Financial Economics - Asset Pricing and Portfolio Selection
Lecture 6

Hongtao Zhou *

6 Beta Pricing Models

In this chapter, we will derive and discuss formulas for expected returns of assets in terms of their covariances or betas with some random variables. These random variables have the interpretation of being the risk factors that are "priced." The formula most widely used in financial practice - specially in corporate finance - is the Capital Asset Pricing Model (CAPM) which asserts that the priced factor is the return on the market portfolio. The model most widely used currently in academic empirical work is the Fama-French-Carhart model (Fama and French (1993), Carhart (1997)) which uses the market (stock index) return and three excess returns as the factors. The last section of the chapter presents the Arbitrage Pricing Theory (APT) which shows that an approximate beta pricing model holds if returns have a factor structure (or an approximate factor structure) and a stochastic discount factor exists.

6.1 Beta Pricing

We say there is a single-factor beta pricing model with factor $\tilde{f}$ if there exist constants α and λ such that for each return $\tilde{R}$ we have

$$E[\tilde{R}] = \alpha + \lambda \frac{\text{cov}(\tilde{f}, \tilde{R})}{\text{var}(\tilde{f})}. \tag{6.1}$$

The ratio $\text{cov}(\tilde{f}, \tilde{R})/\text{var}(\tilde{f})$ is the "beta" of $\tilde{R}$ with respect to $\tilde{f}$ - it is the coefficient in the regression (orthogonal projection) of $\tilde{R}$ on a constant and $\tilde{f}$. If there is a risk-free asset, then taking $\tilde{R} = R_f$ in (6.1) shows that $\alpha = R_f$. More generally, α is the expected value of a return $\tilde{R}$ satisfying $\text{cov}(\tilde{f}, \tilde{R}) = 0$. This return is called a "zero-beta return," and α is the expected zero-beta return. In the absence of a risk-free asset, different beta pricing models can have different expected zero-beta returns.

---

*These notes are based on *Asset Pricing and Portfolio Choice Theory* by Kerry Back. We thank Professor Back for his kindly and enormous support.

We say there is a multi-factor beta pricing model with factors $\tilde{f}_1, ..., \tilde{f}_k$ if there exist constants α and λ, with λ being a k-dimensional vector, such that for each return $\tilde{R}$ we have

$$E[\tilde{R}] = \alpha + \lambda' \Sigma_F^{-1} \mathrm{Cov}(\tilde{F}, \tilde{R}), \tag{6.2}$$

where Σ_F is the (assumed to be invertible) covariance matrix of the vector $\tilde{F} = (\tilde{f}_1, ..., \tilde{f}_k)'$ and $\mathrm{Cov}(\tilde{F}, \tilde{R})$ denotes the column vector of dimension k with i-th element $\mathrm{cov}(\tilde{f}_i, \tilde{R})$. The vector

$$\beta = \Sigma_F^{-1} \mathrm{Cov}(\tilde{F}, \tilde{R})$$

is the vector of multiple regression betas of the return $\tilde{R}$ on the factors.

We say that λ in (6.1) is the "factor risk premium." It defines the extra expected return an asset earns for each unit increase in its beta. Assuming $\lambda > 0$, an investor gets compensated for holding extra risk in the form of a higher expected return when we measure risk by the beta with respect to $\tilde{f}$. Likewise, any λ_j in (6.2), for $j \in \{1, ..., k\}$, is the "risk premium of factor j."

We can always write beta pricing models in terms of covariances instead of betas: (6.1) is equivalent to

$$E[\tilde{R}] = \alpha + \theta \mathrm{cov}(\tilde{f}, \tilde{R}), \tag{6.3}$$

where

$$\theta = \frac{\lambda}{\mathrm{var}(\tilde{f})}, \tag{6.4}$$

and (6.2) is equivalent to

$$E[\tilde{R}] = \alpha + \theta' \mathrm{Cov}(\tilde{F}, \tilde{R}), \tag{6.5}$$

where

$$\theta = \Sigma_F^{-1} \lambda. \tag{6.6}$$

The number of factors in a beta pricing model is not uniquely determined. Given a k-factor model, we can always use $\lambda' \Sigma_F^{-1} \tilde{F}$ as a single factor. Furthermore, we can always use any stochastic discount factor $\tilde{m}$ as the single systematic risk factor, provided $E[\tilde{m}] \neq 0$. This was shown in (2.19) when there is a risk-free asset, and the same reasoning in general leads to

$$E[\tilde{R}] = \frac{1}{E[\tilde{m}]} - \frac{1}{E[\tilde{m}]} \mathrm{cov}(\tilde{m}, \tilde{R}) \tag{6.7}$$

for any return $\tilde{R}$. Of course, if there is a risk-free asset, then $1/E[\tilde{m}] = R_f$.

6.2 Single-Factor Models with Returns as Factors

If a factor is a return, then its factor risk premium is its ordinary risk premium, treating α as a proxy risk-free return. To see this, suppose there is a single-factor beta pricing model with the factor being a return $\tilde{R}_*$. Then we have

$$E[\tilde{R}] = \alpha + \lambda \frac{\text{cov}(\tilde{R}_*, \tilde{R})}{\text{var}(\tilde{R}_*)}$$

for each return $\tilde{R}$. Because this relation must hold for $\tilde{R} = \tilde{R}_*$, we have

$$E[\tilde{R}_*] = \alpha + \lambda \frac{\text{var}(\tilde{R}_*)}{\text{var}(\tilde{R}_*)} = \alpha + \lambda,$$

implying

$$\lambda = E[\tilde{R}_*] - \alpha. \tag{6.8}$$

There is a beta pricing model with a return as the single factor if and only if the return is on the mean-variance frontier and not equal to (i) the global minimum variance return if there is no risk-free asset, or (ii) the risk-free return if there is a risk-free asset. Thus, there exists α such that

$$E[\tilde{R}] = \alpha + \frac{\text{cov}(\tilde{R}_*, \tilde{R})}{\text{var}(\tilde{R}_*)}(E[\tilde{R}_*] - \alpha). \tag{6.9}$$

for each return $\tilde{R}$ if and only if $\tilde{R}_*$ is on the mean-variance frontier and not equal to the minimum variance return (in particular, not equal to R_f if there is a risk-free asset).

The equivalence between beta pricing and frontier returns is established below using the "orthogonal projections" characterization of the frontier. The formulas from the "calculus approach" can be used instead. The vector of covariances of the return of a portfolio π with the asset returns is $\Sigma\pi$. If there is a risk-free asset, then the frontier portfolios are given in (5.9) as $\pi = \delta\Sigma^{-1}(\mu - R_f\mathbf{1})$. Therefore, for a portfolio π on the frontier, the vector of covariances is $\Sigma\pi = \delta(\mu - R_f\mathbf{1})$; i.e., for each asset i,

$$\text{cov}(\tilde{R}_i, \tilde{R}_*) = \delta(\mu_i - R_f),$$

where $\tilde{R}_*$ is the return of π. The return $\tilde{R}_*$ is different from R_f if and only if $\delta \neq 0$, in which case (6.3) holds with $\theta = 1/\delta$. The "calculus approach" to the frontier can also be used to show the converse (beta pricing with a return as the factor implies the return is on the frontier) and to show the equivalence of beta pricing and being on the mean-variance frontier in the absence of a risk-free asset. See Problems 6.1 and 6.2.

Let $\tilde{R}_*$ be a frontier return not equal to the minimum variance return, i.e., $\tilde{R}_* = \tilde{R}_p + b\tilde{e}_p$

with $b \neq b_m$. For any return $\tilde{R}$, we have

$$\begin{aligned}
\text{cov}(\tilde{R}, \tilde{R}_p) &= \frac{1}{E[\tilde{m}_p^2]}\text{cov}(\tilde{R}, \tilde{m}_p) \\
&= \frac{1 - E[\tilde{m}_p]E[\tilde{R}]}{E[\tilde{m}_p^2]} \qquad (6.10) \\
&= E[\tilde{R}_p^2] - E[\tilde{R}_p]E[\tilde{R}], \qquad (6.11)
\end{aligned}$$

using the definition of $\tilde{R}_p$ for the first equality, (6.7) - the beta pricing model with $\tilde{m}_p$ as the factor - for the second, and the definition of $\tilde{R}_p$ and fact 18 from Section 5.10 for the third. Also,

$$\begin{aligned}
\text{cov}(\tilde{R}, \tilde{e}_p) &= \text{cov}(\tilde{R}_p, \tilde{e}_p) + \text{cov}(\tilde{R} - \tilde{R}_p, \tilde{e}_p) \\
&= -E[\tilde{R}_p]E[\tilde{e}_p] + E[\tilde{R} - \tilde{R}_p] - E[\tilde{R} - \tilde{R}_p]E[\tilde{e}_p] \\
&= E[\tilde{R}] - E[\tilde{R}_p] - E[\tilde{R}]E[\tilde{e}_p], \qquad (6.12)
\end{aligned}$$

using for the second equality the orthogonality of $\tilde{R}_p$ and $\tilde{e}_p$ (fact 8 from Section 5.7) and the fact that $\tilde{e}_p$ represents the expectation operator on the space of excess returns. Combining (6.11) and (6.12) gives

$$\text{cov}(\tilde{R}_*, \tilde{R}) = E[\tilde{R}_p^2] - bE[\tilde{R}_p] + (b - bE[\tilde{e}_p] - E[\tilde{R}_p])E[\tilde{R}],$$

and rearranging this yields

$$E[\tilde{R}] = \frac{bE[\tilde{R}_p] - E[\tilde{R}_p^2]}{b - bE[\tilde{e}_p] - E[\tilde{R}_p]} + \frac{1}{b - bE[\tilde{e}_p] - E[\tilde{R}_p]}\text{cov}(\tilde{R}_*, \tilde{R}).$$

Thus, there is a beta pricing model with $\tilde{R}_*$ as the factor, provided

$$b - bE[\tilde{e}_p] - E[\tilde{R}_p] \neq 0,$$

which is equivalent to $b \neq b_m$.

Now suppose there is a beta pricing model with a return $\tilde{R}_* = \tilde{R}_p + b\tilde{e}_p + \tilde{\epsilon}$ as the factor. We will show that $\tilde{\epsilon} = 0$, implying $\tilde{R}_*$ is a frontier return, and $b \neq b_m$. The meaning of a beta pricing model is of course that there is some α and θ such that, for any return $\tilde{R}$,

$$E[\tilde{R}] = \alpha + \theta\text{cov}(\tilde{R}_*, \tilde{R}). \qquad (6.13)$$

Applying (6.13) successively with $\tilde{R} = \tilde{R}_*$, $\tilde{R} = \tilde{R}_p + b\tilde{e}_p$ and $\tilde{R} = \tilde{R}_p$ gives

$$
\begin{aligned}
E[\tilde{R}_p] + bE[\tilde{e}_p] &= \alpha + \theta\text{var}(\tilde{R}_p + b\tilde{e}_p) + \theta\text{var}(\tilde{\epsilon}), && (6.14)\\
E[\tilde{R}_p] + bE[\tilde{e}_p] &= \alpha + \theta\text{var}(\tilde{R}_p + b\tilde{e}_p), && (6.15)\\
E[\tilde{R}_p] &= \alpha + \theta\text{var}(\tilde{R}_p) + \theta b\text{cov}(\tilde{R}_p, \tilde{e}_p). && (6.16)
\end{aligned}
$$

We used $E[\tilde{\epsilon}] = 0$ for the first equality and $\text{cov}(\tilde{R}_p, \tilde{\epsilon}) = \text{cov}(\tilde{e}_p, \tilde{\epsilon}) = 0$ (facts 11-13 from Section 5.7) for all three. Subtracting (6.15) from (6.14) yields $\text{var}(\tilde{\epsilon}) = 0$; therefore, $\tilde{\epsilon} = 0$. Subtracting (6.16) from (6.15) yields

$$
\begin{aligned}
bE[\tilde{e}_p] &= \theta b\text{cov}(\tilde{R}_p, \tilde{e}_p) + \theta b^2\text{var}(\tilde{e}_p)\\
&= -\theta bE[\tilde{R}_p]E[\tilde{e}_p] + \theta b^2 E[\tilde{e}_p](1 - E[\tilde{e}_p]),
\end{aligned}
$$

using facts 2 and 17 from Sections 5.7 and 5.8 for the second equality. Now, either $b = 0$ or

$$1 = -\theta(E[\tilde{R}_p] - b(1 - E[\tilde{e}_p])),$$

implying $E[\tilde{R}_p] - b(1 - E[\tilde{e}_p]) \neq 0$. In either case, $b \neq b_m$.

6.3 The Capital Asset Pricing Model

The CAPM states that (6.9) holds for $\tilde{R}_*$ equal to the market return $\tilde{R}_m$. Specifically, the CAPM states that there exists a constant α such that

$$E[\tilde{R}] = \alpha + \frac{\text{cov}(\tilde{R}, \tilde{R}_m)}{\text{var}(\tilde{R}_m)}(E[\tilde{R}_m] - \alpha) \tag{6.17}$$

for each return $\tilde{R}$. As established in the previous section, this is equivalent to the market return being on the mean-variance frontier and not equal to (i) the global minimum variance return if there is no risk-free asset, or (ii) the risk-free return if there is a risk-free asset.

The market return $\tilde{R}_m$ will be on the mean-variance frontier if each investor's optimal portfolio is on the mean-variance frontier. Moreover, each investor will choose a portfolio on the mean-variance frontier if he has no end-of-period endowment and if either (i) he has quadratic utility or (ii) returns belong to the class of separating distributions, which includes elliptical distributions and in particular normal distributions. See Sections 2.6 and 3.8.

Regarding the caveats (i) and (ii) above: Clearly, the market return cannot equal the risk-free return if the market portfolio is risky. We can also rule out the market portfolio being the global minimum variance portfolio when there is no risk-free asset. First, note that if all investors hold portfolios on the efficient part of the mean-variance frontier, then the market portfolio, being a (wealth-weighted) convex combination of investors' portfolios, must be on the efficient part of the frontier. It can equal the global minimum variance portfolio only if all investors hold the global minimum variance portfolio. However, the mean-variance tradeoff is infinite at the global minimum variance portfolio, so an investor will not hold the global minimum variance portfolio unless he is infinitely risk averse.

If investors have end-of-period endowments that are spanned by the asset payoffs, then the CAPM is still true, with an appropriate definition of the market return, when investors have quadratic utility or the returns belong to the class of separating distributions. Being spanned by the asset payoffs means that the endowment $\tilde{y}_h$ of each investor h satisfies

$$\tilde{y}_h = \sum_{i=1}^{n} \psi_{hi} \tilde{x}_i$$

for some ψ_{hi}. This model is equivalent to one in which each investor has no end-of-period endowment but is endowed with an additional ψ_{hi} shares of asset i. If the CAPM holds in the equivalent economy, then it holds in the economy with end-of-period endowments, interpreting the market return as $\tilde{R}_m = \tilde{w}_m / p_m$, where

$$\begin{aligned} \tilde{w}_m &= \sum_{i=1}^{n} \sum_{h=1}^{H} (\bar{\theta}_{hi} + \psi_{hi}) \tilde{x}_i = \sum_{i=1}^{n} \bar{\theta}_i \tilde{x}_i + \sum_{h=1}^{H} \tilde{y}_h, \\ p_m &= \sum_{i=1}^{n} \sum_{h=1}^{H} (\bar{\theta}_{hi} + \psi_{hi}) p_i. \end{aligned}$$

Thus, the market return is total end-of-period wealth, including endowments, divided by the date 0 value of total end-of-period wealth. Of course, this market return is not directly observable, which makes it difficult to empirically test the model.

If investors have end-of-period endowments that are not spanned by the asset payoffs, which is surely the case in reality, then the CAPM will still hold, in a modified form, if (i) investors have quadratic utility or (ii) endowments and asset payoffs are joint normally distributed. The modified version of the CAPM is that there exist constants α and λ such that

$$E[\tilde{R}] = \alpha + \lambda \frac{\text{cov}(\tilde{R}, \tilde{R}_m)}{\text{var}(\tilde{R}_m)} \tag{6.18}$$

for each return $\tilde{R}$, where $\tilde{w}_m$ is end-of-period market wealth, i.e.,

$$\tilde{w}_m = \sum_{i=1}^{n} \bar{\theta}_i \tilde{x}_i + \sum_{h=1}^{H} \tilde{y}_h.$$

Again, α is called the expected zero-beta return and, by the same reasoning as before, it must equal the risk-free return if there is a risk-free asset. Condition (6.18) is established below.

If some of the individual endowments $\tilde{y}_h$ are not spanned by the assets, it could still happen that $\sum_{h=1}^H \tilde{y}_h$ is spanned, in which case $\tilde{w}_m$ is spanned, i.e., is the payoff of some portfolio. Assuming the law of one price, this portfolio has a unique cost p_m, and (6.18) implies (6.17) for the market return $\tilde{R}_m = \tilde{w}_m/p_m$. Of course, this return is again not directly observable. Moreover, if some of the $\tilde{y}_h$ are not spanned, then the usual circumstance will be that $\tilde{w}_m$ is not spanned. In this case, there is no unique price p_m for $\tilde{w}_m$, the return $\tilde{R}_m$ is not well defined, and (6.18) is the most that can be said.

We will show that (6.18) holds when investors have quadratic utility or when endowments and asset payoffs are joint normally distributed. Suppose each investor h has quadratic utility $u_h(w) = \zeta_h w - \frac{1}{2}w^2$ and positive marginal utility at the optimal wealth $\tilde{w}_h$, i.e., $\tilde{w}_h < \zeta_h$ in each state of the world. The first-order condition for portfolio choice is that marginal utility at the optimal wealth is proportional to a stochastic discount factor, so

$$\zeta_h - \tilde{w}_h = \gamma_h \tilde{m}_h$$

for some constant γ_h and stochastic discount factor $\tilde{m}_h$, where $\tilde{w}_h$ denotes the optimal end-of-period wealth of investor h. Adding across investors gives

$$\zeta - \tilde{w}_m = \gamma \tilde{m},$$

where

$$\zeta = \sum_{h=1}^H \zeta_h, \quad \gamma = \sum_{h=1}^H \gamma_h, \quad \tilde{m} = \frac{\sum_{h=1}^H \gamma_h \tilde{m}_h}{\sum_{h=1}^H \gamma_h}.$$

Moreover, $\tilde{m}$ is a stochastic discount factor. The assumption of strictly positive marginal utility implies $\tilde{m} > 0$ in each state of the world, so $E[\tilde{m}] \neq 0$. Substituting $\tilde{m} = (\zeta - \tilde{w}_m)/\gamma$ in (6.7) produces (6.18).

Now consider the joint normality hypothesis. Assume investors' utility functions are strictly monotone and twice continuously differentiable. Substituting $u_h'(\tilde{w}_h) = \gamma_h \tilde{m}_h$ in (6.7) implies

$$E[\tilde{R}] = \frac{\gamma_h}{E[u_h'(\tilde{w}_h)]} - \frac{1}{E[u_h'(\tilde{w}_h)]}\text{cov}(u_h'(\tilde{w}_h), \tilde{R}).$$

Stein's Lemma [needs a reference] and the joint normality hypothesis imply

$$\text{cov}(u_h'(\tilde{w}_h), \tilde{R}) = E[u_h''(\tilde{w}_h)]\text{cov}(\tilde{w}_h, \tilde{R}).$$

Therefore,

$$E[\tilde{R}] = \alpha_h + \theta_h \text{cov}(\tilde{w}_h, \tilde{R}). \tag{6.19}$$

where

$$\alpha_h = \frac{\gamma_h}{E[u_h'(\tilde{w}_h)]} \quad \text{and} \quad \theta_h = -\frac{E[u_h''(\tilde{w}_h)]}{E[u_h'(\tilde{w}_h)]}.$$

Dividing both sides of (6.19) by θ_h and adding over investors produces

$$\left(\sum_{h=1}^{H} \frac{1}{\theta_h}\right) E[\tilde{R}] = \sum_{h=1}^{H} \frac{\alpha_h}{\theta_h} + \text{cov}(\tilde{w}_m, \tilde{R}),$$

implying

$$E[\tilde{R}] = \alpha + \theta \text{cov}(\tilde{w}_m, \tilde{R}), \tag{6.20}$$

where

$$\alpha = \sum_{h=1}^{H} \frac{\alpha_h}{\theta_h} \bigg/ \sum_{h=1}^{H} \frac{1}{\theta_h} \quad \text{and} \quad \theta = 1 \bigg/ \sum_{h=1}^{H} \frac{1}{\theta_h}.$$

6.4 Returns and Excess Returns as Factors

It was shown in Sect. 6.2 that, in a single-factor model, if the factor is a return, then its factor risk premium is its ordinary risk premium, treating α as a proxy risk-free return. The same is true of any factor in a multi-factor model. If a factor is an excess return, then the factor risk premium is simply the expected value of the factor.

Consider a k-factor model and suppose for any $j \in \{1, ..., k\}$ that factor j is a return. Substituting $\tilde{R} = \tilde{f}_j$ in (6.2) gives us

$$E[\tilde{f}_j] = \alpha + \lambda' \Sigma_F^{-1} \text{Cov}(\tilde{F}, \tilde{f}_j). \tag{6.21}$$

The vector $\text{Cov}(\tilde{F}, \tilde{f}_j)$ is the j-th column of the matrix Σ_F. Therefore,

$$\Sigma_F^{-1} \text{Cov}(\tilde{F}, \tilde{f}_j)$$

is j-th column of the identity matrix, meaning that it has a 1 in the j-th place and 0 elsewhere. Hence, $\lambda' \Sigma_F^{-1} \text{Cov}(\tilde{F}, \tilde{f}_j) = \lambda_j$, and (6.21) shows that $\lambda_j = E[\tilde{f}_j] - \alpha$.

Now suppose that factor j is an excess return. For any return $\tilde{R}$, $\tilde{R} + \tilde{f}_j$ is also a return, so we have

$$\begin{aligned} E[\tilde{R}] &= \alpha + \lambda' \Sigma_F^{-1} \text{Cov}(\tilde{F}, \tilde{R}), \\ E[\tilde{R} + \tilde{f}_j] &= \alpha + \lambda' \Sigma_F^{-1} \text{Cov}(\tilde{F}, \tilde{R}) + \lambda' \Sigma_F^{-1} \text{Cov}(\tilde{F}, \tilde{f}_j). \end{aligned}$$

Subtracting the first equation from the second gives

$$E[\tilde{f}_j] = \lambda' \Sigma_F^{-1} \text{Cov}(\tilde{F}, \tilde{f}_j),$$

which means that $E[\tilde{f}_j] = \lambda_j$.

6.5 Projecting Factors on Returns and Excess Returns

In any beta pricing model, any or all of the factors can be replaced by returns or excess returns or a combination thereof. For each factor, the return (or excess return) that can be used as a substitute for the factor is the one that has maximum correlation with the factor, and it is obtained by orthogonal projection of the factor on the space of returns (or the space of excess returns) and a constant. It is called the "factor-mimicking return" (or "factor-mimicking excess return").

Substituting the factor-mimicking return does not change the expected zero-beta return, but substituting the factor-mimicking excess return does generally change the expected zero-beta return, when there is no risk-free asset. Either substitution will generally change all of the multiple regression betas of a return on the set of factors and change all of the factor risk premia. If the factor model is written in terms of covariances instead of betas, as in (6.5), then substituting the factor-mimicking return for a factor will change only the coefficient θ_j on the factor $\tilde{f}_j$ being replaced. Substituting the factor-mimicking excess return will generally change both the coefficient θ_j and the expected zero-beta return.

Consider a k-factor model, and let

$$\tilde{x} = \gamma + \beta \tilde{R}_*$$

denote the orthogonal projection of a factor $\tilde{f}_j$ on the span of a constant and the returns. For any return $\tilde{R}$, we have

$$\text{cov}(\tilde{R}, \tilde{f}_j) = \text{cov}(\tilde{R}, \tilde{x}) = \beta \text{cov}(\tilde{R}, \tilde{R}_*),$$

so we can substitute in (6.5) to obtain

$$E[\tilde{R}] = \alpha + \theta_j \beta \text{cov}(\tilde{R}, \tilde{R}_*) + \sum_{i \neq j} \theta_i \text{cov}(\tilde{R}, \tilde{f}_i)$$

for any return $\tilde{R}$. This shows that there is a k-factor beta pricing model with $\tilde{R}_*$ replacing $\tilde{f}_j$.

The absolute value of the correlation of the return $\tilde{R}_*$ with $\tilde{f}_j$ is the maximum over all returns (this is a generic property of orthogonal projections). We can see this by computing the correlation of any return $\tilde{R}$ with $\tilde{f}_j$ as

$$\begin{aligned}
\text{corr}(\tilde{f}_j, \tilde{R}) &= \frac{\text{cov}(\tilde{f}_j, \tilde{R})}{\text{stdev}(\tilde{f}_j)\text{stdev}(\tilde{R})} \\
&= \frac{\beta\text{cov}(\tilde{R}_*, \tilde{R})}{\text{stdev}(\tilde{f}_j)\text{stdev}(\tilde{R})} \\
&= \frac{\beta\text{corr}(\tilde{R}_*, \tilde{R})\text{stdev}(\tilde{R}_*)}{\text{stdev}(\tilde{f}_j)}.
\end{aligned}$$

Thus, the correlation of $\tilde{f}_j$ with $\tilde{R}$ depends on $\tilde{R}$ only via the correlation of $\tilde{R}$ with $\tilde{R}_*$ and the absolute value is maximized at $\text{corr}(\tilde{R}_*, \tilde{R}_*) = 1$.

Substituting excess returns for factors is very similar. Now let

$$\tilde{x} = \gamma + \beta\tilde{e}_*$$

denote the projection of a factor $\tilde{f}_j$ on a constant and the space of excess returns. By the same reasoning as in the previous paragraph, $\tilde{e}_*$ is the excess return having maximum correlation with $\tilde{f}_j$. The residual $\tilde{f}_j - \tilde{x}$ is orthogonal to a constant (has zero mean) and to each excess return, so it is uncorrelated with each excess return. This implies

$$\text{cov}(\tilde{e}, \tilde{f}_j) = \text{cov}(\tilde{e}, \tilde{x}) = \beta\text{cov}(\tilde{e}, \tilde{e}_*)$$

for each excess return $\tilde{e}$. Choose an arbitrary return and call it $\tilde{R}_0$. Write any return $\tilde{R}$ as $\tilde{R} = \tilde{R}_0 + (\tilde{R} - \tilde{R}_0)$. Then we have

$$\begin{aligned}
\text{cov}(\tilde{R}, \tilde{f}_j) &= \text{cov}(\tilde{R}_0, \tilde{f}_j) + \text{cov}(\tilde{R} - \tilde{R}_0, \tilde{f}_j) \\
&= \text{cov}(\tilde{R}_0, \tilde{f}_j) + \beta\text{cov}(\tilde{R} - \tilde{R}_0, \tilde{e}_*) \\
&= \text{cov}(\tilde{R}_0, \tilde{f}_j) + \beta\text{cov}(\tilde{R}, \tilde{e}_*) - \beta\text{cov}(\tilde{R}_0, \tilde{e}_*) \\
&= \text{cov}(\tilde{R}_0, \tilde{f}_j - \tilde{x}) + \beta\text{cov}(\tilde{R}, \tilde{e}_*).
\end{aligned}$$

Thus, for any return $\tilde{R}$, we have a k-factor beta pricing model:

$$E[\tilde{R}] = \alpha + \text{cov}(\tilde{R}_0, \tilde{f}_j - \tilde{x}) + \theta_j\beta\text{cov}(\tilde{R}, \tilde{e}_*) + \sum_{i \neq j} \theta_i \text{cov}(\tilde{R}, \tilde{f}_i).$$

This shows that the expected zero-beta return is altered by the covariance of the residual $\tilde{f}_j - \tilde{x}$ with a return $\tilde{R}_0$. This covariance is the same for every return $\tilde{R}_0$, due to the residual being orthogonal to excess returns.

If there is a risk-free asset, then we can take $\tilde{R}_0 = R_f$, which shows that $\text{cov}(\tilde{R}_0, \tilde{f}_j - \tilde{x}) = 0$, which must be the case, because the unique expected zero-beta return is R_f when there is a risk-free asset.

6.6 Beta Pricing and Stochastic Discount Factors

There is a beta pricing model with respect to some factors with the expected zero-beta return being nonzero if and only if there is a stochastic discount factor $\tilde{m}$ that is an affine function of the factors with $E[\tilde{m}] \neq 0$. This is true in both single-factor and multi-factor models.

In a single-factor model (6.3) with $\alpha \neq 0$,

$$\tilde{m} = \frac{1}{\alpha} - \frac{\theta}{\alpha}(\tilde{f} - E[\tilde{f}]). \tag{6.22}$$

is a stochastic discount factor. As an example, one can compute a stochastic discount factor supporting the CAPM by taking $\tilde{f}$ to be the market return $\tilde{R}_m$ and $\theta = (E[\tilde{R}_m] - \alpha)/\text{var}(\tilde{R}_m)$ in (6.22).

Notice that $\tilde{m}$ in (6.22) can be negative if $\tilde{f}$ takes certain values. For example, if $\alpha > 0$ and $\theta > 0$, then

$$\tilde{f} > E[\tilde{f}] + \frac{1}{\theta} \quad \Longrightarrow \quad \tilde{m} < 0.$$

There are in general three possibilities: (i) $\tilde{f}$ takes values such that $\tilde{m} \leq 0$ with zero probability; thus, $\tilde{m}$ is strictly positive, (ii) this $\tilde{m}$ is zero or negative with positive probability, but there exists another stochastic discount factor that is strictly positive, or (iii) there is an arbitrage opportunity. If markets are complete, then there is a unique stochastic discount factor, so option (ii) is ruled out. If we want to assume markets are complete and there are no arbitrage opportunities, then we must assume that the range of $\tilde{f}$ is restricted appropriately.

This observation about a single factor $\tilde{f}$ having a restricted range when there are complete markets and no arbitrage opportunities applies also to multiple factors. When a stochastic discount factor is an affine function of $\tilde{f}_1, ..., \tilde{f}_k$, it will be zero or negative for certain values of $\tilde{f}_1, ..., \tilde{f}_k$. If these occur with positive probability, then the stochastic discount factor is not strictly positive. However, if markets are incomplete, there may be another stochastic discount factor that is strictly positive.

If the factors are returns and a stochastic discount factor is an affine function of the returns, then the stochastic discount factor is by definition spanned by a constant and the returns. This implies that the stochastic discount factor is one of the projections $\tilde{m}_{\nu p}$ defined in Section 4.6. If there is a risk-free asset, then, as remarked before, $\tilde{m}_{\nu p}$ equals $\tilde{m}_p$ defined in Section 4.5.

Let $\tilde{F} = (\tilde{f}_1, ..., \tilde{f}_k)'$ and suppose that $\tilde{m} = a + b'\tilde{F}$ is a stochastic discount factor for some constant a and constant vector b. Assume $E[\tilde{m}] \neq 0$. Then (6.7) implies

$$
\begin{aligned}
E[\tilde{R}] &= \frac{1}{E[\tilde{m}]} - \frac{1}{E[\tilde{m}]}\text{cov}(b'\tilde{F}, \tilde{R}) \\
&= \alpha + \theta' \text{Cov}(\tilde{F}, \tilde{R}),
\end{aligned}
$$

where $\alpha = 1/E[\tilde{m}] \neq 0$ and $\theta = \alpha b$. Therefore there is a k-factor beta pricing model with $\tilde{F}$ as the vector of factors.

To establish the converse, suppose there is a k-factor beta pricing model (6.5) with $\tilde{F} = (\tilde{f}_1, ..., \tilde{f}_k)'$ as the vector of factors and $\alpha \neq 0$. Define

$$
\tilde{m} = \frac{1}{\alpha} - \frac{1}{\alpha}\theta'(\tilde{F} - E[\tilde{F}]). \tag{6.23}
$$

For any return $\tilde{R}$, we have

$$
\begin{aligned}
E[\tilde{m}\tilde{R}] &= \frac{1}{\alpha}E[\tilde{R}] - \frac{1}{\alpha}(E[\theta'\tilde{F}\tilde{R}] - E[\theta'\tilde{F}]E[\tilde{R}]) \\
&= \frac{1}{\alpha}E[\tilde{R}] - \frac{1}{\alpha}\text{cov}(\theta'\tilde{F}, \tilde{R}) \\
&= 1,
\end{aligned}
$$

using the definition of covariance for the second equality and the beta pricing model (6.5) for the third. Thus, $\tilde{m}$ is a stochastic discount factor.

6.7 Arbitrage Pricing Theory

We have used the term "factor" in two different senses in this chapter: in "stochastic discount factor" and for the factors in a beta pricing model. We have shown that these two concepts are closely related. Now, we introduce a third related concept: "factor structure" (or "factor model"). The idea behind a factor model is that common exposure to some systematic risk sources is what causes asset returns to be correlated. The risk of each asset return is assumed to consist of a systematic component and an idiosyncratic component, and the idiosyncratic components are assumed to be uncorrelated across assets. In a diversified portfolio, the risk contributed by the idiosyncratic components should be negligible, due to a law of large numbers effect. Intuitively, investors should hold diversified portfolios and hence only be exposed to the systematic risk sources. It then seems sensible that they would require compensation for holding risk (a risk premium) only for the systematic risks. Hence, the risk premium of each asset should depend only on the asset's exposure to the common risk sources and not on its idiosyncratic risk. Thus, an assumption about the correlations of assets implies a conclusion about the pricing of assets. This is the "Arbitrage Pricing Theory," or "APT."

Let $\tilde{F} = (\tilde{f}_1, ..., \tilde{f}_k)'$ and assume it has a nonsingular covariance matrix Σ_F. Consider returns $\tilde{R}_1, ..., \tilde{R}_n$. By orthogonal projection on the span of the $\tilde{f}_j$ and a constant, we have, for each $i = 1, ..., n$,

$$\tilde{R}_i = E[\tilde{R}_i] + \text{Cov}(\tilde{F}, \tilde{R}_i)'\Sigma_F^{-1}(\tilde{F} - E[\tilde{F}]) + \tilde{\epsilon}_i,$$

where $E[\tilde{\epsilon}_i] = 0$ and $\text{cov}(\tilde{f}_j, \tilde{\epsilon}_i) = 0$ for $j = 1, ..., k$. We say that the returns have a factor structure with $\tilde{f}_1, ..., \tilde{f}_k$ as the factors if $\text{cov}(\tilde{\epsilon}_i, \tilde{\epsilon}_l) = 0$ for $i, l = 1, ..., n$ and $i \neq l$. The part

$$\text{Cov}(\tilde{F}, \tilde{R}_i)'\Sigma_F^{-1}(\tilde{F} - E[\tilde{F}])$$

of the return $\tilde{R}_i$ is called its "systematic risk," and the residual $\tilde{\epsilon}_i$ is called its "idiosyncratic risk." Thus, the definition of a factor structure is that the idiosyncratic risks are uncorrelated across assets.

The Arbitrage Pricing Theory (APT) asserts that when returns have a factor structure, there is (at least an approximate) beta pricing model with $\tilde{f}_1, ..., \tilde{f}_k$ as the factors. Thus, "systematic-risk factors" are "beta-pricing factors." Unlike, for example, the CAPM, which is derived from equilibrium considerations - investor optimization and market clearing - the APT is derived from the factor structure of returns and the absence of arbitrage opportunities. To be more precise, it is derived from the factor structure of returns and the existence of a stochastic discount factor. It does not depend on there being a strictly positive stochastic discount factor.

To gain an understanding of the APT, it is useful to consider first the very special case in which the idiosyncratic risks $\tilde{\epsilon}_i$ are all zero (the returns are spanned by a constant and the factors). In this case, for any stochastic discount factor $\tilde{m}$ with $E[\tilde{m}] \neq 0$, we have

$$\begin{aligned} 1 = E[\tilde{m}\tilde{R}_i] &= E[\tilde{R}_i]E[\tilde{m}] + \text{Cov}(\tilde{F}, \tilde{R}_i)'\Sigma_F^{-1}(E[\tilde{m}\tilde{F}] - E[\tilde{m}]E[\tilde{F}]) \\ &= E[\tilde{R}_i]E[\tilde{m}] + \text{Cov}(\tilde{F}, \tilde{m})'\Sigma_F^{-1}\text{Cov}(\tilde{F}, \tilde{R}_i). \end{aligned} \tag{6.24}$$

Rearranging gives

$$E[\tilde{R}_i] = \frac{1}{E[\tilde{m}]} - \frac{1}{E[\tilde{m}]}\text{Cov}(\tilde{F}, \tilde{m})'\Sigma_F^{-1}\text{Cov}(\tilde{F}, \tilde{R}_i). \tag{6.25}$$

Thus, there is a beta pricing model with

$$\lambda = -\frac{1}{E[\tilde{m}]}\text{Cov}(\tilde{F}, \tilde{m})$$

as the vector of factor risk premia.

Now consider the more interesting case in which $\tilde{\epsilon}_i$ is nonzero. This adds the term $E[\tilde{m}\tilde{\epsilon}_i]$ to the right-hand side of (6.24). Recall that $\tilde{\epsilon}_i$ has mean zero. If it also has a price of zero, in the sense then $E[\tilde{m}\tilde{\epsilon}_i] = 0$, then we obtain (6.25) just as when $\tilde{\epsilon}_i = 0$. In general, by following the algebra above, we deduce from (6.24) that

$$E[\tilde{R}_i] = \frac{1}{E[\tilde{m}]} - \frac{1}{E[\tilde{m}]}\text{Cov}(\tilde{F}, \tilde{m})'\Sigma_F^{-1}\text{Cov}(\tilde{F}, \tilde{R}_i) - \frac{E[\tilde{m}\tilde{\epsilon}_i]}{E[\tilde{m}]}. \tag{6.26}$$

Based on a comparison of (6.25) and (6.26), the term $-E[\tilde{m}\tilde{\epsilon}_i]/E[\tilde{m}]$ is called the "pricing error." Denote it by δ_i.

Why should the pricing errors or equivalently the prices $E[\tilde{m}\tilde{\epsilon}_i]$ be zero? The answer is that the $\tilde{\epsilon}_i$ represent risks that can be "diversified away," because they are uncorrelated with each other and with the factors. Consider a diversified portfolio, for example a portfolio that has $1/n$ of its value in each of the n assets. The variance of the portfolio return has a part coming from the variances and covariances of the factors and a part coming from the variances of the $\tilde{\epsilon}_i$. The latter part is

$$\text{var}\left(\sum_{i=1}^n \frac{1}{n}\tilde{\epsilon}_i\right) = \frac{1}{n^2}\sum_{i=1}^n \text{var}(\tilde{\epsilon}_i) \leq \frac{1}{n} \times \max_{i=1,\ldots,n} \text{var}(\tilde{\epsilon}_i).$$

Thus, the total idiosyncratic risk will be near zero when n is large and the $\tilde{\epsilon}_i$ are bounded risks (say, $\text{var}(\epsilon_i) \leq \sigma^2$ for a constant σ and all i). It seems plausible in this circumstance that an asset's expected return (or risk premium) should not depend on its idiosyncratic risk, meaning $E[\tilde{m}\tilde{\epsilon}_i] = 0$. Equivalently, it seems plausible that, if the $\tilde{\epsilon}_i$ are unimportant in this sense, then there should be a stochastic discount factor $\tilde{m}$ that depends only on (is an affine function of) the systematic risks $\tilde{f}_1, ..., \tilde{f}_k$.

There are two problems with the above argument. First, it may not be possible for all investors to hold well-diversified portfolios (portfolios with negligible idiosyncratic risk) because the market portfolio may not be well diversified - e.g., the first asset may represent a large part of the total market. Second, with only finitely many assets, the idiosyncratic risk of a well-diversified portfolio may be small but it is not zero; thus, there could still be some small risk premia associated with the idiosyncratic risks of assets.

As a result of these issues, the conclusion of the APT is only that if there is a "large" number of assets, then "most" of the pricing errors are "small" and therefore, (6.25) is approximately true for most assets. Somewhat more formally, the APT is as follows. Consider an infinite sequence of returns $\tilde{R}_1, \tilde{R}_2, ...$ Consider a sequence of markets indexed by n with the returns in market n being $\tilde{R}_1, ..., \tilde{R}_n$. Suppose there is a stochastic discount factor $\tilde{m}$ with $E[\tilde{m}] \neq 0$ which is a stochastic discount factor for each market $n = 1, 2, ...$ Then for any real number $\delta > 0$, there are only finitely many assets with pricing errors δ_i for which $|\delta_i| \geq \delta$. Any finite subset of assets is a "small" subset of an infinite set. This is the sense in which "most" assets have small pricing errors (smaller than any arbitrary $\delta > 0$).

A proof of the APT is given below. The assumptions are stronger than necessary. This is discussed further in the end-of-chapter notes. One will observe that the proof is purely mechanical. The economics of the problem is embedded in the assumption that a stochastic discount factor exists for the infinite sequence of returns. The intuition for the APT in terms of the residual risks being diversifiable away and hence earning negligible risk premia really does not appear in the proof. Instead, the condition used is that the variance of the residual risk of a portfolio is

$$\sum_{i=1}^{n} w_i^2 \text{var}(\tilde{\epsilon}_i)$$

where w_i denotes the weight on asset i, and this is bounded by $\sigma^2 \sum_{i=1}^{n} w_i^2$ if $\text{var}(\tilde{\epsilon}_i) \leq \sigma^2$ for each i.

Suppose there is a stochastic discount factor $\tilde{m}$ having a finite variance and $E[\tilde{m}] \neq 0$. Assume $\text{var}(\tilde{\epsilon}_i) \leq \sigma^2$ for each i and $\text{cov}(\tilde{\epsilon}_i, \tilde{\epsilon}_j) = 0$ for $i, j = 1, 2, ...$ and $i \neq j$.

Let ℓ^2 denote the space of sequences $x = (x_1, x_2, ...)$ such that $\sum_{i=1}^{\infty} x_i^2 < \infty$. Define the norm of $x \in \ell^2$ to be

$$\|x\| = \sqrt{\sum_{i=1}^{\infty} x_i^2} \tag{6.27}$$

and the inner product of x and w in ℓ^2 to be

$$\langle x, w \rangle = \sum_{i=1}^{\infty} x_i w_i.$$

With these definitions, ℓ^2 is a Hilbert space, and (see Section 4.9*)

$$\|x\| = \max_{\|w\|=1} |\langle x, w \rangle|. \tag{6.28}$$

Fix for the moment an integer n, and let x denote the sequence given by $x_i = E[\tilde{m}\tilde{\epsilon}_i] = -\delta_i E[\tilde{m}]$ for $i = 1, ..., n$ and $x_i = 0$ for $i > n$. The definition (6.27) gives us

$$\sqrt{\sum_{i=1}^{n} \delta_i^2} = \frac{\|x\|}{|E[\tilde{m}]|}.$$

From (6.28), we have

$$\begin{aligned}
\|x\| &= \max_{\|w\|=1} |\langle x, w\rangle| \\
&= \max_{\|w\|=1} \left| \sum_{i=1}^{n} w_i E[\tilde{m}\tilde{\epsilon}_i] \right| \\
&= \max_{\|w\|=1} \left| E\left[\tilde{m} \sum_{i=1}^{n} w_i \tilde{\epsilon}_i \right] \right| \\
&\leq \sqrt{E[\tilde{m}^2]} \max_{\|w\|=1} \sqrt{E\left[\left(\sum_{i=1}^{n} w_i \tilde{\epsilon}_i \right)^2 \right]} \\
&= \sqrt{E[\tilde{m}^2]} \max_{\|w\|=1} \sqrt{\sum_{i=1}^{n} w_i^2 \operatorname{var}(\tilde{\epsilon}_i)} \\
&= \sqrt{E[\tilde{m}^2]} \max_i \sqrt{\operatorname{var}(\tilde{\epsilon}_i)} \\
&\leq \sigma \sqrt{E[\tilde{m}^2]},
\end{aligned}$$

using the Cauchy-Schwartz inequality in L^2 for the first inequality in the string above and the boundedness of the variances of the $\tilde{\epsilon}_i$ for the second. Because this bound is independent of n, we conclude that $\sum_{i=1}^{\infty} \delta_i^2 < \infty$.