Overview
This chapter explores in greater depth than chapter 4 the study of portfolio decisions when investors act to optimize a mean-variance objective function.

In addition to its significance as a testable theory of asset demand, mean-variance analysis plays two other roles: (a) it provides a method for the practical construction of portfolios; and (b) it forms the foundation for the capital asset pricing model, the subject of chapter 6.

Although mean-variance analysis and the CAPM are close relatives, it is important to distinguish between the two. Mean-variance analysis provides a theory of individual behaviour regardless of whether the market, as a whole, is in equilibrium. The CAPM, building on mean-variance analysis, provides a theory of asset prices in market equilibrium.¹ This chapter addresses only the former problem – of individual behaviour – and is silent about the implications of market equilibrium for asset prices.

The analysis in this chapter proceeds in a sequence of steps, each of which builds on the previous one. The steps are summarized as follows.

1. A review of the basic concepts of mean-variance analysis.
2. The choice between two risky assets: the objective here is to construct a frontier between the expected rate of return on each portfolio and the portfolio’s standard deviation of return. No risk-free asset is available.
3. The choice among many risky assets, again excluding risk-free lending or borrowing. The portfolio frontier is found to take the same form as in the two-asset case, and an important proposition – the first mutual fund theorem – is explained.

¹ Market equilibrium refers here to an equality between the total of investors’ demands to hold assets and the aggregate stocks available to be held.
4. The choice among many assets when a risk-free asset is available. In this case, the portfolio frontier takes a simple linear form, hence justifying the second mutual fund theorem.

5. Optimal portfolio selection: here the selection of an optimal portfolio from among the efficient portfolios is examined.

Risk. In this chapter, the standard deviation (or, equivalently for this purpose, variance) of an asset’s rate of return is used to measure the risk associated with holding the asset. As noted in chapter 4, this association should be treated as provisional. Subsequent chapters argue that it needs to be qualified: there may exist other, more suitable, measures of an asset’s risk than its standard deviation of return.

Absence of market frictions. In common with much of financial theory, it is assumed that market frictions (transaction costs and institutional restrictions on trades) can be ignored. Just how restrictive this assumption is, of course, depends on the severity of the frictions in any particular application. The treatment here requires, at least, that market frictions do not impinge in a significant way on the portfolio selection decisions of investors.

5.1 Mean-variance analysis: concepts and notation

5.1.1 The mean-variance objective

Each investor who acts according to a mean-variance objective is assumed to choose a portfolio that maximizes

\[ G = G(\mu_p, \sigma_p^2) \]

subject to the constraint that the total value of assets (calculated at initial prices) does not exceed initial wealth. The expected (or mean) rate of return on the portfolio is denoted by \( \mu_p \). Risk is measured by \( \sigma_p \), the standard deviation of the rate of return on the portfolio. Pairs of \( \mu_p \) and \( \sigma_p \) for which \( G(\mu_p, \sigma_p^2) \) is constant define indifference curves (examined later, in section 5.5; see also chapter 4, page 103).

In this chapter, the investor is assumed to make exactly one portfolio decision; a decision that (for whatever reason) remains unchanged for the whole of the time period being studied. What happens after the decision has been made is ignored. (More complicated decisions involving revisions to portfolios – and other relevant choices – are explored in chapter 11.)

As explained in chapter 4, a mean-variance investor’s optimal portfolio selection can be split logically into two steps. First, the portfolio frontier, comprising
those portfolios for which $\sigma_P^2$ is minimized for each $\mu_P$, is constructed. Second, a choice is made from among the frontier portfolios so as to maximize the objective, $G$, in accordance with preferences, expressed by the investor’s own, personal $G(\cdot, \cdot)$ function. Assuming that the objective function is increasing in expected return and decreasing in risk, only a portion of the portfolio frontier – the efficient portfolio set – is relevant in the second step.

The expected return and standard deviation of portfolio return encapsulate the investor’s beliefs about the rates of return on individual assets: by varying the composition of the portfolio, the investor effectively chooses $\mu_P$ and $\sigma_P$. The constraint on the investor’s portfolio choices is called the portfolio frontier; it is expressed in terms of $\mu_P$ and $\sigma_P$, rather than directly in terms of the amount of each asset held.

The mean-variance model acquires its practical relevance because means, covariances and variances of rates of return can be estimated from past observations on asset prices or from other relevant information. That is, experience (typically, price observations) can be used to represent an investor’s beliefs in a way that involves standard statistical methods – methods that would not necessarily be applicable in other models of portfolio selection. Armed with estimates of means and variances, it is a routine matter to calculate the portfolio frontier.

In what follows, a distinction should be drawn between the theoretical concepts (of means and variances) and the estimates that are made of them. They are different things: the estimates are observable counterparts of the unobservable theoretical concepts. In practical calculations of portfolio frontiers, it is necessary to use numerical values (i.e. estimates) of means and variances. These can be obtained using a variety of methods. Calculations based on past data may be the most convenient way to estimate the means and variances. But individuals may differ in their beliefs (perhaps they differ in their available information, or in how they process it) and there is nothing that compels individuals to use one method rather than another in forming their estimates. Indeed, from the standpoint of economic analysis, investors may be assumed to act only as if they choose according to a mean-variance criterion; there is no reason why they should be consciously aware of means and variances.

If all investors have access to the same information, and if this information is the set of past rates of return on the assets, then, arguably, all investors should possess common beliefs and, hence, should arrive at the same estimates of the means and variances. (More precisely, what is important is whether investors act

---

2 The diagrammatic treatment is given in terms of $\sigma_P$ rather than $\sigma_P^2$. Given that the standard deviation is the positive square root of the variance, minimizing one is equivalent to minimizing the other. Hence, it is a matter of convenience about which to minimize.

3 From now on, ‘variances’ should be understood to include covariances, unless indicated otherwise.
as if they agree. Whether they could or would acknowledge that they agree is another matter, irrelevant in this context.)

To postulate that investors agree about means and variances is restrictive, however, and is not needed in this chapter; the focus of attention is on the decisions of a single investor. The question of whether investors act as if they share the same beliefs about the means and variances of assets’ returns becomes important in the theory of market equilibrium, involving as it does interactions among all investors.

5.1.2 Notation

The following notation is used throughout the remainder of this chapter, and also in chapter 6.

\[ r_j = \text{rate of return on asset } j = 1, 2, \ldots, n \text{ (each } r_j \text{ is a random variable).} \]

\[ \mu_j = \mathbb{E}[r_j], \text{expected rate of return on asset } j. \]

\[ \sigma_{ij} = \text{cov}(r_i, r_j) \equiv \mathbb{E}[(r_i - \mu_i)(r_j - \mu_j)], \text{covariance between the rates of return } r_i \text{ and } r_j. \]

\[ \sigma_j = +\sqrt{\sigma_{jj}}, \text{standard deviation of return on } j, \text{where } \sigma_{jj} = \text{var}(r_j) \equiv \mathbb{E}[(r_j - \mu_j)^2]. \]

\[ \rho_{ij} = \frac{\sigma_{ij}}{(\sigma_i \sigma_j)}, \text{correlation coefficient between returns on assets } i \text{ and } j. \]

\[ a_j = p_j x_j / A, \text{proportion of portfolio invested in asset } j, \text{with } \sum_j a_j = 1. \]

\[ r_0 = \text{rate of return on risk-free asset, } \mu_0 \equiv r_0. \]

In the following, a subscript \( i \) or \( j \) refers to a single asset while an upper-case subscript (e.g. \( P \) or \( Z \)) refers to a portfolio of several assets. A portfolio is defined as a vector, or list, of asset holdings, \( x_0, x_1, x_2, \ldots, x_n \), chosen subject to the constraint \( \Sigma p_j x_j = A \), where \( A \) denotes initial wealth. The portfolio is more conveniently written in terms of proportions: \( a_0, a_1, a_2, \ldots, a_n \). In this representation asset prices and initial wealth are hidden in the background. The expected rate of return and variance of the rate of return on any portfolio, \( P \), take the form

\[ \mu_P = \sum_j a_j \mu_j \]

and

\[ \sigma_P^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij} \]

The range of \( j \) in the above summation for \( \mu_P \) is \( j = 1, 2, \ldots, n \) when there is no risk-free asset, and \( j = 0, 1, 2, \ldots, n \) when there is a risk-free asset. It should be clear from the context which is intended.
5.2 Portfolio frontier: two risky assets

The case of two risky assets (with no risk-free asset) is a handy building block for the general case of $n$ assets. In figure 5.1 the two end points mark the expected returns and standard deviations of the two assets; the curved line joining them is the portfolio frontier. Note that the frontier must pass through the two dots marked ‘Asset 1’ and ‘Asset 2’ because, with just two assets to choose between, both are on the frontier: depending on the investor’s preferences, total wealth could be devoted entirely to one of the assets. The goal here is to understand how the frontier is constructed. Having grasped the two-asset model, the general case of $n > 2$ assets can be understood with little extra effort.

Define $a = a_1$ so that $(1 - a) = a_2$. Then, from the definitions of expectations and variances,

$$
\mu_p = a\mu_1 + (1 - a)\mu_2
$$

$$
\sigma_p^2 = a^2\sigma_{11} + 2a(1 - a)\sigma_{12} + (1 - a)^2\sigma_{22}
= a^2\sigma_1^2 + 2a(1 - a)\rho_{12}\sigma_1\sigma_2 + (1 - a)^2\sigma_2^2
$$

Note that $\sigma_{11} = \sigma_1^2$, $\sigma_{22} = \sigma_2^2$, $\sigma_{12} = \rho_{12}\sigma_1\sigma_2$.

Consider the special cases for which the correlation between the two rates of return takes on extreme values: $\rho_{12} = \pm 1$:

$$
\sigma_p^2 = a^2\sigma_1^2 \pm 2a(1 - a)\sigma_1\sigma_2 + (1 - a)^2\sigma_2^2
= (a\sigma_1 \pm (1 - a)\sigma_2)^2
\sigma_p = |a\sigma_1 \pm (1 - a)\sigma_2|
$$

Hence, bearing in mind that the standard deviation, $\sigma_p$, must be non-negative (by definition),

$$
\rho_{12} = +1 \implies \sigma_p = a\sigma_1 + (1 - a)\sigma_2
$$

$$
\rho_{12} = -1 \implies \sigma_p = (a\sigma_1 - (1 - a)\sigma_2) \geq 0 \text{ for } a \geq \frac{\sigma_2}{\sigma_1 + \sigma_2}
$$

$$
\rho_{12} = -1 \implies \sigma_p = -(a\sigma_1 - (1 - a)\sigma_2) > 0 \text{ for } a < \frac{\sigma_2}{\sigma_1 + \sigma_2}
$$

The expressions above trace out lines in the $(\mu_p, \sigma_p)$ plane after eliminating $a$, using the definition $\mu_p = a\mu_1 + (1 - a)\mu_2$. The three lines are depicted in figure 5.2.
Fig. 5.1. The efficiency frontier with two assets

The portfolio frontier depicts the minimum $\sigma_p$ for each level of $\mu_p$. It is shown here for non-negative combinations of two assets. For a zero proportion of asset 2 in the portfolio, the frontier is located at the point labelled ‘Asset 1’. Similarly, for a zero proportion of asset 1 in the portfolio, the frontier is located at the point labelled ‘Asset 2’.

Fig. 5.2. The efficiency frontier with two assets and $\rho_{12} = \pm 1$

The portfolio frontier depends, among other things, on the correlation between the assets’ rates of return. At one extreme, $\rho_{12} = +1$, the frontier is the line segment joining points labelled ‘Asset 1’ and ‘Asset 2’. At the other extreme, $\rho_{12} = -1$, the frontier consists of two line segments, from ‘Asset 2’ to $V$, and from $V$ to ‘Asset 1’.
Fig. 5.3. The efficiency frontier allowing for short-sales

The frontier $FF$ allows for all possible combinations of assets in the portfolio (such that their proportions sum to unity) including a negative proportion (short-sale) of one or other asset. The positively sloped portion of $FF$ approaches but never reaches the ray $BC$. The negatively sloped portion of $FF$ approaches but never reaches the ray $BD$. The point marked $MRP$ (minimum risk portfolio) identifies the portfolio for which $\sigma_P$ is smallest for all possible values of expected return (for this portfolio, $\mu_{mrp} = B$).

Remarks

1. For all values of the correlation coefficient, $\rho_{12}$, strictly greater than $-1$ and strictly less than $+1$, the frontier is non-linear. For non-negative holdings of both assets it is the line depicted in figure 5.1, lying within the triangular region bounded by the lines in figure 5.2.

2. Extension of the portfolio frontier beyond the points given by $a = 1$ or $a = 0$ involves the short-sale of one of the assets (the resulting funds being used to purchase the other).

3. It can be shown that the relationship between the expected return on the portfolio, $\mu_P$, and the variance, $\sigma_P^2$, is a parabola. Usually, as here, the graph is drawn using the standard deviation $\sigma_P$, in which case the relationship between $\mu_P$ and $\sigma_P$ is a hyperbola.

This hyperbola takes the form depicted in figure 5.3 on page 120. Note that the frontier approaches, but never intersects, the two rays $BC$ and $BD$ as $\sigma_P$ tends to
infinity. The point \( MRP \) on the frontier, for which \( \sigma_p \) is at a minimum when ranging over all expected return values (i.e. not for a given \( \mu_p \)), is called the **minimum risk portfolio**.\(^4\) The expected rate of return corresponding to \( MRP \) is denoted by point \( B \) in figure 5.3. For later reference, let \( \mu_{\text{mrp}} \) denote the expected rate of return on the minimum risk portfolio. Also for later reference, note that the MRP is optimal only for investors whose preferences, expressed by \( G(\mu_p, \sigma^2_p) \), focus entirely on risk and give zero weight to expected return. Every investor who is prepared to tolerate higher risk for a higher expected rate of return would choose a portfolio with greater risk than that of the MRP.

4. The upward-sloping arm of the frontier defines the set of efficient portfolios – i.e. an efficient portfolio is one for which \( \mu_p \) is maximized for a given \( \sigma_p \). As already noted, if \( G(\mu_p, \sigma^2_p) \) is increasing in expected return and decreasing in risk, the choice of an optimal portfolio will always be made from the efficient set.

5.3 Portfolio frontier: many risky assets and no risk-free asset

Suppose that there are \( n > 2 \) risky assets (with no risk-free asset). The shape of the portfolio frontier is the same as for the two-asset case (it is a hyperbola). Including additional assets allows for increased diversification – i.e. the attainment of at least as low a level of risk for each level of expected return. The frontier with a larger number of assets is located to the left of the frontier with fewer assets. (More precisely, the frontier is nowhere to the right when additional assets are available. This allows for the possibility that the additional assets have returns that are perfectly correlated with combinations of existing assets, and hence do not affect the trade-off between \( \mu_p \) and \( \sigma_p \).)

It is assumed from now on that the \( n \) assets are ‘genuinely different’, in the sense that the return on no asset can be formed as a linear combination of the returns on other assets. **Composite** assets created as portfolios (linear combinations) of existing assets play an important part in what follows, but for convenience they are excluded from the list of \( n \) underlying assets.

To gain some intuition for the multiple-asset case, imagine starting with two assets. Then form an efficient portfolio of the two. This portfolio has a random return (with an expectation and a variance) that depends on the proportions of the two assets held in it. Now treat this portfolio as if it is a single asset. This composite asset is rather like a simple ‘closed-end mutual fund’ in the United States or ‘investment trust’ in Britain. The composite asset can be identified with a point on the efficient frontier for the two underlying assets, the location of the point being determined by the asset proportions in the portfolio.

---

\(^4\) This terminology is convenient but could be misleading: all points on the frontier depict points of minimum risk for a given expected return, \( \mu_p \). The **minimum risk portfolio** is defined without holding \( \mu_p \) given at a specified level. The \( MRP \) identifies a portfolio with global minimum risk.
If a third asset becomes available, it is possible to construct a frontier \textit{as if} for two assets, one of which is a composite of assets 1 and 2 (such as marked by point $C$), the other being asset 3. The line $FF$ depicts the overall portfolio frontier only if asset 3 on its own happens to be the portfolio with minimum $\sigma_p$ when $\mu_p$ is set equal to $\mu_3$, the expected rate of return on asset 3. (In general, frontier portfolios contain non-zero proportions of at least two assets so that ‘Asset 3’ lies strictly inside, not on, the boundary of the frontier.)

By including a third asset along with the first two, the portfolio selection problem can be analysed just as if the choice is between the composite asset and the third asset. All of the analysis of the previous section applies, the result being depicted in figure 5.4.

Beware. While the analysis so far is suggestive, it is incomplete and could be misleading, for the line $FF$ in figure 5.4 is \textit{not} necessarily the frontier in the three-asset case. The reason is that $FF$ denotes the frontier only for cases in which a single asset – asset 3 in figure 5.4 – lies on the frontier. In general, the portfolio frontier lies to the left of $FF$: an individual asset, while on its own constituting a \textit{feasible} portfolio (all initial wealth could be invested in a single asset), is typically not on the frontier, with $n > 2$ assets.

Despite this warning, it is reasonable to suppose that a pair of composite assets – mutual funds – can be constructed from subsets of any number $n > 2$ of assets, and that these two composite assets can be used to trace out a frontier. There are many such pairs, each of which generates its own frontier. The portfolio
frontier for the $n$ asset problem is that which is the ‘furthest to the left’ in the
$\mu_p, \sigma_p$ plane. The upshot, explored further below, is that it is possible to trace
out the portfolio frontier as if there are exactly two assets, which are themselves
composites of the individual assets. (Note that both composite assets may need
to include a non-zero amount of every individual asset.)

The formal optimization problem from which the portfolio frontier is con-
structed is as follows: for a given value of $\mu_p$, choose portfolio proportions,
$a_1, a_2, \ldots, a_n$, to

\[
\begin{align*}
\text{minimize } & \sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij} \\
\text{subject to } & \mu_p = \sum_{j=1}^n a_j \mu_j \text{ and } \sum_{j=1}^n a_j = 1
\end{align*}
\]

A separate minimization is carried out for each given value of $\mu_p$ so that, as
$\mu_p$ is changed, the frontier is traced out. (Appendix 5.2 offers a formal analysis.)

The description of the portfolio frontier above hints at the first mutual fund
theorem (or first separation theorem) of portfolio analysis.\(^5\) A precise statement
of the theorem is: given the existence of $n \geq 2$ assets, the random rates of returns
on which can be expressed entirely by their means and variances, there exist two
mutual funds (composite assets) such that the expected rate of return and variance
of every frontier portfolio can be obtained by holding only the two mutual funds.
That is, for any arbitrary frontier portfolio, the two mutual funds alone can be
combined to form a portfolio that has the same expected rate of return and the
same variance as the arbitrarily chosen frontier portfolio.

In order to gain some intuition for the importance of the first mutual fund
theorem, consider an investor whose beliefs can be expressed in terms of means
and variances and who seeks to choose a portfolio that maximizes $G(\mu_p, \sigma_p^2)$. The
investor’s beliefs are expressed by

\[
\begin{bmatrix}
\mu_1 & \mu_2 & \mu_3 & \ldots & \mu_n \\
\sigma_{11} & \sigma_{12} & \sigma_{13} & \ldots & \sigma_{1n} \\
\sigma_{22} & \sigma_{23} & \ldots & \sigma_{2n} \\
\sigma_{33} & \ldots & \sigma_{3n} \\
\vdots & & & & \\
\sigma_{nn}
\end{bmatrix}
\]

\(^5\) Although the mutual fund theorems of portfolio analysis are most commonly found in mean-variance
analysis, they appear in more general treatments involving broader classes of preferences and distributions
of asset returns.
Suppose that the number of assets, \( n \), is very large (say, 1000) and that the investor finds it too complicated and wearisome to choose among the \( n \) assets in maximizing \( G(\mu_p, \sigma^2_p) \).

Thus, suppose that the information about beliefs (means, variances and covariances) is handed to an expert. The expert does some calculations and replies (correctly) that the investor can achieve a maximum of \( G(\mu_p, \sigma^2_p) \) by choosing between just two assets – say, \( A \) and \( B \) – that are themselves portfolios (mutual funds) of the original \( n \) assets. The investor’s portfolio choice problem – maximization of \( G(\mu_p, \sigma^2_p) \) – is thus dramatically simplified. Note the following.

1. The expert needs to know \textit{nothing} about the investor’s preferences, \( G(\mu_p, \sigma^2_p) \). Hence, the same pair of assets, \( A \) and \( B \), could be used to locate the optimum portfolio for \textit{every} investor who shares the same beliefs (means, variances and covariances).

2. The first mutual fund theorem guarantees that the expert is right to claim that the problem simplifies to the choice between two assets (so long as the expert makes no mistakes in the calculations).

3. The expert creates the two assets as follows. (a) For any level of portfolio expected return, \( \mu_p \), choose the portfolio that minimizes \( \sigma^2_p \). By construction, the solution pair \((\mu_p, \sigma_p)\) lies on the portfolio frontier. (b) Hence construct two portfolios, \( A \) and \( B \), on the frontier corresponding to \textit{any} two different levels of \( \mu_p \).

4. With a knowledge of \( \mu_A, \mu_B, \sigma^2_A, \sigma^2_B, \sigma_{AB} \) (means, variances and covariance of returns for \( A \) and \( B \)), the investor can construct exactly the same portfolio frontier as for the original \( n \) assets. An optimal portfolio comprising just \( A \) and \( B \) is then chosen to maximize \( G(\mu_p, \sigma^2_p) \) from among the frontier portfolios.

While the discussion above is intended to motivate the mutual fund theorem, it is not a proof. A sketch of a proof is as follows. \textit{First}, it can be shown that there is a unique portfolio of the \( n \) assets corresponding to each point on the frontier.\footnote{Uniqueness follows from the assumption that all \( n \) assets are ‘genuinely different’ in the sense that there are no linear dependences among their random rates of return.} \textit{Second}, choose any distinct pair of frontier portfolios: these correspond to the mutual funds referred to in the theorem. \textit{Third}, the expected return on any arbitrary frontier portfolio can be expressed as a portfolio constructed from the mutual funds.\footnote{For example, suppose that the expected returns on the mutual funds are 10 per cent and 20 per cent, respectively. Let the expected return on some other portfolio be \( \mu_p \). Then the portfolio of mutual funds needed to obtain \( \mu_p \) is found by solving: \( \mu_p = 0.10\theta + 0.20(1 - \theta) \) for \( \theta \). In this example, \( \theta = 2 - 10\mu_p \).} \textit{Fourth}, it is possible to show that the newly constructed portfolio satisfies exactly the same first-order (variance-minimizing) conditions as the arbitrarily chosen portfolio. \textit{Fifth}, because the constructed portfolio satisfies the same conditions, it defines the same point on the frontier and hence has the same variance. \textit{Finally}, as far as mean and variance are concerned, the arbitrary portfolio and the portfolio constructed from the mutual funds are identical. Hence,
the first mutual fund theorem holds: any frontier portfolio can be constructed from the two mutual funds.

5.4 Portfolio frontier: many risky assets with a risk-free asset

5.4.1 Efficient portfolios

In the presence of a risk-free asset and any number of risky assets, the set of efficient portfolios is a straight line. In figure 5.5 the efficient portfolios lie along the line $r_0ZE$, where $r_0$ is the risk-free rate of return and $Z$ is the point of tangency between a ray from $r_0$ and $FF$ (the frontier for portfolios of risky assets only). The set of efficient portfolios is obtained by minimizing the portfolio risk, $\sigma_p$, for each given expected portfolio return, $\mu_P$.

The point of tangency depicted by $Z$ in figure 5.5 identifies the efficient portfolio for which the proportion of the risk-free asset is zero. Efficient portfolios to the left of $Z$, along $r_0Z$, include a positive proportion of the risk-free asset, while those to the right, along $ZE$, involve a negative proportion (i.e. borrowing to finance the purchase of risky assets).

The formal optimization problem from which the set of efficient portfolios is constructed is as follows: for a given value of $\mu_P$, choose portfolio proportions, $a_0, a_1, a_2, \ldots, a_n$, to

$$\text{minimize } \sigma^2_P = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij}$$

subject to $\mu_P = \sum_{j=1}^{n} a_j \mu_j + a_0 r_0$ and $a_0 + \sum_{j=1}^{n} a_j = 1$

A separate minimization is carried out for each given value of $\mu_P$ so that, as $\mu_P$ is changed, the efficient set is traced out. (See appendix 5.3 for a formal analysis.)

Remarks

1. *Why is the set of efficient portfolios a straight line?* An intuitive argument is as follows. First, construct the frontier, $FF$, for risky assets alone, as described in the previous section. Next, plot the rate of return, $r_0$, for the risk-free asset on the vertical axis of the $(\mu_P, \sigma_P)$ diagram and connect this with a ray to any point on the frontier $FF$. Points along this ray depict the expected return and risk for portfolios comprising the risk-free asset and the portfolio of risky assets given by the chosen point on the $FF$ frontier. Lower levels of risk can be attained (for each level of expected return) by pivoting the ray through $r_0$ to higher and higher points along $FF$, until the ray is tangential with $FF$. No further reduction in risk is feasible (for a given expected return and the frontier, $FF$, that is). The set of efficient portfolios for all assets including the risk-free asset is then the ray formed from this tangency – i.e. the line $r_0ZE$ in figure 5.5.

2. *Must such a tangency (on the positively sloped portion of $FF$) always exist?* No. A necessary and sufficient condition for the existence of the tangency depicted in
When a risk-free asset, with rate of return $r_0$, is available the portfolio frontier becomes the ray $r_0ZE$, starting at $r_0$ and tangential to $FF$ at $Z$. To the left of $Z$, along $r_0Z$, efficient portfolios contain a non-negative proportion of the risk-free asset. To the right of $Z$, along $ZE$, efficient portfolios involve borrowing at the risk-free rate.

Figure 5.5 is that $r_0 < \mu_{mep}$; i.e. the risk-free interest rate must be less than the expected rate of return on the portfolio of risky assets with minimum risk. (See figure 5.3 on page 120, where point $B$ corresponds to $\mu_{mep}$.)

What happens if $r_0 \geq \mu_{mep}$? Suppose that $r_0 > \mu_{mep}$. In this case, it can be shown that the investor would choose to short-sell a portfolio of risky assets and invest the proceeds in the risk-free asset. The set of efficient portfolios remains a positively sloped straight line through $r_0$, the risk originating in the payoffs from the assets that have been short-sold.

Suppose that $r_0 = \mu_{mep}$. In this case, it can be shown that the investor could choose a portfolio of risky assets such that $\sum_{j=1}^{n} a_j = 0$ (i.e. with some assets short-sold and some positive holdings) to yield an expected return equal to $r_0$ and zero risk, $\sigma_p = 0$. There is no optimal solution in this case; the investor perceives that there exists a portfolio of risky assets that has precisely the same mean-variance properties as the risk-free asset.

---

8 The portfolio that would be short-sold can be identified with the point of tangency between a ray from $r_0$ and the lower, negatively sloped, arm of the $FF$ frontier. For a detailed exposition, see Huang and Litzenberger (1988, pp. 78–80).

9 It is tempting to interpret the basket of risky assets in question as being an arbitrage portfolio: a zero outlay of initial capital on risky assets yields a risk-free return. As the analysis in chapter 7 shows, this temptation should be resisted, for here the absence of risk is conditional upon the investor’s perceptions of the means and variances of assets’ returns. Arbitrage in the strict sense of chapter 7 is not conditional on this information.
While \( r_0 \leq \mu_{mrp} \) could conceivably hold for an individual investor, market equilibrium is inconsistent with this inequality holding for every investor; some investor has to be prepared to hold the assets sold by others. In summary: the circumstances for which \( r_0 \geq \mu_{mrp} \) are pathological – unusual, to say the least. From now on it is assumed that \( r_0 < \mu_{mrp} \).

3. **Second mutual fund theorem (second separation theorem) of portfolio analysis.** Under the same conditions as for the first separation theorem and in the presence of a risk-free asset, any efficient portfolio can be attained by holding at most two assets, one of which is the risk-free asset and the other is a mutual fund. Not just any mutual fund will do. It must be a portfolio chosen from those in the efficient set (any one of these will do).

   To understand why the theorem holds, suppose that the mutual fund is chosen to be the efficient portfolio with expected return \( \mu_Z \) and risk \( \sigma_Z \) (point \( Z \) in figure 5.5). Now any efficient portfolio (i.e. along the line \( r_0ZE \)) can be formed as a combination of this mutual fund and the risk-free asset.

4. Note that the efficient set is a straight line only if the investor can borrow and lend at the same risk-free rate, \( r_0 \). Suppose, instead, that the investor can lend at a rate \( r_0^L \), lower than the rate, \( r_0^B \), at which funds can be borrowed. Now the set of efficient portfolios has three segments (see figure 5.6).

   For low levels of expected return and risk (such that the investor holds a positive amount of the risk-free asset), the efficient portfolios are located on the ray from \( r_0^L \) tangent to \( FF \) at point \( Y \). Points beyond \( Y \) are irrelevant because, by assumption, the investor cannot borrow at rate \( r_0^L \).

   For high levels of expected return and risk (such that the individual borrows in order to invest in risky assets), the efficient portfolios are located on the line segment to the right of \( Z \). Formally, \( Z \) is defined by the ray through \( r_0^B \) that is tangent to \( FF \). Points between \( r_0^B \) and \( Z \) are irrelevant because, by assumption, the investor cannot lend at rate \( r_0^B \).

   At intermediate levels of expected return and risk, the investor neither borrows nor lends (initial wealth being invested entirely in risky assets) and the efficient portfolios are located along \( FF \) between \( Y \) and \( Z \). Thus, the whole set of efficient portfolios is depicted by the connected line segments, \( r_0^L Y, YZ \) and \( ZE \).

### 5.4.2 The trade-off between expected return and risk

In mean-variance analysis, the investor chooses between **efficient portfolios**, where an efficient portfolio is one for which expected return is maximized for a given level of risk (standard deviation of portfolio return, \( \sigma_P \)). The following paragraphs show how the efficient portfolios can be characterized in terms of assets’ means and variances. This is accomplished by deriving expressions for the increment to expected return and risk in response to incremental variations in the amount of an asset in the portfolio.
Fig. 5.6. Efficient portfolios with different lending and borrowing rates

If the rate at which funds can be borrowed, $r_B^0$, exceeds the rate at which funds can be lent, $r_L^0$, the frontier comprises three segments: $r_L^0 Y$ (some funds are lent), $YZ$ (neither lending nor borrowing), $ZE$ (funds are borrowed for investment in risky assets).

Suppose that an investor holds a portfolio labelled by $P$; this portfolio need not be efficient, it is just any feasible portfolio of assets. Denote its expected return and risk by $\mu_P$ and $\sigma_P$, respectively. Now suppose that a small increase is made in the holding of asset $j$, the necessary funds being borrowed at the risk-free interest rate. The quantity of each of the other assets remains unchanged. Denote the increase in the proportion of $j$ in the portfolio by $\Delta a_j$ (and, by construction, $\Delta a_0 = -\Delta a_j$).

**Expected return.** The change in the portfolio’s expected return is the difference between the expected return on asset $j$ and the risk-free rate of interest, $r_0$, multiplied by the change in the proportion of asset $j$:

$$\Delta \mu_P = (\mu_j - r_0) \Delta a_j$$  (5.1)

**Risk (standard deviation).** Calculation of the change in the risk of return on the portfolio, $\Delta \sigma_P$, involves more effort. The derivation in appendix 5.4 shows that

$$\Delta \sigma_P = \beta_{jp} \sigma_P \Delta a_j \quad \text{where} \quad \beta_{jp} = \frac{\sigma_{jp}}{\sigma_P^2}$$  (5.2)
(Strictly, (5.2) is an approximation that approaches an equality in the limit as $\Delta a_j$ tends to zero.) The symbol $\sigma_{jp}$ denotes the covariance of the return on asset $j$ with the whole portfolio. Formally, $\sigma_{jp} = \sum_{i=1}^{n} a_i \sigma_{ij}$. Equation (5.2) shows that $\beta_{jp}\sigma_P$ can be interpreted as the increment to overall portfolio risk resulting from an increment to the proportion of asset $j$ held in the portfolio – i.e. $\Delta \sigma_P / \Delta a_j$.

The $\beta_{jp}$ term plays a central role in the capital asset pricing model (chapter 6), which is why equation (5.2) is written the way it is. In words, $\beta_{jp}$ captures the relationship between variations in the rate of return on asset $j$ and the rate of return on the whole portfolio, $P$. Note especially that the change in $\sigma_P$ associated with a change in asset $j$ is not equal to $\sigma_j$ (the standard deviation of the rate of return on asset $j$). Why not? Because asset returns may be correlated with one another and these correlations must be taken into account. The influence of the correlations is encapsulated in $\beta_{jp}$.

Efficient portfolios. It is possible now to obtain a necessary condition that must be satisfied if $P$ is to represent an efficient portfolio. If $P$ is efficient, it must be the case that a small change in the portfolio proportion of any asset must disturb the expected return per unit of risk by the same amount for all assets. That is, $\Delta \mu_P / \Delta \sigma_P$ must be equal for all assets. Taking the ratio of (5.1) and (5.2), it follows that $(\mu_j - r_0) / \beta_{jp}\sigma_P$ must be equal for all assets $j = 1, 2, \ldots, n$. Otherwise, it would be possible to obtain a higher expected return with no higher risk, contradicting the hypothesis that $P$ is efficient.\(^{10}\)

Finally, note that the slope of the trade-off between the expected return and risk for the whole of any efficient portfolio, $P$, is equal to $(\mu_P - r_0) / \sigma_P$. The reason for this is that the portfolio $P$ can itself be interpreted as a single composite asset for which the necessary condition must hold. The rate of return on the whole portfolio must, by definition, always be perfectly positively correlated with itself, so that $\beta_{PP} = 1$, and the result follows.

The above analysis can be summarized by writing down the conditions that must necessarily hold for every efficient portfolio, $P$:

$$\frac{\mu_1 - r_0}{\beta_{1P}\sigma_P} = \frac{\mu_2 - r_0}{\beta_{2P}\sigma_P} = \ldots = \frac{\mu_n - r_0}{\beta_{nP}\sigma_P} = \frac{\mu_P - r_0}{\sigma_P}$$ \hspace{1cm} (5.3)

Notice that the $\sigma_P$ terms in the denominator could be cancelled out in (5.3).

\(^{10}\) It is, of course, not permissible to divide by $\beta_{jp}$ if $\beta_{jp} = 0$. This case (which is of some interest in chapter 6) can be handled without difficulty because it can be shown that, if $\beta_{jp} = 0$, the expected rate of return on asset $j$ equals the risk-free rate, $\mu_j = r_0$. Also ignored are ‘corner solutions’ that arise when a zero quantity of some asset is held in an efficient portfolio. These cases – which can occur if short-sales (i.e. negative asset holdings) are prohibited – involve replacing the first-order equalities with inequalities.
Equation (5.3) holds for every efficient portfolio. Hence, it holds for $Z$, the portfolio comprising only risky assets. Equation (5.3) then becomes

$$
\frac{\mu_1 - r_0}{\beta_{1Z}} = \frac{\mu_2 - r_0}{\beta_{2Z}} = \cdots = \frac{\mu_n - r_0}{\beta_{nZ}} = \mu_Z - r_0 \tag{5.4}
$$

Note that the common term, $\sigma_Z$, has been cancelled out in going from (5.3) to (5.4). This is just for convenience. Conditions (5.4) are very important. They lie at the heart of the CAPM. But there is more to the CAPM than the equalities of (5.4). The extra conditions are studied in chapter 6.

### 5.4.3 The Sharpe ratio and risk-adjusted performance

The Sharpe ratio (named after its originator, William Sharpe) for any asset, or portfolio of assets, $j$, is defined by

$$
s_j = \frac{\mu_j - r_0}{\sigma_j}
$$

In words, $s_j$ denotes the expected excess return on asset $j$ normalized by its standard deviation. In practice, the Sharpe ratio would be measured by substituting the sample mean rate of return for $\mu_j$ and the sample standard deviation for $\sigma_j$. It provides a way of comparing assets with differing expected returns and risks (risk being identified with the value of $\sigma_j$).

For a graphical interpretation, figure 5.7 reproduces figure 5.5, with the FF frontier omitted. Consider any asset, say asset 1, and draw a ray from $r_0$ on the vertical axis to the point labelled $A_1$ $(\mu_1, \sigma_1)$. The slope of this line equals the Sharpe ratio for asset 1, $s_1 = (\mu_1 - r_0)/\sigma_1$. The point $A_2$ identifies a second asset with a lower Sharpe ratio.

Suppose that $s_e$ denotes the Sharpe ratio for an efficient portfolio. All efficient portfolios share the same Sharpe ratio, which equals the slope of the line segment $r_0E$ in figure 5.7. From the diagram it can be seen that $s_P \leq s_e$ for any asset or portfolio of assets, $P$, whether or not $P$ is efficient.

The risk-adjusted performance (RAP) is derived from the Sharpe ratio (see Modigliani and Modigliani (1997)). To define the RAP, suppose that a ‘benchmark portfolio’ is identified. In principle, this can be any portfolio, but typically it denotes a portfolio composed of a broad range of assets. Let $\sigma_B$ denote its standard deviation of return. The RAP for any asset $j$ is defined as

$$
RAP_j = r_0 + \frac{\sigma_B}{\sigma_j} (\mu_j - r_0)
$$
or $RAP_j = r_0 + \sigma_B s_j$, substituting $j$’s Sharpe ratio. In words, $RAP_j$ would be the expected rate of return on asset $j$ if its risk equalled that of the benchmark but its Sharpe ratio remained unchanged. Notice, for the illustration in figure 5.7, the Sharpe ratio for asset 1 exceeds that of asset 2 even though asset 2 has a higher expected rate of return.

In practice, sample means and variances are substituted for the theoretical values and RAP values are calculated to compare the performance of assets (or portfolios of assets) relative to the benchmark portfolio.

### 5.5 Optimal portfolio selection in the mean-variance model

What we have done so far is to trace out the portfolio frontier and identify the set of efficient portfolios. With agreement about means and variances, the efficient set of portfolios is the same for each investor. There is still scope, however, for different investors to choose different portfolios. The reason is that investors can differ in their preferences, commonly interpreted (in portfolio selection) as reflecting different investors’ attitudes to risk.

In view of the mutual fund theorems outlined in previous sections, it is sufficient to assume that the portfolio is selected as a combination of exactly two assets.
The indifference curves express the investor's preferences among different values of $\mu_p$ and $\sigma_p$. The most preferred (i.e. optimal) portfolio lies on the highest indifference curve that is attainable subject to being feasible (i.e. on or below the $r_0ZE$ line, denoting the efficient set). For the preferences shown here, the optimum is at point $O$.

Consider the case in which a risk-free asset is available. In figure 5.5 the set of efficient portfolios is represented by the straight line passing through $r_0$ and tangent, at $Z$, to the portfolio frontier (for risky assets). See figure 5.8, in which the efficient portfolio set is depicted by the line $r_0ZE$ tangent to the frontier $FF$ at $Z$, the efficient portfolio comprising only risky assets.

Any efficient portfolio can be constructed as a combination of the risk-free asset and the portfolio of risky assets depicted by the point $Z$. Given the indifference curves shown in figure 5.8, the optimum portfolio is depicted by point $O$.

The trade-off between $\mu_p$ and $\sigma_p$, expressed by the line $r_0ZE$ in figure 5.8, can be written as

$$\mu_p = r_0 + \frac{\mu_Z - r_0}{\sigma_Z} \sigma_p$$

That is, the relationship between $\mu_p$ and $\sigma_p$ is linear, with intercept $r_0$ and slope equal to $(\mu_Z - r_0)/\sigma_Z$.

Suppose now that preferences can be expressed as $G(\mu_p, \sigma_p^2) = \mu_p - \alpha \sigma_p^2$, where $\alpha > 0$ expresses attitude to risk and, hence, can differ between investors. (Recall the discussion in chapter 4, pages 103–04.)
Assuming that preferences take this form, it can be shown that the proportion of initial wealth invested in the risky asset, \( q \), is given by

\[
q = \frac{\mu_Z - r_0}{2\alpha\sigma^2_Z} \tag{5.5}
\]

(See appendix 5.5 for the derivation.)

Although this formula holds only for the objective function \( G(\mu_p, \sigma^2_p) = \mu_p - \alpha\sigma^2_p \), it is useful in practical applications. The result confirms intuition that: (a) the greater the excess expected return, \( \mu_Z - r_0 \), the greater the holding of the risky asset; (b) the riskier the risky asset (i.e. the higher the level of \( \sigma^2_Z \)), the lower the holding of the risky asset; and (c) the greater the risk tolerance (i.e. the smaller is \( \alpha \)), the higher the holding of the risky asset.

5.6 Summary

This chapter has shown that the following implications can be derived for investors whose beliefs about asset returns are expressed in terms of means and variances.

1. The portfolio frontier for risky assets alone is obtained by minimizing risk, \( \sigma_p \) or \( \sigma^2_p \), for each level of expected return, \( \mu_p \), and is a hyperbola in the \((\mu_p, \sigma_p)\) plane.

Consider any two distinct portfolios, \( A \) and \( B \), on the frontier. Then every portfolio on the frontier can be expressed as a combination of \( A \) and \( B \). The efficient portfolios are frontier portfolios for which expected return is maximized at a given level of risk.

2. When a risk-free asset is available, the set of efficient portfolios is described by a straight line, with intercept at the risk-free rate and tangential to the risky asset frontier. Every efficient portfolio can be expressed as a linear combination of any two distinct efficient portfolios. The conditions that characterize the set of efficient portfolios form the basis for the CAPM, studied in chapter 6.

3. The set of efficient portfolios depends only on investors’ beliefs about rates of return (expressed by means and variances), not on preferences with respect to risk. Investors with different risk preferences select different portfolios but always choose a member of the efficient set.

Further reading

There exist many expositions of mean-variance analysis at various levels and directed towards different audiences. For readers seeking a more extended, management-oriented approach, the following are representative: Grinblatt and Titman (2001, chaps. 4 & 5); Sharpe, Alexander and Bailey (1999, chaps. 7–9). An excellent exposition appears in Elton, Gruber, Brown and Goetzmann (2003, chaps. 4–6). At a more advanced level, a rigorous treatment can be found in
Huang and Litzenberger (1988, chap. 4). This reference is particularly useful for sorting out awkward special cases. An early and not very well-known analysis of the portfolio frontier, still worth attention, is provided by Merton (1972). The most comprehensive and detailed examination of mutual fund theorems remains that of Cass and Stiglitz (1970).

Appendix 5.1: Numerical example: two risky assets

The following example corresponds to section 5.2 and focuses on the special cases of perfect positive and negative correlations between assets’ rates of return. Suppose that \( \mu_1 = 0.25, \mu_2 = 0.50, \sigma_1 = 0.20 \) and \( \sigma_2 = 0.40 \). Hence

\[
\mu_p = 0.25a + 0.50(1 - a) = 0.50 - 0.25a
\]

The special cases, \( \rho_{12} = \pm 1 \), result in

\[
\rho_{12} = +1 \implies \sigma_p = 0.20a + 0.40(1 - a) \\
\rho_{12} = -1 \implies \sigma_p = (0.20a - 0.40(1 - a)) \geq 0 \\
\text{for } a \geq \frac{0.40}{0.20 + 0.40} = \frac{2}{3}
\]

\[
\rho_{12} = -1 \implies \sigma_p = -(0.20a - 0.40(1 - a)) \geq 0 \\
\text{for } a \leq \frac{0.40}{0.20 + 0.40} = \frac{2}{3}
\]

For \( \rho_{12} = +1 \)

\[
\sigma_p = 0.20a + 0.40(1 - a) = 0.40 - 0.20a \\
a = \frac{0.40 - \sigma_p}{0.20}
\]

\[
\mu_p = 0.50 - 0.25 \left( \frac{0.40 - \sigma_p}{0.20} \right) = 1.25\sigma_p
\] (5.6)

For \( \rho_{12} = -1 \) and \( a \geq 2/3 \)

\[
\sigma_p = 0.20a - 0.40(1 - a) = 0.60a - 0.40 \\
a = \frac{\sigma_p + 0.40}{0.60}
\]

\[
\mu_p = 0.50 - 0.25 \left( \frac{\sigma_p + 0.40}{0.60} \right) = \frac{1}{3} - \frac{5}{12}\sigma_p
\] (5.7)
For $\rho_{12} = -1$ and $a \leq 2/3$

$$\sigma_p = -0.20a + 0.40(1 - a) = 0.40 - 0.60a$$

$$a = \frac{0.40 - \sigma_p}{0.60}$$

$$\mu_p = 0.50 - 0.25 \left( \frac{0.40 - \sigma_p}{0.60} \right) = \frac{1}{3} + \frac{5}{12} \sigma_p$$ (5.8)

The equations (5.6), (5.7) and (5.8) define the three sides of the triangle in a diagram such as figure 5.2 on page 119.

**Appendix 5.2: Variance minimization: risky assets only**

This appendix explores the properties of the portfolio frontier when the investor chooses among $n$ risky assets, beginning in A.5.2.1 with the formal minimization of portfolio variance. Appendix section A.5.2.2 offers a graphical interpretation, while A.5.2.3 describes a special case (relevant for the capital asset pricing model, chapter 6) that compares portfolios, the rates of return on which are uncorrelated.

**A.5.2.1: The portfolio frontier**

Form the Lagrangian

$$\mathcal{L} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij} + \gamma \left( \mu_p - \sum_{j=1}^{n} \mu_j a_j \right) + \lambda \left( \sum_{j=1}^{n} a_j - 1 \right)$$

where $\gamma$ and $\lambda$ are Lagrange multipliers.

Noting that the minimization is carried out for each value of $\mu_p$, the first-order conditions are found by partially differentiating $\mathcal{L}$ with respect to $a_1, a_2, \ldots, a_n, \gamma, \lambda$, and setting the resulting expressions to zero:

$$\frac{\partial \mathcal{L}}{\partial a_j} = 2 \sum_{i=1}^{n} a_i \sigma_{ij} - \gamma \mu_j + \lambda = 0 \quad j = 1, 2, \ldots, n$$ (5.9)

$$\frac{\partial \mathcal{L}}{\partial \gamma} = \mu_p - \sum_{j=1}^{n} \mu_j a_j = 0$$ (5.10)

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{j=1}^{n} a_j - 1 = 0$$ (5.11)
The second-order conditions are messy and not very revealing. Given the quadratic minimand (i.e. the variance, $\sigma^2_P$) and the linear constraints, it is tedious, though not difficult, to check that the second-order conditions are indeed satisfied.

The optimum portfolio proportions can be found by solving the $n + 2$ first-order conditions for the portfolio proportions and the values of $\gamma$ and $\lambda$. Instead of deriving the explicit solution, it is more instructive to interpret the first-order conditions. Multiply through (5.9) by $a_j$ and sum over $j$ to give

$$2 \sum_{j=1}^{n} \sum_{i=1}^{n} a_j a_i \sigma_{ij} - \gamma \sum_{j=1}^{n} a_j (\mu_j - \omega) = 0$$

(5.12)

where $\omega \equiv \lambda / \gamma$.

Substituting from (5.10) and (5.11) into (5.9) and (5.12) and rearranging gives

$$\sigma^2_P = (\gamma/2)(\mu_P - \omega)$$

(5.13)

$$\sigma_{jp} = (\gamma/2)(\mu_j - \omega) \quad \text{for } j = 1, 2, \ldots, n$$

(5.14)

where $\sigma_{jp} = \sum_{i=1}^{n} a_i \sigma_{ij}$ is the covariance of the return from asset $j$ with portfolio $P$. Combining (5.13) and (5.14) to eliminate $\gamma$ yields a condition that must hold for all portfolios on the frontier:

$$\frac{\mu_j - \omega}{\sigma_{jp}} = \frac{\mu_P - \omega}{\sigma^2_P} \quad \text{for } j = 1, 2, \ldots, n$$

Given that the first-order conditions are linear in $a_j$, the solution for the variance-minimizing value of each $a_j$ is unique. A sketch of this result is as follows. From (5.14) solve for the $a_j$ as functions of $\sigma_{ij}, \mu_j$ ($i, j = 1, 2, \ldots, n$), $\gamma$ and $\omega$. That this solution is unique follows from the linearity of the equations and the assumption that there are no exact linear dependencies among the returns on the $n$ assets (i.e. the assets are genuinely different from one another). Next, substituting the $a_j$ into the two constraints and rearranging yields a unique solution for each of $\gamma$ and $\omega$ as a function of the $\sigma_{ij}, \mu_j$ and $\mu_P$.

Finally, substitute the values of $\gamma$ and $\omega$ back into the initial solutions for $a_j$, which now become functions of the $\sigma_{ij}, \mu_j$ and $\mu_P$ as required (i.e. $\gamma$ and $\omega$ have now been eliminated). The resulting expressions for $a_j$ are messy and not otherwise needed here; hence, they are omitted.

**A.5.2.2: A graphical interpretation**

An interpretation of $\omega$ can be given with reference to figure 5.9. Choose an efficient portfolio and denote it by $P$. The slope of the frontier, $d\mu_P/d\sigma_P$, at
Choose any portfolio on the frontier – say, $P$ – and draw a straight line tangent to $FF$. The line intersects the vertical axis at a point such as $\omega$. All feasible portfolios (i.e. on or within $FF$) with an expected return equal to $\omega$ (i.e. along the horizontal line starting at $W$) have rates of return that are uncorrelated with the return to portfolio $P$.

$P$ is equal to $(\mu_P - \omega)/\sigma_P$. This result can be obtained by, first, invoking the envelope theorem to obtain $d\sigma_P^2/d\mu_P = \gamma$.\(^{11}\) Hence,

$$
\frac{d\sigma_P^2}{d\mu_P} = \frac{d\sigma_P^2}{d\sigma_P} \frac{d\sigma_P}{d\mu_P} = 2\sigma_P \frac{d\sigma_P}{d\mu_P} = \gamma
$$

(5.15)

Rearranging gives

$$
\frac{d\mu_P}{d\sigma_P} = \sigma_P \frac{2}{\gamma} = \sigma_P \frac{\mu_P - \omega}{\sigma_P^2} = \frac{\mu_P - \omega}{\sigma_P}
$$

(5.16)

where (5.13) has been substituted to eliminate $2/\gamma$.

\(^{11}\) In this context the envelope theorem implies that $d\sigma_P^2/d\mu_P = \partial \mathcal{L}/\partial \mu_P$, where the total derivative serves as a reminder that the minimizing portfolio proportions and the Lagrange multipliers change in response to an incremental change in $\mu_P$. A derivation from first principles – tantamount to proof of the envelope theorem – involves totally differentiating $\mathcal{L}$ with respect to $\mu_P$, allowing for changes in the optimal portfolio proportions in response to the change in $\mu_P$. Substitution from the first-order conditions then yields $d\sigma_P^2/d\mu_P = \gamma$. For treatments of the envelope theorem, see Samuelson (1947, pp. 34–6), or The New Palgrave Dictionary of Money and Finance (Newman, Milgate and Eatwell, 1992, Vol. II, pp. 158–9).
A.5.2.3: Portfolios with uncorrelated returns

In the capital asset pricing model (see chapter 6, section 6.6) it is necessary to identify portfolios, the rates of return on which have a zero correlation with any given frontier portfolio. The following derivation shows that all such portfolios have an expected return equal to ω. (While the value of ω depends on the frontier portfolio, this dependence is not made explicit in the notation.)

Consider any portfolio (it need not be on the frontier) denoted by W with asset proportions \( w_1, w_2, \ldots, w_n \), rate of return \( r_W = \sum w_j r_j \) and expected rate of return \( \mu_W = \sum w_j \mu_j \) (see figure 5.9). The asset proportions in frontier portfolio P are denoted by \( a_1, a_2, \ldots, a_n \), with rate of return \( r_P = \sum a_j r_j \) and expected rate of return \( \mu_P = \sum a_j \mu_j \).

The covariance between \( r_P \) and \( r_W \), \( \text{cov}(r_P, r_W) \), is given by

\[
\text{cov}(r_P, r_W) = E[(r_P - \mu_P)(r_W - \mu_W)] = E\left[ \sum_{i=1}^{n} a_i (r_i - \mu_i) \sum_{j=1}^{n} w_j (r_j - \mu_j) \right] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i w_j \sigma_{ij} = \sum_{j=1}^{n} w_j \sum_{i=1}^{n} a_i \sigma_{ij} = \sum_{j=1}^{n} w_j \sigma_{jp} \text{ (5.17)}
\]

where the last equality follows from (5.14) because \( P \) denotes a frontier portfolio. If \( r_P \) and \( r_W \) are uncorrelated, \( \text{cov}(r_P, r_W) = 0 \). It then follows from (5.17) that \( \mu_W = \omega \) (because \( \sum w_j (\mu_j - \omega) = \mu_W - \omega \sum w_j \), and portfolio proportions sum to unity). Referring back to figure 5.9, the expected return on any portfolio uncorrelated with \( P \) is equal to \( \omega \). These portfolios are located along the horizontal line beginning at point \( W \).

It is now possible to demonstrate that the expected return, \( \mu_{mrrp} \), on the portfolio with global minimum risk (depicted by \( MRP \) in figure 5.9) must satisfy \( \mu_{mrrp} > \omega \), given that \( \mu_P > \mu_{mrrp} \) (because \( P \) is on the upward-sloping portion of \( FF \)). Construct a new portfolio with proportion \( \theta \) of the portfolio denoted by \( P \) and
(1 − θ) of W. From the first mutual fund theorem, any frontier portfolio can be constructed by an appropriate choice of θ. Let S denote the variance of the rate of return on this portfolio, so that

\[ S = \theta^2 \sigma_P^2 + (1 - \theta)^2 \sigma_W^2 \]

where \( \sigma_P^2 \) and \( \sigma_W^2 \) are the variances corresponding to \( P \) and \( W \), respectively. Note that no covariance term appears because, by construction, the returns on \( P \) and \( W \) are uncorrelated.

The variance on the MRP portfolio is found by choosing \( \theta \) to minimize \( S \). Differentiating with respect to \( \theta \) and setting the resulting expression to zero provides the first-order condition

\[ \frac{dS}{d\theta} = 2\theta \sigma_P^2 - 2(1 - \theta)\sigma_W^2 = 0 \]

The second-order condition for a minimum, \( d^2S/d\theta^2 = 2(\sigma_P^2 + \sigma_W^2) > 0 \), is certainly satisfied. The first-order condition can readily be solved to give \( \theta_{mrp} = \sigma_W^2/(\sigma_P^2 + \sigma_W^2) \). Therefore, \( 0 < \theta_{mrp} < 1 \). Hence, it follows that \( \mu_{mrp} = \theta_{mrp} \mu_P + (1 - \theta_{mrp}) \omega > \omega \) as claimed, and as depicted in figure 5.9.

**Appendix 5.3: Variance minimization with a risk-free asset**

Although the analysis in the presence of a risk-free asset closely resembles that in its absence, the outline below is self-contained as far as it goes.\(^\text{12}\) It is instructive to recognize that \( \omega \) and \( r_0 \) play exactly analogous roles (see appendix 5.2).

Note that \( a_0 = 1 - \sum_{j=1}^{n} a_j \), so that

\[ \mu_p = \sum_{j=1}^{n} \mu_j a_j + a_0 r_0 \]

\[ = \sum_{j=1}^{n} \mu_j a_j + \left( 1 - \sum_{j=1}^{n} a_j \right) r_0 \]

\[ = \sum_{j=1}^{n} (\mu_j - r_0) a_j + r_0 \]

Form the Lagrangian

\[ \mathcal{L} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij} + \gamma \left( \mu_p - \sum_{j=1}^{n} (\mu_j - r_0) a_j - r_0 \right) \]

\(^\text{12}\) The second-order conditions are ignored for exactly the same reason as in appendix 5.2.
where $\gamma$ is a Lagrange multiplier. The value of $\mu_P$ is a parameter in this optimization – i.e. a separate minimization is carried out for each value of $\mu_P$.

The first-order conditions are found by partially differentiating the Lagrangian, $L$, with respect to $a_1, a_2, \ldots, a_n, \gamma$ and setting the resulting expressions to zero:

$$\frac{\partial L}{\partial a_j} = 2 \sum_{i=1}^{n} a_i \sigma_{ij} - \gamma (\mu_j - r_0) = 0 \quad j = 1, 2, \ldots, n \quad (5.18)$$

$$\frac{\partial L}{\partial \gamma} = \mu_P - \sum_{j=1}^{n} (\mu_j - r_0) a_j - r_0 = 0 \quad (5.19)$$

The optimum portfolio proportions (one set for each value of $\mu_P$) are found by solving these $n+1$ conditions for the portfolio proportions and the value of $\gamma$. Multiply through (5.18) by $a_j$ and sum over $j$ to give

$$2 \sum_{j=1}^{n} \sum_{i=1}^{n} a_j a_i \sigma_{ij} - \gamma \sum_{j=1}^{n} a_j (\mu_j - r_0) = 0 \quad (5.20)$$

Substituting from (5.19), equations (5.20) and (5.18) can be written as

$$\sigma_P^2 = (\gamma/2)(\mu_P - r_0) \quad (5.21)$$

$$\sigma_{jp} = (\gamma/2)(\mu_j - r_0) \quad \text{for } j = 1, 2, \ldots, n \quad (5.22)$$

where $\sigma_{jp} = \sum_{i=1}^{n} a_i \sigma_{ij}$ is the covariance of the return on asset $j$ with the return on the portfolio $P$.

Elimination of $\gamma$ from (5.21) and (5.22) yields

$$\frac{\mu_j - r_0}{\sigma_{jp}} = \frac{\mu_P - r_0}{\sigma_P^2} \quad \text{for } j = 1, 2, \ldots, n \quad (5.23)$$

Inserting the definition of $\beta_{jp} \equiv \sigma_{jp}/\sigma_P^2$, equations (5.23) provide the necessary conditions that hold for all efficient portfolios (see (5.3) on page 129 in section 5.4).

**Appendix 5.4: Derivation of $\Delta \sigma_P = \beta_{jp} \sigma_P \Delta a_j$**

The standard deviation of the return on any portfolio is defined by

$$\sigma_P = \left[ \sigma_P^2 \right]^{\frac{1}{2}} = \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij} \right]^{\frac{1}{2}}$$
Portfolio selection: the mean-variance model

141

Partially differentiate with respect to \( a_j \) to give

\[
\frac{\partial \sigma_p}{\partial a_j} = \frac{1}{2} \left( \sum \sum a_i a_j \sigma_{ij} \right)^{\frac{1}{2}-1} \left[ 2 \sum_{i=1}^{n} a_i \sigma_{ij} \right] = \left[ \sigma_p^2 \right]^{\frac{1}{2}} \sigma_{jp} = \frac{\sigma_{jp}}{\sigma_p} = \frac{\sigma_{jp}}{\sigma_p^2} \sigma_p
\]

\[
= \beta_{jp} \sigma_p
\]

(5.24)

Therefore, using \( \Delta \) to denote small, discrete, changes, (5.24) can be approximated by \( \Delta \sigma_p \approx \beta_{jp} \sigma_p \Delta a_j \), which appears as an equality – expression (5.2) on page 128 in section 5.4.2.

Appendix 5.5: The optimal portfolio with a single risky asset

Suppose that the investor chooses a risk-free asset and a single risky asset with expected return \( \mu_Z \) and variance of return \( \sigma_Z^2 \). (Typically, the risky asset will itself be a portfolio of individual risky assets.) Let \( q \) denote the proportion of the portfolio invested in the risky asset. It follows immediately that the expectation \( \mu_P \) and variance \( \sigma_P^2 \) of returns on the whole portfolio are

\[
\mu_P = q \mu_Z + (1-q)r_0 \quad \text{and} \quad \sigma_P^2 = q^2 \sigma_Z^2
\]

(5.25)

because the variance of the risk-free rate of return is zero.

In this case it is possible to obtain a very simple form for the trade-off between \( \mu_P \) and \( \sigma_P \). From (5.25), \( \sigma_P = q \sigma_Z \) and \( \mu_P = r_0 + (\mu_Z - r_0)q \). Eliminating \( q \) gives

\[
\mu_P = r_0 + \frac{\mu_Z - r_0}{\sigma_Z} \sigma_P
\]

(5.26)

as described above in section 5.5. Note that the investor’s optimization problem can be viewed as one of choosing \( \mu_P \) and \( \sigma_P \) to maximize \( G(\mu_P, \sigma_P^2) \) subject to the constraint (i.e. the trade-off) given by (5.26).

Given the functional form \( G(\mu_P, \sigma_P^2) = \mu_P - \alpha \sigma_P^2 \), the maximization is achieved by eliminating \( \mu_P \) and \( \sigma_P \) using (5.25) – that is

\[
G = q \mu_Z + (1-q)r_0 - \alpha q^2 \sigma_Z^2
\]

Differentiating with respect to the investor’s choice variable, \( q \), and setting the resulting expression to zero yields the first-order condition

\[
\mu_Z - r_0 - 2\alpha q \sigma_Z^2 = 0
\]

(5.27)

The second-order condition for a maximum, \( d^2 G/dq^2 = -2\alpha \sigma_Z^2 < 0 \), is satisfied if \( \alpha > 0 \), as assumed.

Rearranging the first-order condition, (5.27), to solve for \( q \) results in equation (5.5) on page 133.


