Pricing by the absence of arbitrage or by using a SDF is an elegant theory, but it is difficult in practice to apply this theory unless the number of states is not large and the relation between the assets’ cash flows in those states can be established. This is where the Capital Asset Pricing Model enters the picture. In the CAPM, the states need not be identified, and finding the relation between cash flows or returns is simplified to computing statistical covariances.

Mean-variance analysis and the CAPM are certainly two of the best-known and most often used tools of modern finance theory. Mean-variance analysis uses variance as the single measure of risk. This is consistent with expected utility theory only under limited conditions, but these limitations are more than offset by the simplicity of use and the powerful intuitions it provides.

Mean-variance theory is based on the idea that the expected utility of any portfolio can be completely described by its mean and variance. The optimal portfolio can then be chosen based on a tradeoff between the two. There are two distinct models for which this is true. If utility is quadratic, then the expected utility depends only on mean and variance regardless of the distribution, provided expected utility is defined. The other model restricts the random variables allowed so that the distribution of any portfolio’s return is completely described by the mean and variance.

### Quadratic Utility

Any cardinal utility function can be changed to another equivalent one by a positive linear transformation. So distinct quadratic utility functions are characterized by a single parameter as $u(W) = W - bW^2$ with no loss of generality. The constant and the coefficient on $W$ have been eliminated by the linear transformation leaving only the parameter $b$ which must be positive for risk aversion. For an investor with quadratic utility, expected utility is

$$
\mathbb{E}[u(\tilde{W})] = \mathbb{E}[\tilde{W} - b\tilde{W}^2] = \mathbb{E}[\tilde{W}] - b(\text{var}[\tilde{W}] + \mathbb{E}^2[\tilde{W}]).
$$

This depends only on the mean and variance, and higher variance is disliked for $b > 0$.

There are two problems with this approach. First marginal utility is only positive for outcomes less than $W_{\text{max}} \equiv (2b)^{-1}$. Second the Arrow-Pratt measure of risk aversion for quadratic utility is $A(W) = 2b/(1 - 2bW) = 1/(W_{\text{max}} - W)$. This is negative for wealth above $W_{\text{max}}$, but those outcomes have already been excluded. This problem could be avoided by setting $b$ so that the maximum wealth level is far removed from current wealth. Suppose $b$ is set so that $W_{\text{max}} = kW_0$. Then relative risk aversion is $W_0A(W_0) = (k - 1)^{-1}$. Estimates of relative risk aversion vary widely, but most asset based studies indicate that it lies somewhere in the range of 1 to 5. Those numbers would restrict the maximum wealth to be considered to 1.2 to 2 times current wealth.

Furthermore, even if such a restriction were acceptable absolute risk aversion is increasing for quadratic utility. This is at odds with the evidence that indicates decreasing absolute risk aversion.

Mean-variance analysis based on quadratic utility is used sometimes as an approximation for small risks as shown in the Further Notes section at the chapter end. It is also a basis for analysis in continuous-time models as seen in later chapters. However, for single-period models, the usual justification for mean-variance analysis is based on the distribution of returns.

### Elliptical Distributions

One obvious necessary condition for mean-variance analysis to be valid is that the distribution must be described by just two parameters that can be mapped to mean and variance. There are many such distributions like the normal, the uniform, the gamma, and the Laplace. Some other
distributions have only a single parameter, like the exponential, or more than two, like the Weibull distribution. Tobin (1958) conjectured that any two parameter distribution provided indirect preferences over means and variances, but this is not sufficient. It is not fixed gambles that are being compared, but portfolios so combinations of the distributions must still fall within in the same two-parameter class if valid comparisons are to be made. This immediately eliminates some distributions. For example, combinations of two independent uniformly distributed variables have a triangular density while combinations of three independent uniformly distributed variables have a density that is three quadratic curves pieced together. So combinations of uniformly distributed variables cannot be compared knowing just the mean and variance. What is required for portfolio analysis is that the return distribution be unchanged under addition. The general class of independently distributed random variables with this additive property is the Lévy stable distributions. Of these, only the normal distribution has a finite variance.\footnote{The stable distributions are discussed more at the end of this chapter.}

Of course, there is no need for the individual asset returns to be mutually independent; in fact, they definitely are not so they should not be modeled that way. Correlation could be introduced with a factor structure; for example, \( \tilde{r} = a + b\tilde{\varepsilon} + \tilde{e} \) with \( \tilde{\varepsilon} \) and \( \tilde{e} \) having a joint stable distribution. However, there is a better more-inclusive choice, the elliptical distributions.

Elliptical distributions were introduced into portfolio analysis by Chamberlain (1983) and Owen and Rabinovitch (1983). These distributions get their name from the ellipsoidal shape of their isoprobability manifolds. The best-known example of an elliptical distribution is the multivariate normal. But other than familiarity, there is little advantage in assuming normality rather than a more general elliptical distribution in portfolio analysis.

Elliptical distributions are characterized by a probability density, \( f_{\psi} \), (if it exists) and a characteristic function, \( \Psi \),

\[
f(r) = k |\Sigma|^{-1/2} g\left((r - \mu)\Sigma^{-1}(r - \mu)\right)
\]

\[
\Psi(\theta) \equiv \mathbb{E}[e^{\theta^T r}] = e^{\theta^T \mu + \frac{1}{2} \theta^T \Sigma \theta}
\]

where \( \mu \) is the vector of means, \( \Sigma \) is the covariance matrix, and \( k \) is a scaling factor that makes the density integrate to 1.\footnote{Elliptical distributions include all the symmetric stable distributions and like them, can be fat-tailed with undefined means or variances. However, elliptical distributions are symmetric so \( \mu \) is always the vector of medians and \( \Sigma \) is some general co-dispersion matrix. The discussion below remains valid for such elliptical distributions and is consistent with expected utility maximization provided that expected utility is defined for the distribution in question.} The covariance matrix, \( \Sigma \), is positive definite so there are no linear redundancies among the assets. A risk-free asset can be included as a degenerate case with one row and column of \( \Sigma \) all zeros, but is generally handled separately to keep the matrix positive definite.

The function \( g \) can be essentially any continuous function from the nonnegative reals to the nonnegative reals for which the density can be normalized to unit mass over the relevant domain. The best-known elliptical distribution is the multivariate normal. Other examples include the multivariate Student-t distribution, the multi-variate Laplace distribution, and variance-subordinated normal distributions. It is obvious from (2) that all elliptical distributions are symmetric about their point of means just like the multivariate normal.

The normal and Laplace distributions will be used later in some examples. Their densities and characteristic functions are

\[
\begin{align*}
\text{Normal:} & \quad g_N(\cdot) = \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) & \psi_N(\theta) = \exp(i\mu\theta - \frac{1}{2}\sigma^2\theta^2) \\
\text{Laplace:} & \quad g_L(\cdot) = \exp\left(-\sqrt{2} |x - \mu| / \sigma \right) & \psi_L(\theta) = \exp(i\mu\theta)(1 + \frac{1}{2}\sigma^2\theta^2)^{-1}
\end{align*}
\]
Stability under addition or portfolio combinations can be verified immediately from (2). The distribution of the rate of return on a portfolio with weights \( w \) has a mean \( \mu_p = w^t \mu \), a variance \( \sigma_p^2 = w^t \Sigma w \), and a characteristic function of \( E[\exp(i \theta^t \tilde{r})] \). Define the vector \( \theta = \theta^t w \), then

\[
E[e^{i \theta^t \tilde{r}}] = E[e^{i \theta^t \psi(\theta^t \tilde{\Sigma} \theta)}] = e^{i \theta^t \mu_p \psi(\theta^t \Sigma w)} = \exp(i \theta^t \mu_p \psi(\theta^t \sigma_p^2))
\]

(4)

Obviously, the characteristic functions of all portfolios come from the same distribution differing only in mean and variance, \( \mu_p \) and \( \sigma_p^2 \). Therefore, expected utilities can be computed from just the expectation and variance of their outcomes and, of course, the common basic distribution defined by \( \psi \).

Just as with the normal distribution, any elliptical asset’s or portfolio’s return can be expressed as a translated and scaled variable \( \tilde{r}_p = \mu_p + \sigma_p \tilde{\nu} \) where \( \tilde{\nu} \) is a standardized elliptical variable with zero mean and unit variance and standardized density \( f(\nu) = kg(\nu^2) \). All of this remains true if there is a risk-free asset. The portfolio weights, \( w \), are simply not constrained to sum to one. The risk-free asset changes the mean of the portfolio return to \( \mu_w = (1 - 1^t w) \mu_f + w^t \mu \) and scales the standard deviation by \( 1^t w \).

**Minimum-Variance and Mean-Variance Efficient Portfolios**

Because all portfolios are completely characterized by their mean and standard deviation, the expected utility of any portfolio can be described by the derived utility function

\[
V(\mu_w, \sigma_w^2) \equiv E[u(W_0(1 + \tilde{r}_w))] = \int_{-\infty}^{\infty} u(W_0(1 + \mu_w + \sigma_w \nu)) g(\nu) d\nu.
\]

(5)

The expectation has been computed in the definition so \( V \) is an ordinal rather than cardinal utility function. Positive marginal utility and risk aversion of the direct utility function translate to the indirect utility function as a preference for higher mean and lower variance.

At the margin an increase in mean return holding variance fixed increases utility by

\[
\frac{\partial V(\mu_w, \sigma_w^2)}{\partial \mu_w} = \frac{\partial}{\partial \mu_w} \int_{-\infty}^{\infty} u(W_0(1 + \mu_w + \sigma_w \nu)) g(\nu) d\nu = W_0 \int_{-\infty}^{\infty} u'(W_0(1 + \mu_w + \sigma_w \nu)) g(\nu) d\nu > 0.
\]

(6)

This must be positive because both marginal utility and the probability density are positive. The preference for the highest expected rate of return at a given variance shows that the set of maximum-mean portfolios must include all the optimal or efficient portfolios. In fact, the sets are identical as shown later so the maximum-mean portfolios are generally called the mean-variance efficient portfolios.

For many purposes, it is simpler to work with more inclusive set, namely the minimum-variance portfolios. The elements of this set are the unique portfolio with smallest possible variance at each mean. This set also partially characterizes efficiency because at a given mean, the portfolio with the lowest variance is most preferred

\[
\frac{\partial V(\mu_w, \sigma_w^2)}{\partial \sigma_w} = \frac{\partial}{\partial \sigma_w} \int_{-\infty}^{\infty} u(W_0(1 + \mu_w + \sigma_w \nu)) g(\nu) d\nu = \int_{-\infty}^{\infty} \rho W_0 u'(W(1 + \mu_w + \sigma_w \nu)) g(\nu) d\nu = \int_{0}^{\infty} \nu W_0 \left[u'(W_0(1 + \mu_w + \sigma_w \nu)) - u'(W_0(1 + \mu_w - \sigma_w \nu))\right] g(\nu) d\nu < 0.
\]

(7)

The second equality in (7) follows because elliptical densities are symmetric about 0 with \( g(-\nu) =
g(υ). This integral is negative when utility is strictly concave because the marginal utility being subtracted exceeds the first and the other factors are positive.

Thus, at a given variance, investors always prefer the portfolio with the highest mean and at a given mean, and the lowest variance at a given mean. However, these sets are not identical. The set of mean-variance efficient portfolios is only a subset of the set of minimum-variance portfolios.

\[ \mu = \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} + \eta [\mathbf{w} - \mathbf{w}^T \mu] + \gamma (1 - \mathbf{1}' \mathbf{w}) \]

\[ 0 = \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \Sigma \mathbf{w} - \eta \mu - \gamma \mathbf{1}. \]

If the covariance matrix is nonsingular, the minimum-variance portfolio at the expected rate of return of \( \hat{\mu} \) is \( \hat{\mu} \Sigma^{-1} \mu + \gamma \Sigma^{-1} \mathbf{1} \). The covariance matrix is nonsingular provided there is no risk-free asset and there are no arbitrages or redundant assets among the risky assets. Using the constraints to determine the Lagrange multipliers, the minimum variance set is

\[
\begin{align*}
\mathbf{w}_\mu &= \eta \mu \Sigma^{-1} \mathbf{1} + \gamma \Sigma^{-1} \mu \\
&= \frac{C \hat{\mu} - A \Sigma^{-1} \mu + B - A \hat{\mu}}{D} \Sigma^{-1} \mu + \frac{B - A \hat{\mu}}{D} \Sigma^{-1} \mathbf{1} \\
\text{where } &A \equiv \mathbf{1}' \Sigma^{-1} \mu, \quad B \equiv \mu' \Sigma^{-1} \mu > 0, \quad C \equiv \mathbf{1}' \Sigma^{-1} \mathbf{1} > 0, \quad D \equiv BC - A^2 > 0.
\end{align*}
\]

From the first order condition and the definitions of \( \eta \) and \( \gamma \) in (9), the variance of this portfolio is

\[
\sigma^2 = w'_\mu \Sigma w_\mu = w'_\mu (\eta \mu + \gamma \mu) \mathbf{1} = \eta \hat{\mu} + \gamma \mu = \frac{C \hat{\mu} - A \hat{\mu} + B - A \hat{\mu}}{D} = \frac{C \hat{\mu}^2 - 2 A \hat{\mu} + B}{D}. \]

So the minimum-variance frontier is a parabola in mean-variance space which is a hyperbola in

3 Because an increase in variance always decreases \( V \), and therefore expected utility, variance is a complete ordering of riskiness in a Rothschild-Stiglitz sense among distributions of a given elliptical type.

4 \( B \) and \( C \) are positive because they are quadratic forms of a positive definite matrix. \( D \) is positive by the Cauchy Schwart inequality. \( A \) can have either sign.
Equation (9) identifies two particular portfolios, with weights proportional to $\Sigma^{-1}\mu$ and to $\Sigma^{-1}1$. These portfolios are labeled as $\tau \equiv \Sigma^{-1}\mu/A$ and $g \equiv \Sigma^{-1}1/C$ in the figure; the denominators serve only to normalize the portfolio weights to sum to 1. For portfolio $g$, $\eta = 0$ so the constraint on the expected rate of return is not active. Therefore, portfolio $g$ is the global minimum variance portfolio as shown. This is also obvious because $g = (1\Sigma^{-1}1)^{-1}\Sigma^{-1}1$ is the only portfolio on the minimum variance frontier whose weights are independent of the expected rates of return which must be true of the global minimum variance portfolio.

The expected rate of return and variance of the global minimum-variance portfolio are $\mu_g = g^T\mu = A/C$ and $\sigma_g^2 = g^T\Sigma g = 1/C$. From (10), $\sigma_g^2 \sim C\hat{\mu}^2/D$ for large $\hat{\mu}$ so the asymptotes of the hyperbola are $\mu = \mu_g \pm \sigma\sqrt{D/C}$. Portfolio $\tau$ lies at the point that a line through the origin is tangent to the hyperbola; although this characterization is unimportant. The expected rate of return and variance of portfolio $\tau$ are $\mu_\tau = \tau^T\mu = B/A$ and $\sigma_\tau^2 = \tau^T\Sigma\tau = B/A^2$. The sign of $\mu_\tau - \mu_g$ is

$$\text{sgn}(\mu_\tau - \mu_g) = \text{sgn}\left(\frac{B}{A} - \frac{A}{C}\right) = \text{sgn}\left(\frac{BC - AC^2}{AC^2}\right) = \text{sgn}(A) = \text{sgn}(\mu_g)$$

so either $\mu_\tau > \mu_g > 0$ or $\mu_\tau < \mu_g < 0$. The former is the standard case.

The precise location of portfolio $\tau$ is unimportant. What is important is that the entire minimum-variance set is spanned by these two portfolios, $W_{mv} = \alpha\tau + (1 - \alpha)g$ as shown in (9). Because the minimum variance set includes all of the mean-variance efficient portfolios, that set, too, is spanned by these two portfolios or any other two portfolios in the set.

If there is a risk-free asset, then the covariance matrix is singular and (9) does not apply. In this case the minimization problem can be restated in excess return form with the risk-free asset eliminated. The constraint in excess return form is $w'(\mu - r_f 1) = 0$. The portfolio weights are investments in the risky assets only so they need not sum to 1. If $1'w < 1$, then the residual amount is invested in the risk-free asset. If $1'w > 1$, then the investor is borrowing to buy on margin. The problem and solution are

$$L = \frac{1}{2} w'\Sigma w + \lambda[x - w'(\mu - r_f 1)]$$

$$0 = \frac{\partial L}{\partial w} = \Sigma w - \lambda(\mu - r_f 1) \quad \Rightarrow \quad w_i = \lambda_i \Sigma^{-1}(\mu - r_f 1).$$

Figure 5.2: Minimum-Variance Portfolios and Spanning

This figure illustrates the location of portfolios $\tau$ and $g$ that together span all of the minimum variance portfolios.
Solving for constraint, \( w'_x (\mu - r) = x \), determines \( \lambda = x/a \) and

\[
 w_x = \frac{x}{b} \Sigma^{-1}(\mu - r, 1) = \frac{xa}{b} t
\]

where \( b \equiv (\mu - r, 1)' \Sigma^{-1}(\mu - r, 1) = B - 2r, A + r^2 C \)

\[
a \equiv 1' \Sigma^{-1}(\mu - r, 1) = A - r, C \quad t \equiv \frac{\Sigma^{-1}(\mu - r, 1)}{a} .
\]

The expected excess rate of return and variance on portfolio \( t \) are \( \mu_t - r = t'(\mu - r, 1) = b/a \) and \( \sigma_t^2 = t' \Sigma t = b/a^2 \).

The weights of portfolio \( t \) are

\[
 t = \frac{A \tau - r, C g}{b} = \frac{A \tau - r, C g}{A - r, C} .
\]

Therefore, \( t \) lies on the hyperbola of minimum-variance portfolios. Combinations of any portfolio with the risk-free asset trace out a straight line in \( \mu-\sigma \) space, so portfolio \( t \) must be at the point of tangency between the hyperbola and a line through the origin in excess return space. It cannot lie on a shallower line as it would not then be a minimum variance portfolio. It is called the tangency portfolio for this reason.

The tangent line represents portfolio combinations of the risk-free asset and portfolio \( t \). This line is called the borrowing-lending line, for an obvious reason, or the capital market line. Because the risk-free asset contributes nothing to risk or expected excess rate of return, both scale in proportion in these combinations. The standard deviation as a function of the excess expected rate of return \( x \) is

\[
 \sigma_x = \sqrt{w_x' \Sigma w_x} = \frac{|x|}{b} \sqrt{(\mu - r, 1)' \Sigma^{-1} \Sigma^{-1}(\mu - r, 1)} = \frac{|x|}{\sqrt{b}} .
\]

So two rays originating at the origin have slopes of \( \pm \sqrt{b} \) as shown in the figure. The efficient portfolios form the upper ray. The risky-asset-only global minimum-variance portfolio has an expected excess rate of return of \( \mu_g = g' \mu = (1' \Sigma^{-1} 1)^{-1} 1' \Sigma^{-1} \mu = A/C \). If this exceeds \( r \), the tangent portfolio is on the upper ray as shown. This is the standard case in the CAPM equilibrium.
discussed below. As before, the entire mean-variance efficient set is spanned by any two portfolios in it, any two portfolios lying on the upper ray. In this case there are two obvious portfolios to choose, the risk free asset and portfolio \( t \).

Another way to describe the tangency portfolio is as the portfolio of risky-assets that maximizes the Sharpe ratio. The Sharpe ratio of any portfolio is the slope of the line connecting it with the origin, its expected excess rate of return divided by its standard deviation. All combinations of any portfolio with the risk free asset trace out a ray in the figure (or two rays with opposite slopes if shorting the risky-asset portfolio is allowed). Therefore, if there were a portfolio with a higher Sharpe ratio than \( t \), it could not be mean-variance efficient.

The Sharpe ratio maximization problem is

\[
\max \frac{w'\left(\mu - r_f\right)}{(w'\Sigma w)^{1/2}} \quad \text{subject to} \quad 1'w = 1.
\]

However, the constraint need not be applied because the Sharpe ratio is homogeneous of degree zero in the portfolio weights. If \( w \) maximizes (16), then so does \( kw \) for any \( k \neq 0 \). This means we can ignore the constraint in the maximization and then normalize the weights to sum to 1.

The first-order condition is

\[
0 = \frac{\partial S}{\partial w} = \frac{(\mu - r_f)'}{(w'\Sigma w)^{1/2}} \cdot \frac{w'\left(\mu - r_f\right)}{(w'\Sigma w)^{1/2}} \Sigma w.
\]

Fortunately, we need not solve (17) but only verify that \( t \), or more simply, its unnormalized equivalent, \( t_u = \Sigma^{-1}(\mu - r_f) \), is a solution.

\[
\frac{(t_u'\Sigma t_u)(\mu - r_f)' - t_u'(\mu - r_f)\Sigma t_u}{(t_u'\Sigma t_u)^{1/2}} = \frac{a(\mu - r_f)' - a(\mu - r_f)}{(t_u'\Sigma t_u)^{1/2}} = 0
\]

verifying that \( t_u \), and therefore \( t \), is the Sharpe ratio maximizing portfolio. The maximized Sharpe ratio is

\[
S_{\max} = S_t = \sqrt{(\mu - r_f)' \Sigma^{-1}(\mu - r_f)} = \sqrt{b} = \frac{\mu_t - r_f}{\sigma_t}
\]

which is the slope of the ray from 0 to the tangency portfolio, \( t \).

This maximum Sharpe ratio is sometimes called the price of risk because it is the best possible increase in the expected rate of return per unit increase in standard deviation. But the ratio \((\mu_t - r_f) / \sigma_t^2\) is also called the price of risk. It is the increase in expected rate of return per unit increase in portfolio variance and also the increase in expected rate of return per unit increase in covariance as described below.

The analysis here is true for any set of random variables with a non-singular covariance matrix. What makes elliptical variables important is that only for elliptical variables is it true that the mean and variance completely describe the probability distribution of all portfolios.

**The Capital Asset Pricing Model (CAPM)**

The tangency portfolio has the property that each asset’s expected excess rate of return is

\[^5\text{If } \mu_g > r_f, \text{ then the tangency portfolio is on the lower ray. All mean-variance-efficient portfolios short the tangency portfolio and lend. If } \mu_g = r_f, \text{ then the two rays of the mean-variance efficient set are both asymptotes hyperbola, and the only “tangency” is at } \sigma = \infty. \text{ Neither of the latter cases will arise in equilibrium. Some older texts illustrate the minimum risk set as having two tangencies, one on each limb. This is impossible.}\]
proportional to the covariance of its return with the return on the tangency portfolio. The vector of covariances of the assets’ returns with any portfolio, \( \mathbf{w} \), is \( \mathbf{\psi}_w = \mathbf{\Sigma} \mathbf{w} \). So for the tangency portfolio, as defined in (13) the covariances are

\[
\mathbf{\psi}_t = \mathbf{\Sigma} \mathbf{t} = \mathbf{\Sigma} \mathbf{a}^{-1} \mathbf{\Sigma}^{-1} (\mathbf{\mu} - r_f \mathbf{1}) = \mathbf{a}^{-1} (\mathbf{\mu} - r_f \mathbf{1}) \equiv (\mathbf{\mu}_t - r_f \mathbf{1}) \quad \Rightarrow \quad (\mathbf{\mu}_t - r_f \mathbf{1}) = \mathbf{a} \mathbf{\psi}_t = \mathbf{\psi}_t \frac{\mathbf{\mu}_t - r_f \mathbf{1}}{\sigma_t^2}.
\]

The final equality is true because the relation must hold for every portfolio including the tangency portfolio itself, so \( \sigma_t^2 = t' \mathbf{\psi}_t = \mathbf{a} \sigma_t^2 = t' (\mathbf{\mu} - r_f \mathbf{1}) \equiv (\mathbf{\mu}_t - r_f \mathbf{1}) \), and \( \mathbf{a} = (\mathbf{\mu}_t - r_f \mathbf{1})/\sigma_t^2 \). This is a useful property because once the tangency portfolio is identified, covariances with it explain all expected rates of return.

This last result is a mathematical tautology about minimizing variance. As such there is no economic content. We can introduce economic content by arguing: if (i) all investors have homogeneous beliefs and the same investment horizon, (ii) the assets are all infinitely divisible and can be traded without other frictions, and (iii) investors can borrow and lend unlimited amounts at the risk-free rate, then all investors will solve the same minimum-variance problem. Therefore, all of their portfolio demands are some combination of the tangency portfolio and borrowing or lending. So aggregate demand for the risky assets must be proportional to the tangency portfolio. In equilibrium demand equals supply so the tangency portfolio must be the market portfolio, and expected rate of return are linearly related to betas with the market portfolio

\[
(\mathbf{\mu}_t - r_f \mathbf{1}) = \mathbf{\psi}_m \frac{\mathbf{\mu}_m - r_f \mathbf{1}}{\sigma_m^2} = (\mathbf{\mu}_m - r_f \mathbf{1}) \mathbf{\beta}_m.
\]

\( \mathbf{\beta}_m \equiv \mathbf{\psi}_m / \sigma_m^2 \) is the vector of covariances of the returns on the stocks with the returns on the market portfolio divided by the market’s variance. In other words, it is the vector of regression coefficients of assets’ returns on the market’s return.

This equation embodies the important lesson of the CAPM. There are some risks that cannot be reduced by diversification. If there are common factors tend to affect asset prices in the same way, then those risk will remain even in a diversified portfolio. With everyone trying to eliminate risks, it is the risk of the market portfolio that remains, and that risk must receive compensation. Further in this chapter and in later ones we will see that other risks might be important and receive compensation as well.

The Zero-Beta CAPM

If there is no risk-free asset, there is no tangency portfolio, and it obviously cannot be the market portfolio. However, similar reasoning still applies. As before denote by \( \mathbf{\psi}_{mv} \), the vector of covariances of the assets’ returns with the return on a given minimum-variance portfolio. As seen previously, all these portfolios are spanned by \( \mathbf{\tau} \) and \( \mathbf{g} \) so for \( \mathbf{w}_{mv} = \alpha \mathbf{\tau} + (1 - \alpha) \mathbf{g} \),

\[
\mathbf{\psi}_{mv} = \mathbf{\Sigma} \mathbf{w}_{mv} = \mathbf{\Sigma} [\alpha \mathbf{\tau} + (1 - \alpha) \mathbf{g}] = \alpha \frac{\mathbf{\Sigma} \mathbf{\Sigma}^{-1} \mathbf{\mu}}{1' \mathbf{\Sigma}^{-1} \mathbf{\mu}} + (1 - \alpha) \frac{\mathbf{\Sigma} \mathbf{\Sigma}^{-1} \mathbf{1}}{1' \mathbf{\Sigma}^{-1} \mathbf{1}} = c_1 \mathbf{\mu} + c_2 \mathbf{1}.
\]

\[6\] More precisely, it is the market portfolio of risky assets only. If the risk-free asset (e.g., government bonds) is in a positive net supply, it is part of the market, but part of the market portfolio of risky assets. All statements about the market portfolio in the CAPM may be interpreted to mean the market portfolio of risky assets only.

\[7\] Any portfolio except the global minimum variance can be used. The latter cannot be used because all portfolios and assets have the same covariance with \( \mathbf{g} \); \( \sigma_{w_g} = \mathbf{w} \mathbf{\Sigma} \mathbf{g} = \mathbf{w} \mathbf{\Sigma} \mathbf{\Sigma}^{-1} \mathbf{1}/C = 1/C = \sigma_g^2 \).
It would be circular reasoning to use portfolio $\tau$ to describe expected rates of return because it is defined using $\mu$. However, as before, all investors want to hold combinations of $\tau$ and $g$ so the market portfolio must be a combination as well. Therefore, it is a minimum-variance portfolio. Neither the market $m$ nor portfolio $g$ depend on knowing the expected rates of return. So (22) gives

$$\sigma_m^2 = m^\prime \psi_m = c_1 \mu_m + c_2 \quad \sigma_{mg} = g^\prime \psi_m = c_1 \mu_g + c_2 \ .$$

These two equations can be solved for $c_1$ and $c_2$. When introduced back into (22), we have

$$\mu = \frac{\mu_m - \mu_g}{\sigma_m^2 - \sigma_{mg}} \psi_m - \frac{\mu_m \sigma_{mg} - \mu_g \sigma_m^2}{\sigma_m^2 - \sigma_{mg}} \mathbf{1} \ .$$

This equation shows that covariances with the market still explain all differences in expected rates of return, $\mu_i - \mu_j \propto \sigma_{mi} - \sigma_{mj}$, though the relation is not nearly as simple as the CAPM.

There is another minimum-variance portfolio that can be identified without using expected rates of return, and the equilibrium relation is simpler with it. This is the minimum-variance portfolio that is uncorrelated with the market — the zero-beta portfolio. The zero-beta portfolio is the solution to

$$\min_w \frac{1}{2} w^\prime \Sigma w + \eta(0 - m^\prime \Sigma w) + \gamma(1 - 1^\prime w) \ .$$

The first order conditions are $\mathbf{0} = \Sigma w - \eta \Sigma m - \gamma \mathbf{1}$. The constants can be determined by applying $1^\prime z = 1$ and $m^\prime \Sigma z = 0$. The zero-beta portfolio holdings are

$$z = \frac{m - \sigma_m^2 \Sigma^{-1} 1}{1 - \sigma_m^2 \Sigma^{-1} 1} \ .$$

The relation in (24) still holds but now $\sigma_{mg}$ is replaced by $\sigma_{mz} = 0$ so (24) simplifies to

$$\mu - \mu_z \mathbf{1} = \frac{\mu_m - \mu_z}{\sigma_m^2} \psi_m = (\mu_m - \mu_z) \beta_m \ .$$

This is exactly the CAPM beta pricing relation with the expected rate of return on the zero-beta portfolio replacing the risk-free rate. This relation is generally known as the (Fischer) Black or zero-beta CAPM.

The one additional requirement for this model is that short sales be unrestricted. This assumption replaces the assumption of unrestricted borrowing and lending. Short sales of the risky assets are not necessary in the original CAPM because all investors hold the market and the market portfolio obviously has no short positions so no one will wish to short any risky asset. But here short positions are required to create the portfolios with very high expected rates of return and possibly the portfolios near the global minimum variance portfolio.

The same pricing relation holds if investors can lend but not borrow at a risk-free rate or if there are different interest rates for lending, $r_L$, and for borrowing, $r_B > r_L$. This can be seen easily in the figure. Investors’ optimal portfolios will be of three types: combinations of portfolio $A$ with lending, portfolios on the minimum-variance hyperbola located between $A$ and $B$, or holdings of portfolio $B$ levered up with borrowing. If no risk-free borrowing is allowed, then there are no optimal portfolios in this last set. Instead long positions in portfolio $B$ with short positions in portfolio $A$ will trace out the remainder of the minimum-variance hyperbola above $B$. In this case, there is no special portfolio $B$; any portfolio above $A$ can be used. Once again each investor holds some combination of the minimum-variance portfolios $A$ and $B$, and the zero-beta version of the CAPM will obtain. When borrowing is possible, the market will be a minimum-variance portfolio somewhere between $A$ and $B$. The borrowing and lending rates play no role in the equilibrium.
except that the zero-beta rate must be between the two. It coincides with the lending rate if all investors lend and it coincides with the borrowing rate if all investors borrow. Whether or not borrowing without lending or vice versa is possible depends on the market environment. In either case, there must be some source of borrowing or lending outside of the market considered. Government bonds could be an obvious outside source that permits investors to lend if they are not considered part of total wealth. Bank deposits by investor who do not participate in the stock market could be a source to provide borrowing to investors without lending.

The CAPM Equilibrium

There are four distinct portions of the derivation of the CAPM pricing result, (i) verifying that mean and variance provide a complete description of all portfolios’ returns distributions, (ii) determining the relation between each asset’s expected rate of return and its covariance with the tangency portfolio, (iii) proving that all investors wish to hold mean-variance efficient portfolios, and (iv) demonstrating that the tangency portfolio is the market portfolio.

The CAPM is not a general equilibrium. It assumes that demand is equal to supply, but it makes no attempt to describe the supply functions of the assets. In addition, the distribution of returns is taken as given. There is no discussion about how those particular distributions arise. In fact, there is no proof that any equilibrium exists at all. If there is an equilibrium, then the market portfolio must be the tangency portfolio, but the usual stated assumptions are neither sufficient nor necessary to show that an equilibrium exists.

In particular, investors need not be risk averse for the CAPM to hold. If all investors have increasing utility, whether or not utility is concave, they will all wish to hold a portfolio with the highest possible mean at any given variance. The minimum-variance set is a bit of a red herring. It is true that risk averse investors do want portfolios from the minimum-variance set. But all investors with increasing utility want portfolios from the maximum-mean set, which is the upper limb of the minimum-variance hyperbola or ray, whether they are risk averse or not. This alone is sufficient to ensure that aggregate demand is a mean-variance efficient portfolio.

---

Figure 5.4: Minimum-Variance Portfolios with Different Borrowing and Lending Rates

This figure illustrates the minimum-variance set when investors borrow at a higher rate than that at which they lend. The minimum-variance set consists of the line segment from $r_L$ that is tangent at $A$, the hyperbola portion from $A$ to $B$, and the ray emanating from $r_B$ beyond $B$ that is tangent to the hyperbola.
Unfortunately, there is often no optimal portfolio for investors who are not risk averse. Indeed, investors with globally convex utility have infinite long and short demands. The general notion is that risk aversion will ensure that the optimal demands are finite. But a dislike of variance is not sufficient to guarantee this.

Suppose an investor’s mean-standard-deviation indifference curves are characterized by

$$\mu(v, \sigma) = v + \delta \sigma - \left(\gamma / \eta\right)(1 - e^{-\eta \sigma})$$

with all parameters positive. This equation gives the expected rate of return \(\mu\) that must be earned by a portfolio with risk \(\sigma\) to have utility \(v\). The slope of the indifference curves is \(\partial \mu / \partial \sigma = \delta - \gamma e^{-\eta \sigma}\). Provided \(\delta > \gamma\), this is positive for all \(\sigma\). Therefore, the investor always prefers lower variance holding \(\mu\) fixed. However, the slope of any indifference curve is never greater than \(\delta\) so if the Sharpe ratio of the market is higher than \(\delta\), the investor desires infinite leverage, and there is no equilibrium even though all investors dislike variance. Of course, this demand would presumably increase either the prices of the risky assets or the interest rate or both leading to a lower Sharpe ratio thereby allowing an equilibrium.

So a dislike of variance is neither necessary nor sufficient to guarantee the CAPM equilibrium. Additional assumptions are required to prove this, and exogenously specifying the means and variances is not satisfactory.

Of course, the CAPM is not the only model for which this is true. In many pricing models in finance, the equivalent of steps (iii) and (iv) above are simply assumed; it is asserted that there is a representative investor who holds the market portfolio and that his utility function is known. This assumption greatly simplifies the derivation of pricing results, but may be difficult to justify or even be demonstratively false. In the context of elliptical distributions, these last steps are trivial.

The flavor of these other models’ development can be demonstrated as follows. A representative investor who uses mean-variance analysis is assumed with no specific justification. As he is representative, he must hold the market. If the investments in assets \(i\) and \(j\) are changed by \(\omega_i\) and \(\omega_j\) financed by borrowing or lending, the mean and variance of the portfolio become

$$\begin{align*}
E[\tilde{r}_{m+\omega}] &= E[\tilde{r}_m] + \omega_i E[\tilde{r}_i - r_f] + \omega_j E[\tilde{r}_j - r_f] \\
\text{var}[\tilde{r}_{m+\omega}] &= \sigma_m^2 + 2\omega_i \sigma_m + 2\omega_j \sigma_{jm} + \omega_i^2 \sigma_i^2 + 2\omega_i \omega_j \sigma_{ij} + \omega_j^2 \sigma_j^2 .
\end{align*}$$

(28)

At the margin the changes are

$$\begin{align*}
dE[\tilde{r}_{m+\omega}]|_{\omega=0} &= E[\tilde{r}_i - r_f]d\omega_i + E[\tilde{r}_j - r_f]d\omega_j \\
d\text{var}[\tilde{r}_{m+\omega}]|_{\omega=0} &= 2\sigma_m d\omega_i + 2\sigma_{jm} d\omega_j .
\end{align*}$$

(29)

Choose the alteration, \(d\omega_j = (-\sigma_m / \sigma_{jm}) d\omega_i\) that leaves the variance of the portfolio unchanged. The change in the portfolio’s expected rate of return is

$$dE[\tilde{r}_m] = \left(E[\tilde{r}_i - r_f] - \frac{\sigma_m}{\sigma_{jm}} E[\tilde{r}_j - r_f]\right)d\omega_i .$$

(30)

The term in parentheses must be zero because otherwise choosing \(d\omega_i\) to have the same sign would create a portfolio with the same variance and a higher expected rate of return meaning the original portfolio could not have been mean-variance optimal as assumed. This means that the expected excess rate of return on every asset must be proportional to its covariance with the return on the market portfolio. As this is true for all assets it must be true for every portfolio so the expected excess rate of return on market must be in the same proportion to its covariance with itself

$$\frac{E[\tilde{r}_i] - r_f}{\sigma_m} = \frac{E[\tilde{r}_m] - r_f}{\sigma_m^2} \Rightarrow E[\tilde{r}_i] - r_f = \frac{\sigma_m}{\sigma_m^2} \left(E[\tilde{r}_m] - r_f\right) .$$

(31)

The same is true for the zero-beta version of the CAPM. This is demonstrated with a
slightly different proof to illustrate an alternate technique. The representative investor has the derived utility function, \( V(\mu, \sigma^2) \). Now consider a small increase, \( \omega \), in the holding of some asset \( i \) financed by shorting the zero-beta portfolio. This will alter the portfolio’s expected rate of return and variance to \( \mu_p = \mu_m + \omega(\mu_i - \mu_m) \) and \( \sigma_p^2 = \sigma_m^2 + 2\omega(\sigma_{mi} - \sigma_{mx}) + \omega^2(\sigma_i^2 - \sigma_{xi} + \sigma_{x}^2) \). As the market portfolio is assumed optimal, a small increase or decrease in the holding of any asset must leave expected utility unchanged. So

\[
0 = \frac{\partial V(\mu, \sigma^2)}{\partial \omega} \bigg|_{\omega=0} = V_1 \frac{\partial \mu}{\partial \omega} + V_2 \frac{\partial \sigma^2}{\partial \omega} \bigg|_{\omega=0} = V_1(\mu_i - \mu_x) + V_2 2\sigma_{mi}.
\]

The terms in the derivative of the variance that still contain \( \omega \) are zero when evaluated at \( \omega = 0 \), and \( \sigma_{mx} = 0 \) by definition. Therefore, the expected excess rate of return on any asset is proportional to its covariance with the market portfolio, \( \mu_i - \mu_x = \lambda \sigma_{mi} \) where \( \lambda = -V_1/(2V_2) \). As this is true for all assets, it must be true for all portfolios. So \( \lambda \) must be \((\mu_m - \mu_x)/\sigma_{m}^2 \). The CAPM relation (21) follows immediately.

**Individually Optimal Portfolios**

In the CAPM, every investor holds the market portfolio in some combination with borrowing or lending. The particular elliptical distribution does not alter the makeup of the tangency portfolio, which depends only on the mean vector and covariance matrix. It does, however, affect the borrowing lending decision. Typically, distributions with fatter tails will induce more lending. To illustrate, consider an investor with exponential utility, \(-e^{-aw}\), who faces a market with either a normal or Laplace distribution.

The expected utility for any portfolio can be computed for either the normal or Laplace distribution by using \( \theta = ia \) in the characteristic functions in (3)

\[
\mathbb{E}[-e^{-aw}] = \begin{cases} 
-\exp(-a\mu W_0 + \frac{1}{2}a^2\sigma^2 W_0^2) & \text{Normal} \\
-\exp(-a\mu W_0)(1 - \frac{1}{2}a^2\sigma^2 W_0^2)^{-1} & \text{Laplace}.
\end{cases}
\]

As the expectation has already been determined, an equivalent ordinal utility function can be derived with the monotonic transformation

\[
V(\mu, \sigma^2) \equiv \frac{-\ell n \left(-\mathbb{E}[-e^{-aw}]\right)}{a W_0} = \begin{cases} 
\mu - \frac{1}{2}(\text{RRA})\sigma^2 & \text{Normal} \\
\mu + \left(a W_0\right)^{-1} \ell n \left(1 - \frac{1}{2}(\text{RRA})^2 \sigma^2\right) & \text{Laplace}.
\end{cases}
\]

where \( a W_0 \) is the relative risk aversion.

The indifference curves for the normal distribution are parabolas \( \mu_{\infty}(V, \sigma) = V + \frac{1}{2}a W_0 \sigma^2 \). For the Laplace distribution, the indifference curves can determined using the Taylor expansion \( \ell n(1-x) = -\sum_{n=1}^{\infty} n^{-1} x^n \)

\[
\mu_{\infty}(V, \sigma) = V - (a W_0)^{-1} \ell n \left(1 - \frac{1}{2}(a W_0)^2 \sigma^2\right) = V + \frac{1}{2}a W_0 \sigma^2 + \frac{1}{2} \sum_{n=2}^{\infty} n^{-1} (a W_0)^n \sigma^{2n}.
\]

Clearly the indifference curves for the Laplace distribution lie above those for the normal distribution. This means the investor requires a higher expected rate of return to compensate for a given increase in risk. The slopes of Laplace indifference curves are also larger at every \( \sigma \) so the capital market line is tangent to one of them at a smaller standard deviation. That is, the investor facing assets with a Laplace distribution holds more of the risk-free asset and less of the risky assets than an investor facing assets with normal distributions with the same means and covariance matrix.
This can be verified directly. A portfolio with a fraction \( w \) in the market tangency portfolio and \( 1-w \) in lending has a mean and variance for wealth of 
\[ [1 + r_f + w(\mu_m - r_f)]W_0 \] 
and 
\[ w^2 \sigma_m^2 W_0^2 \]. So the first-order condition and optimal holding under the normal distribution are
\[
0 = \frac{\partial V}{\partial w} = (\mu_m - r_f)W_0 - aW_0^2 \sigma_m^2 w \quad \Rightarrow \quad w^*_N = \frac{\mu_m - r_f}{aW_0 \sigma_m^2}.
\] (36)

For the Laplace distribution, they are
\[
0 = \frac{\partial V}{\partial w} = (\mu_m - r_f)W_0 - \frac{aW_0^2 \sigma_m^2 w}{1 - \frac{1}{2} a^2 W_0^2 \sigma_m^2 w^2} \quad \Rightarrow \quad w^*_L = \frac{1}{aW_0} \sqrt{1 + 2(\mu_m - r_f)^2 / \sigma_m^2 - 1} \frac{\mu_m - r_f}{\mu_m - r_f}.
\] (37)

Under both distributions, the demand is inversely proportional to relative risk aversion. Using an exact two-term Taylor expansion for the square root function
\[
w^*_N = \frac{1}{aW_0} \left[ \frac{\mu_m - r_f}{\sigma_m^2} - \xi \left( \frac{\mu_m - r_f}{2 \sigma_m^2} \right)^3 \right]
\] (38)

for some \( \xi \in (0, 1) \). The first term is the demand under a normal distribution and the second term is negative assuming a positive risk premium on the market so the demand under the Laplace distribution is indeed smaller than the demand under the normal distribution.

**State Prices and the SDF in the CAPM**

What state prices are consistent with the CAPM? For a state-space probability density \( \pi(s) \), the expected excess rate of return and covariance with the market for asset \( i \) are
\[
\mu_i - r_f = \int \pi(s)[r_i(s) - r_f] ds
\]
\[
cov[\tilde{r}_i, \tilde{r}_m] = \cov[\tilde{r}_i - r_f, \tilde{r}_m] = \int \pi(s)[r_i(s) - r_f][r_m(s) - \mu_m] ds.
\] (39)

From the pricing relation we know these are proportional so 
\[ \mu_i - r_f = \lambda \cov[\tilde{r}_i, \tilde{r}_m], \] where \( \lambda \) is the price of risk, 
\[ \lambda \equiv (\mu_m - r_f) / \sigma_m. \] So
\[
0 = \mu_i - r_f - \lambda \cov[\tilde{r}_i, \tilde{r}_m] = \int \pi(s)[r_i(s) - r_f] \left( 1 - \lambda [r_m(s) - \mu_m] \right) ds.
\] (40)

We also know that the state price density prices all rates of return at zero
\[
0 = \int q(s)[r_i(s) - r_f] ds.
\] (41)

Together (40) and (41) would seem to imply that the state-s price is 
\[ q(s) = \pi(s)(1 - \lambda [r_m(s) - \mu_m]). \]

The existence of any set of state prices proves that the Law of One Price holds and that there is no risk-free arbitrage. However, it does not show that there is no arbitrage as some of these state prices could be negative. Indeed, for a normal distribution (or any other unbounded elliptical distribution), the market return can be arbitrarily large, so some of the state prices suggested by (41) must be negative. Therefore, we cannot conclude there is no arbitrage. On the other hand, there cannot arbitrage in a market where all risky-assets have unbounded distributions because then no portfolio (except the risk-free asset) can be constructed that has only nonnegative outcomes.9

---

9 Arbitrage is possible for elliptical distributions with bounded domains. For example, there could be two assets whose outcomes do not overlap. Shorting the lower to purchase the higher is an arbitrage, though it is not risk-free.
Therefore, there must be a set of strictly positive state prices. Obviously Arrow-Debreu securities do not have elliptically distributed returns so the market in a CAPM model cannot be complete if all assets have elliptical returns. In an incomplete market, state prices are not unique so there must be some other set which is strictly positive.

The first-order condition for an optimal portfolio is

\[
\mathbb{E}[u'(\tilde{W}^*)(\tilde{r} - r_f)] = 0,
\]

where \( \tilde{W}^* \) random wealth generated by the portfolio optimal for the given utility function. Under the CAPM, this portfolio must be the market portfolio levered. Choose exponential utility with that level of risk aversion for which the investor optimally holds the market with no leverage. Then

\[
0 = \mathbb{E}[\exp(-a\tilde{W}^*)(\tilde{r} - r_f)] = \int \pi(s) \exp\left(-aW_0[1+r_m(s)](r_f(s) - r_f)\right)
\]

\[
\Rightarrow q(s) = \pi(s) \exp\left(-aW_0[1+r_m(s)]\right) > 0.
\]

We know from (36) that the investor holds the market unlevered if \( aW_0 = (\mu_m - r_f)/\sigma_m^2 \) so the state-s state price can be expressed as \( q(s) = \pi(s) \exp\left[-(\mu_m - r_f)[1+r_m(s)]/\sigma_m^2\right] \). These state prices are positive verifying that there is no arbitrage. The SDF is the state price per unit probability so

\[
\tilde{m} = \exp\left[-(\mu_m - r_f)(1+r_m)^2/\sigma_m^2\right].
\]

If the market’s distribution is not normal, then (43) is still valid; however, the investor who holds the market changes. For a Laplace distribution, the representative investor has a relative risk aversion of

\[
aW_0 = \sqrt{1 + 2(\mu_m - r_f)^2/\sigma_m^2} - 1 \]

so that expression for \( aW_0 \) replaces \( (\mu_m - r_f)/\sigma_m^2 \) in (44).

As with any SDF, a tradable SDF can be constructed by projection into the space of returns. Comparing the CAPM pricing relation to the portfolio version of SDF pricing in equation (36) in Chapter 3,

\[
\text{Portfolio SDF: } \mathbb{E}[\tilde{r}] - r_f 1 = \frac{\text{cov} (\tilde{r}_m, \tilde{r})}{\text{var}(\tilde{r}_m)} \mathbb{E}[\tilde{r}_m - r_f] \quad \text{CAPM: } \mathbb{E}[\tilde{r}] - r_f 1 = \frac{\text{cov} (\tilde{r}_m, \tilde{r})}{\text{var}(\tilde{r}_m)} \mathbb{E}[\tilde{r}_m - r_f]
\]

it might seem obvious that the market portfolio is the SDF portfolio, but this is not the case. Recall that the random realization of the SDF is the state price per unit probability which is positively proportional to marginal utility. But marginal utility is decreasing in final wealth while the market portfolio’s return is obviously increasing. So the market portfolio cannot be the SDF portfolio as they are negatively related.

In fact, the CAPM equilibrium has virtually no relation to the SDF portfolio. The existence of the latter depends only on the absence of risk-free arbitrage. The “CAPM” relation in (46) holds for the tangency portfolio and all other minimum-variance portfolios whether or not the conditions for a CAPM equilibrium are valid. This means the portfolio SDF like the tangency portfolio must lie on the borrowing-lending line. We will see shortly that it always lies on the extension of the borrowing-lending line below the risk-asset only hyperbola.

To determine the exact position of the SDF portfolio, write the gross return on any minimum-variance portfolio as combination of the gross returns on the SDF portfolio and the risk-free asset,

\[
1 + \tilde{r}_m = \alpha(1 + r_f) + (1 - \alpha)(1 + \tilde{r}_m) = 1 + \tilde{r}_m - \alpha(\tilde{r}_m - r_f).
\]

This is possible because the SDF portfolio is a minimum-variance portfolio, as shown, and all minimum-variance portfolios are
Mean-Variance Analysis and the CAPM — © Jonathan Ingersoll 15 version: September 11, 2019

The first equation of (47) is a simple expansion. The second equality applies only for SDFs. The second term is zero because any SDF prices all excess returns at zero. The right-hand side of this second equality obviously has a minimum value when \( \alpha = 0 \); therefore, the SDF portfolio is the minimum variance portfolio with the smallest mean squared gross return. This verifies that the SDF portfolio is unique as previously shown in Chapter 3.

This derivation is illustrated in the Figure. The mean squared return of any portfolio its variance plus the square of its expected return, \( \mathbb{E}[(1 + r_m)^2] = \sigma^2 + (1 + \mu)^2 \). The right-hand side of this equation is the formula of a circle in mean-standard-deviation space with a center at \( \sigma = 0 \) and \( \mu = -100\% \). So the SDF portfolio lies at the point where a circle centered there is tangent to the extended borrowing lending line.

To determine the mean and variance of the SDF portfolio, we can minimize the mean squared gross return of combinations of the risk-free asset and the tangency portfolio. The return on such a combinations is \( \tilde{r} = r_f + \theta(r_t - r_f) \) so \( \sigma^2 = \theta^2 \sigma_t^2 \) and \( \mu = r_f + \theta(\mu_t - r_f) \). The first order condition of the minimization and its solution are

\[
0 = \frac{\partial}{\partial \theta} \left[ (1 + r_f + \theta(\mu_t - r_f))^2 + \theta^2 \sigma_t^2 \right] = 2[1 + r_f + \theta(\mu_t - r_f)](\mu_t - r_f) + 2\theta \sigma_t^2
\]

\[
\Rightarrow \theta^* = -\frac{(\mu_t - r_f)(1 + r_f)}{(\mu_t - r_f)^2 + \sigma_t^2}.
\]

As shown graphically, the SDF portfolio holds a short position in the tangency portfolio (assuming \( \mu_t > r_f \) so that the tangency portfolio is on the upper limb). Its expected rate of return and standard deviation are

---

Figure 5.4: The Mean-Variance SDF

This figure illustrates the location of the asset based SDF. It is on the extension of the capital market line below the risk-free rate.
where $S_t$ is the Sharpe ratio of the tangency portfolio. The Sharpe ratio of the SDF portfolio is the negative of the Sharpe ratio of the tangency portfolio (again assuming $\mu > r_f$). This further demonstrates it is on the lower extension of the borrowing lending line.

A derivation similar to that in (48) and (49) can be used to determine the expected rate of return and variance of the SDF portfolio in the absence of a risk-free asset. We can adapt equation (21) in Chapter 3, to give the holdings in the SDF portfolio:

$$\begin{align*}
\eta &= \Sigma^{-1}1 - (1 + \mu' \Sigma^{-1}1)(\mu' \Sigma^{-1}1)^{-1} \Sigma^{-1}1 = \Sigma^{-1}1 - (1 + B)^{-1} A \Sigma^{-1}1.
\end{align*}$$

Without loss of generality, the assets’ share sizes have been defined so that all asset prices are 1; that is $p \to 1$. This means that the expected payoffs on each asset are the same as the expected returns per dollar so $p$ as used in Chapter 2 has the same meaning as used here. It is clear from (50) that the SDF portfolio is a combination of the global minimum variance portfolio and portfolio $\tau$. The SDF portfolio must still have the minimum mean squared return so it is at the tangent point of a circle centered at $\mu = -100\%$, $\sigma = 0$ and the lower limb of the minimum-variance hyperbola as shown in the second figure. This can be verified by the same analytical reasoning using the global minimum variance portfolio in place of the risk-free asset. All minimum-variance portfolios are combinations of any two, $\tilde{r}_m = \tilde{r}_m + \alpha(\tilde{r}_g - \tilde{r}_m)$. So

$$\begin{align*}
\mathbb{E}[(1+r_m)^2] &= \mathbb{E}[(1+r_m)^2] + 2\alpha \mathbb{E}[(1+r_m)(\tilde{r}_g - \tilde{r}_m)] + \alpha^2 \mathbb{E}[(\tilde{r}_g - \tilde{r}_m)^2] \\
&= \mathbb{E}[(1+r_m)^2] + \alpha^2 \mathbb{E}[(\tilde{r}_g - \tilde{r}_m)^2].
\end{align*}$$

Once again the cross-expectation term is zero because a SDF prices any difference in returns at zero, and $\alpha = 0$ corresponds to both the SDF portfolio and the portfolio that has a minimum value of $\mathbb{E}[(1+r)^2]$. The SDF portfolio is at the tangency of the minimum variance hyperbola and a circle centered at $\mu = -100\%$, $\sigma = 0$. This result is shown in the figure.

The SDF portfolio must be a combinations of portfolios $\tau$ and $g$ as is any minimum variance portfolio, $\tilde{r}_m = \tilde{r}_g + \theta (\tilde{r}_g - \tilde{r}_m)$. The minimum mean squared return is
\[
\min_0 [1 + \mu_g + \theta(\mu_r - \mu_g)]^2 + \theta^2 \sigma_r^2 + 2\theta(1 - \theta)\sigma_{rg} + (1 - \theta)^2 \sigma_g^2 \\
= \min_0 [1 + \mu_g + \theta(\mu_r - \mu_g)]^2 + \theta^2 \sigma_r^2 + (1 - \theta^2)\sigma_g^2. \tag{52}
\]

The simplification on the right-hand side follows because \(\sigma_{rg} = \sigma_g^2\) as shown in footnote 7. The first-order condition and solution are
\[
0 = 2[1 + \mu_g + \theta(\mu_r - \mu_g)](\mu_r - \mu_g) + 2\theta(\sigma_r^2 - \sigma_g^2) \\
\Rightarrow \theta = -\frac{(\mu_r - \mu_g)(1 + \mu_g)}{(\mu_r - \mu_g)^2 + \sigma_r^2 - \sigma_g^2}. \tag{53}
\]

The expected rate of return and standard deviation of the SDF portfolio are
\[
\mu_m - \mu_g = \theta(\mu_r - \mu_g) = -\frac{(\mu_r - \mu_g)(1 + \mu_g)}{(\mu_r - \mu_g)^2 + \sigma_r^2 - \sigma_g^2} \\
\sigma_m^2 = \theta^2 \sigma_r^2 + (1 - \theta^2)\sigma_g^2 = \sigma_g^2 + \frac{(\mu_r - \mu_g)^2(1 + \mu_g)^2(\sigma_r^2 - \sigma_g^2)}{[(\mu_r - \mu_g)^2 + \sigma_r^2 - \sigma_g^2]^2}. \tag{54}
\]

Of course with no risk-free asset, Sharpe ratios are not defined, but if we define a Sharpe-like ratio as \(S'_m = (\mu_m - \mu_g)/(\sigma_m^2 - \sigma_g^2)^{1/2}\), then a similar result does obtain
\[
S'_m = \frac{\mu_m - \mu_g}{\sqrt{\sigma_m^2 - \sigma_g^2}} = -\frac{\mu_r - \mu_g}{\sqrt{\sigma_r^2 - \sigma_g^2}} = -S'_r. \tag{55}
\]

If there is a risk-free asset, then it is the global minimum variance portfolio with a variance of zero and an expected rate of return of \(r_f\). So this result includes \(S_m = -S_r\) as a special case.

As mentioned, all of this analysis and discussion remains true even when mean-variance analysis is not consistent with EUT so that mean-variance efficient portfolios are not necessarily optimal. The minimum-variance portfolios are described in terms of means and covariances, and their relation to expected rates of return is a tautology which has nothing to do with the CAPM equilibrium involving the market. All that is required is no risk-free arbitrage opportunities.

**The Roll and Other Critiques**

Richard Roll has criticized the CAPM model as essentially untestable. His criticism consists of two parts. First the mean-variance pricing relation in (20) and (22) are tautologies. Any mean-variance efficient portfolio can be used to price all the assets. Furthermore, in any sample of data sufficiently large to generate a non-singular covariance matrix, there is always a tangency portfolio (or spanning pairs of minimum-variance portfolios if there is no risk-free asset). The tangency portfolio prices all the assets exactly in sample; the pricing is tautological and requires no further model assumptions. So the only empirically testable statement of the CAPM is that the market portfolio is mean-variance efficient.

His second point is that the market portfolio is unobservable. The market portfolio should include all available assets and not just stocks or other financial assets. It should include real estate, human capital, precious metals, collectibles, and anything that has value and can be saved. The market portfolio of stocks might be close to the complete market portfolio, but that doesn’t necessarily imply that it prices assets with little error. Indeed, it is not even obvious how we might measure how close two portfolios are. One possibility is R-squared, another possibility is the difference in their Sharpe ratios. Still another is the sum of squared differences in their portfolio
weights. It seems reasonable that a portfolio close to the tangency portfolio in one of these measures should be close in the other two and could price assets with little error, but whether or not it does depends on the covariance matrix.

This criticism can be countered to some extent by arguing that the model applies not to all assets but to the subset of assets actually included in the test. This can be described as a narrow framing of the portfolio problem. Narrow framing is applied in many behavioral models, though it is not consistent with expected utility maximization as originally formulated.

Another criticism is that the CAPM is easily rejected as it predicts not only what expected returns should be, but also that every investor holds the market portfolio. This latter prediction is clearly false. It could and should be argued that we don’t expect models to be an exact description of reality. As long as most investors hold reasonably diversified portfolios, deviations will average out and the market portfolio should price assets approximately. However, we do not know how close the approximation is. We would need to know how well portfolios that are close to mean-variance efficient price assets. This is a topic addressed in Chapter 10.

Another problem with the testability of the CAPM is that both expected rates of return and betas are unobservable; they need to be estimated. Of course, this is true of many model predictions; however, there is a particular difficulty for the CAPM. Usually expected rates of return are estimated by sample averages and betas are estimated by regression coefficients. In the simplest procedures, the returns are assumed to be independent and identically distributed over time. However, if the expected rates of return and covariance matrix are constant, then the tangency portfolio has a constant weighting, but this is certainly not true of the market portfolio. In fact, it could be true only if dividends and net share repurchases exactly offset the realized differences in returns.

Jensen’s Alpha

Despite Roll’s criticism, tests of or based on the CAPM are often conducted using some proxy for the market portfolio to determine betas. The tests are typically in cross-sectional regressions of the form

\[
\text{avg} \{ \tilde{r}_i - r_f \} = \alpha + \gamma \beta_i + \sum \delta_j x_{ji} + \varepsilon_i. \tag{56}
\]

The extra independent regressors, \(x_{ji}\), are variables thought to possibly explain expected returns. Examples include variance, residual variance, skewness, co-skewness, the Fama-French Factors, momentum, and industry factors. To test for nonlinearity, \(\hat{\beta}_i\) or separate \(\beta\)’s base on increases and decreases in the market can be used.

Because the independent variable, \(\beta_i\), as well as many of the \(x_{ji}\)’s must be estimated rather than measured, these tests must be conducted with care. A standard method is the Fama-Macbeth (1973) two-step regression which is based on the assumption that there is zero autocorrelation in returns. If the CAPM is correct, the test should show that \(\gamma = \text{avg} \{ \tilde{r}_m - r_f \}\) while \(\alpha\) and all the \(\delta\)’s are zero. In these tests, the intercepts are known as Jensen’s alphas. When other factors are used and the \(\delta\)’s are allowed to be non-zero, the intercepts are referred to with names like three (or more) factor alphas.

Conditional CAPM

The CAPM, as developed here, is a single period model. However, it is often used as the basis for an intertemporal model applying it period-by-period. When the CAPM is tested, it is virtually always treated that way implicitly. Betas and expected rates of return are estimated from a time series. Typically, the return distribution is assumed to be independently and identically distributed over time or that there are some simple time series properties. But if these values are changing over time, then the single-period model cannot be applied directly to the data even if the
CAPM is valid period-by-period. Suppose the CAPM is valid each period based on the information available to market participants at the start of the period, $\Phi_t$, then

$$
\mathbb{E}[(\tilde{r}_i - r_f) | \Phi_t] = \frac{\text{cov}[(\tilde{r}_i - r_f), (\tilde{r}_m - r_f) | \Phi_t]}{\text{var}[(\tilde{r}_m - r_f) | \Phi_t]} \mathbb{E}[(\tilde{r}_m - r_f) | \Phi_t] \equiv \beta_i \times \gamma_t
$$

(57)

where $(\tilde{r}_i - r_f)$ is the return in excess of the interest rate at time $t$ measured forward from time $t$, and $\gamma_t$ is the forward-looking market risk premium at time $t$.\(^{10}\) This relation is the conditional CAPM. If the returns are independent and identically distributed over time, then the betas and $\gamma_t$ are constant and the unconditional or static CAPM is valid as well. Otherwise, the unconditional expectation of (57) is

$$
\mathbb{E}[(\tilde{r}_i - r_f) = \mathbb{E}\{\tilde{\beta}_i \tilde{\gamma}_t\} = \mathbb{E}\{\tilde{\beta}_i\} \times \mathbb{E}[\tilde{\gamma}_t] + \text{cov}(\tilde{\beta}_i, \tilde{\gamma}_t) \quad \text{“error” term}
$$

(58)

where all the expectations are unconditional. One problem is immediately obvious. If the betas and the market’s risk premium change over time in a correlated fashion, there is an error in the static CAPM prediction. This error is in addition to any sampling error in the estimation of betas and the market’s risk premium.

To illustrate consider the following example. There are two states. In the first state, the market risk premium is 10% and the betas of two stocks are $\beta_1 = 0.5$, $\beta_2 = 1.5$. In the second state, the market risk premium is 20%, and the betas of two stocks are $\beta_1 = 1.5$, $\beta_2 = 0.5$. Each stock has an average beta of 1, and the market has an average risk premium of 15%, so a direct application of the static CAPM would say that each stock should also have a risk premium of 15%. But if the conditional CAPM holds, the risk premiums on the stock 1 in the two states are $0.5 \cdot 10\% = 5\%$ and $1.5 \cdot 20\% = 30\%$ for an average of 17.5%. The two risk premiums on the stock 2 are $1.5 \cdot 10\% = 15\%$ and $0.5 \cdot 20\% = 10\%$ for an average of 12.5%. The static version of the CAPM is rejected, even though the conditional CAPM in (57) holds. The “error” term in (58) explains the difference.

$$
\text{cov}[\tilde{\beta}_i, \tilde{\gamma}_t] = 0.025, \quad \text{cov}[\tilde{\beta}_2, \tilde{\gamma}_t] = -0.025
$$

$$
\Rightarrow \quad \mathbb{E}[\tilde{r}_i - r_f] = 0.15 + 0.025 = 17.5\% \quad \mathbb{E}[\tilde{r}_2 - r_f] = 0.15 - 0.025 = 12.5\%.
$$

(59)

In general (58) can be re-expressed as

$$
\mathbb{E}[(\tilde{r}_i - r_f) = \mathbb{E}\{\tilde{\beta}_i\} \times \mathbb{E}[\tilde{\gamma}_t] + \frac{\text{cov}(\tilde{\beta}_i, \tilde{\gamma}_t)}{\text{var}[\tilde{\gamma}_t]} \mathbb{E}[\tilde{\gamma}_t].
$$

(60)

The fraction in the second term is also a regression coefficient, namely that of the beta regressed on the market's risk premium. So the conditional CAPM has the appearance of a two factor model.

Another problem that arises in the static CAPM is that even with no time series covariation between $\beta_i$ and $\gamma_t$, standard econometric techniques do not necessarily produce an estimate for $\beta_i$ that is equal to its expectation $\tilde{\beta}_i \equiv \mathbb{E}[\tilde{\beta}_i]$. This can be illustrated with a simple example. Let $\tilde{x}_t \equiv \tilde{r}_t - r_f$ be the excess return on asset $i$ over the period $t$ to $t+1$, then

$$
\tilde{x}_m = \tilde{\gamma}_t + \tilde{\epsilon}_m = \tilde{\gamma} + \tilde{\epsilon}_t + \tilde{\epsilon}_m
$$

where $\tilde{\gamma} \equiv \mathbb{E}[\tilde{\gamma}]$ and $\tilde{\epsilon}_t \equiv \tilde{\gamma}_t - \tilde{\gamma}

(61)

\(^{10}\) Doing the analysis in terms of the excess rates of return avoids dealing with the stochastic properties of the interest rate. Only the properties of the excess return are relevant. The notation $(\tilde{r}_t - r_f)$ serves as a reminder of this. The same is true for the zero-beta version of the CAPM, though that requires knowing the zero-beta rate; estimating it can introduce noise and possible bias.
where $\overline{\gamma} \equiv \mathbb{E}[\gamma]$ is the average risk premium in the time series, and $\tilde{\gamma}_t \equiv \gamma_t - \overline{\gamma}$ is the variation in the risk premium around its long-term average. With no loss of generality, the realized excess return on any asset is

$$\tilde{x}_t = \alpha + \beta x_{m,t} + \tilde{\epsilon}_t = \alpha + (\tilde{\beta} + \tilde{\epsilon}_t) \tilde{x}_{m,t} + \tilde{\epsilon}_t \quad \text{with} \quad \text{cov}[\tilde{\epsilon}_t, \tilde{x}_{m,t}] = 0 \quad (62)$$

and $\tilde{\epsilon}_{m,t}, \tilde{\epsilon}_t, \tilde{\epsilon}_t$, and $\tilde{\epsilon}_t$ mutually uncorrelated in the cross section and over time.\(^\text{11}\)

If the conditional CAPM holds each period, then $\alpha \equiv 0$. The standard time-series estimated beta has an expected value of

$$\mathbb{E}[\hat{\beta}] = \frac{\text{cov}[\tilde{x}_t, \tilde{x}_{m,t}]}{\text{var}[\tilde{x}_{m,t}]} = \frac{\text{cov}[(\tilde{\beta} + \tilde{\epsilon}_t) \tilde{x}_{m,t} + \tilde{\epsilon}_t, \tilde{x}_{m,t}]}{\text{var}[\tilde{x}_{m,t}]} \frac{\mathbb{E}[(\tilde{\beta} + \tilde{\epsilon}_t) \tilde{x}_{m,t}^2] - \mathbb{E}[(\tilde{\beta} + \tilde{\epsilon}_t) \tilde{x}_{m,t}] \mathbb{E}[\tilde{x}_{m,t}]}{\text{var}[\tilde{x}_{m,t}]}$$

$$= \frac{\tilde{\beta} \mathbb{E}[\tilde{x}_{m,t}^2] + \mathbb{E}[\tilde{\epsilon}_t \tilde{x}_{m,t}^2] - \tilde{\beta} \mathbb{E}[\tilde{x}_{m,t}]^2 - \mathbb{E}[\tilde{\epsilon}_t \tilde{x}_{m,t}] \mathbb{E}[\tilde{x}_{m,t}]}{\text{var}[\tilde{x}_{m,t}]}$$

$$= \frac{\tilde{\beta} \mathbb{E}[\tilde{x}_{m,t}^2] + \mathbb{E}[\tilde{\epsilon}_t \tilde{x}_{m,t}^2]}{\text{var}[\tilde{x}_{m,t}]} = \tilde{\beta} + \frac{\mathbb{E}[\tilde{\epsilon}_t \tilde{x}_{m,t}^2]}{\text{var}[\tilde{x}_{m,t}]} \quad (63)$$

The final term is zero because $\tilde{\epsilon}_t$ is uncorrelated with $\tilde{\epsilon}_t$ and $\tilde{\epsilon}_{m,t}$ which make up $\tilde{x}_{m,t}$. The first and third term sum to $\tilde{\beta}$. So the time-series beta estimate is unbiased only when it is uncorrelated with the squared market return. If $\beta_t$ tends to be higher when the market is more variable, then the estimated will be biased high and vice versa.

\(^\text{11}\)The idiosyncratic risks, $\tilde{\epsilon}_t$ and $\tilde{\epsilon}_{m,t}$ must be mutually uncorrelated by construction and be uncorrelated with past realizations of $\tilde{\epsilon}_t$ and $\tilde{\epsilon}_{m,t}$ because, by definition, they are not forecastable as of time $t$. The random components of betas and risk premium, $\tilde{\epsilon}_t$ and $\tilde{\epsilon}_{m,t}$, must be uncorrelated to eliminate the error term in (58) though they could be correlated with any past realizations. This would further affect the estimation of the betas.


Arthur Bowley (1937) *Wages and Income in the United Kingdom since 1860*.

Alternate Graphical Representation of the Mean-Variance Problem

The mean-variance problem in this chapter was illustrated in mean-standard-deviation space. That is the usual representation. Some texts prefer to illustrate the problem in mean-variance space. The difference can cause confusion if the axes are not clearly labeled.

![Figure 5.6: The Zero-Beta Portfolio](image)

This figure illustrates the location of the zero-beta portfolio in mean-variance space.

In mean-standard-deviation space a line from through any portfolio and the risk-free asset traces out the possible mean and standard deviations that can be created by combining that portfolio with borrowing or lending. In mean-variance space, the borrowing lending “line” is a parabola opening to the right with its vertex on the mean axis at the risk-free rate. So the line tangent to the minimum-variance hyperbola intersects the mean axis at the expected rate of return of zero beta portfolio for the portfolio at the tangency spot. This is true for any minimum variance portfolio and not just the tangency portfolio and is true regardless of the existence of a risk-free asset.

In mean-variance space, the expected rate of return on the zero beta portfolio for any minimum variance portfolio is found at the intercept of a line through that portfolio and the global minimum variance portfolio. That is, the slopes of the line segments are equal

\[
\frac{\mu_{mv} - \mu_g}{\sigma_{mv}^2 - \sigma_g^2} = \frac{\mu_g - \mu_{z(mv)}}{\sigma_g^2}.
\]

(64)

To verify this claim, we need to determine the expected rate of return on a portfolio uncorrelated with a parabola portfolio. From equation (9), the covariance between any portfolio, \( \bf{w} \), and a minimum-variance portfolio, \( \bf{w}_{mv} \), is

\[
\text{cov}(\tilde{r}_w, \tilde{r}_{mv}) = \bf{w} \Sigma \bf{w}_{mv} = \bf{w} \Sigma [\eta(\mu_w) \Sigma^{-1} \mu + \gamma(\mu_w) \Sigma^{-1} 1] = \eta(\mu_{mv}) \mu_w + \gamma(\mu_{mv})
\]

where

\[
\eta(\mu_{mv}) \equiv \frac{C \mu_{mv} - A}{D} \quad \gamma(\mu_{mv}) \equiv \frac{B - A \mu_{mv}}{D}
\]

(65)

\[
A \equiv 1' \Sigma^{-1} \mu \quad B \equiv \mu' \Sigma^{-1} \mu > 0 \quad C \equiv 1' \Sigma^{-1} 1 > 0 \quad D \equiv BC - A^2 > 0.
\]

This can be simplified as follows
\[ \text{cov} (\tilde{r}_w, \tilde{r}_{\text{mv}}) = D^{-1} [(C \mu_{\text{mv}} - A)\mu_w + B - A \mu_{\text{mv}}] = D^{-1} [C \mu_{\text{mv}} \mu_w - A(\mu_w + \mu_{\text{mv}}) + B] \\
= D^{-1} [C \mu_{\text{mv}} \mu_w - A(\mu_w + \mu_{\text{mv}}) + A^2/C + B - A^2/C] \\
= D^{-1} [C \mu_{\text{mv}} \mu_w - A(\mu_w + \mu_{\text{mv}})/C + A^2/C^2] + (AC - A^2)/CD \\
= D^{-1} C(\mu_{\text{mv}} - \mu_g )(\mu_w - \mu_g) + 1/C. \]  

(66)

The second equality adds and subtracts \( A^2/C \) to complete the square. The last equality uses the definition \( D \equiv BC - A^2 \) and the previously derived relation \( \mu_g = A/C \). This holds for any portfolio \( w \) so the expected rate of return a portfolio uncorrelated any parabola portfolio (except the global minimum variance portfolio) satisfies \( (\mu_{\text{mv}} - \mu_g ) (\mu_{\text{z(mv)}} - \mu_g ) = -D/C^2 \). Solve this for \( \mu_g - \mu_{\text{z(mv)}} \) then substitute along with \( \sigma_g^2 = 1/C \) as previously determined into the right hand side of (64) giving

\[ \frac{\mu_g - \mu_{\text{z(mv)}}}{\sigma_g^2} = \frac{D/[C^2 (\mu_{\text{mv}} - \mu_g )]}{1/C} = \frac{D}{C(\mu_{\text{mv}} - \mu_g )} = \frac{D}{C\mu_{\text{mv}} - A}. \]  

(67)

For the left-hand side of (64), substitute \( \sigma_g^2 = D^{-1} (C \mu_{\text{mv}}^2 - 2A \mu_{\text{mv}} + B) \) from (10), \( \mu_g = A/C \), and \( \sigma_g^2 = 1/C \) as before

\[ \frac{\mu_{\text{mv}} - \mu_g}{\sigma_{\text{mv}}^2 - \sigma_g^2} = \frac{D C(\mu_{\text{mv}} - A/C)}{C(C \mu_{\text{mv}}^2 - 2A \mu_{\text{mv}} + B) - D} = \frac{D(\mu_{\text{mv}}^2 - A)}{C(C \mu_{\text{mv}}^2 - 2A \mu_{\text{mv}} + B) - BC + A^2}. \]  

(68)

So (64) is confirmed.

**More on the Zero-Beta CAPM**

There are a couple of features of the zero-beta CAPM that are seldom noted. The first involves the budget constraint. If there is free-disposal the budget constraint should properly be applied as an inequality rather than an equality. The reason is that free disposal is a risk-free asset with an expected rate of return of \(-100\% \) that can be used only for “lending”. Provided the expected rate of return on the global minimum variance portfolio exceeds \(-100\%, \) lending even at an interest rate of \(-100\% \) is a better mean-variance choice than holding some portfolios close to the

**Figure 5.7: The Mean-Variance SDF**

This figure illustrates the mean-variance efficient set when there is no risk-free asset, but there is free disposal.
global minimum variance portfolio. This is illustrated in the figure.

Returns are normally distributed, and the minimum variance hyperbola is characterized by 
\( \mu_T = 0.15, \sigma_T^2 = 0.04, \mu_g = -0.5, \sigma_g^2 = \sigma_{gt} = 0.01 \). The optimal portfolio for an exponential utility investor with a relative risk aversion of \( aW_0 = 100 \) holds \( w_T = 0.217 \) and \( w_g = 0.783 \) achieving an expected rate of return of \(-0.36\), a standard deviation of 0.107, and an expected utility of 0.251.\(^{12}\) If discarding wealth is allowed, the investor discards 28.6% of his wealth and holds the remainder in a portfolio with \( w_T = 0.433 \) and \( w_g = 0.567 \) achieving an expected rate of return of \(-0.41\) and a standard deviation of 0.089. Because this investor is very risk averse, his indifference curves are quite steep, and the decrease in the variance more than offsets the decrease in the expected rate of return. Expected utility increases to 0.279.

In order for such a portfolio to be optimal three conditions must be met: (i) there must be no advantageous trade-off with time-0 consumption. (ii) There must be no limited liability assets, (iii) the investor must be sufficient risk averse.

If the investor can increase time-0 consumption, is not satiated, and increasing time-0 consumption does not decrease the utility of later consumption,\(^{13}\) then consuming more at time 0 always dominates the disposal of wealth. If there is an asset whose payoffs are never negative, then buying such an asset always increases consumption at time 1 and its contribution to expected utility. Again this dominates disposal. Finally, as illustrated in the figure, the investors risk aversion must be large enough to have steep indifference curves.

A second peculiarity of the zero-beta CAPM is that minimum-variance portfolios on the lower limb of the hyperbola can sometimes be optimal, though by definition they are not mean-variance efficient. It should be no surprise that mean-variance efficiency does not coincide with optimality. What is surprising is that this can be true even when mean-variance analysis fully describes the portfolio problem. A simple example is given here. This topic is addressed in more detail in the next chapter.

There are two assets whose returns are independently distributed each with two equally likely outcomes. Asset one has a rate of return of either 0.6 or \(-0.2\). Asset two’s rates of return are 0.3 and \(-0.1\). The expected rates of return are 0.2 and 0.1, and the variances are 0.16 and 0.04. The global minimum-variance portfolio holds \( w = \sigma_2^2/(\sigma_1^2 + \sigma_2^2) = 0.2 \) in asset one. It has four equally likely outcomes, 0.36, 0.2, 0.04, \(-0.12\) with an expected rate of return of 0.12 and variance of 0.032. Because there are only two assets, every possible portfolio is fully characterized by its mean and variance; in fact, every possible portfolio is characterized by its mean alone. Nevertheless, some portfolios below the global minimum variance portfolio are optimal.

The minimum return on the global minimum-variance portfolio is less than the minimum return on asset two. In fact, the portfolio invested 100% in asset two has the highest possible minimum return. So an infinitely risk averse investor would invest entirely in asset two, and an extremely risk averse investor would hold a portfolio with nearly 100% in asset two. But the expected rate of return of asset two is less than the expected rate of return on the global minimum variance portfolio so it is on the lower limb of the hyperbola. Exponential utility investors with \( aW_0 > 11.484 \) hold portfolios optimally choose lower limb portfolios.

\(^{12}\) Both expected utilities are given for the transformed utility \( \mu - \frac{1}{2} aW_0 \sigma^2 \).

\(^{13}\) Time-0 consumption can decrease the utility of consumption in later periods in some models with habits. These models are discussed in Chapter __.
Stable Distributions and the Mandelbrot Hypothesis — an Historical Aside

The importance of the stable distributions\(^{14}\) in the history of Finance was two-fold. First they have fat tails, and the evidence found by the early 1970s was making it increasingly obvious that the changes in the prices of stocks and other assets had too many outliers compared to normal distributions. Second, the sum of two stable variables of the same type was also a stable variable, so stable variables were amenable to portfolio analysis just like elliptical variables.

In fact, this latter property essentially defines stable variables. A simple definition originally due to Khinchin is: If \(x\) and \(y\) are independent random variables with the same probability distribution, then the distribution is a stable distribution if and only if for all positive \(a\) and \(b\),\(^{15}\) \(ax + by\) has the same distribution as \(kx + d\) for some \(k > 0\) and some \(d\). In other words, two independent random variables are from a stable distribution if positively weighted sums of those variables have the same distribution after scaling and relocation. Of course, independence across asset returns wasn’t desirable, but correlation could be introduced with a factor structure.

Khinchin and Lévy (1936) showed that all symmetric stable distributions have characteristic functions of the form\(^{16}\)

\[
\psi(\theta) = \mathbb{E}[e^{i\theta X}] = \exp(i\mu\theta - c^{\alpha}\theta^{\alpha})
\]

The resemblance of the characteristic function to the elliptical is obvious, and the sum property is equally obvious. The constant \(\alpha\) (\(0 < \alpha \leq 2\)) is the characteristic exponent; it describes the behavior of the tails of the distribution. The smaller is \(\alpha\), the more weight there is in the tails. The normal or Gaussian distribution with \(\alpha = 2\) has the lightest tails. For other values of \(\alpha\), the density function decreases like \(|x|^{-\alpha+1}\). This means that the moments \(\mathbb{E}[|X|^\gamma]\) are defined only for \(\gamma < \alpha\) when \(\alpha < 2\). The constants, \(\mu\) and \(c > 0\) are called the location and scale. They correspond to mean and standard deviation. Large values of \(c\) give a more dispersed distribution, but for \(\alpha \neq 2\), \(c\) must be interpreted in terms of some other measure of dispersion like mean absolute deviation as the variance is undefined.

The problem with the symmetric stable distributions are several. First the density functions can be expressed with simple functions only for the normal and the Cauchy (\(\alpha = 1\)) distributions. For the latter case, the density is\(^{17}\)

\[
f(x) = \frac{1}{\pi c [1 + (x - \mu)^2/c^2]}
\]

This makes modelling difficult. Second the data that was that was used by Mandelbrot (1963) and Fama (1963) to fit the stable distributions was log returns. This doesn’t adapt to portfolio analysis which combines raw returns. The subsequent development of other ways to model the fat tails of stock returns has largely removed the stable distributions from further consideration in Finance.

Risk Parity

Risk parity is a portfolio formation technique that is based on the allocation of risk rather than the allocation of capital. Usually risk is defined as standard deviation so it falls in the mean-

---

\(^{14}\) More formally, the stable distributions are called Lévy alpha-stable distributions, Pareto-Lévy distributions, or stable Paretoian distributions. This class of distributions serves as a generalization of the normal distribution for a Generalized Central Limit Theorem that allows infinite variances; the limiting sum of normalized independent variables converges to one of these distributions.

\(^{15}\) If the distributions are symmetric, then negative weights are permissible as well.

\(^{16}\) The character function for all asymmetric stable distributions is \(\exp(i\mu\theta - c^{\alpha}\theta^{\alpha}) [1 - i\beta \text{sgn}(\theta) \tan(\frac{1}{2} \alpha \pi)]\).

\(^{17}\) The densities can be written using special functions like the hypergeometric for \(\alpha = 1/3, 2/3, 4/3, 1/2, \) and 3/2 as well.
variance class, even though it is not directly concerned with means. The first risk-parity fund, the All Weather fund was created in 1996, though the term risk parity itself only dates back to its use by PanAgora Asset Management in 2005. Risk parity has become quite popular in the years since the financial crisis of the late 2000s with over $100 billion in assets.

Risk parity is typically applied to the allocation among classes of assets, but it can be used for forming portfolios of stocks as well. Risk parity seeks to equalize the contribution of each asset (or class) to a portfolio’s risk. The volatility of any portfolio is $\sigma(w) = \sqrt{w'\Sigma w}$. Because $\sigma(w)$ is homogeneous of degree one in the portfolio weights, it follows from Euler's Theorem that a portfolio’s standard deviation is $\sigma(w) = w'(\partial \sigma(w)/\partial w)$. This means the partial derivatives are the average as well as the marginal contribution of each asset to the portfolio’s standard deviation.\(^\text{18}\)

The total contribution of asset $i$ to portfolio risk is therefore $w_i \cdot \partial \sigma(w)/\partial w_i$. Risk parity is achieved when each asset contributes equally to total risk. That is,

$$\frac{\partial \sigma(w)}{\partial w_i} = \frac{\sigma(w)}{n} \quad \text{(71)}$$

Another way to describe the risk-parity portfolio is

$$\frac{\sigma(w)}{n} = w_j \frac{\partial \sigma(w)}{\partial w_j} = w_i \frac{\partial \sqrt{w'\Sigma w}}{\partial w_i} = w_i \frac{(\Sigma w)_i}{\sigma(w)} \quad \Rightarrow \quad w_i = \frac{\sigma^2(w)}{n(\Sigma w)_i}. \quad \text{(72)}$$

That is each asset’s weight is inversely proportional to the asset’s beta with the risk-parity portfolio. However, this is only a characterization rather than a solution because the betas depend on the construction of the portfolio. In general, (72) forms a set of simultaneous quadratic equations with no analytical solution except in special cases.

One solvable case is when all correlations are equal; $\sigma_{ij} = \rho \sigma_i \sigma_j \ \forall i \neq j$. In this case,

$$w_i (\Sigma w)_i = w_i^2 \sigma_i^2 + w_i \sum_{j \neq i} w_j \rho \sigma_i \sigma_j = w_i \sigma_i \left( w_i \sigma_i + \rho \sum_{j \neq i} w_j \sigma_j \right) = k \forall i. \quad \text{(73)}$$

It is clear by inspection that $w_i \propto \sigma_i^{-1}$ is the solution so the risk parity weights in this case are

$$w_i = \frac{\sigma_i^{-1}}{\sum \sigma_i^{-1}}. \quad \text{(74)}$$

The weights of the risk-parity portfolio are inversely proportional to each asset’s standard deviation and do not depend on the common correlation coefficient. The variance of this portfolio is

$$\sum_i \sum_j w_i w_j \sigma_{ij} = \sum_i w_i^2 \sigma_i^2 + \rho \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j = \frac{n + \rho n(n-1)}{\left( \sum \sigma_i^{-1} \right)^2} = \frac{\rho + \frac{1}{n} (1-\rho)}{\left( \frac{1}{n} \sum \sigma_i^{-1} \right)^2}, \quad \text{(75)}$$

which is proportional to the square of the harmonic mean of the standard deviations.

A risk parity portfolio is a type of low risk portfolio. Its risk falls somewhere between that of the global minimum variance portfolio and the equally weighted portfolio. A description of the portfolio weights for these three portfolios is given in the table below.

<table>
<thead>
<tr>
<th></th>
<th>equally-weighted</th>
<th>risk-parity</th>
<th>global minimum-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_i$</td>
<td>$w_j$</td>
<td>$w_i \frac{\partial \sigma(w)}{\partial w_i} = w_j \frac{\partial \sigma(w)}{\partial w_j}$</td>
<td>$w_i \frac{\partial \sigma(w)}{\partial w_i} = \frac{\partial \sigma(w)}{\partial w_i}$</td>
</tr>
</tbody>
</table>

so the risk-parity portfolio can be seen to lie between the other two in some sense.

\(^{18}\) Risk parity can be used with any other risk measure that is homogeneous of degree one in the portfolio weights.
This property can be verified as follows. Consider the minimization problem

\[
\min \sqrt{w'\Sigma w} \quad \text{subject to} \quad 1'w = 1 \quad \text{and} \quad \sum_{i=1}^{n} \ell i n w_i \geq c.
\]  

(77)

Note that the weights have all been implicitly constrained to be positive by the constraint on the sum of the logarithms. The problem in (77) is a convex program with a lower bound of zero, so the solution is unique provided the final constraint can be satisfied so that there are any solutions. The average portfolio weight is \( \bar{w} = n^{-1} \). Because \( \ell n \) is a concave function, from Jensen's inequality \( n^{-1} \sum \ell n w_i \leq \ell n \bar{w} = -\ell n n \). This means that the only feasible solution to (77) with \( c = -n \ell n n \) is \( w = n^{-1}1 \), the equally weighted portfolio. Solutions to (77) for any smaller \( c \) are possible, and there are no solutions for \( c > -n \ell n n \).

The achieved minimum is obviously weakly increasing in \( c \) because the constraint tightens as \( c \) grows. If \( c = -\infty \), there is no constraint so the result is the global minimum variance portfolio when weights are restricted to be nonnegative. Clearly the variance of this restricted portfolio is no less than the unrestricted global minimum variance.

So if the risk parity portfolio is a solution to (77), that portfolio has a variance between the other two. To confirm the risk-parity portfolio is a solution, ignore the \( 1'w = 1 \) constraint for the moment and call the unrestricted weights \( \omega \). The first-order conditions for this problem are

\[
0 = (\Sigma \omega)_{i} - \lambda \omega_{i}^{-1} \Rightarrow \omega_{i} (\Sigma \omega)_{i} = \lambda.
\]  

(78)

This is exactly the equation for the risk-parity portfolio with the weights not yet normalized. Define \( w \equiv \omega/\Omega \) where \( \Omega \equiv 1'\omega \). Then the weights in \( w \) do sum to one. It is the risk-parity portfolio and it satisfies (77) with \( c' \equiv c - n \Omega \) because

\[
\sum \ell n w_i = \sum \ell n \omega_i - n \Omega \geq c - n \Omega = c'.
\]  

(79)