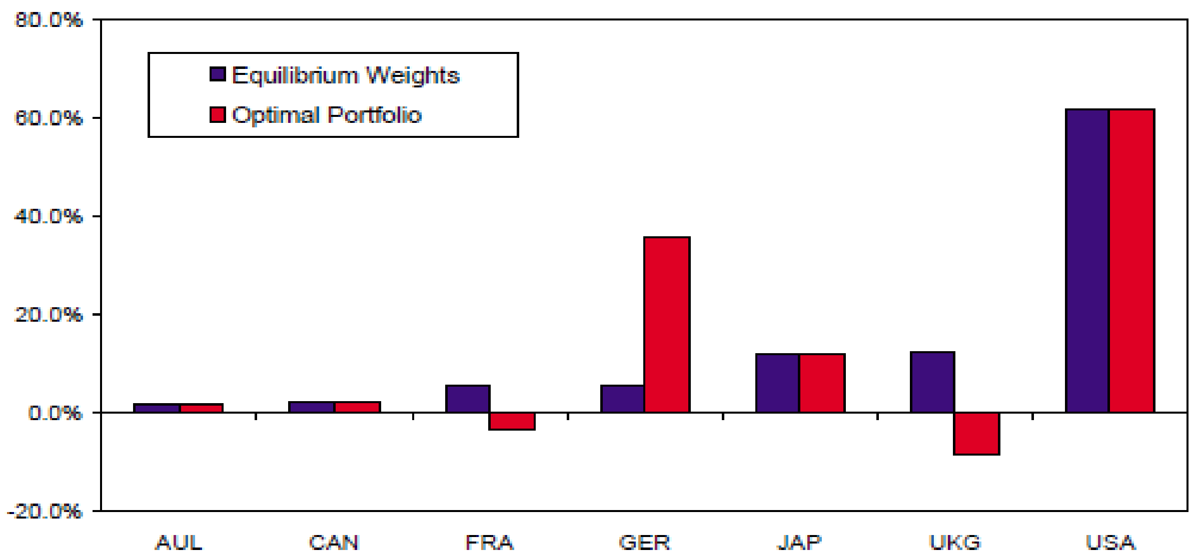
$$\mathbf{w}_{\mathbf{q}} = [(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^{\mathbf{T}}\mathbf{\Omega}^{-1}\mathbf{P}]^{-1}\mathbf{P}^{\mathbf{T}}\mathbf{\Omega}^{-1}\mathbf{P}$$

Thus the weights depend on the confidence of the estimates of the market equilibrium returns (represented by τ) and the views (represented by $\mathbf{\Omega}$) respectively.

Another form is $E[\mathbf{R}] = \mathbf{\Pi} + \tau\mathbf{\Sigma}\mathbf{P}^{\mathbf{T}}[\mathbf{\Omega} + \mathbf{P}\tau\mathbf{\Sigma}\mathbf{P}^{\mathbf{T}}]^{-1}[\mathbf{Q} - \mathbf{P}\mathbf{\Pi}]$

which explicitly shows the tilt away from the equilibrium.

2.5.2.5. An example from the Black-Litterman model



Referring to the example in section 2.5.1, the results from the Black-Litterman model is shown in the diagram above (He and Litterman (1999)). Compared to the previous diagram, it is seen that only weights that have been shifted are the weights in three countries in which the returns are altered, and all of them changed in the expected direction – the weight of Germany go up, while the weights of France and England go down. The weights of other countries are unaffected. This kind of result is more intuitive compared to the "black-box" approach from a straightforward implementation of MPT.

2.5.2.6. Some properties and remarks

More details can be found in He and Litterman (1999). To summarize briefly, if no further constraints are added,

- The optimal portfolio is the market equilibrium portfolio plus a weighted sum of the portfolios about which the investor has views.
- The weight on the portfolio increases as the expected return of the view increases, as well as when the investor becomes more confident about the view.
- The weights will not deviate from the market capitalization weighted portfolio on the assets which the investor does not have views.
- The weight for a view is positive when the view is more bullish than that implied by the CAPM.
- However, in the presence of other constraints (e.g. risk or beta limitations), the returns of all assets will change when some inputs are modified, irrespective of whether a view has been expressed or not.

There are some hurdles in implementing the model. Firstly, note that each view is supposed to be unique and uncorrelated with the other views, and this is difficult to achieve in practice. Secondly, estimation of the "variances of the views" is not a trivial task: one possible method is to define some kind of confidence interval around the mean.

In some scenarios, a ranking of assets is produced as the view, and thus the view matrix becomes a scalar with weights of either +1 or –1, representing assets which should be long or short in the portfolio respectively.

Finally, let's briefly recap the advantages of this model:

- There is no need for the investors to have a view on every available asset. Investors can specify their views on the assets that they have strong opinions, and rely on equilibrium returns as the starting points for most other assets.
- The model can incorporate "relative" views in addition to "absolute" views, which is a common type of opinion among investment professionals.
- One can have control over the confidence level of views (although it is difficult to specify this in a meaningful way).
- The optimization results are more intuitive and less sensitive to changes in input returns.

On top of some difficulties noted above, we must realize that the model is still sensitive to assumptions, e.g. the model assumes that views are independent of each other. Also, it must be understood that the final result is not really a portfolio in the "optimal" sense. Having said that, the model is still a step forward within quantitative investment management.

Additional References

Black, Fischer, and Robert Litterman. "Global Portfolio Optimization." *Financial Analysts Journal*, September/October 1992, pp. 28-43.

He, Guangliang and Robert Litterman. "The Intuition Behind Black-Litterman Model Portfolios." *Goldman Sachs Investment Management*, December 1999.

Idzorek, Thomas M. "A Step-by-Step Guide to the Black-Litterman Model." *Ibbotson Associates*, April, 2005.

Satchell, S. and A. Scowcroft. "A Demystification of the Black-Litterman Model: Managing Quantitative and Traditional Construction." *Journal of Asset Management*, September 2000, pp.138-150.