CHAPTER 6

Risk Aversion and Investment Decisions, Part II: Modern Portfolio Theory

Chapter Outline

Intermediate Financial Theory.

6.1 Introduction

In the context of the previous chapter, we encountered the following canonical portfolio problem:

$$\max_a EU(\tilde{Y}_1) = \max_a EU[Y_0(1+r_\text{f}) + a(\tilde{r} - r_\text{f})] \tag{6.1}$$

Here the portfolio choice is limited to allocating investable wealth, Y_0, between a risk-free and a risky asset, a being the amount invested in the latter.

More generally, we can admit N risky assets, with returns $(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_N)$, as in the Cass–Stiglitz theorem. The above problem in this case becomes

$$\max_{\{a_1,a_2,\ldots,a_N\}} EU\left(Y_0(1+r_\text{f}) + \sum_{i=1}^{N} a_i(\tilde{r}_i - r_\text{f})\right) = \max_{\{w_1,w_2,\ldots,w_N\}} EU\left(Y_0(1+r_\text{f}) + \sum_{i=1}^{N} w_i Y_0(\tilde{r}_i - r_\text{f})\right) \tag{6.2}$$

Equation (6.2) reexpresses the problem with $w_i = (a_i/Y_0)$, the proportion of wealth invested in the risky asset i, being the key decision variable rather than a_i, the amount of money invested.

The latter expression may further be written as

$$\max_{\{w_1,w_2,\ldots,w_N\}} EU\left\{ Y_0\left[(1+r_\text{f}) + \sum_{i=1}^{N} w_i(\tilde{r}_i - r_\text{f})\right]\right\} = EU\{Y_0[1+\tilde{r}_P]\} = EU\{\tilde{Y}_1\} \tag{6.3}$$

where $\tilde{Y}_1$ denotes the end-of-period wealth and $\tilde{r}_P$ the rate of return on the overall portfolio of assets held.

Modern portfolio theory (MPT) explores the details of portfolio choice such as Problem (6.3), (i) under the mean–variance utility hypothesis and (ii) for an arbitrary number of risky investments, with or without a risk-free asset.[1] The goal of this chapter is to review the fundamentals underlying this theory. We first draw the connection between the mean–variance utility hypothesis and our earlier utility development.

6.2 More About Utility Functions and Return Distributions

What provides utility? As noted in Chapter 3, financial **theory** assumes that the ultimate source of a consumer's satisfaction lies in consuming the goods and services he is able to

[1] Markowitz (1952) developed the theory for portfolios exclusively composed of risky assets. Tobin (1958) added a risk-free asset.

purchase.[2] Preference relations and utility functions are accordingly defined on bundles of consumption goods (recall Theorem 3.1):

$$u(c_1, c_2, \ldots, c_M) \tag{6.4}$$

where the indexing $i = 1, \ldots, M$ is across date-state (contingent) commodities: goods are characterized not only by their identity as a product or service but also by the time and state in which they may be consumed. States of nature, however, are mutually exclusive. For each date and state of nature (θ), there is a traditional budget constraint

$$p_{1\theta}c_{1\theta} + p_{2\theta}c_{2\theta} + \cdots + p_{M\theta}c_{M\theta} \leq Y_\theta \tag{6.5}$$

where the indexing runs across goods for a given state θ; in other words, the M quantities $c_{i\theta}$, $i = 1, \ldots, M$, and the M prices $p_{i\theta}$, $i = 1, \ldots, M$ correspond to the M goods available in state of nature θ, while Y_θ is the ("end-of-period") wealth level available in that same state. We quite naturally assume that the number of goods available in each state is constant.[3]

In this context, and in some sense summarizing what we discussed in Chapter 5, it is quite natural to think of an individual's decision problem as being undertaken sequentially, in three steps.

Step 1, The consumption–savings decision. Here, the issue is deciding how much to consume versus how much to save today: how to split period zero income Y_0 between current consumption spending C_0 and saving S_0 for consumption in the future where

$$C_0 + S_0 = Y_0$$

Step 2, The portfolio problem. At this second step, the problem is to choose assets in which to invest one's savings so as to obtain the desired pattern of end-of-period wealth across the various states of nature. This means, in particular, allocating $(Y_0 - C_0)$ between the risk-free and the N risky assets with $\left(1 - \sum_{i=1}^{N} w_i\right)(Y_0 - C_0)$ representing the investment in the risk-free asset, and $(w_1(Y_0 - C_0), w_2(Y_0 - C_0), \ldots, w_N(Y_0 - C_0))$, representing the vector of investments in the various risky assets.

Step 3, Tomorrow's consumption choice. Given the realized state of nature and the level of wealth obtained, there remains the issue of choosing consumption bundles to maximize the utility function (Expression (6.4)) subject to Constraint (6.5) where

[2] Of course this does not mean that nothing else in life provides utility or satisfaction (!) but the economist's inquiry is normally limited to the realm of market phenomena and economic choices.

[3] This is purely formal: if a good is not available in a given state of nature, it is understood nevertheless to exist but with a total, tradable economy-wide endowment of the good being zero.

$$Y_\theta = (Y_0 - C_0)\left[(1 + r_f) + \sum_{i=1}^{N} w_i(r_{i\theta} - r_f)\right]$$

and $r_{i\theta}$ denotes the *ex post* return to asset i in state θ.

In such problems, it is fruitful to work via backward induction, starting from the end (step 3). Step 3 is a standard microeconomic problem, and for our purpose its solution can be summarized by a utility-of-money function $U(Y_\theta)$ representing the (maximum) level of utility that results from optimizing in step 3 given that the wealth available in state θ is Y_θ.

In other words, we define $U(Y_\theta)$ as

$$U(Y_\theta) \equiv \max_{(c_{1\theta},\ldots,c_{M\theta})} u(c_{1\theta},\ldots,c_{M\theta})$$

$$\text{s.t. } p_{1\theta}c_{1\theta} + \cdots + p_{M\theta}c_{M\theta} \leq Y_\theta$$

Naturally enough, maximizing the expected utility of Y_θ across all states of nature becomes the objective of step 2:

$$\max_{\{w_1,w_2,\ldots,w_N\}} EU(\tilde{Y}) = \sum_\theta \pi_\theta U(Y_\theta)$$

Here π_θ is the probability of state of nature θ. The end-of-period wealth (a random variable) can now be written as $\tilde{Y} = (Y_0 - C_0)(1 + \tilde{r}_P)$, with $(Y_0 - C_0)$ the initial wealth net of date 0 consumption and $\tilde{r}_P = r_f + \sum_{i=1}^{N} w_i(\tilde{r}_i - r_f)$ the rate of return on the portfolio of assets in which $(Y_0 - C_0)$ is invested. This brings us back to Eq. (6.3):

$$\max_{w_1,\ldots,w_N} EU(\tilde{Y}) = \max_{w_1,\ldots,w_N} EU((Y_0 - C_0)(1 + \tilde{r}_P))$$

In general, the optimal portfolio allocation decision and the optimal consumption savings decision must be considered jointly. In the case that $U(Y)$ has the constant relative risk aversion (CRRA) form (and, more generally, if $U(Y)$ is homogeneous[4]), however, these decisions can be separated and undertaken sequentially with the former preceding the latter.[5] To confirm this assertion, observe that if

$$U(Y) = \frac{1}{1-\gamma}(Y)^{1-\gamma}$$

[4] A utility function $U(Y)$ is said to be homogeneous if and only if it has the property that for any $\Delta > 0$, $U(\Delta Y) = \Delta^v U(Y)$ for some v, a real number. As such, $U(Y)$ is said to be homogeneous of degree v.

[5] Footnote 1 of Chapter 4 details the form of $u(c_1,\ldots, c_M)$ such that $U(Y)$ is CRRA.

$$\begin{aligned}
\max_{w_1,\ldots,w_N} EU(\tilde{Y}) &= \max_{w_1,\ldots,w_N} E\left(\frac{1}{1-\gamma}((Y_0-C_0)(1+\tilde{r}_P))^{1-\gamma}\right) \\
&= \max_{w_1,\ldots,w_N} E\left((Y_0-C_0)^{1-\gamma}\left(\frac{1}{1-\gamma}\right)(1+\tilde{r}_P)^{1-\gamma}\right) \\
&= \left\{(Y_0-C_0)^{1-\gamma}\cdot \max_{w_1,\ldots,w_N} E\hat{U}(1+\tilde{r}_P)\right\}
\end{aligned} \tag{6.6}$$

where $\hat{U}(1+r_P) = ((1+r_P)^{1-\gamma})/(1-\gamma)$. Although seemingly defined over rates of return, $\hat{U}(\ )$ technically remains a utility-of-money function because its domain represents the wealth accruing to the investor in period 1 if he saves and invests \$1.00 in period zero at the rate r_P. Note also that decomposition (6.6) allows the investor to complete, first, step 2 of his overall decision problem and then step 1 as per backward induction. Step 1 requires that the investor choose his level of savings $(Y_0 - C_0)$ so as optimally to trade off utility at $t = 0$ against expected utility at $t = 1$ given that each unit of savings is invested optimally, i.e., in a portfolio whose proportions are the solution to Problem (6.6).

Accordingly, the investor's comprehensive savings–portfolio composition problem becomes

$$\max_{C_0, w_1,\ldots,w_N} \frac{(C_0)^{1-\gamma}}{1-\gamma} + \delta\left\{(Y_0-C_0)^{1-\gamma} \max_{w_1,\ldots,w_N} E\hat{U}(1+\tilde{r}_P)\right\}$$

as per our discussion in Section 5.6.3. In particular, the investor's optimal portfolio proportions will be the same irrespective of the amount he decides to invest.

The remainder of this chapter will focus exclusively on the determination of optimal portfolio proportions, i.e., on step 2 above (for step 1, see Chapter 5). Accordingly, we will assume $U(Y)$ is homogeneous, drop the $U(Y)$, $\hat{U}(r_P)$ distinction, and consider utility functions $U(\ )$ defined interchangeably over r_P or Y. We next take advantage of a very useful mathematical approximation.

Using a simple Taylor series approximation, one can see that the mean and variance of an investor's wealth distribution are the critical elements to the determination of his expected utility *for any distribution* and any concave utility function. Let $\tilde{Y}$ denote an investor's uncertain end period wealth, and $U(\cdot)$ as his utility-of-money function. The Taylor series approximation for the investor's utility of wealth $U(Y)$ around $E(\tilde{Y})$ yields

$$U(Y) = U\left[E(\tilde{Y})\right] + U'\left[E(\tilde{Y})\right]\left[Y - E(\tilde{Y})\right] + \frac{1}{2}U''\left[E(\tilde{Y})\right][Y-E(\tilde{Y})]^2 + H_3 \tag{6.7}$$

where $\tilde{H}_3(Y) = \sum_{j=3}^{\infty}\frac{1}{j!}U^{(j)}\left[E(\tilde{Y})\right][Y-E(\tilde{Y})]^j$.

Now let us compute expected utility using this approximation:

$$EU(\tilde{Y}) = U[E(\tilde{Y})] + U'[E(\tilde{Y})]\,\underbrace{[E(\tilde{Y}) - E(\tilde{Y})]}_{=0}$$
$$+\frac{1}{2}U''\left[\mathrm{E}(\tilde{Y})\right]\,\underbrace{E[\tilde{Y}-E(\tilde{Y})]^2}_{=\sigma^2(\tilde{Y})} + E\tilde{H}_3$$
$$= U\left[E(\tilde{Y})\right] + \frac{1}{2}U''\left[E(\tilde{Y})\right]\sigma^2(\tilde{Y}) + E\tilde{H}_3$$

If $E\tilde{H}_3$ is small, $E(\tilde{Y})$ and $\sigma^2(\tilde{Y})$ become central to determining $EU\,(\tilde{Y})$, at least to a first approximation.[6] Since $\tilde{Y}$ inherits the form of the distribution on $\tilde{r}_P(\tilde{Y} = Y_0(1 + \tilde{r}_P))$, μ_P and σ_P are, by extension, central components of $\tilde{r}_P$.[7] In other words, we could substitute r_P for Y in the discussion above and arrive at the analogous conclusion. It remains to dispense with the approximation error term $E\tilde{H}_3$ and in this regard there are two routes open to us.

First, if $U(Y)$ is a quadratic function, U'' is a constant and, as a result, $H_3 \equiv 0$, so $E(\tilde{Y})$ and $\sigma^2(\tilde{Y})$ are all that matter.[8] It follows by equivalence that μ_P and σ_P alone determine $EU(\tilde{r}_P)$. Assuming the utility function is quadratic, however, is not fully satisfactory since the preference representation would then possess an attribute we earlier deemed fairly implausible: increasing absolute risk aversion (IARA) (see Chapter 4). On this ground, supposing all or most investors have a quadratic utility function is very restrictive.

Second, we may straightforwardly assume that individual asset returns are normally distributed: as a weighted average of normal random variables, $\tilde{r}_P$ will also be normally distributed. It follows immediately that we may assert $\mathbb{U}(\tilde{r}_P) = EU(\tilde{r}_P) \equiv \mathbb{U}(\mu_P, \sigma_P)$

To confirm this assertion, observe that

$$\tilde{z} = \frac{\tilde{r}_P - \mu_P}{\sigma_P} \sim N(0, 1),$$

if $\tilde{r}_P$ is distributed normally. As a result

$$EU(\tilde{r}_P) \equiv EU(\mu_P + \tilde{z}\sigma_P)$$

so that $EU(\tilde{r}_P)$ depends only on μ_P and σ_P, and we may write $\mathbb{U}(\tilde{r}_P)$as $\mathbb{U}(\mu_P, \sigma_P)$. This latter formulation allows us to compute investor indifference curves in $\mu_P \times \sigma_P$ space easily.

---

6 Since $U''(E(\tilde{Y})) < 0$, by concavity, and H_3 is small, the above equation confirms Jensen's inequality.

7 If $\tilde{Y}$ is normally distributed (and thus $\tilde{r}_P$), H_3 can be expressed in terms of $E(\tilde{Y})$ and $\sigma^2(\tilde{Y})$, so, again, $E(\tilde{Y})$ and $\sigma^2(\tilde{Y})$ alone determine $EU(\tilde{Y})$; similarly μ_P and σ_P alone determine $\tilde{r}_P$.

8 These well-known assertions are detailed in Appendix 6.1 where it is also shown that, under either of the above hypotheses, indifference curves in the mean–variance space are increasing and convex to the origin.

Accordingly, our operating assumptions will be expanded to include the assumed normality of asset return distributions.

The normality assumption on the rate of return processes for individual stocks and stock indices is fairly robust empirically but it is not satisfied exactly: there is typically too much probability in the tails of the return distributions (see Section 6.3). Option-based instruments, which are increasingly prevalent in investor portfolios, are also characterized by rate of return probability distributions that are far from normal. These remarks taken together suggest that the analysis to follow must be viewed as a useful and productive approximation, and not more.

Let us summarize our setting going forward. First, the investor has a one-period investment horizon and a homogeneous utility function. Second, we assume the investor knows and takes as exogenously given the vector of return random variables $(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_N)$; in particular, he knows returns are normally distributed, and he knows each security's μ and σ. Together these assumptions allow us to infer his preferences have the VNM utility form $\mathbb{U}(\mu_P, \sigma_P)$, where $\mathbb{U}_1(\mu_P, \sigma_P) > 0$, and $\mathbb{U}_2(\mu_P, \sigma_P) < 0$ (see Appendix 6.2 for details).

We have not yet discussed how these rate of return distributions arise. We are at the stage of identifying demand curves, and not yet attempting to describe how equilibrium prices or returns are determined.

6.3 Refining the Normality-of-Returns Assumption

Before we embark upon a (μ_P, σ_P)-based theory of portfolio formation, we must first address two final objections to the normality-of-returns assumption. First, the assumption that period returns (e.g., daily, monthly, annual) are normally distributed is inconsistent with the limited liability feature of most financial instruments, i.e., $\tilde{r}_i \geq -1$ for most securities i and certainly for stocks: the worst an equity investor can do is to lose his initial investment. Normality presents a further problem for the discrete compounding of cumulative returns: the product of normally distributed random variables (returns) is not itself normally distributed.

As first suggested in Box 3.1, both of these objections are made moot if we assume that all returns are continuously compounded and define a stock's rate of return as $\tilde{r}_{i,t} = \ln((\tilde{q}^e_{i,t+1} + \widetilde{\mathrm{div}}_{i,t+1})/(q^e_{i,t}))$.

First and foremost, continuous compounding preserves limited liability since $Y_0 e^r \geq 0$ for any $r \in (-\infty, +\infty)$. It has the added feature that compounding preserves normality since the sum of normally distributed random variables is normally distributed (see again the comments in Box 3.1).

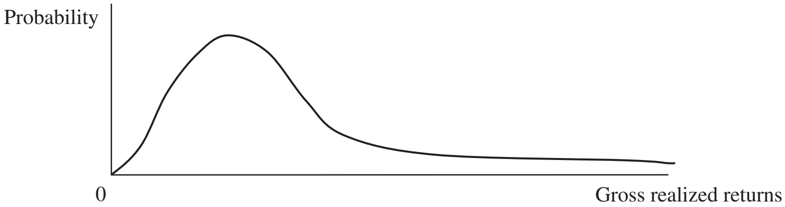

Figure 6.1
Lognormal probability density functions for gross realized returns.

Accordingly, the working assumption in empirical financial economics is that equity returns are continuously compounded and normally distributed in any period of time; in other words, for any stock i and any time t,

$$\tilde{r}_{i,t} = \ln\left(\frac{\tilde{q}^e_{i,t+1} + \widetilde{\mathrm{div}}_{i,t+1}}{q^e_{i,t}}\right) \sim N(\mu_i, \sigma_i)$$

With $\tilde{r}_{i,t} \sim N(\mu_i, \sigma_i)$ the corresponding probability density function for gross realized returns, $(\tilde{q}^e_{i,t} + \widetilde{\mathrm{div}}_{i,t})/(q^e_{i,t-1})$, has the form represented in Figure 6.1 with a mean of $\mathrm{e}^{\mu_i + (1/2)\sigma_i^2}$ and a variance of $\mathrm{e}^{(2\mu_i + \sigma_i)}(e^{\sigma_i^2} - 1)$.

Extending the idea a bit, we can say more generally that if an investor's wealth in period t is Y_t, then his wealth in period $t+1$, $\tilde{Y}_{t+1} = \mathrm{e}^{\tilde{r}^P_t} Y_t$, has a probability density function of the same form. By way of language, we say that realized gross equity returns are lognormally distributed because their logarithm is normally distributed.

There is substantial statistical evidence to support the normality assumption, subject to two very important qualifications:

1. While the normal distribution is perfectly symmetric about its mean, individual daily stock returns are frequently skewed to the right. Conversely, the returns to certain stock indices appear skewed to the left.[9]
2. Sample daily return distributions for most individual stocks exhibit "excess kurtosis" or "fat tails," i.e., there is more probability in the tails than would be justified by the normal distribution.[10] In plain language this fact means that extreme events—very positive or very negative returns—are observed with greater frequency than the normal distribution

[9] Skewness: The extent to which a probability density is "pushed to the left or right" is measured by the skewness statistic $S(\tilde{r}_{it})$, defined by $S(\tilde{r}_{it}) = E((r_{it} - \mu_i)^3/\sigma_1^3)$. $S(\tilde{r}_{it}) \equiv 0$ if $\tilde{r}_{it}$ is normally distributed. $S(\tilde{r}_{it}) > 0$ suggests a rightward bias, and conversely if $S(\tilde{r}_{it}) < 0$.

[10] Kurtosis is measured as the normalized fourth moment: $K(\tilde{r}_{it}) = E((r_{it} - \mu_i)^4/\sigma_1^4)$. If $\tilde{r}_{it}$ is normal, then $K(\tilde{r}_{it}) = 3$, but fat-tailed distributions with extra probability weight in the tail areas have higher kurtosis measures.

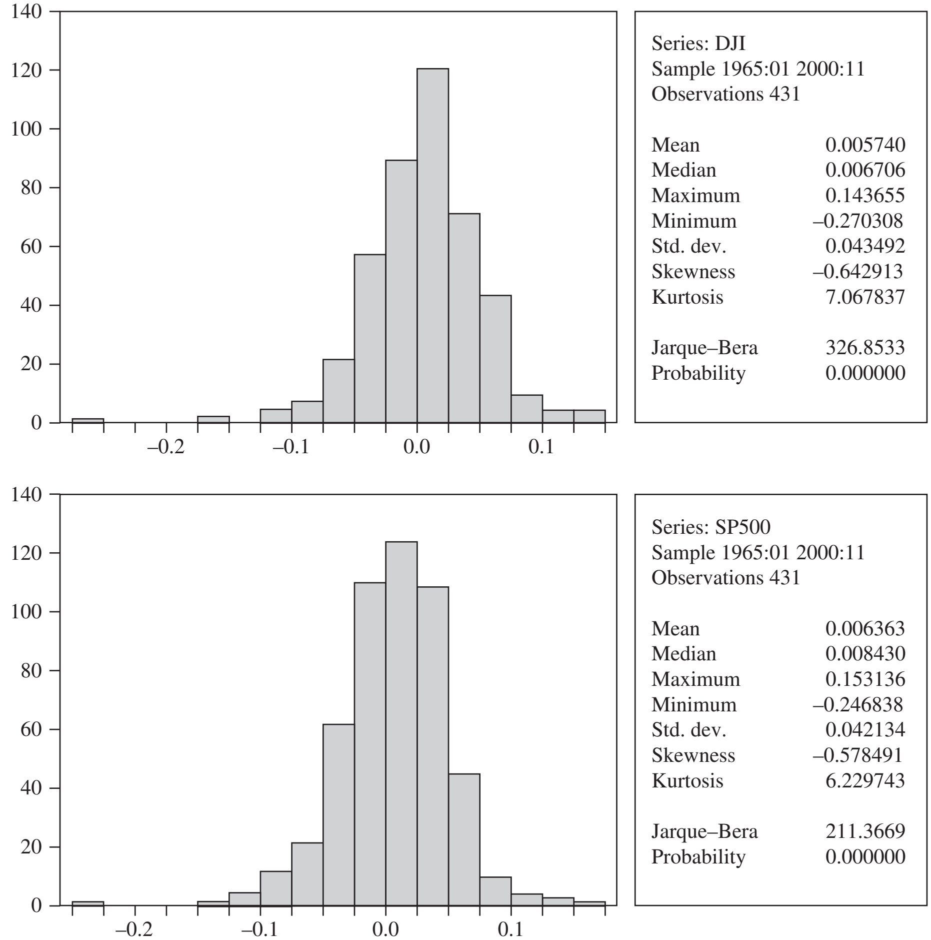

Figure 6.2
Empirical return distributions: Dow Jones and S&P$_{500}$.

would predict. The same is true of stock indices. The extent of this excess kurtosis diminishes substantially, however, when monthly data is used, but it is not eliminated.

Figure 6.2 illustrates the returns on the Dow Jones and the S&P 500. Both indices display negative skewness and a significant degree of kurtosis (Box 6.1).[11]

[11] Note on the Jarque–Bera statistic: This is a test statistic for normality. It is based on the skewness and kurtosis of the sample data. For a large sample, say $n \geq 100$, with skewness S and kurtosis K, the Jarque–Bera statistic, $(n/6)(S^2 + ((K-3)^2)/4)$, follows a Chi Square distribution with 2 degrees of freedom if the underlying distribution is normal. For each of these samples, n is 431 and the reported measures of skewness and kurtosis imply that the likelihood of the distribution being strictly normal is very low ($\approx$ 0.00), although the deviation from normality does not appear to be too extreme (the Jarque–Bera values of truly nonnormal distributions may exceed 10,000).

BOX 6.1 Connection to Factor Models

Our analysis so far has characterized the benefits of diversification in terms of the mean–variance efficient frontier of portfolio returns. Factors, however, underlie returns. Recall in Chapter 2 where we hypothesized the existence of multiple factors $\tilde{F}^1, \tilde{F}^2, \ldots, \tilde{F}^J$ for which, for any asset i,

$$\tilde{r}_i = \alpha_i + \beta_i^1 \tilde{F}^1 + \beta_i^2 \tilde{F}^2 + \ldots + \beta_i^J \tilde{F}^J + \tilde{\varepsilon}_i$$

with different stocks, via the β_i^js, displaying different factor sensitivities, and $\text{cov}(\tilde{\varepsilon}_i, F^j) = 0$ for all j. In this context, risk reduction via diversification is concerned with the $\tilde{\varepsilon}_i$ terms, their correlations across different securities, etc. The risks associated with the factors themselves cannot be diversified away as these risks affect all stock returns similarly. Accordingly, for any asset i,

$$\mu_i = \alpha_i + \beta_i^1 E\tilde{F}^1 + \cdots + \beta_i^J E\tilde{F}^J$$

and $\sigma_i^2 = \sigma^2_{\tilde{F}^1,\ldots,\tilde{F}^J} + \sigma^2_{\varepsilon_i}$, where $\sigma^2_{\tilde{F}^1,\ldots,\tilde{F}^J}$ denotes aggregate factor risk.

There is one further complication. Even if individual stock returns are lognormally distributed, the returns to a portfolio of such stocks need not be lognormal because the log of a sum is not equal to the sum of the logs:

$$\ln(1 + \tilde{r}_P) = \ln(1 + w_1\tilde{r}_1 + \cdots + w_N\tilde{r}_1) \neq w_1 \ln(1 + \tilde{r}_1) + \cdots + w_N \ln(1 + \tilde{r}_N)$$

The extent of the error introduced by assuming lognormal portfolio returns is usually quite small, however, if the return period is short (e.g., daily) so that $\ln(1 + \tilde{r}_i) \approx \tilde{r}_i$.

Let us pause at this point and review the sum total of all our underlying assumptions:

1. For all risky assets under consideration for portfolio inclusion, $\tilde{r}_{i,t} \sim N(\mu_i, \sigma_i)$. This is a reasonable first approximation.
2. The investor's utility-of-money function $U(\ )$ is homogeneous so that the same optimal portfolio proportions, once determined, apply for all investable wealth levels.
3. Individual asset returns and portfolio returns are continuously compounded.
4. As a result of Assumptions 1–3, the investor's VNM utility function $\mathbb{U}(\ )$, defined over portfolio return distributions $\tilde{r}_P$, can be expressed in the form $\mathbb{U}(\mu_P, \sigma_P)$.

With these four requirements in mind, let us proceed to the central concept of this chapter.

6.4 Description of the Opportunity Set in the Mean–Variance Space: The Gains from Diversification and the Efficient Frontier

The main idea of this chapter is as follows: The expected return to a portfolio is the *weighted* average of the expected returns of the assets composing the portfolio. The same result is not generally true for the variance. The variance of a portfolio is generally *smaller* than the weighted average of the variances of individual asset returns corresponding to this portfolio. Therein lie the gains from diversification.

Let us illustrate this assertion, starting with the case of a portfolio of two assets only. The typical investor's objective is to maximize a function $U(\mu_R, \sigma_P)$, where $U_1 > 0$ and $U_2 < 0$: the investor likes expected return (μ_P) and dislikes standard deviation (σ_P). In this context, one recalls that an asset (or portfolio) A is said to **mean−variance dominate** an asset (or portfolio) B if $\mu_A \geq \mu_B$ and simultaneously $\sigma_A < \sigma_B$, or if $\mu_A > \mu_B$ while $\sigma_A \leq \sigma_B$. We can then define the **efficient frontier** as the locus of all nondominated portfolios in the mean−standard deviation space. By definition, no ("rational") mean−variance investor would choose to hold a portfolio not located on the efficient frontier. The shape of the efficient frontier is thus of primary interest.

Next consider the efficient frontier in the two-asset case for a variety of possible asset return correlations. The basis for the results of this section is the formula for the variance of a portfolio of two assets, 1 and 2, defined by their respective expected returns, μ_1, μ_2, standard deviations, σ_1 and σ_2, and their correlation $\rho_{1,2}$:

$$\sigma_P^2 = w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2 + 2w_1(1-w_1)\sigma_1\sigma_2\rho_{1,2}$$

where w_i is the proportion of the portfolio allocated to asset i. The following results, detailed in Appendix 6.2, are of special importance.

Case 1 (Reference): In the case of two risky assets with perfectly positively correlated returns, the efficient frontier is linear. In that extreme case, the two assets are essentially identical, there is no gain from diversification, and the portfolio's standard deviation is nothing other than the average of the standard deviations of the component assets:

$$\sigma_R = w_1\sigma_1 + (1-w_1)\sigma_2$$

As a result, the equation of the efficient frontier is

$$\mu_R = \mu_1 + \frac{\mu_2 - \mu_2}{\sigma_2 - \sigma_1}(\sigma_R - \sigma_1)$$

as depicted in Figure 6.3. It assumes that positive amounts of both assets are held.

Case 2: In the case of two risky assets with imperfectly correlated returns, the standard deviation of the portfolio is necessarily smaller than it would be if the two component assets were perfectly correlated. By the previous result, one must have $\sigma_P < w_1\sigma_1 + (1-w_1)\sigma_2$, provided the proportions are not 0 or 1. Thus, the efficient frontier must stand left of the straight line in Figure 6.3. This is illustrated in Figure 6.4 for different values of $\rho_{1,2}$.

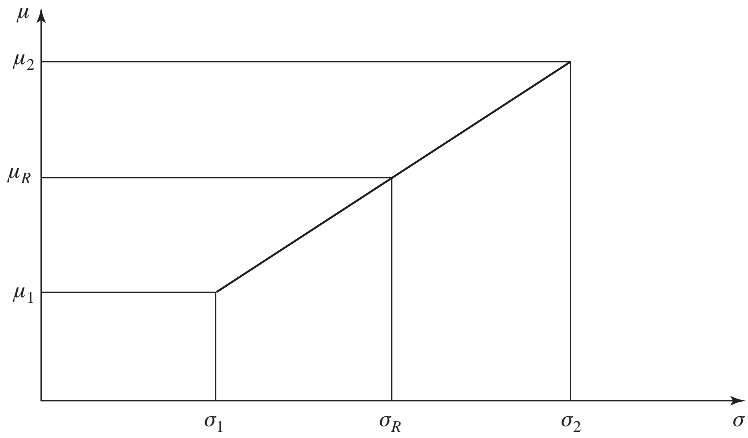

Figure 6.3
The efficient frontier: two perfectly correlated risky assets.

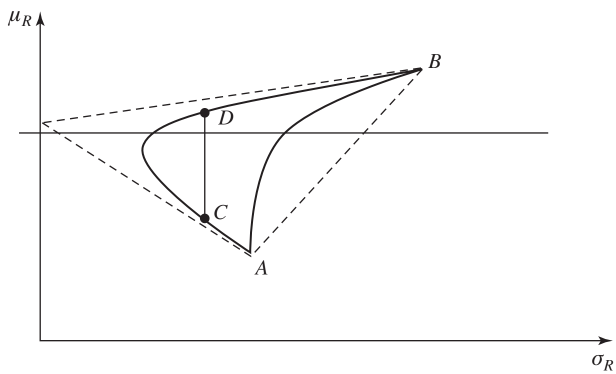

Figure 6.4
The efficient frontier: two imperfectly correlated risky assets.

The smaller the correlation (the further away from +1), the more to the left is the efficient frontier as demonstrated formally in Appendix 6.2. Note that the diagram makes clear that in this case, some portfolios made up of assets 1 and 2 are, in fact, dominated by other portfolios. Unlike in Case 1, not all portfolios are efficient. In view of future developments, it is useful to distinguish the **minimum variance frontier** from the efficient frontier. In the present case, all portfolios between *A* and *B* belong to the minimum variance frontier, i.e., they correspond to the combination of assets with minimum variance for all arbitrary levels of expected returns. However, certain levels of expected returns are not efficient targets since higher levels of returns can be obtained for identical levels of risk. Thus, portfolio *C* has minimum variance, but it is

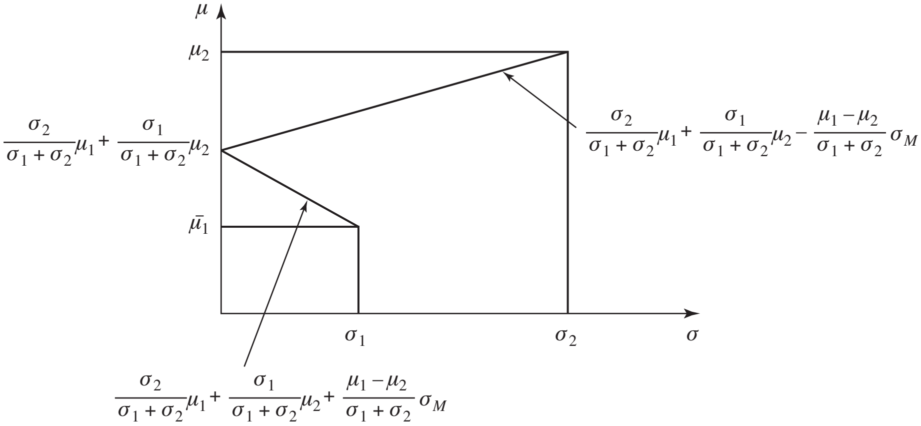

Figure 6.5
The efficient frontier: two perfectly negatively correlated risky assets.

not efficient, being dominated, for instance, by portfolio D. Figure 6.4 again assumes positive amounts of both assets (A and B) are held.

Case 3: If the two risky assets have returns that are perfectly negatively correlated, one can show that the minimum variance portfolio is risk free while the frontier is once again linear. Its graphical representation in that case is in Figure 6.5, with the corresponding demonstration found in Appendix 6.2.

Case 4: If one of the two assets is risk free, then the efficient frontier is a straight line originating on the vertical axis at the level of the risk-free return. In the absence of a short sales restriction, i.e., if it is possible to borrow at the risk-free rate to leverage one's holdings of the risky asset, then, intuitively enough, the overall portfolio can be made riskier than the riskiest among the existing assets. In other words, it can be made riskier than the one risky asset, and it must be that the efficient frontier is projected to the right of the (μ_2, σ_2) point (defining asset 1 as the risk-free asset). This situation is depicted in Figure 6.6, with the corresponding results demonstrated in Appendix 6.2.

Case 5 (*n* risky assets): It is important to realize that a portfolio is also an asset, fully defined by its expected return, its standard deviation, and its correlation with other existing assets or portfolios. Thus, the previous analysis with two assets is more general than it appears: It can easily be repeated, with one of the two assets being a portfolio. In that way, one can extend the analysis from two to three assets, from three to four, and so on. If there are n risky, imperfectly correlated assets, then the efficient frontier will have the bullet shape of Figure 6.7. Adding an extra asset to the two-asset framework implies that the diversification possibilities are improved and that, in principle, the efficient frontier is displaced to the left.

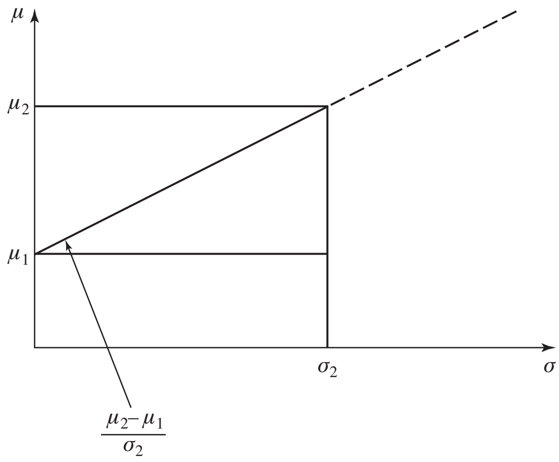

Figure 6.6
The efficient frontier: one risky and one risk-free asset.

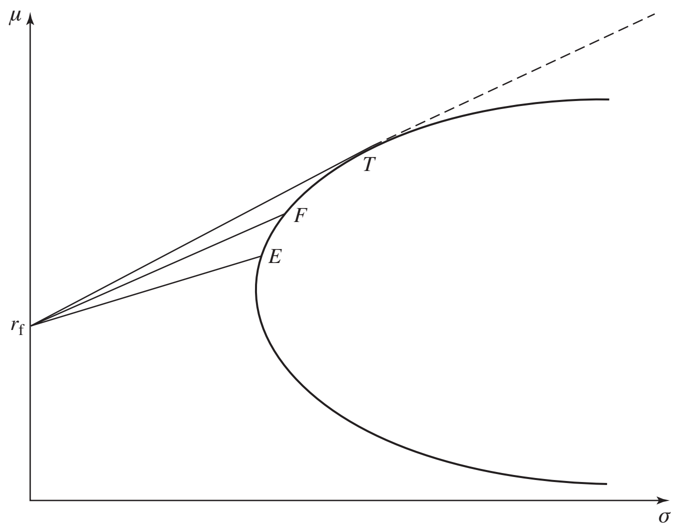

Figure 6.7
The efficient frontier: one risk-free and *n* risky assets.

Case 6: If there are *n* risky assets and a risk-free one, the efficient frontier is a straight line once again. To arrive at this conclusion, let us arbitrarily pick one portfolio on the efficient frontier when there are *n* risky assets only, say portfolio *E* in Figure 6.7, and make up all possible portfolios combining *E* and the risk-free asset.

What we learned above tells us that the set of such portfolios is the straight line joining the point $(0, r_f)$ to E. Now we can quickly check that all portfolios on this line are dominated by those we can create by combining the risk-free asset with portfolio F. Continuing our reasoning in this way and searching for the highest similar line joining $(0, r_f)$ with the risky asset bullet-shaped frontier, we obtain, as the truly efficient frontier, the straight line originating from $(0, r_f)$ that is tangent to the risky asset frontier. Let T be the tangency portfolio. As before, if we allow a short position in the risk-free asset, the efficient frontier extends beyond T; it is represented by the broken line in Figure 6.7.

Formally, with n assets (possibly one of them risk free), the efficient frontier is obtained as the relevant (nondominated) portion of the minimum variance frontier, the latter being the solution, for all possible expected returns μ, to the following quadratic program (QP):

$$
\text{(QP)} \qquad \begin{aligned} &\min_{w_i\text{'s}} \sum_i \sum_j w_i w_j \sigma_{ij} \\ &\text{s.t.} \sum_i w_i \mu_i = \mu \\ &\qquad \sum_i w_i = 1 \end{aligned}
$$

In (QP) we search for the vector of weights that minimizes the variance of the portfolio (verify that you understand the writing of the portfolio variance in the case of n assets) under the constraint that the expected return on the portfolio must be μ. This defines one point on the minimum variance frontier. One can then change the fixed value of μ, equating it successively to all plausible levels of portfolio expected return; in this way one effectively draws the minimum variance frontier.[12] Program (QP) is the simplest version of a family of similar quadratic programs used in practice. This is because (QP) includes the minimal set of constraints. The first is only an artifice in that it defines the expected return to be reached in a context where μ is a parameter; the second constraint is simply the assertion that the vector of w_i's defines a portfolio (and thus that they add up to one).

Many other constraints can be added to customize the portfolio selection process without altering the basic structure of problem (QP). Probably the most common implicit or explicit constraint for an investor involves limiting her investment universe. As mentioned earlier (Chapter 3, Box 3.2), the well-known *home bias puzzle* reflects the difficulty in explaining, from the MPT viewpoint, why investors do not invest a larger fraction of their portfolios in stocks quoted "away from home," i.e., in international, or emerging markets. This can be viewed as the result of an unconscious limitation of the investment universe considered by the investor. Self-limitation may also be fully conscious and explicit as in the case of

[12] While in principle one could as well maximize the portfolio's expected return for given levels of standard deviation, it turns out to be more efficient computationally to do the reverse.

"ethical" mutual funds that exclude arms manufacturers or companies with a tarnished ecological record from their investment universe. These constraints are easily accommodated in our setup, as they simply appear or do not appear in the list of the N assets under consideration.

Other common constraints are nonnegativity constraints ($w_i \geq 0$), indicating the impossibility of short selling some or all assets under consideration. Without nonnegativity constraints, mean−variance analysis often leads to efficient portfolios with large short positions in some assets.[13] Short selling may be impossible for feasibility reasons (exchanges or brokers may not allow it for certain instruments) or, more frequently, for regulatory reasons applying to specific types of investors, e.g., pension funds.

An investor may also wish to construct an efficient portfolio subject to the constraint that his holdings of some stocks should not, in value terms, fall below a certain level (perhaps because of potential tax liabilities or because ownership of a large block of this stock affords some degree of managerial control). This requires a constraint of the form

$$w_j \geq \frac{V_j^*}{V_P}$$

where V_j^* is the postulated lower bond on the value of his holdings of stock j and V_P is the overall value of his portfolio.

Other investors may wish to obtain the lowest risk subject to a required expected return constraint and/or be subject to a constraint that limits the number of stocks in their portfolio (in order, possibly, to economize on transaction costs). An investor may, for example, wish to hold at most 3 out of a possible 10 stocks, yet to hold those 3 that give the minimum risk subject to a required return constraint. With certain modifications, this possibility can be accommodated into (QP) as well. Appendix 6.3 details how Microsoft Excel® can be used to construct the portfolio efficient frontier under these and other constraints.

6.5 The Optimal Portfolio: A Separation Theorem

The optimal portfolio is naturally defined as that portfolio maximizing the investor's (mean−variance) utility; in other words, that portfolio for which the investor is able to reach the highest indifference curve, which we know to be increasing and convex to the origin. If the efficient frontier has the shape described in Figure 6.6, i.e., if there is a risk-free asset, then all tangency points must lie on the same efficient frontier, regardless of the rate of risk aversion of the investor. Let there be two investors sharing the same perceptions as to expected returns, variances, and return correlations but differing in their willingness to

[13] An illustration of this particular phenomenon is presented in Chapter 7.

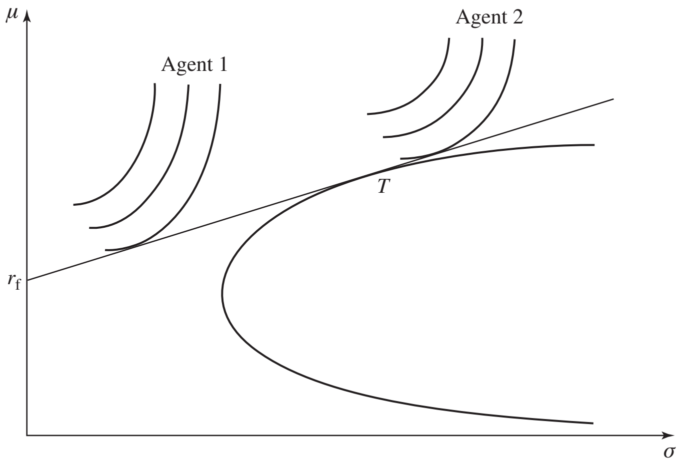

Figure 6.8
The optimal portfolios of two differently risk-averse investors.

take risks. The relevant efficient frontier will be identical for these two investors, although their optimal portfolios will be represented by different points on the same line: with differently shaped indifference curves the tangency points must differ. See Figure 6.8.

Nevertheless, it is a fact that our two investors will invest in the same two "funds," the risk-free asset on the one hand, and the risky portfolio (T) identified by the tangency point between the straight line originating from the vertical axis and the bullet-shaped frontier of risky assets, on the other. This is the **two-fund theorem**, also known as the **separation theorem**, because it implies that the optimal portfolio of risky assets can be identified separately from an investor's knowledge of his risk preference. This result will play a significant role in Chapter 8 when we construct the capital asset pricing model.

6.6 Stochastic Dominance and Diversification

In this concluding section, we propose to deal with a typical question facing an investor: should a particular asset be added to the investor's portfolio? More precisely, if the particular asset in question is added, what characteristics must it have in order that the investor can be assured that his expected utility of wealth will increase?

To be more precise, suppose the investor's current portfolio P is characterized by (μ_P, σ_P). He is considering adding some of asset A, characterized by $(\mu_A, \sigma_A, \rho_{AP})$ to his portfolio. Although the strict attainment of mean–variance efficiency may well require it, he does not plan at this juncture to alter the relative composition of the assets he already has. Indeed, to implement substantial changes in his portfolio's existing composition is likely to be a costly proposition, and let us assume also that the avoidance of this cost is presently dominant in his decision making. These considerations mean that if the investor decides to allocate some of his wealth to A, his new portfolio will then be composed of two assets, some units of the original portfolio P and some amount of money invested in asset A. We denote this new portfolio P'.[14]

Let us go on and further assume that by undertaking this investment, the investor hopes to improve the overall return on his wealth. Since the expected return on a portfolio is simply the weighted average of the expected returns of the assets contained in the portfolio, the only way for the investor to increase his original portfolio's expected returns is to add assets for which $\mu > \mu_P$. Accordingly, we will assume this is true for the asset in question: $\mu_A > \mu_P$ (e.g., we may think of A as itself a portfolio of emerging markets stocks which had high returns in the 1980–2000 period). Few benefits in life are totally free, however, which means that $\sigma_A^2 > \sigma_P^2$ as well (this fact also characterized emerging markets stocks relative to the $S\&P_{500}$ in the 1980s and 1990s). If $\rho_{AP} \approx 1$, so that asset A and the investor's existing portfolio P have identical return patterns, the investor will be confronted by a difficult choice: by allocating some of his cash to A, his portfolio's μ will certainly rise, but its risk will almost surely increase as well. If $\rho_{AP} < 1$, however, it is possible that the investor's overall portfolio risk would fall with the inclusion of A, i.e., A and P's return patterns could be sufficiently different so as to allow risk to decline. How different would the return patterns have to be for this "diversification" effect to dominate? Would a VNM-expected utility maximize necessarily agree to include asset A in that case? These are the questions and context under consideration in this final chapter topic. We formalize them as follows:

> Question 1: Under what conditions will the addition of an asset to a portfolio (in the aforementioned sense) allow the portfolio's expected return to increase and its standard deviation to decline? This is one characterization of the gains to diversification in a mean/variance environment.
>
> Question 2: Will a VNM-expected utility maximizing investor also prefer to be more highly diversified in the sense of Question 1? To say it differently, although the inclusion of a higher expected return, higher standard deviation asset may make sense from a mean–variance perspective, would a typical VNM-expected utility investor always agree?

[14] There are a number of scenarios to which this problem description will apply. For one such scenario, suppose an investor is confronted with an investment opportunity (A) when his savings are invested in an ETF. If he sells some units of his ETF to finance the purchase of A he is not altering the proportions of the other securities he implicitly owns by virtue of owning units of the ETF. A second scenario is as follows: An investor owns a portfolio (P) with a number N of specific assets. He has accumulated some cash to invest, either by increasing the assets in (P) in a way that leaves the proportions unchanged (given the data available to him it is on the efficient frontier for those assets), or he can invest the same cash in a new asset (A) which has presented itself. Under either scenario, the analysis will apply.

If their respective returns patterns differ, we know that some of A's return variation (risk) can be canceled out by variation in the return of P, and that the extent of this "cancelation effect" increases the lower the ρ_{AP}. Accordingly, Question 1 can be forthrightly restated as follows: How low must ρ_{AP} be in order that the inclusion of A into portfolio P will reduce overall portfolio risk?[15] The answer is provided in Theorem 6.1.

Theorem 6.1 Consider an N-asset portfolio P characterized by (μ_P, σ_P) to which an investor is considering adding an asset A where (i) $\mu_A > \mu_P$ and (ii) $\rho_{AP} < (\sigma_P/\sigma_A)$. Then there exists a portfolio P^* such that $\mu_{P^*} > \mu_P$ and $\sigma_{P^*} < \sigma_P$ where P^* contains A in positive amount.

Proof Consider a portfolio P' defined by proportions $w_A \geq 0$ in A and $1 - w_A$ in P. Then,

$$\sigma^2_{P'} = (1-w_A)^2\sigma^2_P + w_A^2\sigma_A^2 + 2\, w_A(1-w_A)\,\text{cov}\,(\tilde{r}_A, \tilde{r}_P)$$
$$\frac{\partial\sigma^2_{P'}}{\partial w_A} = -\,2(1-w_A)\sigma^2_P + 2\, w_A\sigma^2_A + 2\,\text{cov}(\tilde{r}_A, \tilde{r}_P) - 4\, w_A\text{cov}(\tilde{r}_A, \tilde{r}_P)$$

Thus, $(\partial\sigma^2_{P'}/\partial w_A) = 2\,\text{cov}\,(\tilde{r}_A, \tilde{r}_P) - 2\sigma^2_P$ for $w_A = 1$. Thus

$(\partial\sigma^2_{P'}/\partial w_A) < 0$ for $w_A = 1$ if and only if $\text{cov}(\tilde{r}_A, \tilde{r}_P) < \sigma^2_P$, which is true if and only if $\rho_{AP} < (\sigma_P/\sigma_A)$, as assumed.

By the continuity of $(\partial\sigma^2_{P'}/\partial w_A)$ as a function of w_A, there exists a $\hat{w}_A > 0$ such that for all $0 \leq w_A < \hat{w}_A$

$\mu_{P'} = \pi_A\mu_A - (1-\pi_A)\mu_P > \mu_P$ (by assumption (i)), and
$\sigma_{P'} < \sigma_P$ (by the above computation).

Pick one $\hat{w}_A > 0$ with corresponding portfolio P^*. Then

$$\mu_{P^*} > \mu_P \tag{6.8a}$$

$$\sigma_{P^*} < \sigma_P \tag{6.8b}$$

Portfolio P^* thus strictly dominates P in a mean–variance framework.

Portfolio P^* as constructed in the Lemma may not be efficient: further gains may be obtainable by altering the composition of the fraction $(1 - w_A)$ of P^* that is composed of the original P. Our investor accepts this possibility, however, because of his interest in minimizing his transactions costs.

[15] To clarify this language, we will take "inclusion" to mean that the investor's original portfolio with $(w_P = 1, w_A = 0)$ becomes one where $(w_P < 1, w_A > 0)$.

It is also worth noting that condition (ii) of the Lemma is always satisfied provided $\rho_{AP} = 0$, i.e., when the return on the asset under consideration for portfolio inclusion is statistically independent of the returns to the portfolio to which it may be "joined." It follows that if enough independent (of one another) assets are included in a portfolio (at least with equal proportions), its risk can be made arbitrarily small. This assertion encapsulates the insurance principle.[16] For equity investors, it is essentially impossible to find numerous assets with this property, however, as our discussion in Chapter 7 will confirm.

At this juncture, we have given an answer to Question 1: if conditions (i) and (ii) of Proposition 6.1 are satisfied, then the investor can increase his portfolio's overall expected return and reduce its risk by adding some of asset A. To answer Question 2, however, additional results are needed. They are as follows.

Theorem 6.2 Consider two return distributions $\tilde{r}_1$ and $\tilde{r}_2$ with, respectively, cumulative distribution functions $F_{\tilde{r}_1}$ and $F_{\tilde{r}_2}$, where $\tilde{r}_1 \sim N(\mu_1, \sigma_1)$ and $\tilde{r}_2 \sim N(\mu_2, \sigma_2)$. Then $F_{\tilde{r}_1}$ FSD $F_{\tilde{r}_2}$ if and only if the following hold:

a. $\mu_1 > \mu_2$
b. $\sigma_1 = \sigma_2$.

Proof See Levy (2006), Theorem 6.1

Theorem 6.3 Consider two return distributions $\tilde{r}_1$ and $\tilde{r}_2$ with respective cumulative distribution functions $F_{\tilde{r}_1}$and $F_{\tilde{r}_2}$, where $\tilde{r}_1 \sim N(\mu_1, \sigma_1)$ and $\tilde{r}_2 \sim N(\mu_2, \sigma_2)$. Then $F_{\tilde{r}_1}$ SSD $F_{\tilde{r}_2}$ if and only if the following hold:

a. $\mu_1 \geq \mu_2$
b. $\sigma_1 \leq \sigma_2$.

with at least one strong inequality.

Proof See Levy (2006), Theorem 6.2.

Corollary 6.1 applies these results (Theorem 6.3, in particular) to our investor's situation.

Corollary 6.1 Consider portfolios P and P^* of Proposition 6.1, which satisfy Eqs. (6.8a) and (6.8b). Let their respective cumulative distribution functions be denoted $F_{\tilde{r}_{P*}}$and $F_{\tilde{r}_P}$. Then $F_{\tilde{r}_{P*}}$ SSD $F_{\tilde{r}_P}$.

[16] Insurance companies write policies to many insurers (car, fire, etc.) whose risks to the company are independent, e.g., the risks to a 64-year-old driver in New York are independent of the risks of a similarly aged driver in Illinois.

Proof Application of Theorem 6.3.

We are getting closer to answering our investor's second question but we are not yet quite there. It would be convenient to answer this question with an immediate appeal to Theorem 4.3, but that theorem is expressed in terms of payoffs while Corollary 6.1 is expressed in terms of returns.[17] This (small) inconsistency, however, can be easily overcome as follows.

Let $F_{\tilde{y}}$ denote the cumulative distribution function associated with some random variable $\tilde{y}$. By Corollary 6.1, we know that $F_{\tilde{r}_{P^*}}$ SSD $F_{\tilde{r}_P}$. It follows immediately that $F_{1+\tilde{r}_{P^*}}$ SSD $F_{1+\tilde{r}_P}$ and thus, by Theorem 4.2,

$$EU(1+\tilde{r}_{P^*}) > EU(1+\tilde{r}_P)$$

for any increasing, concave utility function $U(\ )$, where these expectations are "linear in the probabilities" as per VNM-expected utility theory.

If we further assume $U(\ )$ is homogeneous of degree ν and that the start-of-period wealth is Y_0, then

$$\begin{aligned} Y_0^{\nu} EU(1+\tilde{r}_{P^*}) &> Y_0^{\nu} EU(1+\tilde{r}_P) \\ EU(Y_0(1+\tilde{r}_{P^*})) &> EU(Y_0(1+\tilde{r}_P)) \end{aligned} \tag{6.9}$$

for any increasing, concave utility function $U(\ )$. If our investor's utility function is increasing, concave, and homogeneous, his expected utility will thus increase with the inclusion of A.

Note that expectations in Eq. (6.9) are taken with respect to the investor's actual future wealth payoffs. We summarize these thoughts in Theorem 6.4.

Theorem 6.4 Consider a universe of assets whose returns are normally distributed and consider greater diversification in the sense of Theorem 6.1 as reflected in portfolios P^* and P. Then for any VNM-expected utility investor with an increasing, concave, and homogeneous utility function, it will be welfare improving to include some positive amount of asset A.

Proof The discussion above.

As a conclusion to our discussion, Theorem 6.4 is a bit unsatisfying, however, because it assumes that the investor's next-period wealth distribution is normally distributed, which is certainly not the case if the assets in the portfolio are equities, as we have assumed. Rather, the investor's wealth distribution will, in fact, be lognormal. To modify our results to accommodate this additional fact, let us first denote $\tilde{Y}_1^P$ and $\tilde{Y}_1^{P^*}$ as the next-period random

[17] These quantities are of course equivalent if our investor's wealth is \$1 (1 CHF, etc.), but we need greater generality.

wealth levels if Y_0, the investor's present wealth, is invested, respectively, in P or P^*. Under continuous compounding,

$$E\tilde{Y}_1^P = EY_0\,\mathrm{e}^{\tilde{r}_P} = Y_0\,\mathrm{e}^{\mu_P + 1/2\sigma_P^2}$$
$$\sigma^2_{\tilde{Y}_1^P} = Y_0^2\,\mathrm{e}^{2\mu_P+\sigma_P^2}\left(\mathrm{e}^{\sigma_P^2} - 1\right)$$
$$E\tilde{Y}_1^{P^*} = EY_0\,\mathrm{e}^{\tilde{r}_{P^*}} = Y_0\,\mathrm{e}^{\mu_{P^*} + 1/2\sigma_{P^*}^2}$$
$$\sigma^2_{\tilde{Y}_1^{P^*}} = Y_0^2\,\mathrm{e}^{2\mu_{P^*}+\sigma_{P^*}^2}\left(\mathrm{e}^{\sigma_{P^*}^2} - 1\right)$$

For the relationships above, since $\mu_{P^*} > \mu_P$ and $\sigma_{P^*} < \sigma_P$ it is not guaranteed that $E\tilde{Y}_1^{P^*} > E\tilde{Y}_1^P$ because $\sigma_{P^*}^2 < \sigma_P^2$! The dependence of $E\tilde{Y}_1^{P^*}$ on σ_{P^*} (also for P) follows from the asymmetric nature of the lognormal distribution (which governs both $\tilde{Y}_1^P$ and $\tilde{Y}_1^{P^*}$): a decrease in the variance (which applies to $\tilde{r}_{P^*}$ relative to $\tilde{r}_P$) must be accompanied by a shift in probability toward the left tail which inevitably diminishes the mean as well. To obtain results analogous to Theorem 6.5, the following result is needed.

Theorem 6.5 Let $\tilde{Y}_1$ and $\tilde{Y}_2$ be two lognormally distributed random variables with cumulative distribution functions $F_{\tilde{Y}_1}$ and $F_{\tilde{Y}_2}$, respectively. Then, $F_{\tilde{Y}_1}$ SSD $F_{\tilde{Y}_2}$ if and only if

a. $E\tilde{Y}_1 \geq E\tilde{Y}_2$
b. $\sigma_{\tilde{Y}_1}/E\tilde{Y}_1 \leq \sigma_{\tilde{Y}_2}/E\tilde{Y}_2$,

with at least one strict inequality.

Proof Levy (2006), Theorem 6.5

Our final result is a direct application of Theorem 6.5.

Theorem 6.6 Let $\tilde{Y}_1^P$ and $\tilde{Y}_1^{P^*}$ be as in the discussion above. Assume that Eqs. (6.8a), (6.8b), and (6.10)

$$\mu_{P^*} + (1/2)\sigma_{P^*}^2 > \mu_P + (1/2)\sigma_P^2 \tag{6.10}$$

hold. Then $F_{Y_1^{P^*}}$ SSD $F_{Y_1^P}$, and a VNM-expected utility investor with an increasing and concave utility-of-money function prefers greater diversification in the sense of adding asset A to his portfolio.

Proof A direct application of Theorem 6.5 is as follows. By assumption (6.10), $EY_1^{P^*} \geq EY_1^P$ which satisfies condition (a) of Theorem 6.5. Furthermore,

$$\left(\frac{\sigma_{\tilde{Y}_1^{P*}}}{EY_1^{P*}}\right)^2 = \left(\frac{Y_0^2\,\mathrm{e}^{2\mu_{P*}+\sigma_{P*}^2}}{\left[Y_0\,\mathrm{e}^{\mu_{P*}+1/2\sigma_{P*}}\right]^2}\right)\left(\mathrm{e}^{\sigma_{P*}^2}-1\right)$$

$$= \left(\mathrm{e}^{\sigma_{P*}^2}-1\right) < \left(\mathrm{e}^{\sigma_P^2}-1\right) = \left(\frac{\sigma_{\tilde{Y}_1^P}}{EY_1^P}\right)^2$$

by Eqs. (6.8a) and (6.8b).

Thus, $(\sigma_{\tilde{Y}_1^{P*}}/EY_1^{P*}) < (\sigma_{\tilde{Y}_1^P}/EY_1^P)$ and (b) of Theorem 6.5 is satisfied.

What is to be learned from this discussion? First and foremost, we learned that adding an asset A to a portfolio P for which $\mu_A > \mu_P$ and $\rho_{AP} < (\sigma_P/\sigma_A)$ will create a second-order stochastically dominating distribution relative to that of the original portfolio P. In general, however, this is not enough to guarantee that the investor's expected utility of wealth will be enhanced by its inclusion. That requires the new portfolio (P^*, the portfolio with some A included) to satisfy $\mu_{P*} + (1/2)\sigma_{\tilde{r}_{P*}} > \mu_P + (1/2)\sigma_{\tilde{r}_P}$. This amounts to requiring that the asset up for inclusion should have a significantly higher expected return than the portfolio into which it may be placed. This is an illustration where "means matter more than variances" (and, as we shall see, means are also more difficult to estimate precisely).

We now conclude this chapter.

6.7 Conclusions

First, it is important to keep in mind that everything said so far in this chapter applies to the *subjective* expectations regarding future return distributions that the investor may hold. There is no requirement that these distributions be objective or "rational."

Second, although initially conceived in the context of descriptive economic theories, the success of portfolio theory arose primarily from the possibility of giving it a normative interpretation, i.e., of seeing the theory as providing a guide on how to proceed to identify a potential investor's optimal portfolio. In particular, it points to the information requirements to be fulfilled (ideally). Under the formal restrictions spelled out in this chapter, one cannot identify an optimal portfolio without estimates of mean returns, standard deviations of returns, and correlations among returns. As Chapter 7 will make clear, this is no easy task empirically. Investors may thus simply fall back on simpler allocation strategies such as an equally weighted portfolio. As we will also see in Chapter 7, there is a substantial body of empirical evidence that argues for this alternative.

One can view the role of the financial analyst as providing plausible figures for the relevant statistics or offering alternative scenarios for consideration to the would-be investor. This is the first absolutely critical step in the search for an optimal portfolio.

The computation of the (subjective) efficient frontier is the second step, and it essentially involves solving the QP problem possibly in conjunction with constraints specific to the investor. The third and final step consists of defining, at a more or less formal level, the investor's risk tolerance and, on that basis, identifying his optimal allocation between risk free and risky assets.

References

Levy, H., 2006. Stochastic Dominance: Investment Decision Making Under Uncertainty. Springer Science and Business Media, New York.

Markowitz, H.M., 1952. Portfolio selection. J. Finan. 7, 77–91.

Tobin, J., 1958. Liquidity preference as behavior towards risk. Rev. Econ. Stud. 26, 65–86.

Appendix 6.1: Indifference Curves Under Quadratic Utility or Normally Distributed Returns

In this appendix, we demonstrate more rigorously that if an investor's utility function is quadratic or if returns are normally distributed, the investor's expected utility of the portfolio's rate of return is a function of the portfolio's mean return and standard deviation only (Part I). We subsequently show that in either case, investor indifference curves are convex to the origin (Part II).

Part I

If the utility function is **quadratic**, it can be written as

$$U(r_P) = a + b\, r_P + c\, r_P^2$$

where r_P denotes a portfolio's rate of return. Let the constant $a = 0$ in what follows since it does not play any role. For this function to make sense, we must have $b > 0$ and $c < 0$. The first and second derivatives are, respectively,

$$U'(r_P) = b + 2c\, r_P \text{ and } U''(r_P) = 2c < 0$$

Expected utility is then of the following form:

$$E(U(\tilde{r}_P)) = bE(\tilde{r}_P) + c(E(\tilde{r}_P^2)) = b\mu_P + c\mu_P^2 + c\sigma_P^2$$

that is of the form $g(\sigma_P, \mu_P)$. As shown in Figure A6.1, this function is strictly concave. But it must be restricted to ensure positive marginal utility: $\tilde{r}_P < -b/2c$. Moreover, the coefficient of absolute risk aversion is increasing ($R_A' > 0$). These two characteristics are unpleasant, and they prevent a more systematic use of the quadratic utility function.

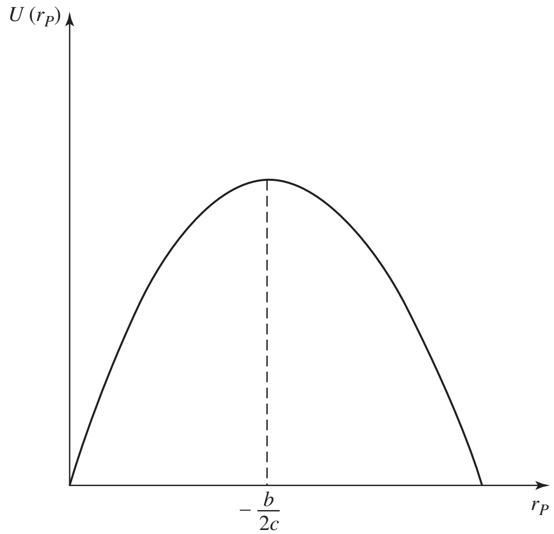

Figure A6.1
The graph of a quadratic utility function.

Alternatively, if the individual asset returns $\tilde{r}_i$ are normally distributed, $\tilde{r}_P = \sum_i w_i \tilde{r}_i$ is normally distributed as well. Let $\tilde{r}_P$ have density $f(\tilde{r}_P)$, where

$$f(\tilde{r}_P) = N(\tilde{r}_P; \mu_P, \sigma_P)$$

The standard normal variate $\tilde{Z}$ is defined by

$$\tilde{Z} = \frac{r_P - \mu_P}{\sigma_P} \sim N(\tilde{Z}; 0, 1)$$

$$\text{Thus,} \quad \tilde{r}_P = \sigma_P \tilde{Z} + \mu_P \tag{6.11}$$

$$\mathbb{U}(\tilde{r}_P) = E(U(r_P)) = \int_{-\infty}^{+\infty} U(r_P) f(r_P) \mathrm{d}r_P = \int_{-\infty}^{+\infty} U(\sigma_P Z + \mu_P) N(Z; 0, 1) \mathrm{d}Z \tag{6.12}$$

The quantity $E(U(r_P))$ is again a function of σ_P and μ_P only. Maximizing $E(U(\tilde{r}_P))$ amounts to choosing w_i so that the corresponding σ_P and μ_P maximize the integral (6.12).

Part II

Construction of indifference curves in the mean−variance space. There are again two cases.

U Is Quadratic

An indifference curve in the mean−variance space is defined as the set: $\{(\sigma_P, \mu_P)|E(U(\tilde{r}_P)) = b\mu_P + c\mu_P^2 + c\sigma_P^2 = k\}$, for some utility level k.

This can be rewritten as

$$\sigma_P^2 + \mu_P^2 + \frac{b}{c}\mu_P + \frac{b^2}{4c^2} = \frac{k}{c} + \frac{b^2}{4c^2}$$

$$\sigma_P^2 + \left(\mu_P + \frac{b}{2c}\right)^2 = \frac{k}{c} + \frac{b^2}{4c^2}$$

This equation defines the set of points (σ_P, μ_P) located in the circle of radius $\sqrt{\frac{k}{c} + \frac{b^2}{4c^2}}$ and of center $(0, -(b/2c))$ as in Figure A6.2.

In the relevant portion of the (σ_P, μ_P) space, indifference curves thus have positive slope and are convex to the origin.

The Distribution if R Is Normal

One wants to describe

$$\left[(\sigma_P, \mu_P)\middle|\int_{-\infty}^{+\infty} U(\sigma_P Z + \mu_P)N(Z; 0, 1)\mathrm{d}Z = \overline{U}\right]$$

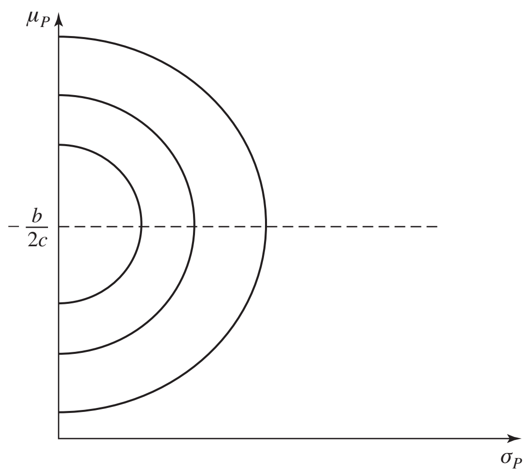

Figure A6.2
The indifference curves of a quadratic utility agent.

Differentiating totally yields:

$$0 = \int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P)(Z\mathrm{d}\sigma_P + \mathrm{d}\mu_P)N(Z;0,1)\mathrm{d}Z, \quad \text{or}$$

$$\frac{\mathrm{d}\mu_P}{\mathrm{d}\sigma_P} = -\frac{\displaystyle\int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P)ZN(Z;0,1)\mathrm{d}Z}{\displaystyle\int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P)N(Z;0,1)\mathrm{d}Z}$$

If $\sigma_P = 0$ (at the origin),

$$\frac{\mathrm{d}\mu_P}{\mathrm{d}\sigma_P} = -\frac{\displaystyle\int_{-\infty}^{+\infty} ZN(Z;0,1)\mathrm{d}Z}{\displaystyle\int_{-\infty}^{+\infty} N(Z;0,1)\mathrm{d}Z} = 0$$

If $\sigma_P > 0$, $\mathrm{d}\mu_P/\mathrm{d}\sigma_P > 0$.

Indeed, the denominator is positive since $U'(\cdot)$ is positive by assumption, and $Z \sim N(0, 1)$ is a probability density function, hence it is always positive.

The expression $\int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P)ZN(Z,0,1)\mathrm{d}Z$ is negative under the hypothesis that the investor is risk averse—in other words, that $U(\cdot)$ is strictly concave. If this hypothesis is verified, the marginal utility associated with each negative value of Z is larger than the marginal utility associated with positive values. Since this is true for all pairs of $\pm Z$, the integral on the numerator is negative. See Figure A6.3 for an illustration.

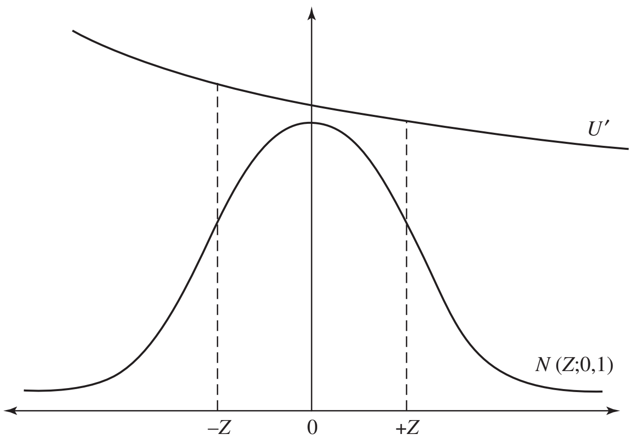

Figure A6.3
The marginal utility for negative values of Z higher than for positive ones.

Proof of the Convexity of Indifference Curves

Let two points (σ_P, μ_P) and $(\sigma_{P'}, \mu_{P'})$ lie on the same indifference curve offering the same level of expected utility $\overline{U}$. Let us consider the point $(\sigma_{P''}, \mu_{P''})$, where $\sigma_{P''} = \alpha\sigma_P + (1-\alpha)\sigma_{P'}$ and $\mu_{P''} = \alpha\mu_P + (1-\alpha)\mu_{P'}$.

One would like to prove that
$E(U(\sigma_{P''}\tilde{Z} + \mu_{P''})) > \alpha E(U(\sigma_P\tilde{Z} + \mu_P)) + (1-\alpha)E(U(\sigma_{P'}\tilde{Z}\_\mu_{P'})) = \overline{U}$.

By the strict concavity of U, the inequality

$$U(\sigma_{P''}Z + \mu_{P''}) > \alpha U(\sigma_P Z + \mu_P) + (1-\alpha)U(\sigma_{P'}Z + \mu_{P'})$$

is verified for all (σ_P, μ_P) and $(\sigma_{P'}, \mu_{P'})$ and any Z value.

One may thus write

$$\int_{-\infty}^{+\infty} U(\sigma_{P''}Z + \mu_{P''})N(Z;0,1)\mathrm{d}Z >$$
$$\alpha\int_{-\infty}^{+\infty} U(\sigma_P Z + \mu_P)N(Z;0,1)\mathrm{d}Z + (1-\alpha)\int_{-\infty}^{+\infty} U(\sigma_{P'}Z + \mu_{P'})N(Z;0,1)\mathrm{d}Z, \text{ or}$$
$$E(U(\sigma_{P''}Z + \mu_{P''})) > \alpha E(U\sigma_P Z + \mu_P)) + (1-\alpha)E(U(\sigma_{P'}Z + \mu_{P'})), \text{ or}$$
$$E(U(\sigma_{P''}Z + \mu_{P''})) > \alpha\overline{U} + (1-\alpha)\overline{U} = \overline{U}$$

See Figure A6.4 for an illustration.

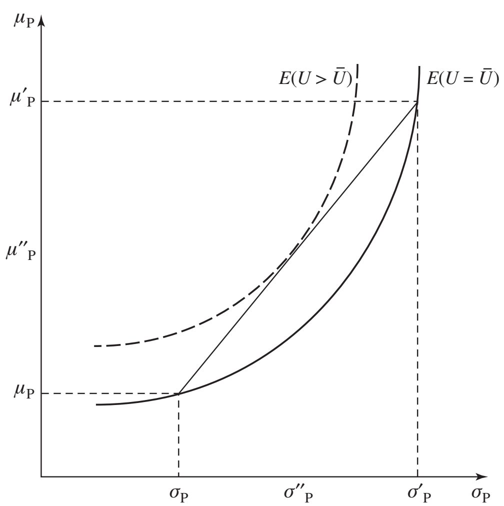

Figure A6.4
The indifference curves are convex-shaped.

Appendix 6.2: The Shape of the Efficient Frontier; Two Assets; Alternative Hypotheses

Perfect Positive Correlation (Figure 6.3)

$$\rho_{12} = 1$$

$\sigma_P = w_1\sigma_1 + (1 - w_1)\sigma_2$, the weighted average of the standard deviations of individuals asset returns

$$\begin{aligned}
\rho_{1,2} &= 1 \\
\mu_P &= w_1\mu_1 + (1 - w_1)\mu_2 = \mu_1 + (1 - w_1)(\mu_2 - \mu_1) \\
\sigma_P^2 &= w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho_{1,2} \\
&= w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2 \\
&= (w_1\sigma_1 + (1 - w_1)\sigma_2)^2 \text{ [perfect square]}
\end{aligned}$$

$$\sigma_P = \pm(w_1\sigma_1 + (1 - w_1)\sigma_2) \Rightarrow w_1 = \frac{\sigma_P - \sigma_2}{\sigma_1 - \sigma_2}; 1 - w_1 = \frac{\sigma_1 - \sigma_P}{\sigma_1 - \sigma_2}$$

$$\mu_P = \mu_1 + \frac{\sigma_1 - \sigma_P}{\sigma_1 - \sigma_2}(\mu_2 - \mu_1) = \mu_1 + \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}(\sigma_P - \sigma_1)$$

Imperfectly Correlated Assets (Figure 6.4)

$$-1 < \rho_{12} < 1$$

Reminder: $\mu_P = w_1\mu_1 + (1 - w_1)\mu_2$

$$\sigma_P^2 = w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2 + 2w_1(1 - w_1)\sigma_1\sigma_2\rho_{1,2}$$

Thus,

$$\frac{\partial\sigma_P^2}{\partial\rho_{1,2}} = 2w_1(1 - w_1)\sigma_1\sigma_2 > 0$$

which implies $\sigma_P < w_1\sigma_1 + (1 - w_1)\,\sigma_2$. σ_P is smaller than the weighted average of the σ's, there are gains from diversifying.

Fix μ_P, hence w_1, and observe: as one decreases $\rho_{1,2}$ (from $+1$ to -1), ${\sigma^2}_P$ diminishes (and thus also σ_P). Hence, the opportunity set for $\rho = \overline{\rho} < 1$ must be to the left of the line AB ($\rho_{1,2} = 1$) except for the extremes.

$$w_1 = 0 \Rightarrow \mu_P = \mu_2 \text{ and } \sigma_P^2 = \sigma_2^2$$
$$w_1 = 1 \Rightarrow \mu_P = \mu_1 \text{ and } \sigma_P^2 = \sigma_1^2$$

Perfect Negative Correlation (Figure 6.5)

$$\rho_{1,2} = -1$$
$$\sigma_P^2 = w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2 - 2w_1w_2\sigma_1\sigma_2 \text{ with } (w_2 = (1-w_1))$$
$$= (w_1\sigma_1 - (1-w_1)\sigma_2)^2 \text{ [perfect square again]}$$
$$\sigma_P = w \pm [w_1\sigma_1 - (1-w_1)\sigma_2] = \pm[w_1(\sigma_1+\sigma_2) - \sigma_2]$$
$$w_1 = \frac{\pm\sigma_P + \sigma_2}{\sigma_1 + \sigma_2}$$
$$\sigma_P = 0 \Leftrightarrow w_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2}$$
$$\mu_P = \frac{\pm\sigma_P + \sigma_2}{\sigma_1 + \sigma_2}\mu_1 + \left(1 - \frac{\pm\sigma_P + \sigma_2}{\sigma_1 + \sigma_2}\right)\mu_2$$
$$= \frac{\pm\sigma_P + \sigma_2}{\sigma_1 + \sigma_2}\mu_1 + \frac{\sigma_1 \pm \sigma_P}{\sigma_1 + \sigma_2}\mu_2$$
$$= \frac{\sigma_2}{\sigma_1 + \sigma_2}\mu_1 + \frac{\sigma_1}{\sigma_1 + \sigma_2}\mu_2 \pm \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2}\sigma_P$$

One Riskless and One Risky Asset (Figure 6.6)

$$\text{Asset } 1: \mu_1, \sigma_1 = 0$$
$$\text{Asset } 2: \mu_2, \sigma_2$$
$$\mu_1 < \mu_2$$

$$\mu_P = w_1\mu_1 + (1-w_1)\mu_2$$
$$\sigma_2^2 = w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2 + 2w_1(1-w_1)\text{cov}_{1,2}$$
$$= (1-w_1)^2\sigma_2^2 \text{ since } \sigma_1^2 = 0 \text{ and } \text{cov}_{1,2} = \rho_{1,2}\sigma_1\sigma_2 = 0; \text{ thus,}$$
$$\sigma_P = (1-w_1)\sigma_2, \text{ and}$$
$$w_1 = 1 - \frac{\sigma_P}{\sigma_2}$$

Appendix 6.3: Constructing the Efficient Frontier

In this appendix, we outline how Excel's SOLVER program may be used to construct an efficient frontier using historical data on returns. Our method does not require the explicit computation of means, standard deviations, and return correlations for the various securities under consideration; they are implicitly obtained from the data directly.

The Basic Portfolio Problem

Let us, for purposes of illustration, assume that we have assembled a time series of four data points (monthly returns) for each of three stocks, and let us further assume that these four realizations fully describe the relevant return distributions. We also assign equal probability to the states underlying these realizations.

Following our customary notation, let w_i represent the fraction of wealth invested in asset i, $i = 1,2,3$, and let r_{P,θ_j} represent the return for a portfolio of these assets in the case of event θ_j, $j = 1,2,3,4$. The Excel formulation analogous to problem (QP) of the text is found in Table A6.1. where ($A1$) through ($A4$) define the portfolio's return in each of the four states; ($A5$) defines the portfolio's average return; ($A6$) places a bound on the expected return; by varying μ, it is possible to trace out the efficient frontier; ($A7$) defines the standard deviation when each state is equally probable; and ($A8$) is the *budget constraint.*

The Excel-based solution to this problem is

$$w_1 = 0.353$$
$$w_2 = 0.535$$
$$w_3 = 0.111$$

Table A6.1: The excel formulation of the (QP) problem

$\min_{\{w_1,w_2,w_3,w_4\}}$SD
(Minimize portfolio standard deviation)
Subject to:
$(A1) r_{P,\theta_1} = 6.23w_1 + 5.10w_2 + 7.02w_3$
$(A2) r_{P,\theta_2} = -0.68w_1 + 4.31w_2 + 0.79w_3$
$(A3) r_{P,\theta_3} = 5.55w_1 - 1.27w_2 - 0.21w_3$
$(A4) r_{P,\theta_4} = -1.96w_1 + 4.52w_2 + 10.30w_3$
$(A5) \mu_P = 0.25r_1^P + 0.25r_2^P + 0.25r_3^P + 0.25r_4^P$
$(A6)\ \mu_P \geq \mu = 3$
$(A7)$ SD = SQRT(SUMPRODUCT($r_{P,\theta_1}, r_{P,\theta_2}, r_{P,\theta_3}, r_{P,\theta_4}$))
$(A8)\ w_1 + w_2 + w_3 = 1$

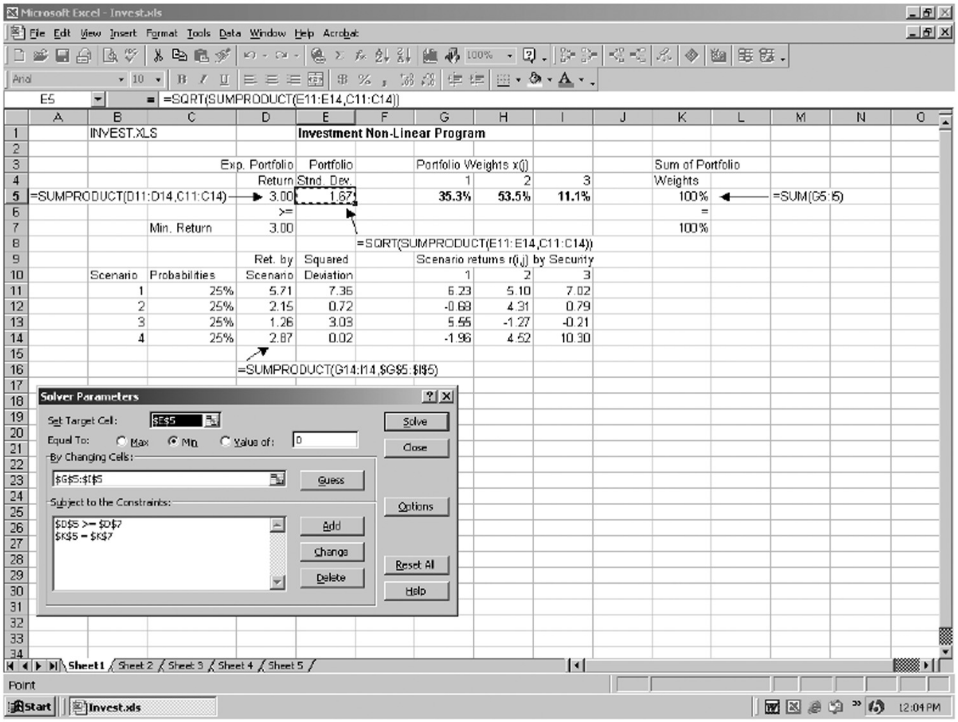

Figure A6.5
Excel Screen 1.

when μ is fixed at $\mu = 3.0\%$. The corresponding portfolio mean and standard deviation are $\mu_P = 3.00$, and $\sigma_P = 1.67$. Screen 1 (Figure A6.5) describes the Excel setup for this case.

Note that this approach does not require the computation of individual security expected returns, variances, or correlations, but it is fundamentally no different from Problem (QP) in the text, which does require them. Note also that by recomputing "min SD" for a number of different values of μ, the efficient frontier can be well approximated.

Generalizations

The approach described above is very flexible and accommodates a number of variations, all of which amount to specifying further constraints.

Nonnegativity Constraints

These amount to restrictions on short selling. It is sufficient to specify the additional constraints

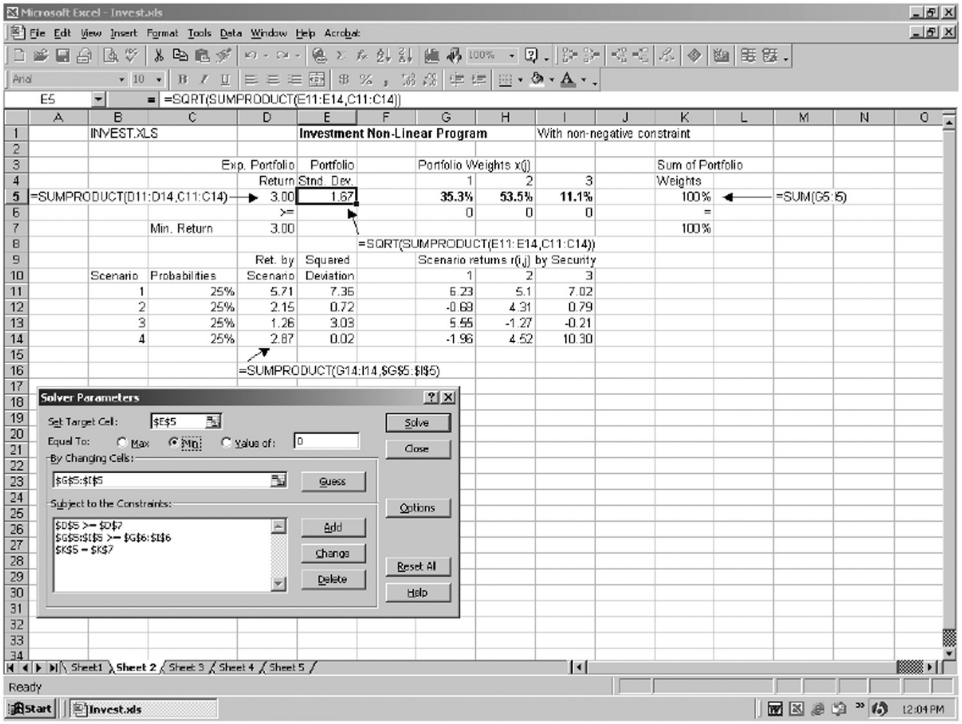

Figure A6.6
Excel Screen 2.

$$w_1 \geq 0$$
$$w_2 \geq 0$$
$$w_3 \geq 0$$

The functioning of SOLVER is unaffected by these added restrictions (although more constraints must be added), and for the example above the solution remains unchanged. (This is intuitive since the solutions were all positive.) See Excel Screen 2 (Figure A6.6).

Composition Constraints

Let us enrich the scenario. Assume that the market prices of stocks 1, 2, and 3 are, respectively, \$25, \$32, and \$17, and that the current composition of the portfolio consists of 10,000 shares of stock 1, 10,000 shares of stock 2, and 30,000 shares of stock 3, with an aggregate market value of \$1,080,000. You wish to obtain the lowest standard deviation for

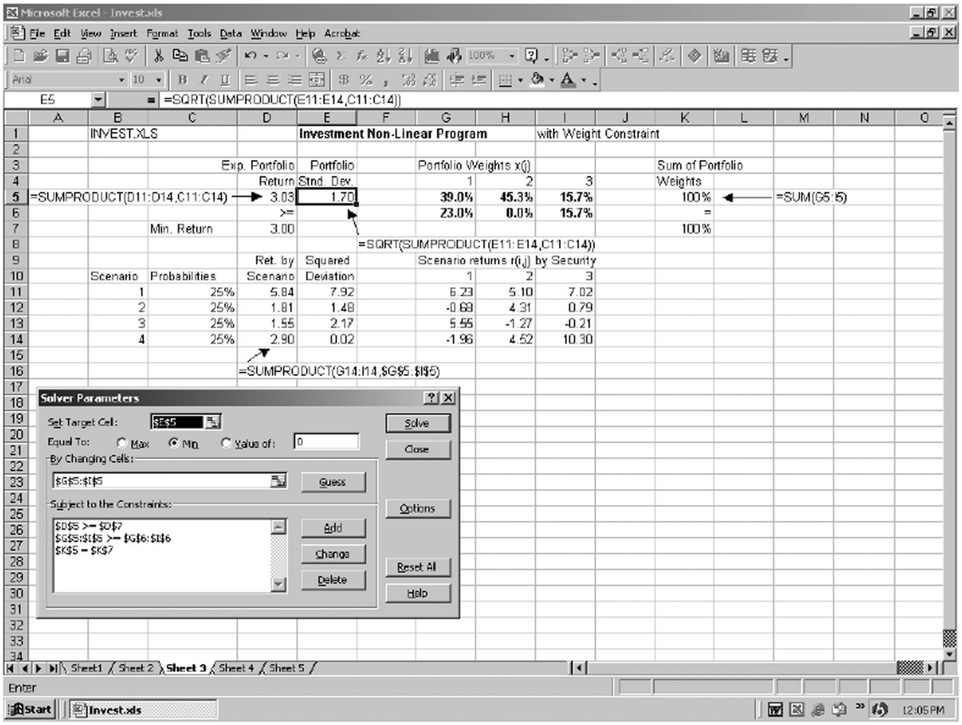

Figure A6.7
Excel Screen 3.

a given expected return subject to the constraints that you retain 10,000 shares of stock 1 and 10,000 shares of stock 3. Equivalently, you wish to constrain portfolio proportions as follows:

$$w_1 \geq \frac{10,000 \times \$25}{\$1,080,000} = 0.23$$

$$w_3 = \frac{10,000 \times \$17}{\$1,080,000} = 0.157$$

while w_2 is free to vary. Again SOLVER easily accommodates this. We find $w_1 = 0.39$, $w_2 = 0.453$, and $w_3 = 0.157$, yielding $\mu_P = 3.03\%$ and $\sigma_P = 1.70\%$. Both constraints are binding. See Excel Screen 3 (Figure A6.7).

Adjusting the Data (Modifying the Means)

On the basis of the information in Table A6.2,

Table A6.2: Hypothetical return data

	Probability	Stock 1 (%)	Stock 2 (%)	Stock 3 (%)
State 1	0.25	6.23	5.10	7.02
State 2	0.25	− 0.68	4.31	0.79
State 3	0.25	5.55	− 1.27	− 0.21
State 4	0.25	− 1.96	4.52	10.30

Table A6.3: Modified return data

	Probability	Stock 1 (%)	Stock 2 (%)	Stock 3 (%)
Event 1	0.25	7.23	6.10	6.02
Event 2	0.25	0.32	5.31	− 0.21
Event 3	0.25	6.55	− 0.27	− 1.21
Event 4	0.25	− 0.96	5.52	9.30

$$\mu_1 = 2.3\%$$
$$\mu_2 = 3.165\%$$
$$\mu_3 = 4.47\%$$

Suppose other information becomes available suggesting that, over the next portfolio holding period, the returns on stocks 1 and 2 would be 1% higher than their historical mean and the return on stock 3 would be 1% lower. This supplementary information can be incorporated into min SD by modifying Table A6.2. In particular, each return entry for stocks 1 and 2 must be increased by 1% while each entry of stock 3 must be decreased by 1%. Such changes do not in any way alter the standard deviations or correlations implicit in the data. The new input table for SOLVER is found in Table A6.3.

Solving the same problem, min SD without additional constraints yields $w_1 = 0.381$, $w_2 = 0.633$, and $w_3 = -0.013$, yielding $\mu_P = 3.84$ and $\sigma_P = 1.61$. See Excel Screen 4 (Figure A6.8).

Constraints on the Number of Securities in the Portfolio

Transactions costs may be substantial. In order to economize on these costs, suppose an investor wished to solve min SD subject to the constraint that his portfolio would contain at most two of the three securities. To accommodate this change, it is necessary to introduce three new binary variables that we will denote x_1, x_2, x_3, corresponding to stocks 1, 2, and 3, respectively. For all x_i, $i = 1, 2, 3$, $x_i \in \{0, 1\}$. The desired result is obtained by adding the following constraints to the problem min SD:

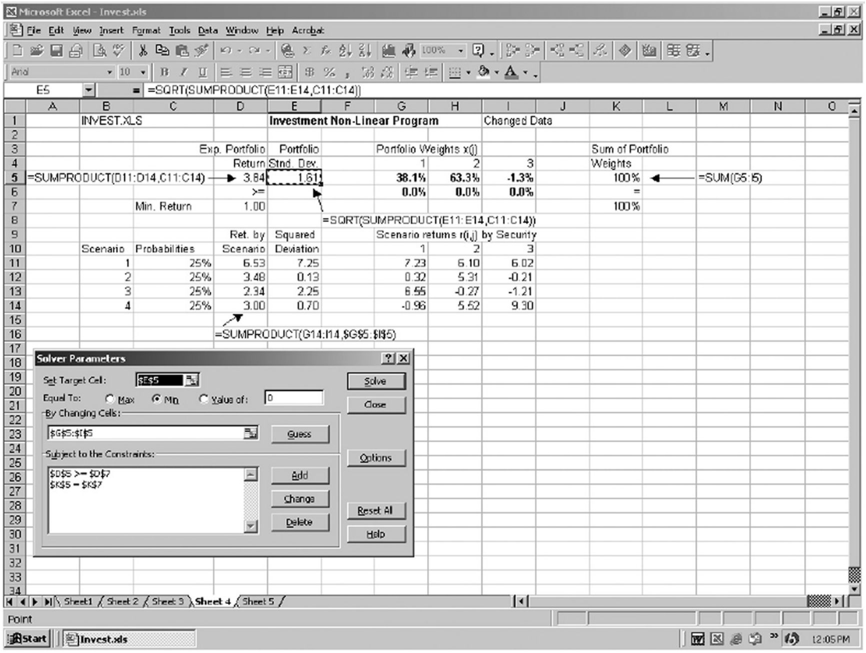

Figure A6.8
Excel Screen 4.

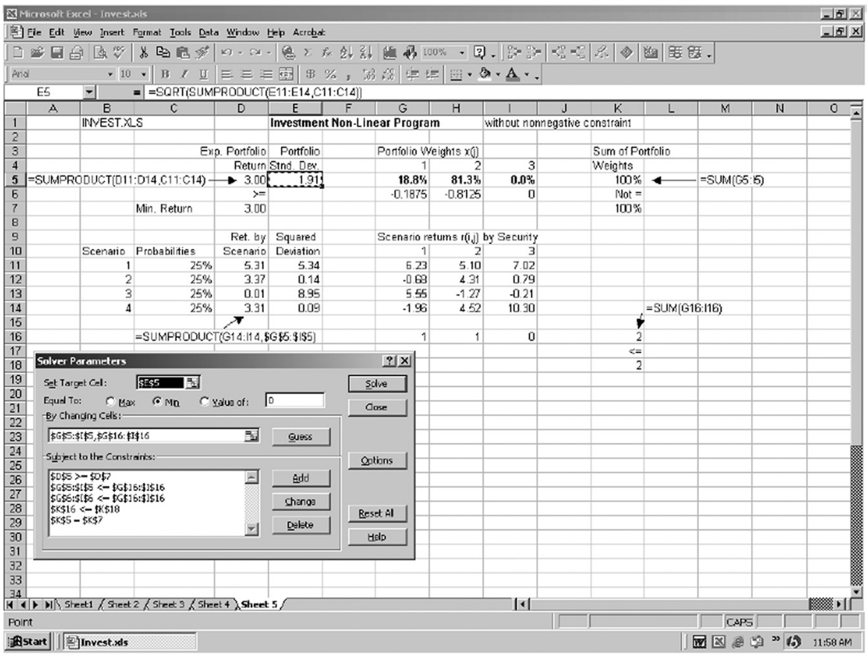

Figure A6.9
Excel Screen 5.

$$
\begin{aligned}
& w_1 \le x_1 \\
& w_2 \le x_2 \\
& w_3 \le x_3 \\
& x_1 + x_2 + x_3 \le 2, \\
& x_1, x_2, x_3 \ \text{are binary}
\end{aligned}
$$

In the previous example, the solution is to include only securities one and two with proportions $w_1 = 0.188$, and $w_2 = 0.812$. See Excel Screen 5 (Figure A6.9).