Chapter 6 — Extensions and Applications of the CAPM

The CAPM with Market Frictions

The CAPM market is a simple one. All investors have the same beliefs, the same investment horizon, and evaluate assets using only means and covariances. They face no restrictions on their ability to hold stocks in any quantity, borrow unlimited amounts or to short any asset, and trading is free of costs and differential taxes. The capital assets are investors’ only assets; they have no human capital, Non-marketable income, or non-trading assets like homes. Ignoring all these complications leads to the simple solution see in the previous chapter. When these frictions are present, the equilibrium will be affected.

It might seem that if some friction created a large disequilibrium cost, there would be a market response to eliminate it. Some institution or intermediary, like a bank, clearing corporation, mutual fund, or insurance company, should arise that would profit from removing or reducing the costs to investors. So frictions, at least those that induce large costs, should effectively disappear.

Of course setting up these intermediaries is costly as well. If eliminating the friction removes the cost to the investors, there may be no profit left for the intermediary. This would mean there was no incentive to create or continue to operate them, which means the frictional costs would still be present. In other words, there is probably some optimal level of disequilibrium costs. Of course, these costs are only disequilibrium in the sense of the incomplete model so it might better be described as an equilibrium level of disequilibrium.

Most of our models continue to make simplifying assumptions and ignore many of the complexities. Merton Miller (1987) once wrote, “That we abstract from all these stories in building our models is not because the stories are uninteresting, but because they may be too interesting and thereby distract us from the pervasive market forces that should be our principal concern.”

This chapter looks at some of those interesting stories. It starts with the CAPM and examines how various frictions distort the equilibrium. Another purpose of this chapter is to illustrate various modeling techniques.

The CAPM with Derivative Contracts

Whether or not an elliptical distribution is a valid or approximate characterization of stock returns; it is certainly not approximately true when derivative assets like options are available. Regardless of the distribution of a stock price, $S$, the return on a call option with payoff $\max(S - K, 0)$ clearly cannot be elliptical as the left tail is truncated while the right is not. So it would seem the presence of derivative contracts like options would eliminate a CAPM equilibrium. This is only partially true.

Suppose that all non-financial assets, those in positive net supply, have returns with a joint elliptical distributed. Further, suppose that sufficient derivative assets are introduced to make the market effectively complete. The efficient set is convex in an effectively complete market with homogeneous beliefs among investors. As the efficient set is convex, the market portfolio, which is a convex combination of efficient portfolios, must be efficient. That means there is a risk-averse representative investor who holds the market. This representative investor optimally chooses not hold the derivative assets so would obviously hold the same portfolio if they were not available but the fundamental assets had the same joint distribution. Because that distribution is elliptical, the efficient market portfolio must be mean-variance efficient among the non-financial assets. Therefore, the CAPM pricing relation holds for those assets. The valuation of the derivative contracts in a CAPM equilibrium is discussed in Chapter 7 in a more general context.
The CAPM with Limited Borrowing

In the previous chapter it was shown that a restriction on borrowing leads to the Black or zero-beta version of the model even in the presence of a risk-free asset. Here that restriction is examined in more detail to pin down the zero-beta rate.

To analyze the effect of borrowing restrictions, assume that each investor can borrow no more than a given fraction (or multiple) of his wealth. The margin restriction means that an investor’s purchase of the risky assets cannot exceed \( \delta \) times his initial wealth, i.e., \( W \mathbf{1}'w \leq \delta W \). If the assets have a multivariate normal distribution, the portfolio problem for a CARA investor with risk tolerance (the reciprocal of risk aversion) of \( T \) is to maximize

\[
\max_{\mathbf{w}} \left[ (1 + r_f) + \mathbf{w}'(\mathbf{\mu} - r_f \mathbf{1}) \right] W - \frac{1}{2} T^{-1}W^2 \mathbf{w}'\mathbf{\Sigma} \mathbf{w} + \lambda (\delta - \mathbf{1}'\mathbf{w}) W.
\]

The final term is zero when the margin constraint is not binding. The first order conditions and solution are

\[
0 = \mathbf{\mu} - r_f \mathbf{1} - T^{-1}W \mathbf{\Sigma} \mathbf{w} - \lambda \mathbf{1} \Rightarrow \begin{cases} \mathbf{w}^* W = T[\mathbf{\Sigma}^{-1}(\mathbf{\mu} - r_f \mathbf{1}) - \lambda \mathbf{\Sigma}^{-1} \mathbf{1}] \\ \lambda \geq 0 \\ \lambda (\delta - \mathbf{1}'\mathbf{w}) = 0. \end{cases}
\]

The last two relations are the standard Kuhn Tucker conditions. The multiplier, \( \lambda \), measures the cost of the constraint and must be nonnegative. It is positive only when the constraint is binding.

The optimal portfolio consists of holdings of the risk-free asset; the tangency portfolio, \( \mathbf{t} \propto \mathbf{\Sigma}^{-1}(\mathbf{\mu} - r_f \mathbf{1}) \); and possibly the global minimum-variance portfolio, \( \mathbf{g} \propto \mathbf{\Sigma}^{-1} \mathbf{1} \). If the leverage constraint is not binding (\( \lambda = 0 \)), the investor holds the tangency portfolio levered up or down proportionally to his risk tolerance. Those investors for whom the borrowing constraint is binding (\( \lambda > 0 \)) combine a short position in the global minimum-variance portfolio with a long position in the tangency portfolio to create a portfolio on the minimum-variance hyperbola, then lever that position to the maximum. When the borrowing constraint is binding, \( \mathbf{1}'\mathbf{w} \) is equal to \( \delta \), the maximum margin. So the Lagrange multiplier is

\[
\lambda = \max \left[ \frac{\mathbf{1}'\mathbf{\Sigma}^{-1}(\mathbf{\mu} - r_f \mathbf{1}) - \delta W/T}{\mathbf{1}'\mathbf{\Sigma}^{-1} \mathbf{1}}, 0 \right] = \max \left[ \frac{a - \delta \cdot \text{(RRA)}}{C}, 0 \right]
\]

where \( a \equiv \mathbf{1}'\mathbf{\Sigma}^{-1}(\mathbf{\mu} - r_f \mathbf{1}) \) and \( C \equiv \mathbf{1}'\mathbf{\Sigma}^{-1} \mathbf{1} \).

---

1 Utility of time-1 wealth is \(-\exp(-W/T)\). Throughout this section \( a \) denotes \( \mathbf{1}'\mathbf{\Sigma}^{-1}(\mathbf{\mu} - r_f \mathbf{1}) \) as defined in (3) rather than the coefficient of absolute risk aversion.
as defined in the previous chapter. \((\text{RRA}) \equiv W/T\) is the investor’s relative risk aversion. The variable \(\lambda\) measures the individual cost of the borrowing constraint. So the cost is smaller the higher is an investor’s allowed margin or the higher is his relative risk aversion because that lessens his desire to lever.

All investors hold either a combination of the tangency portfolio with lending or a combination of some portfolio above \(t\) on the minimum-variance hyperbola with borrowing. The latter investors’ minimum-variance hyperbola portfolios are equivalent to a long position in the tangency portfolio and a short position in the global minimum variance portfolio. As all investors’ portfolios of risky assets are combinations of two minimum-variance hyperbola, the market portfolio (of risky assets) is as well and lies on the minimum variance hyperbola. Therefore, the zero-beta version of the CAPM holds. However, it is more convenient to express the equilibrium in terms of the interest rate rather than the zero-beta rate.

The aggregate demand must equal the aggregate supply which is the market portfolio so

\[
\mathbf{m} \sum_{k} W_k = \sum_{k} \mathbf{w}_k W_k = \Sigma^{-1}(\mathbf{r} - \mathbf{r}_f) \sum_{k} T_k - \Sigma^{-1} \mathbf{1} \sum_{k} \lambda_k T_k
\]

\[
\Rightarrow \mathbf{m} W_{\text{agg}} = \bar{T} \Sigma^{-1}(\mathbf{r} - \mathbf{r}_f) - \bar{\lambda} \Sigma^{-1} \mathbf{1}
\]

where \(W_{\text{agg}} = \sum_{k} W_k \quad T = \sum_{k} T_k \quad \bar{\lambda} = \sum_{k} \lambda_k T_k \).

The “market” risk tolerance, \(\bar{T}\), is the mean of the individual risk tolerances. The market leverage constraint, \(\bar{\lambda}\), is the risk-tolerance weighted average of the individual constraints. As all the constraints and risk tolerances are nonnegative, so is \(\bar{\lambda}\). If the constraint is binding for any investor then \(\bar{\lambda}\) is strictly positive. Solving (4) for the vector of expected rates of return gives

\[
\mathbf{r} - \mathbf{r}_f = \bar{T}^{-1}W_{\text{agg}} \Sigma \mathbf{m} + \bar{\lambda} \mathbf{1}
\]

Premultiplying by \(\mathbf{m}\) gives

\[
\mathbf{r}_m - r_f - \bar{\lambda} = \bar{T}^{-1}W_{\text{agg}} \Sigma^2 \mathbf{m}
\]

so substituting for \(\bar{T}^{-1}W_{\text{agg}}\) leaves

\[
\mathbf{r} - (r_f + \bar{\lambda}) \mathbf{1} = \Sigma \mathbf{m} \frac{\mathbf{r}_m - r_f - \bar{\lambda}}{\Sigma^2 \mathbf{m}} = \beta (\mathbf{r}_m - r_f - \bar{\lambda})
\]

The security market lines has positive, rather than zero, intercept and a slope that is less than the risk premium on the market. The individual assets CAPM alphas are \(\alpha = (1 - \beta) \bar{\lambda}\) which are decreasing in the betas. The zero-beta portfolio has an expected excess rate of return of \(\bar{\lambda}\) rather than 0 even though there is a risk-free asset. The market portfolio is on the minimum-variance hyperbola but at a point beyond the tangency portfolio, and its Sharpe ratio is not the largest possible. The largest Sharpe ratio, as usual, belongs to the tangency portfolio which has a market beta less than one.

That \(S_t > S_m\) is obvious from the figure, but it can also be derived algebraically. From (4), the market portfolio is \(\mathbf{m} = (\text{RRA})\Sigma^{-1}(\mathbf{r} - \mathbf{r}_f) - \bar{\lambda} \Sigma^{-1} \mathbf{1}\) where \((\text{RRA}) = W_{\text{agg}}/T\) is the relative risk aversion of the market. The expected excess rate of return and variance of the market are

\[
\mu_m - r_f = \mathbf{m}'(\mathbf{r} - \mathbf{r}_f) = (\text{RRA})[(\mathbf{r} - \mathbf{r}_f)' \Sigma^{-1}(\mathbf{r} - \mathbf{r}_f) - \bar{\lambda}(\mathbf{r} - \mathbf{r}_f)' \Sigma^{-1} \mathbf{1}] = (\text{RRA})(b - \bar{\lambda} a)
\]

\[
\sigma_m^2 = \mathbf{m}' \Sigma \mathbf{m} = (\text{RRA})^2 [(\mathbf{r} - \mathbf{r}_f)' \Sigma^{-1}(\mathbf{r} - \mathbf{r}_f) - \bar{\lambda}(\mathbf{r} - \mathbf{r}_f)' \Sigma^{-1} \mathbf{1} + \bar{\lambda}^2 \mathbf{1}' \Sigma^{-1} \mathbf{1}]
\]

\[
= (\text{RRA})^2 (b - 2 \bar{\lambda} a + \bar{\lambda}^2 C)
\]

where \(b = (\mathbf{r} - \mathbf{r}_f)' \Sigma^{-1}(\mathbf{r} - \mathbf{r}_f)\)

and \(a\) and \(C\) are defined in (3). The Sharpe ratios of the tangency and market portfolios are
\[ S_t = \sqrt{b} \quad \text{and} \quad S_m = \frac{b - \lambda a}{\sqrt{b - 2\lambda a + \lambda^2 C}}. \]  

The tangency portfolio’s Sharpe ratio is larger because

\[ S_t^2 - S_m^2 = b - \frac{(b - \lambda a)^2}{b - 2\lambda ab + \lambda^2 C} = \frac{b^2 - 2\lambda ab + \lambda^2 C b - b^2 + 2\lambda ab - \lambda^2 a^2}{b - 2\lambda a + \lambda^2 C} = \frac{\lambda^2 (C b - a^2)}{b - 2\lambda a + \lambda^2 C}. \]

The denominator is \((RRA)^2 \sigma_m^2 > 0\), and the numerator is positive by the Cauchy Schwartz inequality so \(S_t > S_m\).

The tangency portfolio’s market beta is

\[ \beta_t = \frac{\sigma_{tm}}{\sigma_m^2} = \frac{\sigma_{tm}^2}{\sigma_m^2} \times \frac{\mu_t - r_f}{\mu_t - r_f} \times \frac{\sigma_t^2}{\sigma_m^2} = \frac{S_m \sigma_t}{S_t \sigma_m} < 1. \]

The third equality follows because all assets are priced by the tangency portfolio tautologically. The inequality follows because the tangency portfolio’s Sharpe ratio is larger as just shown, and the \(\sigma_m > \sigma_t\) because the market portfolio lies further out on the minimum-variance hyperbola.

The effect in this model is an aggregate one. Individual assets have CAPM alphas, \(\alpha = (1 - \beta) \lambda\), but this is due not to any differential information but to a change in the price of risk due to an effective increase in the interest rate of \(\lambda\). This model is the basis for the “Betting Against Beta” strategy used, for example, by AQR hedge funds. These strategies form market neutral, zero-beta funds that are short in high beta assets and long in low beta assets.

**The CAPM with Restricted Stock Holdings**

In many contexts some investors will hold portfolios that deviate from the market portfolio. For example, almost all individual investors have too little wealth to be able to purchase all the stocks in their market proportions. Much real estate is in private hands and cannot be held in others’ portfolios. Some investors may avoid certain stocks because they feel they lack adequate information to invest in them. Residents of a country definitely prefer to invest in companies based in their home country. In the U.S., at least, this is also true to a lesser extent for companies located in the state or locality where the investor resides. This is known as a home bias and may, of course, be due to better information about home-based stocks.

In recent years, ESG investing has become more common. ESG stands for environmental, social, and corporate governance. In the 1970s, there was a movement for many pension and endowment plans to divest their holdings in companies that did business with the apartheid regime in South Africa based on the Sullivan principles developed by the Reverend Leon Sullivan, then a member of the General Motors board of directors. While that was not the first episode of “social” investing, it did seem to spark a renewed interest in corporate social responsibility.

One aspect of ESG investing is to avoid the stocks of companies whose practices are suspect under one or more criteria. So ESG mutual funds tend to avoid the stocks of tobacco companies and to prefer “green” companies like those developing solar panels or wind turbines over traditional electric companies. In 2018, almost half of U.S. endowment funds, foundations, and pension plans considered ESG features in their investment decisions. This was nearly twice as many as five years earlier.

To analyze this behavior, consider an economy in which all investors have homogeneous beliefs, but only some are mean-variance maximizers. The mean-variance maximizers hold the tangency portfolio. The non-mean-variance maximizers’ portfolios differ from the tangency portfolio by a residual portfolio \(p_k\). That is, each investor’s portfolio is \(w = (1 - \theta_k) t + \theta_k p_k\). For
example, portfolio $p_k$ might be long First Solar and short Philip Morris International if the investor wanted to favor investing in solar projects and to shun tobacco companies. For the mean-variance maximizers, $\theta_k = 0$.

Aggregating demands, the market portfolio is $m = (1 - \theta)t + \theta p$ where $\theta = \sum W_k \theta_k / \sum W_k$. The properties of the market portfolio are

$$
\begin{align*}
\mu_m - r &= \theta(\mu_p - r) + (1 - \theta)(\mu_k - r) \\
\sigma_m^2 &= \theta^2 \sigma_p^2 + 2\theta(1 - \theta)\sigma_{p_t} + (1 - \theta)^2 \sigma_t^2
\end{align*}
$$

and

$$
\sigma_m = \theta \sigma_p + (1 - \theta) \sigma_t
$$

where $i$ denotes any asset or portfolio. The relations in (11) can be re-expressed to give the tangency portfolio’s properties

$$
\begin{align*}
\mu_t - r &= \frac{\mu_m - r - \theta(\mu_p - r)}{1 - \theta} \\
\sigma_t^2 &= \frac{\sigma_m^2 - \theta^2 \sigma_p^2 - 2\theta(1 - \theta)\sigma_{p_t} - 2\theta \sigma_{p_m} - 2\theta^2 \sigma_{p_t}^2 - 2\theta \sigma_{p_m}^2}{(1 - \theta)^2}
\end{align*}
$$

The CAPM relation based on the tangency portfolio is exactly correct. It is a tautology independent of any economic assumptions. Using the relations in (12) for any asset or portfolio $i$,

$$
\begin{align*}
\mu_i - r &= \frac{\sigma_{i_m}}{\sigma_i^2}(\mu_i - r) = \frac{\sigma_{i_m} - \theta \sigma_{i_p}}{\sigma_m^2 + 2\theta^2 \sigma_p^2 - 2\theta \sigma_{p_m}}(1 - \theta)(\mu_i - r) \\
&= \frac{\sigma_{i_m} - \theta \sigma_{i_p}}{\sigma_m^2 + 2\theta^2 \sigma_p^2 - 2\theta \sigma_{p_m}}[\mu_m - r - \theta(\mu_p - r)]
\end{align*}
$$

This relation is valid for any asset or portfolio. For portfolio $p$, it gives

$$
\begin{align*}
\mu_p - r &= \sigma_{p_m}^2(2\sigma_{p_p}^2 - 2\theta \sigma_p^2) = (\sigma_{p_m} - \theta \sigma_p^2)[\mu_m - r - \theta(\mu_p - r)] \\
&= (\mu_p - r)(\sigma_{p_m}^2 - \theta \sigma_p^2) = (\sigma_{p_m} - \theta \sigma_p^2)(\mu_m - r) \\
\mu_p - r &= \frac{\sigma_{p_m} - \theta \sigma_p^2}{\sigma_m^2 - \theta \sigma_p^2} \\
&= \frac{\beta_p - \theta \sigma_p^2 / \sigma_m^2}{1 - \theta \beta_p} = \frac{\beta_p - \theta \sigma_p^2 / \sigma_m^2}{1 - \theta \beta_p}
\end{align*}
$$

where $\beta_p$ is the market beta of portfolio $p$ and $s_p^2$ is its residual variance.

Now suppose that mean-variance investors predominate in the market (or the non-mean-variance investors only deviate from the tangency portfolio by a small amount) so that $\theta$ is small. Express the denominator of $\Gamma$ using a second-order exact Taylor expansion

$$
\begin{align*}
\mu_p - r &= \beta_p - \frac{\theta \sigma_p^2}{\sigma_m^2} \left(1 + \theta \beta_p + \bar{\beta}^2 \beta_p^2\right) \\
&= \beta_p - \frac{\theta \sigma_p^2}{\sigma_m^2} \left[1 - \frac{\theta \sigma_p^2 / \sigma_m^2}{1 - \theta \beta_p}\right] - \frac{\theta \sigma_p^2}{\sigma_m^2} \bar{\beta}^2 \beta_p^2 - \frac{\theta \sigma_p^2}{\sigma_m^2} \bar{\sigma}_p^2
\end{align*}
$$

Even if the ESG investors hold only the portfolio $p$ so that their $\theta_k = 1$, the market $\theta$ will be small if the mean-variance investors dominate in the market.

---

2 Even if the ESG investors hold only the portfolio $p$ so that their $\theta_k = 1$, the market $\theta$ will be small if the mean-variance investors dominate in the market.
This derivation is completely general. The market can always be decomposed into the tangency portfolio and some residual portfolio \( p \), and the CAPM relation is always true for the tangency portfolio. This analysis is useful only if the deviation portfolio \( p \) can be identified. In that case, (15) indicates how this deviation affects the relation between beta and expected rates of return.

These results should tend to extend positively (or negatively) to those assets long (or short) in portfolio \( p \). Returning to equation (13) and substituting for \( \mu_p - r \) from (14)

\[
\mu_j - r = \frac{\sigma_m^2 - \theta \sigma_{ip}}{\sigma_m^2 + \theta^2 \sigma_p^2 + 2 \theta \sigma_{pm}}[\mu_m - r - \theta(\mu_p - r)]
\]

\[
= \frac{\sigma_m^2 - \theta \sigma_{ip}}{\sigma_m^2 + \theta^2 (\beta_p \sigma_m^2 + \sigma_p^2) - 2 \theta \beta_p \sigma_m^2} (\mu_m - r)[1 - \theta(\beta_p - \Gamma)]
\]

\[
\Rightarrow \quad \frac{\mu_j - r}{\mu_m - r} = \frac{\beta_j - \theta \sigma_{ip} / \sigma_m^2}{1 + \theta^2 \beta_p^2 - 2 \theta \beta_p + \theta^3 \sigma_p^2 / \sigma_m^2} (1 - \theta \beta_p + \theta \Gamma) 
\]

\[
= \frac{\beta_j - \theta \sigma_{ip} / \sigma_m^2}{(1 - \theta \beta_p)^2 + \theta \Gamma (1 - \theta \beta_p)} = \frac{\beta_j - \theta \sigma_{ip} / \sigma_m^2}{1 - \theta \beta_p}
\]

The covariance \( \sigma_{ip} \) can be expressed as \( \sigma_{ip} = \beta_p \sigma_m^2 + s_p \) where \( s_p \) is the covariance between the market-residual risks of asset \( i \) and portfolio \( p \). Therefore

\[
\frac{\mu_j - r}{\mu_m - r} = \frac{\beta_j - \theta \beta_p \beta_i - \theta s_{ip} / \sigma_m^2}{1 - \theta \beta_p} = \beta_j - \frac{\theta s_{ip}}{(1 - \theta \beta_p) \sigma_m^2}.
\]

This relation shows that a stock whose residual risk covaries positively with the residual risk of portfolio \( p \) will have an expected rate of return smaller than predicted by the CAPM. Obviously, this will tend to be stocks that are held long in \( p \), the “good” ESG stocks. The “bad” ESG stocks will tend to have positive CAPM alphas.

Note that this result does not mean that ESG investing has the opposite of its intended effect. It is true that investors in the “good” ESG stocks earn low returns, but that is only an effect in the secondary market. In the primary market, “bad” firms have atypically high costs of capital so it is more difficult for them to raise money for their activities. That is the desired outcome of ESG investing.

The CAPM with Liquidity Considerations

An asset’s liquidity measures how well it can be quickly sold for its fair value. Stocks and other financial assets are usually fairly liquid. In contrast, houses, other real estate, automobiles, and artwork are less liquid. Cash, of course, is the most liquid asset. Other things being equal, investors would prefer assets with high liquidity so that they could be easily sold in an emergency or if better opportunities arose.

Liquidity is usually measured by the bid-ask spread, the difference between the prices offered for immediate purchase and sale. The bid-ask spread is readily determined for many financial assets, though it may be very difficult to estimate for things like real estate.

The primary cause of illiquidity is differential information. If there is frequent important information that takes some time to become commonly known, then there is more chance that either the potential buyer or seller possesses information that the other does not. This leads to a larger bid-ask spread to protect the other party. For this reason, illiquidity is related to price volatility, but they are not the same. The price can be quite volatile but the asset can remain liquid.
if the information is commonly known. Long-term government bonds are more volatile than short-term bonds, but they are not more liquid.

Liquidity can be incorporated into the CAPM in the following model. At any point each asset has a bid and ask price. It is common to think of the midpoint of the bid and ask as the fair price.\(^3\) The bid, ask, and midpoint price of asset \(i\) are \(B_i, A_i,\) and \(P_i \equiv (B_i + A_i)/2.\) Each asset’s liquidity is measured by half the percentage spread, \(s_i \equiv (A_i - B_i)/(2P_i) = (A_i - P_i)/P_i = (P_i - B_i)/P_i.\)

An outside econometrician only sees realized prices which are just as likely to be bids or asks so on average the prices they see are a series of midpoints prices. Measured returns are the rate of increase in the midpoint. However, any given investor purchases at the ask and sells at the bid so the realized return is 
\[
10 1 0 (1 ) (1 ) (1 ) (1 ) (1 ) (1 ) \frac{BA P s P s r s s s r}{\xi} = \xi
\]
where \(r_\xi\) is the rate or return typically measured by the econometrician, and 
\[
\xi \equiv (1-s)/(1+s) \leq 1.
\]
The spread parameter \(\xi\) is a measure of liquidity; the closer \(\xi\) is to 1, the more liquid is the stock.

Assuming the investor starts with all his wealth in cash and wishes to hold only cash at the end of the single period, the portfolio problem is
\[
\text{max } \mathbb{E}[\tilde{W}_t] - \frac{1}{2} \text{var}[\tilde{W}_t]\ 
\text{subject to } \tilde{W}_t = W_0 \sum_{i=0}^{n} w_i (1 + \tilde{r}_i) \xi_{i} \quad \text{and} \quad \sum_{i=0}^{n} w_i = 1. \quad (18)
\]
The first-order conditions are
\[
\frac{\partial}{\partial w_0} : \ 0 = (1 + r_j) \xi_{j} - \lambda
\]
\[
\frac{\partial}{\partial w_i} : \ 0 = \mathbb{E}[1 + \tilde{r}_i] \xi_{i} - aW_0 \sum_{j=1}^{n} w_j \xi_{j} \text{cov}[\tilde{r}_i, \tilde{r}_j] - \lambda.
\]

Define the spread-adjusted interest rate and rates of return as
\[
\tilde{r}_i = r_i + \xi_{i} - 1
\]
and 
\[
\tilde{r}_i = r_i - \xi_{i} - 1
\]
This is exactly the same as the first-order condition in the standard problem so the same continuing analysis applies. The CAPM holds when applied to the spread-adjusted rates of return
\[
\mathbb{E}[\tilde{r}_i - r_j] = \text{cov}[\tilde{r}_i, \tilde{r}_j] \text{var}[\tilde{r}_i] \mathbb{E}[\tilde{r}_m - r_j]. \quad (20)
\]
The spread parameter \(\xi_{m}\) is a weighted average of the individual spread parameters. Expressing the equilibrium in terms of the “measured” returns
\[
\mathbb{E}[(1 + \tilde{r}_i) \xi_{i} - (1 + r_j) \xi_{j}] = \frac{\xi_{m} \text{cov}[\tilde{r}_i, \tilde{r}_m]}{\text{var}[\tilde{r}_m]} \mathbb{E}[(1 + \tilde{r}_m) \xi_{m} - (1 + r_j) \xi_{j}]
\]

\[
\Rightarrow \mathbb{E}[\tilde{r}_i - r_j] + (1 + r_j) \left(1 - \frac{\xi_{i}}{\xi_{m}}\right) = \frac{\text{cov}[\tilde{r}_i, \tilde{r}_m]}{\text{var}[\tilde{r}_m]} \left[\mathbb{E}[\tilde{r}_m - r_j] + (1 + r_j) \left(1 - \frac{\xi_{i}}{\xi_{m}}\right)\right]. \quad (21)
\]

The difference between this relation and the CAPM is the alpha
\[
\alpha_i = (1 + r_j) \left[\frac{\xi_{i}}{\xi_{m}} - 1 + \beta_i \left(1 - \frac{\xi_{i}}{\xi_{m}}\right)\right]. \quad (22)
\]

The spread parameters \(\xi_{m}\) are all less than or equal to 1. The risk-free asset is likely the most liquid asset with \(\xi_{m} \approx 1\) so the constant part of the alpha should be positive at least for most stocks. That

\(^3\) Models in later chapters will show that the fair price need not be at the average of the bid and ask. Nonetheless, the average is often chosen as the representative price because it is easily determined while the fair price depends on many other features beyond the scope of the current model.
is, an extra return is required to induce the holding of these less liquid assets. Of course, this affects all stocks and the market’s alpha must be zero as shown. So the average stock with a beta of 1 has 
\[
\alpha_i = (1 + r_f)\xi_i - \xi_m = (1 + s_i)(1 + s_i + s_i^2 + \cdots) \cdot \xi_i \n\]
Stocks that are more liquid than average, \(\xi_i > \xi_m\), have negative alphas and smaller than average returns. Smaller companies, stocks with low stock prices, and stocks with higher volatilities all tend to have larger spreads. This model predicts that such stocks should have positive alphas. Holding beta fixed, alphas are increasing and convex is the spread because

\[
(1 + s_i)^2 = (1 + s_i)(1 + s_i + s_i^2 + \cdots) \cdot \xi_i \n\]

Assuming the risk-free asset is more liquid than the average stock, \(\xi_f > \xi_m\), the alphas are decreasing in beta. In the standard CAPM, the compensation for beta risk is \(\mu - r_f\). The spread reduces the expected gross return. When the spread on the risk-free rate is smaller than that on stocks, \(\mu - r_f < \mu - r_f\), the compensation for \(\beta\)-risk is less.

**The CAPM with Non-Marketable Assets or Income**

The three previous models were straightforward extensions to the CAPM. The altered equilibrium could be described by alphas relative to the standard model. The next two models are somewhat different. Rather than deriving “mispricing” alphas, the models introduce a second source of risk. The resulting models are two factor models.

If an investor has another source of income apart from his capital assets, then that will typically affect his investment decision and the resulting equilibrium. Suppose an investor has non-marketable income, \(\bar{y}\), that has a joint elliptical distribution with the rates of return on the marketable assets. For simplicity, we will refer to this as wage income, but it could be income from any nonmarketable source. The portfolio problem facing the investor can be viewed as one with an additional asset, but the holding of this asset is constrained to a particular level. The investor will still wish to hold a portfolio that produces the minimum variance of wealth at some level of expected wealth, but this will typically not be a portfolio that is mean-variance efficient in the absence of the other income.

If investors have homogeneous beliefs, their choices will involve the tangency portfolio. But they will deviate from that portfolio in different ways depending on their extra income so the tangency portfolio will no longer be the market portfolio of risky assets except by coincidence.

To determine the effect of non-marketable wage uncertainty on asset prices, it is convenient to decompose each investor’s wage income into its common and individual components. Let \(\bar{y}\) be the random per capita wage income realized in the economy of \(K\) investors, and express each investor’s wage as a projection onto the per capita wages

\[
\tilde{y}_k = \alpha_k + \gamma_k \bar{y} + \tilde{\epsilon}_k . \tag{23} \n\]

Because \(\sum \tilde{y}_k = K \bar{y}\), clearly \(\sum \alpha_k = 0\), \(\sum \gamma_k = K\), and \(\sum \tilde{\epsilon}_k \equiv 0\), by construction. Each residual term \(\tilde{\epsilon}_k\) is uncorrelated with per capita income, but they are not mutually uncorrelated as they must sum to zero. This is completely general independent of the assumption about elliptical distributions. The vector of the covariances of the investor’s wages with the asset returns is

\[
\text{cov}[\tilde{y}_k, \bar{r}] = \gamma_k \text{cov}[\bar{y}, \bar{r}] + \text{cov}[\tilde{\epsilon}_k, \bar{r}] = \gamma_k \Psi + \Phi_k . \n\]

---

4 Amihud and Meldelson (1986) conclude that expected returns are increasing and concave in spreads. Their model is a multi-period one, and the concavity is due to clientele effects. Investors with longer horizons preferentially hold less liquid stocks to reap the higher returns. They are less concerned with liquidity as the spread applies once per trade rather than once per period.

5 Examples include houses, some retirement plans or trusts, private businesses, and government transfer payments. In a multi-period model \(\bar{y}\) is the payoff plus the present value of future payoffs. For wages, this includes the end of period value of human capital as well as the wages received. These are analogous to the final stock price and dividend paid.
With an elliptical distribution, expected utility can be expressed with the derived utility function, \( V(\mu, \sigma^2) \) defined over the expected wealth \( \mathbb{E}[\tilde{W}] = \mathbb{E}[\tilde{y}_k] + W_0[1 + r_f + w'(\mu - r_f \mathbf{1})] \) and variance of wealth \( \text{var}[\tilde{W}] = \text{var}[\tilde{y}_k] + 2W_0w' (\gamma_k \psi + \phi_k) + W_0^2 \Sigma w \). Here \( W_0 \) is the current value of liquid wealth excluding the value of future wages. The investor’s portfolio problem is

\[
\max_w V(\mathbb{E}[\tilde{y}_k] + W_0[1 + r_f + w'(\mu - r_f \mathbf{1})], \text{var}[\tilde{y}_k] + W_0w'(\gamma_k \psi + \phi_k) + W_0^2 \Sigma w). 
\] (24)

The portfolio weights are unconstrained because they do not include the risk-free asset. The first order conditions are

\[
0 = \frac{\partial V}{\partial w} = V_1(w_0(\mu - r_f \mathbf{1}) + 2V_2(w_0(\gamma_k \psi + \phi_k + W_0 \Sigma w))
\] (25)

with solution

\[
w^* = -\frac{V_1}{2W_0V_2(w_0(\mu - r_f \mathbf{1}) - \frac{\gamma_k \Sigma^{-1} \psi}{W_0} - \frac{1}{W_0} \Sigma^{-1} \phi_k}.
\] (26)

The optimal portfolio for each investor is a mixture of the risk-free asset, the tangency portfolio, \( t \equiv \Sigma^{-1}(\mu - r_f \mathbf{1})/\Sigma^{-1}(\mu - r_f \mathbf{1}) \), and two hedge portfolios, \( h = \Sigma^{-1} \psi/\Sigma^{-1} \psi \) and \( h_k \equiv \Sigma^{-1} \phi_k/\Sigma^{-1} \phi_k \). As before, the tangency portfolio has the highest Sharpe ratio providing the best tradeoff between risk and expected return among the traded assets.

The makeup of the two hedge portfolios does not depend on any of the assets’ expected rates of return. Portfolio \( h \) provides a hedge against a decrease in per capita income. It is the portfolio with the highest squared correlation with changes in aggregate or per capita wages as shown by this maximization

\[
\text{max corr}^2[\tilde{y}, \tilde{r}] = \text{max}_w \frac{\gamma_k^2(w' \psi)^2}{\sigma_y^2 w' \Sigma w}
\] (27)

As in the Sharpe ratio maximization, the constraint \( 1'w \) need not be applied because the correlation is homogeneous of degree zero in the portfolio weights. Again, we need not solve this equation but only verify that the portfolio with weights proportional to \( \Sigma^{-1} \psi \) is a solution. Substituting \( \Sigma^{-1} \psi \) into the last line of (27) we have

\[
(\psi' \Sigma^{-1} \psi \Sigma^{-1} \psi - \psi' \Sigma^{-1} \psi) \Sigma^{-1} \psi = (\psi' \Sigma^{-1} \psi - \psi' \Sigma^{-1} \psi) \psi = 0.
\] (28)

So the hedge portfolio does have the maximum possible squared correlation with per capita wages. Positive or negative correlations between per capita income and portfolio \( h \) are equally useful as the investor can purchase or short the hedge. The covariance of the hedge portfolio with per capita income is \( \gamma_k \psi' \Sigma^{-1} \psi/\Sigma^{-1} \psi \). The sign of the covariance is the same as \( \text{sgn}(\gamma_k/\Sigma^{-1} \psi) \). The investor’s demand for the hedge portfolio is \( -\gamma_k (1' \Sigma^{-1} \psi)/W \), which has the opposite sign to the covariance. So the optimal holding of the hedge portfolio reduces the contribution of income uncertainty to the variance of wealth.

Similarly, portfolio \( h_k \) is maximally correlated with the residual of the wage income of investor \( k \). The residual hedge portfolio is positively correlated with the residual income and investor \( k \) takes a short position if \( 1' \Sigma^{-1} \phi_k > 0 \). Otherwise the portfolio is negatively correlated and the investor takes a long position.

The ratio of the partial derivatives of the utility function, \( T_k \equiv V_1^k / 2V_2^k \), is a measure of
absolute risk tolerance. However, it need not be specified to derive the pricing relation. The aggregate dollar demand must equal the market portfolio, \( m \sum_k W_k \), so

\[
m \sum_k W_k \equiv \sum w_k^* W_k = \Sigma^{-1}(\mu - r_j) 1 \sum T_k - \Sigma^{-1} \psi \sum \gamma_k - \Sigma^{-1} \sum \phi_k .
\] (29)

This equation can be simplified using \( \Sigma \gamma_k = K \) and \( \Sigma \phi_k = \text{cov}[\Sigma_k \bar{e}_k, \bar{r}] = \text{cov}[0, \bar{r}] = 0 \)

\[
m \bar{W} = \Sigma^{-1}(\mu - r_j) \bar{T} - \Sigma^{-1} \psi
\] (30)

where \( \bar{W} \) and \( \bar{T} \) are simple averages across all investors. Solve (30) for \( \Sigma^{-1}(\mu - r_j) \), and substitute that back into the individual demands in (26) giving

\[
W_k w_k^* = \frac{T_k}{\bar{T}} \bar{W} m - \left( \gamma_k - \frac{T_k}{\bar{T}} \right) \Sigma^{-1} \psi - \Sigma^{-1} \phi_k .
\] (31)

Each investor position in the risky assets is now expressed in terms of portfolios, \( m \), \( h \), and \( h_k \). As in the standard CAPM all of these portfolios can be identified without reference to expected rates of return so they can be used to explain them.

An investor who is more risk tolerant than average, \( T_k > \bar{T} \), puts more dollars into the market portfolio undertaking a larger share of the risk in proportion to his wealth exactly as in the standard CAPM. All investors have a “natural” tendency to short the hedge portfolio assuming it is positively correlate with per capita income. Investors whose income is more sensitive to per capita income (higher \( \gamma_k \)) hedge more aggressively, However, hedging is also affected by risk tolerance. The net holding of the hedge portfolio must be zero, and this is the position taken by an investor with average risk tolerance, \( T_k = \bar{T} \), and average income sensitivity, \( \gamma_k = 1 \). Investors with a sufficiently high risk tolerance, \( T_k > \bar{T} \gamma_k \), trade against their natural tendency to hedge. Instead they accept a higher expected rate of return for providing income insurance to others.

The equilibrium expected returns can be determined as in the standard CAPM. Using (30) to evaluate \( \Sigma m \) we have

\[
\begin{align*}
\sigma_m^2 &= m' \Sigma m = \bar{T} \bar{W}^{-1} m' (\mu - r_j) 1 - \bar{W}^{-1} m' \psi = \bar{T} \bar{W}^{-1} (m_m - r_j) - \bar{W}^{-1} \sigma_{m\psi} \\
\sigma_{mh} &= h' \Sigma m = \bar{T} \bar{W}^{-1} h' (\mu - r_j) 1 - \bar{W}^{-1} h' \psi = \bar{T} \bar{W}^{-1} (h_h - r_j) - \bar{W}^{-1} \sigma_{h\psi} .
\end{align*}
\] (32)

Once these two simultaneous equations are solved for \( \bar{T} \) and \( \bar{W} \), the solutions can be introduced back into (30), which can be rearranged to express the assets’ risk premiums

\[
\mu - r_j = \frac{\Sigma m \sigma_{m\psi} - \psi \sigma_{mh}}{\sigma^2_m \sigma_{h\psi} - \sigma_{m\psi} \sigma_{mh}} (m_m - r_j) + \frac{\psi \sigma_{m^2} - \Sigma m \psi \sigma_{m\psi}}{\sigma^2_m \sigma_{h\psi} - \sigma_{m\psi} \sigma_{mh}} (h_h - r_j) .
\] (33)

The two fractions will be recognized as the multiple regression coefficients of the security returns regressed on the returns on the market and portfolio \( h \) using per capita income \( \bar{y} \) as an instrumental variable for the latter. Because the hedge portfolio is constructed to have the maximum possible correlation with per capita income, the covariance of each asset’s return with the hedge portfolio is proportional to its covariances with per capita income,

\[
\Sigma h = \Sigma \Sigma^{-1} \psi = \frac{\psi}{\eta} \Rightarrow \sigma_{m\psi} = \eta \sigma_{mh} \quad \sigma_{h\psi} = \eta \sigma_{h}^2 .
\] (34)

Using (34), the relation with per capita income can all be eliminated, and the expected rate of

---

6 For exponential utility with normally distributions, expected utility is \(-a \mathbb{E}[\bar{W}] + \frac{1}{2} a^2 \text{var}[ar{W}]\) so \( T = 1/a \).
return relation in (33) can be rewritten as

\[
\mu - r_f \mathbf{1} = \frac{\sum \mathbf{m} \sigma_m^2}{\sigma_m^2} (\mu_m - r_f) + \frac{\sum \mathbf{h} \sigma_h^2}{\sigma_h^2} (\mu_h - r_f) \\
= \beta_m (\mu_m - r_f) + \beta_h (\mu_h - r_f)
\]

(35)

where the two betas are now the OLS regression coefficients from the multiple regressions

\[
\tilde{r} = \alpha + \beta_m \tilde{r}_m + \beta_h \tilde{r}_h + \tilde{e}
\]

(36)

with no instrumental variable.

The size of the two portfolio risk premiums can be determined using (32). Rather than solving the two equations to eliminate \(\bar{T}\) and \(\bar{W}\), solve for the risk premiums

\[
\mu_m - r_f = \bar{T}^{-1}[\bar{W} \sigma_m^2 + \sigma_{m^F}] \quad \text{and} \quad \mu_h - r_f = \bar{T}^{-1}[\bar{W} \sigma_h^2 + \sigma_{h^F}].
\]

(37)

These values can be estimated or calibrated for a specific model. Note that the market risk premium could be negative in this model if the return on the market has a strong negative correlation with income. In practice, this correlation is positive. The hedge portfolio’s risk premium can also be negative, but again in practice the premium is positive for a portfolio positively correlated with per capita income.

The market betas in (35) are not the standard CAPM betas. They come from the multiple regression in (36). If per capita income is uncorrelated with the return on the market, then the vector \(\beta_m\) does hold the standard CAPM betas. But even then the CAPM pricing relation will not hold. From (34) if \(\sigma_{m^F} = 0\), then \(\sigma_{mh} = 0\), and the assets’ risk premiums are

\[
\mu - r_f \mathbf{1} = \frac{\sum \mathbf{m}}{\sigma_m^2} (\mu_m - r_f) + \frac{\sum \mathbf{h}}{\sigma_h^2} (\mu_h - r_f).
\]

(38)

So the CAPM pricing result would be valid when the market and per capita income are uncorrelated only if the hedge portfolio has no risk premium as well, and from (37), this is true only if \(\sigma_{m^F} = 0\). Because the hedge portfolio is maximally correlated with per capita income, \(\sigma_{h^F}\) can be zero only if every asset is uncorrelated with per capita income. In this case the CAPM holds because there can be no per capita income hedging by investors. Investors will still hedge the specific residual income risk, but that hedging always nets to zero.

Pricing can also be expressed in terms of the market portfolio and income ignoring the hedge portfolio.\(^7\) Use the first equation in (32) to solve for \(\bar{T}\), introduce that back into (30), and solve for the vector of premiums

\[
(\mu - r_f \mathbf{1}) = [\mathbf{\Sigma} m + \psi \bar{W}^{-1}] \bar{T}^{-1} \bar{W} = \frac{\bar{W} \Sigma m + \psi}{\bar{W} \sigma_m^2 + \sigma_{m^F}} (\mu_m - r_f).
\]

(39)

Per capita wealth appears in this relation due to the asymmetric treatment of returns and income. Returns on assets returns are unit-free while income is measured in dollars. To make the two components comparable, we can measure the return on the nonmarketable assets. Let \(\bar{Y}\) be the value at time 0 of the per capita wages at time 1. The return on the non-marketable wages is \(\bar{Y}/Y\).\(^8\) Equation (39) can now be written in the more symmetric fashion

\[\text{The risk premium on the hedge portfolio is not explicitly required because it is proportional to that on the market. From (37), the relative magnitude of the two risk premiums is } \mu_h - r_f = (\mu_m - r_f)/[(\bar{W} \sigma_{mh} + \sigma_{h^F})/\bar{W} \sigma_m^2 + \sigma_{m^F}]].\]

\[\text{In a multiperiod model, the time-1 payoff includes the time-1 human capital so the rate of return is } (\Delta \bar{Y} + \bar{Y})/\bar{Y}.\]
\[(\mu - r_f) = \frac{\mathbf{W} \Sigma \mathbf{m} + \mathbf{W} \mathbf{\psi}_{\mathbf{m}, \mathbf{r}}}{\mathbf{W} \sigma_m^2 + \mathbf{W} \mathbf{\sigma}_{m, \mathbf{r}}^2}(\mu_m - r_f)\]  \hspace{1cm} (40)

where \(\mathbf{\psi}_{\mathbf{m}, \mathbf{r}}\) is the vector of covariances of asset returns with the return on non-marketable assets and \(\mathbf{\sigma}_{m, \mathbf{r}} = \mathbf{m} \mathbf{\psi}_{\mathbf{m}, \mathbf{r}}\) is the covariance of the market return with the return on non-marketable assets.

Even if the covariances are constant, this measure of risk will change over time when the relative values of physical and human capital do. One of the most valuable non-marketable assets is human capital. Estimation of the stock of human capital is difficult; however, Bowley’s Law\(^9\) states the wages are a constant fraction of income, which might indicate that \(\bar{Y}\) and \(\bar{W}\) are also in constant proportion. Nevertheless, the law is more of a stylized fact that runs counter to the evidence. The Bureau of Labor Statistics shows that the labor share of income in the U.S. has decreased from around 65% in 1950 to about 59% in 2010. On the other hand, the fraction of adults (25 years or older) who had completed high school rose from 34% to over 80% during that same period, and those with college degrees increased from 6% to over 25% so we might expect per capita human capital had increased.

Other state variables that affect utility broadly across most investors can be handled in a similar fashion. A portfolio that hedges risks in this variable will form a portion of the optimal portfolio for those investors and the risk-premium on this portfolio will determine, in part, the risk premiums on all assets. One example is residential real estate. The value of a home has both idiosyncratic and common components. We will see in later multiperiod models that this is also true even for variables that do not enter the direct utility function provided they do affect how an investor trades.

### Pricing Models Based on Higher Moments

Because mean-variance analysis and the CAPM are now so well known it is instructive to look at a generalization to help understand its properties more generally. An obvious extension is to suppose that all portfolios are characterized by their first three moments so that derived utility can be expressed as \(V(\mu, \sigma^2, \kappa^3)\) where \(\kappa^3\) is the third central moment of the portfolio’s return.\(^{10}\)

Start with optimal portfolio, \(p\), for some investor, and consider borrowing a fraction \(w\) of the portfolio value and investing it in asset \(i\). The mean, variance, and the third moment of the new portfolio are now

\[
(\mu, \sigma^2, \kappa^3) = (\mu_p + w(\mu_i - r_f), \sigma_p^2 + 2w\sigma_{ip} + w^2\sigma_i^2, \kappa^3 = \kappa_{ppp} + 3w\kappa_{ip} + 3w^2\kappa_{ip} + w^3\kappa_{iii})
\]

where \(\kappa_{ijk} = \mathbb{E}[(r_i - \bar{r})(r_j - \bar{r})(r_k - \bar{r})]\) is any one of the possible skewness or co-skewness moments. The change in utility for the small addition, \(dw\), of asset \(i\) to the portfolio is

\[
\frac{dV}{dw} = V_1 \frac{\partial \mu}{\partial w} + V_2 \frac{\partial \sigma^2}{\partial w} + V_3 \frac{\partial \kappa^3}{\partial w} = V_1(\mu_i - r_f) + V_2(\sigma_{ip} + w\sigma_i^2) + V_3(3(\kappa_{ipp} + 2w\kappa_{ipp} + w^2\kappa_{iii})).
\]  \hspace{1cm} (42)

This derivative must be zero at the optimum \(w = 0\) so

\[
0 = \left. \frac{dV}{dw} \right|_{w=0} = V_1(\mu_i - r_f) + 2V_2\sigma_{ip} + 3V_3\kappa_{ipp} \Rightarrow \mu_i - r_f = -\frac{2V_2}{V_1}\sigma_{ip} - \frac{3V_3}{V_1}\kappa_{ipp}.
\]  \hspace{1cm} (43)

---

\(^9\) Bowley (1937) asserted the constancy of the wage share. Samuelson (1964, p. 736) attached this name to the assertion. It is at odds with classical economic theory in which the shares of capital, labor and other inputs are flexible.

\(^{10}\) The arguments of derived utility function here are mean, variance, and third moment rather than mean and standard deviation as used previously. This serves only to simplify the analysis and expressions.
Expected rates of return are linear in an asset’s covariance with the optimal portfolio and in one of the two co-third moments. That is, only \(\kappa_{ipp}\) matters; the own third moment \(\kappa_{iii}\) as well as the other co-moment, \(\kappa_{iip}\), are irrelevant just as own variance is irrelevant in the CAPM.

This expected rate of return relation holds for any asset or portfolio including the optimal one so

\[
-\frac{2V_2}{V_1} = \frac{\mu_p - r_f}{\sigma_p^2} + \frac{3V_3\kappa_p^3}{V_1\sigma_p^2} \quad \Rightarrow \quad \mu_i - r_f = \frac{\sigma_{ip}}{\sigma_p^2} \left( \mu_p - r_f \right) + \frac{3V_3\kappa_{ipp}^3}{\sigma_p^2} \left[ \frac{\kappa_{ipp}}{\kappa_{ip}} \right].
\]

The effect of co-skewness on the mean is

\[
\frac{\partial \mu_i}{\partial \kappa_{ipp}} = -\frac{3V_3}{V_1}.
\]

If the representative investor has no taste for or against skewness \((V_3 = 0)\), then mean-variance evaluation holds. More generally \(\kappa_{ipp}\) decreases the risk premium if a higher third moment is preferred \((V_3 > 0)\). This is the opposite of the covariance effect under mean-variance analysis because variance is disliked while positive skewness is preferred. The reason covariance is disliked in mean-variance analysis is that increasing the holding of an asset that covaries positively with the portfolio increases the portfolio’s variance. Does increasing the holding of an asset with a positive \(\kappa_{ipp}\) increase the skewness?

The answer to this question is yes. The co-skewness measure is \(\kappa_{ipp} = \mathbb{E}[ (\tilde{r}_i - \tilde{r})(\tilde{r}_p - \tilde{r}_p)^2 ] \) so if \(\kappa_{ipp}\) is positive, asset \(i\) tends to have a smaller return when the return on portfolio \(p\) is near its mean and a larger return when the return on portfolio \(p\) is in either tail. Therefore, adding asset \(i\) to the portfolio tends to increase the portfolio’s return in both tails and decrease it near the mean. This leads to more skewness in the portfolio. Investors will try to buy more of those assets with positive co-skewness bidding their prices up. In equilibrium this reduces their expected return.

The reason only \(\kappa_{ipp}\) is important, and not \(\kappa_{ipp}\) is that \(\kappa_{ipp}\) describes the change in the distribution when a small amount of asset \(i\) is added to the optimal portfolio. The other co-moment, \(\kappa_{iip}\), describes how adding a small amount of the optimal portfolio to asset \(i\) would change the return distribution, but asset \(i\) is not a relevant portfolio at which to start the comparison. Another way to describe this is that the third moment of the portfolio is the weighted sum of this co-moment, \(\kappa_p^3 = \sum \omega_i \kappa_{ipp}\); just as variance is the weighted sum of the covariances, \(\sigma_p^2 = \sum \omega_i \sigma_{ip}\).

This analysis has been based entirely on individual preferences. However, if the individual is the representative investor, then portfolio \(p\) is the market portfolio, and co-skewness with the market determines the equilibrium expected rates of return.

\[
\mu_i - r_f = \frac{\sigma_{ip}}{\sigma_m} (\mu_m - r_f) + \frac{3V_3\kappa_{ip}^3}{V_1\sigma_m^2} \left[ \frac{\kappa_{ipp}^3}{\kappa_{ip}} \right].
\]

Once the market portfolio introduced, the other utility ratio, \(V_3/V_1\) can also be eliminated using the zero-beta portfolio. From (44), the expected rate of return on the zero-beta portfolio is \(\mu_z - r_f = -3\kappa_{zmm}V_3/V_1\). So assuming \(\kappa_{zmm} \neq 0\), \(3V_3/V_1\) can be replaced by \(- (\mu_z - r_f) / \kappa_{zmm}\), and

\[
\mu_i - r_f = \frac{\sigma_{ip}}{\sigma_m} (\mu_m - r_f) - \frac{\sigma_{ip}}{\sigma_m} (\mu_z - r_f) \frac{\kappa_{m}^3}{\kappa_{zmm}} + \frac{\kappa_{zmm}}{\kappa_{zmm}} (\mu_z - r_f) .
\]
This is a two-factor model like the CAPM with non-marketable income, but it can also be interpreted as CAPM equilibrium with errors. The final term is an idiosyncratic pricing error or alpha. If expected rates of return are negatively related to $\kappa_{\text{imm}}$ due to a preference for positive skewness as (45) indicates, then $(\mu_x - r_f)/\kappa_{\text{zm}}$ must be negative. So the alphas are negatively related to $\kappa_{\text{imm}}$. In addition, the slope of the security market line is off by $- (\mu_x - r_f) \kappa_m^3 / \kappa_{\text{zm}}$. As noted $(\mu_x - r_f) / \kappa_{\text{zm}}$ should be negative so the slope of the security market line is flatter than predicted by the CAPM when the market displays negative skewness which has been typical. This story is consistent with the empirical findings, though this hardly represents an empirical test of the model.

This analysis assumes that a three-moment pricing model consistent with expected utility maximization exists. One condition to make this true is that the investor has a cubic utility function. In this case expected utility is

$$
\mathbb{E}[u(\bar{W})] = \mathbb{E}[\bar{W} - b\bar{W}^2 + c\bar{W}^3] = \mu_w - b(\sigma_w^2 + \mu_w^2) + c(\kappa_w^3 + 3\mu_w\sigma_w^2 + \mu_w^3)
$$

(48)

which depends only on the first three moments. But cubic utility, like quadratic utility, is only increasing and concave for bounded outcomes. The marginals of the derived utility function are

$$
V'_i(\mu, \sigma^2, \kappa^3) = 1 - b\mu + c\mu^2 \quad V''_i(\mu, \sigma^2, \kappa^3) = -b + 3c\mu.
$$

(49)

So for sufficiently high means, cubic utility either has a negative first derivative ($c < 0$) or a positive second derivative ($c > 0$) and, therefore, displays either a preference for less or for risk.

In addition, even if all investors have cubic utility and homogeneous beliefs, there is no guarantee that the market portfolio will be an optimal portfolio. Cubic utility is not sufficient to ensure the convexity of the set of efficient portfolios. This issue is addressed in the next chapter.

As with mean-variance analysis, a restriction of the distribution of returns can justify three-moment pricing. Suppose the rates of return on the risky assets are

$$
\bar{r} = \mu + c \cdot \bar{z} + \bar{\varepsilon} \quad \text{with} \quad \mathbb{E}[\bar{\varepsilon}] = 0, \mathbb{E}[\bar{\varepsilon}^2] = \Omega, \mathbb{E}[\bar{z}] = 0, \mathbb{E}[\bar{z}^2] = 1, \mathbb{E}[\bar{z}^3] \neq 0
$$

(50)

where $\bar{\varepsilon}$ is a vector of elliptical random variables, and $\bar{z}$ is a mean-0, variance-1 random variable that is independent of the ellipticals. The return on any portfolio is $\mu_p + c_p \bar{z} + \bar{\varepsilon}_p$, where $\mu_p = w'\mu, c_p = w'c$, and $\bar{\varepsilon}_p = w'\bar{\varepsilon}$. The latter is a linear combination of elliptical random variable, and therefore also elliptical. So any portfolio is completely characterized by $\mu_p, c_p$, and $\text{var}[\bar{\varepsilon}_p]$. The portfolio’s second and third central moments are $c_p^2 + \text{var}[\bar{\varepsilon}_p]$ and $c_p^3 \mathbb{E}[\bar{z}^3]$, so knowing the mean and second and third central moments completely describes the portfolio.

Any linear combination of $\bar{\varepsilon}$ has mean zero and is independent of $\bar{z}$ so it is a pure addition to risk in a Rothschild-Stiglitz sense. That means for any portfolio it is always best to minimize the $\bar{\varepsilon}$ risk at any given level of skewness and expected excess rate of return. Because the $\bar{\varepsilon}$ are elliptical, their Rothschild-Stiglitz risk can be measured by variance. The minimum-variance portfolio at a given mean excess return and skewness is also the minimum residual-variance portfolio. It is the unique solution to

$$
\min_w \frac{1}{2} w'\Omega w \quad \text{subject to} \quad w'(\mu - r_f 1) = x, w'c = c.
$$

(51)

The portfolio weights are not constrained to sum to one as the position can be financed by borrowing or lending. The first-order conditions are
\[ 0 = \frac{\partial L}{\partial w} = \Omega w - \lambda (\mu - r_f 1) - \gamma c \quad \Rightarrow \quad w^* = \lambda \Omega^{-1} (\mu - r_f 1) + \gamma \Omega^{-1} c \quad (52) \]

The optimal combinations of risky assets are all spanned by two mutual funds, whose holdings are \( w_1 \propto \Omega^{-1} (\mu - r_f 1) \) and \( w_2 \propto \Omega^{-1} c \). The market portfolio of risky assets must be some combination of the two risky-asset mutual funds so it is also efficient, and our original assumption of using it as a starting point for a portfolio is justified.

The co-moment pricing result can be extended to any number of moments. The resulting description of the risk premium in the mean, covariance and remaining \( N-2 \) central comoments that are of interest is

\[ \mu_i - r_f = \frac{\sigma^p_i}{\sigma^p_p} (\mu_p - r_f) + \sum_{n=3}^N \frac{n V_i}{V_i} \kappa_{p,p}^n \left[ \frac{\sigma^p_{i,p}}{\sigma^p_p} - \frac{\kappa_{i,p}^n}{\kappa_{p,p}^n} \right] \]

where \( \kappa_{i,p}^n \equiv \mathbb{E}[(\tilde{r}_i - \mu_i)(\tilde{r}_p - \mu_p)^{n-1}] \).

The comparative static is

\[ \frac{\partial \mu_i}{\partial \kappa_{i,p}^n} = \frac{n V_i}{V_i} \]

In equilibrium, the \( n \)th co-moment reduces expected returns if the \( n \)th moment is liked and increases them if the \( n \)th moment is disliked.

**The Small Firm Effect**

Most of these models have one thing in common, stocks have expected returns that differ from the CAPM prediction; that is, they have alphas. When borrowing is limited, low beta stocks have positive alphas. Stocks that are shunned for ESG or similar reasons have positive alphas. Stocks that are negatively co-skewed with the market or that are positively correlated with non-marketable income also have high alphas. Any deviation from the CAPM is going to give similar results.

The CAPM, with or without alphas can be used to determine initial prices as follows. Let \( \tilde{x}_i \) and \( v_i \) denote the time-1 cash flow and time-0 value of firm \( i \). Then

\[ \frac{\mathbb{E}[\tilde{x}_i]}{v_i} = 1 + r_f + \frac{\mathbb{E}[\tilde{r}_m - r_f]}{\sigma_m^2} \text{cov}[\tilde{x}_i, \tilde{r}_m] + \alpha_i \quad \Rightarrow \quad v_i = \frac{\mathbb{E}[\tilde{x}_i] - \lambda \text{cov}[\tilde{x}_i, \tilde{r}_m]}{1 + r_f + \alpha_i} \quad (55) \]

where \( \lambda = \mathbb{E}[\tilde{r}_m - r_f]/\sigma_m^2 \) is the price of risk.\(^{12}\)

Now consider two similar firms whose cash flows have the same expectation and same covariance with the market. The firm with the higher alpha will have the smaller value at time zero. This means that whenever size is measured by market value rather than cash flows or earnings, the deviations from the CAPM examined here all tend to make small firms have positive alphas holding other things equal. In fact, this is true for virtually any deviation from the CAPM or from any other pricing model. Ignored features tend to lead to the small firm effect.

---

\(^{12}\) This price of risk is different from the more commonly used market Sharpe ratio, \( \mathbb{E}[\tilde{r}_m - r_f]/\sigma_m \).
Further Notes

Alternate Derivation of the Moment Pricing Model

A somewhat different derivation of the moment pricing model is given here. It shows that the model might hold, at least approximately under more general conditions. It starts with the general pricing relation in equation (32) in the next chapter

$$\mathbb{E}[\tilde{r}_i] - r_j = \frac{\text{cov}[v'(\tilde{r}_m), \tilde{r}_j]}{-\mathbb{E}[v'(\tilde{r}_m)]} \equiv \lambda \text{cov}[v'(\tilde{r}_m), \tilde{r}_j]$$

(56)

where $v$ is the derived utility function defined over returns for which the market portfolio is optimal. Expanding the utility function around the expected rate of return on the market

$$v'(\tilde{r}_m) = v'(\bar{r}_m) + \sum_{n=1}^{1} \frac{1}{n!} v^{(n+1)}(\bar{r}_m)(\tilde{r}_m - \bar{r}_m)^n$$

(57)

So

$$\mathbb{E}[\tilde{r}_i] - r_j = \lambda \left\{ \sum_{n=1}^{1} \frac{1}{n!} v^{(n+1)}(\bar{r}_m) \text{cov}[(\tilde{r}_m - \bar{r}_m)^n, \tilde{r}_j] \right\} \equiv \sum_{n=1}^{1} \gamma_n \text{cov}[(\tilde{r}_m - \bar{r}_m)^n, \tilde{r}_j].$$

(58)

The first term is the standard CAPM relation, though $\gamma_1$ is not necessarily the ratio of the expected excess rate of return on the market divided by the market's variance. The later terms are the co-skewness and higher co-moments because

$$\text{cov}[(\tilde{r}_m - \bar{r}_m)^n, \tilde{r}_j] = \text{cov}[(\tilde{r}_m - \bar{r}_m)^n, \tilde{r}_j - \bar{r}_j]$$

$$= \mathbb{E}[(\tilde{r}_m - \bar{r}_m)^n(\tilde{r}_j - \bar{r}_j)] - \mathbb{E}[(\tilde{r}_m - \bar{r}_m)^n] \mathbb{E}[\tilde{r}_j - \bar{r}_j] = \kappa_{i,m(n)} - 0.$$ 

(59)

If a small number of terms provide a sufficient approximation, the multi-moment pricing model is established; however, the risk premiums $\gamma_n$ are not identified.