AT THE END OF CHAPTER 5 we summarized empirical evidence indicating that the CAPM beta does not completely explain the cross section of expected asset returns. This evidence suggests that one or more additional factors may be required to characterize the behavior of expected returns and naturally leads to consideration of multifactor pricing models. Theoretical arguments also suggest that more than one factor is required, since only under strong assumptions will the CAPM apply period by period. Two main theoretical approaches exist. The Arbitrage Pricing Theory (APT) developed by Ross (1976) is based on arbitrage arguments and the Intertemporal Capital Asset Pricing Model (ICAPM) developed by Merton (1973a) is based on equilibrium arguments. In this chapter we will consider the econometric analysis of multifactor models.

The chapter proceeds as follows. Section 6.1 briefly discusses the theoretical background of the multifactor approaches. In Section 6.2 we consider estimation and testing of the models with known factors, while in Section 6.3 we develop estimators for risk premia and expected returns. Since the factors are not always provided by theory, we discuss ways to construct them in Section 6.4. Section 6.5 presents empirical results. Because of the lack of specificity of the models, deviations can always be explained by additional factors. This raises an issue of interpreting model violations which we discuss in Section 6.6.

6.1 Theoretical Background

The Arbitrage Pricing Theory (APT) was introduced by Ross (1976) as an alternative to the Capital Asset Pricing Model. The APT can be more general than the CAPM in that it allows for multiple risk factors. Also, unlike the CAPM, the APT does not require the identification of the market portfolio. However, this generality is not without costs. In its most general form
the APT provides an *approximate* relation for expected asset returns with an unknown number of unidentified factors. At this level rejection of the theory is impossible (unless arbitrage opportunities exist) and as a consequence testability of the model depends on the introduction of additional assumptions.\(^1\)

The Arbitrage Pricing Theory assumes that markets are competitive and frictionless and that the return generating process for asset returns being considered is

\[
R_i = a_i + b_i'f + \epsilon_i \tag{6.1.1}
\]

\[
E[\epsilon_i | f] = 0 \tag{6.1.2}
\]

\[
E[\epsilon_i^2] = \sigma_i^2 \leq \sigma^2 < \infty, \tag{6.1.3}
\]

where \(R_i\) is the return for asset \(i\), \(a_i\) is the intercept of the factor model, \(b_i\) is a \((K \times 1)\) vector of factor sensitivities for asset \(i\), \(f\) is a \((K \times 1)\) vector of common factor realizations, and \(\epsilon_i\) is the disturbance term. For the system of \(N\) assets,

\[
\mathbf{R} = \mathbf{a} + \mathbf{Bf} + \mathbf{\epsilon} \tag{6.1.4}
\]

\[
E[\mathbf{\epsilon} | \mathbf{f}] = \mathbf{0} \tag{6.1.5}
\]

\[
E[\mathbf{\epsilon}' | \mathbf{f}] = \mathbf{\Sigma}. \tag{6.1.6}
\]

In the system equation, \(\mathbf{R}\) is an \((N \times 1)\) vector with \(\mathbf{R} = [R_1 \ R_2 \ \cdots \ R_N]'\), \(\mathbf{a}\) is an \((N \times 1)\) vector with \(\mathbf{a} = [a_1 \ a_2 \ \cdots \ a_N]'\), \(\mathbf{B}\) is an \((N \times K)\) matrix with \(\mathbf{B} = [b_1 \ b_2 \ \cdots \ b_N]'\), and \(\mathbf{\epsilon}\) is an \((N \times 1)\) vector with \(\mathbf{\epsilon} = [\epsilon_1 \ \epsilon_2 \ \cdots \ \epsilon_N]'\). We further assume that the factors account for the common variation in asset returns so that the disturbance term for large well-diversified portfolios vanishes.\(^2\)

This requires that the disturbance terms be sufficiently uncorrelated across assets.

Given this structure, Ross (1976) shows that the absence of arbitrage in large economies implies that

\[
\mu \approx \nu \lambda_0 + \mathbf{B} \lambda_K, \tag{6.1.7}
\]

where \(\mu\) is the \((N \times 1)\) expected return vector, \(\lambda_0\) is the model zero-beta parameter and is equal to the riskfree return if such an asset exists, and \(\lambda_K\) is a \((K \times 1)\) vector of factor risk premia. Here, and throughout the chapter,

\(^1\)There has been substantial debate on the testability of the APT. Shanken (1982) and Dybvig and Ross (1985) provide one interesting exchange. Dhrymes, Friend, Gultekin, and Gultekin (1984) also question the empirical relevance of the model.

\(^2\)A large well-diversified portfolio is a portfolio with a large number of stocks with weightings of order \(\frac{1}{N}\).
let $\lambda$ represent a conforming vector of ones. The relation in (6.1.7) is approximate as a finite number of assets can be arbitrarily mispriced. Because (6.1.7) is only an approximation, it does not produce directly testable restrictions for asset returns. To obtain restrictions we need to impose additional structure so that the approximation becomes exact.

Connor (1984) presents a competitive equilibrium version of the APT which has exact factor pricing as a feature. In Connor's model the additional requirements are that the market portfolio be well-diversified and that the factors be pervasive. The market portfolio will be well-diversified if no single asset in the economy accounts for a significant proportion of aggregate wealth. The requirement that the factors be pervasive permits investors to diversify away idiosyncratic risk without restricting their choice of factor risk exposure.

Dybvig (1985) and Grinblatt and Titman (1985) take a different approach. They investigate the potential magnitudes of the deviations from exact factor pricing given structure on the preferences of a representative agent. Both papers conclude that given a reasonable specification of the parameters of the economy, theoretical deviations from exact factor pricing are likely to be negligible. As a consequence empirical work based on the exact pricing relation is justified.

Exact factor pricing can also be derived in an intertemporal asset pricing framework. The Intertemporal Capital Asset Pricing Model developed in Merton (1973a) combined with assumptions on the conditional distribution of returns delivers a multifactor model. In this model, the market portfolio serves as one factor and state variables serve as additional factors. The additional factors arise from investors' demand to hedge uncertainty about future investment opportunities. Breeden (1979), Campbell (1993a, 1996), and Fama (1993) explore this model, and we discuss it in Chapter 8.

In this chapter, we will generally not differentiate the APT from the ICAPM. We will analyze models where we have exact factor pricing, that is,

\[ \mu = \iota \lambda_0 + B \lambda_K. \quad (6.1.8) \]

There is some flexibility in the specification of the factors. Most empirical implementations choose a proxy for the market portfolio as one factor. However, different techniques are available for handling the additional factors. We will consider several cases. In one case, the factors of the APT and the state variables of the ICAPM need not be traded portfolios. In other cases the factors are returns on portfolios. These factor portfolios are called mimicking portfolios because jointly they are maximally correlated with the factors. Exact factor pricing will hold with such portfolios. Huberman and Kandel (1987) and Breeden (1979) discuss this issue in the context of the APT and ICAPM, respectively.
6.2 Estimation and Testing

In this section we consider the estimation and testing of various forms of the exact factor pricing relation. The starting point for the econometric analysis of the model is an assumption about the time-series behavior of returns. We will assume that returns conditional on the factor realizations are IID through time and jointly multivariate normal. This is a strong assumption, but it does allow for limited dependence in returns through the time-series behavior of the factors. Furthermore, this assumption can be relaxed by casting the estimation and testing problem in a Generalized Method of Moments framework as outlined in the Appendix. The GMM approach for multifactor models is just a generalization of the GMM approach to testing the CAPM presented in Chapter 5.

As previously mentioned, the multifactor models specify neither the number of factors nor the identification of the factors. Thus to estimate and test the model we need to determine the factors—an issue we will address in Section 6.4. In this section we will proceed by taking the number of factors and their identification as given.

We consider four versions of the exact factor pricing model: (1) Factors are portfolios of traded assets and a riskfree asset exists; (2) Factors are portfolios of traded assets and there is not a riskfree asset; (3) Factors are not portfolios of traded assets; and (4) Factors are portfolios of traded assets and the factor portfolios span the mean-variance frontier of risky assets. We use maximum likelihood estimation to handle all four cases. See Shanken (1992b) for a treatment of the same four cases using a cross-sectional regression approach.

Given the joint normality assumption for the returns conditional on the factors, we can construct a test of any of the four cases using the likelihood ratio. Since derivation of the test statistic parallels the derivation of the likelihood ratio test of the CAPM presented in Chapter 5, we will not repeat it here. The likelihood ratio test statistic for all cases takes the same general form. Defining $J$ as the test statistic we have

\[ J = - \left( T - \frac{N}{2} - K - 1 \right) \left[ \log |\hat{\Sigma}| - \log |\hat{\Sigma}^*| \right], \]  

(6.2.1)

where $\hat{\Sigma}$ and $\hat{\Sigma}^*$ are the maximum likelihood estimators of the residual covariance matrix for the unconstrained model and constrained model, respectively. $T$ is the number of time-series observations, $N$ is the number of included portfolios, and $K$ is the number of factors. As discussed in Chapter 5, the statistic has been scaled by $(T' - \frac{N}{2} - K - 1)$ rather than the usual $T$ to improve the convergence of the finite-sample null distribution.
6.2. Estimation and Testing

6.2.1 Portfolios as Factors with a Riskfree Asset

We first consider the case where the factors are traded portfolios and there exists a riskfree asset. The unconstrained model will be a K-factor model expressed in excess returns. Define $\mathbf{Z}_t$ as an $(N \times 1)$ vector of excess returns for $N$ assets (or portfolios of assets). For excess returns, the K-factor linear model is:

$$
\mathbf{Z}_t = \mathbf{a} + \mathbf{B}\mathbf{Z}_K + \mathbf{\epsilon}_t
$$

(6.2.2)

$$
E[\mathbf{\epsilon}_t] = \mathbf{0}
$$

(6.2.3)

$$
E[\mathbf{\epsilon}_t\mathbf{\epsilon}_t'] = \mathbf{\Sigma}
$$

(6.2.4)

$$
E[\mathbf{Z}_Kt] = \mu_K,
\quad E[(\mathbf{Z}_Kt - \mu_K)(\mathbf{Z}_Kt - \mu_K)'] = \Omega_K
$$

(6.2.5)

$$
\text{Cov}[\mathbf{Z}_Kt, \mathbf{\epsilon}_t'] = \mathbf{0}.
$$

(6.2.6)

$\mathbf{B}$ is the $(N \times K)$ matrix of factor sensitivities, $\mathbf{Z}_Kt$ is the $(K \times 1)$ vector of factor portfolio excess returns, and $\mathbf{a}$ and $\mathbf{\epsilon}_t$ are $(N \times 1)$ vectors of asset return intercepts and disturbances, respectively. $\mathbf{\Sigma}$ is the variance-covariance matrix of the disturbances, and $\Omega_K$ is the variance-covariance matrix of the factor portfolio excess returns, while $\mathbf{0}$ is a $(K \times N)$ matrix of zeroes. Exact factor pricing implies that the elements of the vector $\mathbf{a}$ in (6.2.2) will be zero.

For the unconstrained model in (6.2.2) the maximum likelihood estimators are just the OLS estimators:

$$
\hat{\mathbf{a}} = \hat{\mu} - \hat{\mathbf{B}}\hat{\mu}_K
$$

(6.2.7)

$$
\hat{\mathbf{B}} = \left[ \sum_{t=1}^{T} (\mathbf{Z}_t - \hat{\mu})(\mathbf{Z}_Kt - \hat{\mu}_K)' \right]^{-1} \left[ \sum_{t=1}^{T} (\mathbf{Z}_Kt - \hat{\mu}_K)(\mathbf{Z}_Kt - \hat{\mu}_K)' \right]
$$

(6.2.8)

$$
\hat{\mathbf{\Sigma}} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{Z}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{Z}_Kt)(\mathbf{Z}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{Z}_Kt)',
$$

(6.2.9)

where

$$
\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t \quad \text{and} \quad \hat{\mu}_K = \frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_Kt.
$$

$^3$See equation (5.3.41) and Jobson and Korkie (1982).
For the constrained model, with \( a \) constrained to be zero, the maximum likelihood estimators are

\[
\hat{\beta}^* = \left[ \sum_{t=1}^{T} Z_t' Z_{Kt} \right] \left[ \sum_{t=1}^{T} Z_{Kt}' Z_{Kt} \right]^{-1}
\]

(6.2.10)

\[
\hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} (Z_t - \hat{\beta}^* Z_{Kt})(Z_t - \hat{\beta}^* Z_{Kt})'.
\]

(6.2.11)

The null hypothesis \( a = 0 \) can be tested using the likelihood ratio statistic \( J \) in (6.2.1). Under the null hypothesis the degrees of freedom of the null distribution will be \( N \) since the null hypothesis imposes \( N \) restrictions.

In this case we can also construct an exact multivariate F-test of the null hypothesis. Defining \( J_1 \) as the test statistic we have

\[
J_1 = \frac{\left( T - N - K \right)}{N} \left[ 1 + \hat{\mu}' \hat{\Omega}_K^{-1} \hat{\mu} \right]^{-1} \hat{\mu}' \hat{\Sigma}^{-1} \hat{\mu},
\]

(6.2.12)

where \( \hat{\Omega}_K \) is the maximum likelihood estimator of \( \Omega_K \),

\[
\hat{\Omega}_K = \frac{1}{T} \sum_{t=1}^{T} (Z_{Kt} - \hat{\mu}_K)(Z_{Kt} - \hat{\mu}_K)'.
\]

(6.2.13)

Under the null hypothesis, \( J_1 \) is unconditionally distributed central \( F \) with \( N \) degrees of freedom in the numerator and \( (T - N - K) \) degrees of freedom in the denominator. This test can be very useful since it can eliminate the problems that can accompany the use of asymptotic distribution theory. Jobson and Korkie (1985) provide a derivation of \( J_1 \).

### 6.2.2 Portfolios as Factors without a Riskfree Asset

In the absence of a riskfree asset, there is a zero-beta model that is a multifactor equivalent of the Black version of the CAPM. In a multifactor context, the zero-beta portfolio is a portfolio with no sensitivity to any of the factors, and expected returns in excess of the zero-beta return are linearly related to the columns of the matrix of factor sensitivities. The factors are assumed to be portfolio returns in excess of the zero-beta return.

Define \( R_t \) as an \((N \times 1)\) vector of real returns for \( N \) assets (or portfolios of assets). For the unconstrained model, we have a \( K \)-factor linear model:

\[
R_t = a + B R_{Kt} + \epsilon_t
\]

(6.2.14)

\[
E[\epsilon_t] = 0
\]

(6.2.15)

\[
E[\epsilon_t \epsilon_t'] = \Sigma
\]

(6.2.16)
6.2. Estimation and Testing

\[ E[R_{Kt}] = \mu_K, \quad E[(R_{Kt} - \mu_K)(R_{Kt} - \mu_K)'] = \Omega_K \quad (6.2.17) \]
\[ \text{Cov}[R_{Kt}, \epsilon_t] = O. \quad (6.2.18) \]

\( B \) is the \((N \times K)\) matrix of factor sensitivities, \( R_{Kt} \) is the \((K \times l)\) vector of factor portfolio real returns, and \( a \) and \( \epsilon_t \) are \((N \times l)\) vectors of asset return intercepts and disturbances, respectively. \( O \) is a \((K \times N)\) matrix of zeroes.

For the unconstrained model in (6.2.14) the maximum likelihood estimators are

\[ \hat{\alpha} = \hat{\mu} - B\hat{\mu}_K \quad (6.2.19) \]
\[ \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (R_t - \hat{\alpha} - \hat{B}R_{Kt})(R_t - \hat{\alpha} - \hat{B}R_{Kt})'. \quad (6.2.20) \]

where

\[ \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} R_t \quad \text{and} \quad \hat{\mu}_K = \frac{1}{T} \sum_{t=1}^{T} R_{Kt}. \]

In the constrained model real returns enter in excess of the expected zero-beta portfolio return \( \gamma_0 \). For the constrained model, we have

\[ R_i = \nu \gamma_0 + B(R_{Kt} - \nu \gamma_0) + \epsilon_t \quad (6.2.22) \]
\[ = (\nu - B \nu) \gamma_0 + B R_{Kt} + \epsilon_t. \]

The constrained model estimators are:

\[ \hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} [R_t - \nu \hat{\gamma}_0 - \hat{B}^*(R_{Kt} - \nu \hat{\gamma}_0)] \]
\[ \times [R_t - \nu \hat{\gamma}_0 - \hat{B}^*(R_{Kt} - \nu \hat{\gamma}_0)]' \quad (6.2.24) \]

\[ \hat{\gamma}_0 = [(\nu - \hat{B}^* \nu) \hat{\Sigma}^{*-1}(\nu - \hat{B}^* \nu)]^{-1} \]
\[ \times [(\nu - \hat{B}^* \nu) \hat{\Sigma}^{*-1}(\hat{\mu} - \hat{B}^* \hat{\mu}_K)]. \quad (6.2.25) \]
The maximum likelihood estimates can be obtained by iterating over (6.2.23) to (6.2.25). \( \beta \) from (6.2.20) and \( \Sigma \) from (6.2.21) can be used as starting values for \( \beta \) and \( \Sigma \) in (6.2.25).

Exact maximum likelihood estimators can also be calculated without iteration for this case. The methodology is a generalization of the approach outlined for the Black version of the CAPM in Chapter 5; it is presented by Shanken (1985a). The estimator of \( \gamma_0 \) is the solution of a quadratic equation. Given \( \gamma_0 \), the constrained maximum likelihood estimators of \( \beta \) and \( \Sigma \) follow from (6.2.23) and (6.2.24).

The restrictions of the constrained model in (6.2.22) on the unconstrained model in (6.2.14) are

\[
a = (I - \beta \epsilon) \gamma_0.
\] (6.2.26)

These restrictions can be tested using the likelihood ratio statistic \( J \) in (6.2.1). Under the null hypothesis the degrees of freedom of the null distribution will be \( N-1 \). There is a reduction of one degree of freedom in comparison to the case with a riskfree asset. A degree of freedom is used up in estimating the zero-beta expected return.

For use in Section 6.3, we note that the asymptotic variance of \( \hat{\gamma}_0 \) evaluated at the maximum likelihood estimators is

\[
\text{Var}[\hat{\gamma}_0] = \frac{1}{T} \left( 1 + (\mu_K - \hat{\gamma}_0 \epsilon)' \Omega_K^{-1} (\mu_K - \hat{\gamma}_0 \epsilon) \right) \\
\times [(I - \hat{\beta} \epsilon)' \hat{\Sigma}^{-1} (I - \hat{\beta} \epsilon)]^{-1}.
\] (6.2.27)

6.2.3 Macroeconomic Variables as Factors

Factors need not be traded portfolios of assets; in some cases proposed factors include macroeconomic variables such as innovations in GNP, changes in bond yields, or unanticipated inflation. We now consider estimating and testing exact factor pricing models with such factors.

Again define \( R_t \) as an (N x 1) vector of real returns for \( N \) assets (or portfolios of assets). For the unconstrained model we have a K-factor linear model:

\[
R_t = a + \beta \epsilon + \epsilon_t
\] (6.2.28)

\[
E[\epsilon_t] = 0
\] (6.2.29)

\[
E[\epsilon_t \epsilon_t'] = \Sigma
\] (6.2.30)

\[
E[f_{Kt}] = \mu_{fK}, \quad E[(f_{Kt} - \mu_{fK}) (f_{Kt} - \mu_{fK})'] = \Omega_K
\] (6.2.31)

\[
\text{Cov}[f_{Kt}, \epsilon_t'] = O.
\] (6.2.32)
6.2. Estimation and Testing

\( \mathbf{B} \) is the \((N \times K)\) matrix of factor sensitivities, \( \mathbf{f}_{Kt} \) is the \((K \times 1)\) vector of factor realizations, and \( \mathbf{a} \) and \( \mathbf{\epsilon}_t \) are \((N \times 1)\) vectors of asset return intercepts and disturbances, respectively. \( \mathbf{O} \) is a \((K \times N)\) matrix of zeroes.

For the unconstrained model in (6.2.14) the maximum likelihood estimators are

\[
\mathbf{\hat{a}} = \hat{\mu} - \mathbf{B} \hat{\mu}_{fK}, \\
\mathbf{\hat{B}} = \left[ \sum_{t=1}^{T} (\mathbf{R}_t - \hat{\mu}_f) (\mathbf{f}_{Kt} - \hat{\mu}_{fK})' \right] \\
\quad \times \left[ \sum_{t=1}^{T} (\mathbf{f}_{Kt} - \hat{\mu}_{fK}) (\mathbf{f}_{Kt} - \hat{\mu}_{fK})' \right]^{-1}, \\
\mathbf{\hat{\Sigma}} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{R}_t - \mathbf{\hat{a}} - \mathbf{B} \mathbf{f}_{Kt})(\mathbf{R}_t - \mathbf{\hat{a}} - \mathbf{B} \mathbf{f}_{Kt})',
\]

where

\[
\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{R}_t \quad \text{and} \quad \hat{\mu}_{fK} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{f}_{Kt}.
\]

The constrained model is most conveniently formulated by comparing the unconditional expectation of (6.2.28) with (6.1.8). The unconditional expectation of (6.2.28) is

\[
\mu = \mathbf{a} + \mathbf{B} \mu_{fK},
\]

where \( \mu_{fK} = \mathbb{E}[\mathbf{f}_{Kt}] \). Equating the right hand sides of (6.1.8) and (6.2.36) we have

\[
a = \nu \lambda_0 + \mathbf{B}(\lambda_K - \mu_{fK}).
\]

Defining \( \gamma_0 \) as the zero-beta parameter \( \lambda_0 \) and defining \( \gamma_1 \) as \((\lambda_K - \mu_{fK})\) where \( \lambda_K \) is the \((K \times 1)\) vector of factor risk premia, for the constrained model, we have

\[
\mathbf{R}_t = \nu \gamma_0 + \mathbf{B} \gamma_1 + \mathbf{B} \mathbf{f}_{Kt} + \mathbf{\epsilon}_t.
\]

The constrained model estimators are

\[
\mathbf{\hat{B}}^* = \left[ \sum_{t=1}^{T} (\mathbf{R}_t - \nu \hat{\gamma}_0) (\mathbf{f}_{Kt} + \hat{\gamma}_1)' \right] \\
\quad \times \left[ \sum_{t=1}^{T} (\mathbf{f}_{Kt} + \hat{\gamma}_1) (\mathbf{f}_{Kt} + \hat{\gamma}_1)' \right]^{-1}
\]

(6.2.39)
\[ \hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} [(R_t - \mu \hat{y}_0) - \hat{\mathbf{B}}^*(\mathbf{f}_{Kt} + \hat{\gamma}_1)] \times [(R_t - \mu \hat{y}_0) - \hat{\mathbf{B}}^*(\mathbf{f}_{Kt} + \hat{\gamma}_1)]' \]  
\[ \hat{\gamma} = \left[ \mathbf{X}^* \hat{\Sigma}^{*-1} \mathbf{X} \right]^{-1} \left[ \mathbf{X}^* \hat{\Sigma}^{*-1} \left( \hat{\mu} - \hat{\mathbf{B}}^* \hat{\mu}_{K*} \right) \right] \]  
(6.2.40)

where in (6.2.41) \( \mathbf{X} \equiv [\mathbf{L} \hat{\mathbf{B}}^*] \) and \( \mathbf{\gamma} \equiv [\gamma_0 \gamma_1]' \).

The maximum likelihood estimates can be obtained by iterating over (6.2.39) to (6.2.41). \( \mathbf{B} \) from (6.2.34) and \( \hat{\Sigma} \) from (6.2.35) can be used as starting values for \( \mathbf{B} \) and \( \mathbf{\Sigma} \) in (6.2.41).

The restrictions of (6.2.38) on (6.2.28) are

\[ \mathbf{a} = \mu \gamma_0 + \mathbf{B} \gamma_1. \]  
(6.2.42)

These restrictions can be tested using the likelihood ratio statistic \( J \) in (6.2.1). Under the null hypothesis the degrees of freedom of the null distribution is \( N - K - 1 \). There are \( N \) restrictions but one degree of freedom is lost estimating \( \gamma_0 \), and \( K \) degrees of freedom are used estimating the \( K \) elements of \( \lambda_K \).

The asymptotic variance of \( \hat{\gamma} \) follows from the maximum likelihood approach. The variance evaluated at the maximum likelihood estimators is

\[ \hat{\text{Var}}[\hat{\gamma}] = \frac{1}{T} \left( 1 + (\hat{\gamma}_1 + \hat{\mu}_{K*})' \hat{\Omega}_K^{-1} (\hat{\gamma}_1 + \hat{\mu}_{K*}) \right) \left[ \mathbf{X}^* \hat{\Sigma}^{*-1} \mathbf{X} \right]^{-1}. \]  
(6.2.43)

Applying the partitioned inverse rule to (6.2.43), for the variances of the components of \( \hat{\gamma} \) we have estimators

\[ \hat{\text{Var}}[\hat{\gamma}_0] = \frac{1}{T} \left( 1 + (\hat{\gamma}_1 + \hat{\mu}_{K*})' \hat{\Omega}_K^{-1} (\hat{\gamma}_1 + \hat{\mu}_{K*}) \right) \]  
\[ \left[ \mu' \hat{\Sigma}^{*-1} \mu - \mu' \hat{\Sigma}^{*-1} \hat{\mathbf{B}}^* (\hat{\mathbf{B}}^* \hat{\Sigma}^{*-1} \hat{\mathbf{B}}^*)^{-1} \hat{\mathbf{B}}^* \hat{\Sigma}^{*-1} \mu \right]^{-1} \]  
(6.2.44)

\[ \hat{\text{Var}}[\hat{\gamma}_1] = \frac{1}{T} \left( 1 + (\hat{\gamma}_1 + \hat{\mu}_{K*})' \hat{\Omega}_K^{-1} (\hat{\gamma}_1 + \hat{\mu}_{K*}) \right) (\hat{\mathbf{B}}^* \hat{\Sigma}^{*-1} \hat{\mathbf{B}}^*)^{-1} \]
\[ + (\hat{\mathbf{B}}^* \hat{\Sigma}^{*-1} \hat{\mathbf{B}}^*)^{-1} \hat{\mathbf{B}}^* \hat{\Sigma}^{*-1} \mu (\hat{\text{Var}}[\hat{\gamma}_0]) \]
\[ \times \mu' \hat{\Sigma}^{*-1} \hat{\mathbf{B}}^* (\hat{\mathbf{B}}^* \hat{\Sigma}^{*-1} \hat{\mathbf{B}}^*)^{-1}. \]  
(6.2.45)

We will use these variance results for inferences concerning the factor risk premia in Section 6.3.

### 6.2.4 Factor Portfolios Spanning the Mean-Variance Frontier

When factor portfolios span the mean-variance frontier, the intercept term of the exact pricing relation \( \lambda_0 \) is zero without the need for a riskfree asset.
6.2. Estimation and Testing

Thus this case retains the simplicity of the first case with the riskfree asset. In the context of the APT, spanning occurs when two well-diversified portfolios are on the minimum-variance boundary. Chamberlain (1983a) provides discussion of this case.

The unconstrained model will be a K-factor model expressed in real returns. Define \( R \), as an \((N \times 1)\) vector of real returns for \( N \) assets (or portfolios of assets). Then for real returns we have a K-factor linear model:

\[
R_t = a + B R_{Kt} + \epsilon_t \tag{6.2.46}
\]

\[
E[\epsilon_t] = 0 \tag{6.2.47}
\]

\[
E[\epsilon_t \epsilon'_t] = \Sigma \tag{6.2.48}
\]

\[
E[R_{Kt}] = \mu_K, \quad E[(R_{Kt} - \mu_K) (R_{Kt} - \mu_K)'] = \Omega_K \tag{6.2.49}
\]

\[
\text{Cov}[R_{Kt}, \epsilon'_t] = 0. \tag{6.2.50}
\]

\( B \) is the \((N \times K)\) matrix of factor sensitivities, \( R_{Kt} \) is the \((K \times 1)\) vector of factor portfolio real returns, and \( a \) and \( \epsilon_t \) are \((N \times 1)\) vectors of asset return intercepts and disturbances, respectively. \( O \) is a \((K \times N)\) matrix of zeroes. The restrictions on (6.2.46) imposed by the included factor portfolios spanning the mean-variance frontier are:

\[
a = 0 \quad \text{and} \quad B \mu = \mu. \tag{6.2.51}
\]

To understand the intuition behind these restrictions, we can return to the Black version of the CAPM from Chapter 5 and can construct a spanning example. The theory underlying the model differs but empirically the restrictions are the same as those on a two-factor APT model with spanning. The unconstrained Black model can be written as

\[
R_t = a + \beta_{om} R_{ot} + \beta_{m} R_{mt} + \epsilon_t, \tag{6.2.52}
\]

where \( R_{mt} \) and \( R_{ot} \) are the return on the market portfolio and the associated zero-beta portfolio, respectively. The restrictions on the Black model are \( a = 0 \) and \( \beta_{om} + \beta_m = \mu \) as shown in Chapter 5. These restrictions correspond to those in (6.2.51).

For the unconstrained model in (6.2.46) the maximum likelihood estimators are

\[
\hat{a} = \hat{\mu} - \hat{B} \hat{\mu}_K \tag{6.2.53}
\]
6. Multifactor Pricing Models

\[
\hat{\mathbf{B}} = \left[ \sum_{t=1}^{T} (\mathbf{R}_t - \hat{\mathbf{\mu}})(\mathbf{R}_K^t - \hat{\mathbf{\mu}})_t' \right] \\
\times \left[ \sum_{t=1}^{T} (\mathbf{R}_K^t - \hat{\mathbf{\mu}})(\mathbf{R}_K^t - \hat{\mathbf{\mu}})_t' \right]^{-1} 
\]  \hspace{1cm} (6.2.54)

\[
\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{R}_t - \hat{\mathbf{\alpha}} - \hat{\mathbf{B}}\mathbf{R}_K^t)(\mathbf{R}_t - \hat{\mathbf{\alpha}} - \hat{\mathbf{B}}\mathbf{R}_K^t)' , 
\]  \hspace{1cm} (6.2.55)

where

\[
\hat{\mathbf{\mu}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{R}_t \quad \text{and} \quad \hat{\mathbf{\mu}}_K = \frac{1}{T} \sum_{t=1}^{T} \mathbf{R}_K^t.
\]

To estimate the constrained model, we consider the unconstrained model in (6.2.46) with the matrix \( \mathbf{B} \) partitioned into an \((N \times 1)\) column vector \( \mathbf{b}_1 \) and an \((N \times (K - 1))\) matrix \( \mathbf{B}_1 \) and the factor portfolio vector partitioned into the first row \( \mathbf{R}_1^t \) and the last \((K - 1)\) rows \( \mathbf{R}_K^{*t} \). With this partitioning the constraint \( \mathbf{B} \mathbf{\ell} = \mathbf{\ell} \) can be written \( \mathbf{b}_1 + \mathbf{B}_1 \mathbf{\ell} = \mathbf{\ell} \). For the unconstrained model we have

\[
\mathbf{R}_t = a + \mathbf{b}_1 \mathbf{R}_1^t + \mathbf{B}_1 \mathbf{R}_K^{*t} + \epsilon_t. 
\]  \hspace{1cm} (6.2.56)

Substituting \( a = 0 \) and \( \mathbf{b}_1 = \mathbf{\ell} - \mathbf{B}_1 \mathbf{\ell} \) into (6.2.56) gives the constrained model,

\[
\mathbf{R}_t - \ell \mathbf{R}_1^t = \mathbf{B}_1 (\mathbf{R}_K^{*t} - \ell \mathbf{R}_1^t) + \epsilon_t. 
\]  \hspace{1cm} (6.2.57)

Using (6.2.57) the maximum likelihood estimators are

\[
\hat{\mathbf{B}}_1 = \left[ \frac{1}{T} \sum_{t=1}^{T} (\mathbf{R}_t - \ell \mathbf{R}_1^t)(\mathbf{R}_K^{*t} - \ell \mathbf{R}_1^t)' \right] \frac{1}{T} \sum_{t=1}^{T} (\mathbf{R}_K^t - \hat{\mathbf{\mu}})(\mathbf{R}_K^t - \hat{\mathbf{\mu}})_t' 
\]  \hspace{1cm} (6.2.58)

\[
\hat{\mathbf{b}}_1 = \mathbf{\ell} - \hat{\mathbf{B}}_1 \mathbf{\ell} 
\]  \hspace{1cm} (6.2.59)

\[
\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{R}_t - \hat{\mathbf{B}}_1 \mathbf{R}_K^t)(\mathbf{R}_t - \hat{\mathbf{B}}_1 \mathbf{R}_K^t)' . 
\]  \hspace{1cm} (6.2.60)

The null hypothesis \( a = 0 \) can be tested using the likelihood ratio statistic \( J \) in (6.2.1). Under the null hypothesis the degrees of freedom of the null distribution will be \( 2N \) since \( a = 0 \) is \( N \) restrictions and \( \mathbf{B} \mathbf{\ell} = \mathbf{\ell} \) is \( N \) additional restrictions.

We can also construct an exact test of the null hypothesis given the linearity of the restrictions in (6.2.51) and the multivariate normality assumption.
Defining $f_2$ as the test statistic we have

$$f_2 = \frac{(T - N - K)}{N} \left[ \frac{|\hat{\Sigma}^*|}{|\hat{\Sigma}|} - 1 \right].$$

Under the null hypothesis, $f_2$ is unconditionally distributed central F with $2N$ degrees of freedom in the numerator and $2(T - N - K)$ degrees of freedom in the denominator. Huberman and Kandel (1987) present a derivation of this test.

### 6.3 Estimation of Risk Premia and Expected Returns

All the exact factor pricing models allow one to estimate the expected return on a given asset. Since the expected return relation is $\mu = \nu \lambda_0 + B \lambda_K$, one needs measures of the factor sensitivity matrix $B$, the riskfree rate or the zero-beta expected return $\lambda_0$, and the factor risk premia $\lambda_K$. Obtaining measures of $B$ and the riskfree rate or the expected zero-beta return is straightforward. For the given case the constrained maximum likelihood estimator $B^*$ can be used for $B$. The observed riskfree rate is appropriate for the riskfree asset or, in the cases without a riskfree asset, the maximum likelihood estimator $\hat{\gamma}_0$ can be used for the expected zero-beta return.

Further estimation is necessary to form estimates of the factor risk premia. The appropriate procedure varies across the four cases of exact factor pricing. In the case where the factors are the excess returns on traded portfolios, the risk premia can be estimated directly from the sample means of the excess returns on the portfolios. For this case we have

$$\hat{\lambda}_K = \hat{\mu}_K = \frac{1}{T} \sum_{t=1}^{T} Z_{Kt}.$$  \hspace{1cm} (6.3.1)

An estimator of the variance of $\hat{\lambda}_K$ is

$$\text{Var}[\hat{\lambda}_K] = \frac{1}{T} \hat{\Omega}_K = \frac{1}{T^2} \sum_{t=1}^{T} (Z_{Kt} - \hat{\mu}_K) (Z_{Kt} - \hat{\mu}_K)'.$$  \hspace{1cm} (6.3.2)

In the case where portfolios are factors but there is no riskfree asset, the factor risk premia can be estimated using the difference between the sample mean of the factor portfolios and the estimated zero-beta return:

$$\hat{\lambda}_K = \hat{\mu}_K - \nu \hat{\gamma}_0.$$  \hspace{1cm} (6.3.3)

In this case, an estimator of the variance of $\hat{\lambda}_K$ is

$$\text{Var}[\hat{\lambda}_K] = \frac{1}{T} \hat{\Omega}_K + \text{Var}[\hat{\gamma}_0] \nu',$$  \hspace{1cm} (6.3.4)
where \( \overline{\text{Var}}[\hat{\gamma}_0] \) is from (6.2.27). The fact that \( \hat{\mu}_K \) and \( \hat{\gamma}_0 \) are independent has been utilized to set the covariance term in (6.3.4) to zero.

In the case where the factors are not traded portfolios, an estimator of the vector of factor risk premia \( \lambda_K \) is the sum of the estimator of the mean of the factor realizations and the estimator of \( \gamma_1 \),

\[
\hat{\lambda}_K = \hat{\mu}_{fK} + \hat{\gamma}_1. \tag{6.3.5}
\]

An estimator of the variance of \( \lambda_K \) is

\[
\overline{\text{Var}}[\hat{\lambda}_K] = \frac{1}{T} \hat{\Omega}_K + \overline{\text{Var}}[\hat{\gamma}_1], \tag{6.3.6}
\]

where \( \overline{\text{Var}}[\hat{\gamma}_1] \) is from (6.2.45). Because \( \hat{\mu}_{fK} \) and \( \hat{\gamma}_1 \) are independent the covariance term in (6.3.6) is zero.

The fourth case, where the factor portfolios span the mean-variance frontier, is the same as the first case except that real returns are substituted for excess returns. Here \( \hat{\lambda}_K \) is the vector of factor portfolio sample means and \( \lambda_0 \) is zero.

For any asset the expected return can be estimated by substituting the estimates of \( B, \lambda_0, \) and \( \lambda_K \) into (6.1.3). Since (6.1.8) is nonlinear in the parameters, calculating a standard error requires using a linear approximation and estimates of the covariances of the parameter estimates.

It is also of interest to ask if the factors are jointly priced. Given the vector of risk premia estimates and its covariance matrix, tests of the null hypothesis that the factors are jointly not priced can be conducted using the following test statistic:

\[
J_3 = \frac{T - K}{TK} \hat{\lambda}_K' \overline{\text{Var}}[\hat{\lambda}_K]^{-1} \hat{\lambda}_K. \tag{6.3.7}
\]

Asymptotically, under the null hypothesis that \( \lambda_K = 0, J_3 \) has an \( F \) distribution with \( K \) and \( T - K \) degrees of freedom. This distributional result is an application of the Hotelling \( T^2 \) statistic and will be exact in finite samples for the cases where the estimator of \( \lambda_K \) is based only on the sample means of the factors. We can also test the significance of any individual factor using

\[
J_4 = \frac{\hat{\lambda}_{jK}}{\sqrt{v_{jj}}} \overset{a}{\sim} \mathcal{N}(0, 1), \tag{6.3.8}
\]

where \( \hat{\lambda}_{jK} \) is the \( j \)th element of \( \hat{\lambda}_K \) and \( v_{jj} \) is the \( (j,j) \)th element of \( \overline{\text{Var}}[\hat{\lambda}_K] \). Testing if individual factors are priced is sensible for cases where the factors have been theoretically specified. With empirically derived factors, such tests are not useful because, as we explain in Section 6.4.1, factors are identified only up to an orthogonal transformation; hence individual factors do not have clear-cut economic interpretations.
Shanken (1992b) shows that factor risk premia can also be estimated using a two-pass cross-sectional regression approach. In the first pass the factor sensitivities are estimated asset-by-asset using OLS. These estimators represent a measure of the factor loading matrix \( B \) which we denote \( \hat{B} \). This estimator of \( B \) will be identical to the unconstrained maximum likelihood estimators previously presented for jointly normal and IID residuals.

Using this estimator of \( B \) and the \( (N \times 1) \) vector of asset returns for each time period, the \textit{ex post} factor risk premia can be estimated time-period-by-time-period in the second pass. The second-pass regression is

\[
Z_t = \nu \lambda_{0t} + B \lambda_{Kt} + \eta_t. \tag{6.3.9}
\]

The regression can be consistently estimated using OLS; however, GLS can also be used. The output of the regression is a time series of \textit{ex post} risk premia, \( \hat{\lambda}_{Kt}, t = 1, \ldots, T \), and an \textit{ex post} measure of the zero-beta portfolio return, \( \hat{\lambda}_{0t}, t = 1, \ldots, T \).

Common practice is then to conduct inferences about the risk premia using the means and standard deviations of these \textit{ex post} series. While this approach is a reasonable approximation, Shanken (1992b) shows that the calculated standard errors of the means will underestimate the true standard errors because they do not account for the estimation error in \( \hat{B} \). Shanken derives an adjustment which gives consistent standard errors. No adjustment is needed when a maximum likelihood approach is used, because the maximum likelihood estimators already incorporate the adjustment.

### 6.4 Selection of Factors

The estimation and testing results in Section 6.2 assume that the identity of the factors is known. In this section we address the issue of specifying the factors. The approaches fall into two basic categories, statistical and theoretical. The statistical approaches, largely motivated by the APT, involve building factors from a comprehensive set of asset returns (usually much larger than the set of returns used to estimate and test the model). Sample data on these returns are used to construct portfolios that represent factors. The theoretical approaches involve specifying factors based on arguments that the factors capture economy-wide systematic risks.

#### 6.4.1 Statistical Approaches

Our starting point for the statistical construction of factors is the linear factor model. We present the analysis in terms of real returns. The same analysis will apply to excess returns in cases with a riskfree asset. Recall that
for the linear model we have

\[ R_t = a + B f_t + \epsilon_t \]  

(6.4.1)

\[ E[\epsilon_t \epsilon'_t | f_t] = \Sigma, \]  

(6.4.2)

where \( R_t \) is the \((N \times 1)\) vector of asset returns for time period \( t \), \( f_t \) is the \((K \times 1)\) vector of factor realizations for time period \( t \), and \( \epsilon_t \) is the \((N \times 1)\) vector of model disturbances for time period \( t \). The number of assets, \( N \), is now very large and usually much larger than the number of time periods, \( T \). There are two primary statistical approaches, factor analysis and principal components.

Factor Analysis

Estimation using factor analysis involves a two-step procedure. First the factor sensitivity matrix \( B \) and the disturbance covariance matrix \( \Sigma \) are estimated and then these estimates are used to construct measures of the factor realizations. For standard factor analysis it is assumed that there is a strict factor structure. With this structure \( K \) factors account for all the cross covariance of asset returns and hence \( \Sigma \) is diagonal. (Ross imposes this structure in his original development of the APT.)

Given a strict factor structure and \( K \) factors, we can express the \((N \times N)\) covariance matrix of asset returns as the sum of two components, the variation from the factors plus the residual variation,

\[ \Omega = B \Omega_K B' + D, \]  

(6.4.3)

where \( E[f_t f'_t] = \Omega_K \) and \( \Sigma = D \) to indicate it is diagonal. With the factors unknown, \( B \) is identified only up to an orthogonal transformation. All transforms \( B G \) are equivalent for any \((K \times K)\) orthogonal transformation matrix \( G \), i.e., such that \( G G' = I \). This rotational indeterminacy can be eliminated by restricting the factors to be orthogonal to each other and to have unit variance. In this case we have \( \Omega_K = I \) and \( B \) is unique. With these restrictions in place, we can express the return covariance matrix as

\[ \Omega = BB' + D. \]  

(6.4.4)

With the structure in (6.4.4) and the assumption that asset returns are jointly normal and temporally IID, estimators of \( B \) and \( D \) can be formulated using maximum likelihood factor analysis. Because the first-order conditions for maximum likelihood are highly nonlinear in the parameters, solving for the estimators with the usual iterative procedure can be slow and convergence difficult. Alternative algorithms have been developed by Joreskog (1967) and Rubin and Thayer (1982) which facilitate quick convergence to the maximum likelihood estimators.
6.4. Selection of Factors

One interpretation of the maximum likelihood estimator of $B$ given the maximum likelihood estimator of $D$ is that the estimator of $B$ has the eigenvectors of $\hat{D}^{-1}\hat{V}$ associated with the $K$ largest eigenvalues as its columns. For details of the estimation the interested reader can see these papers, or Morrison (1990, chapter 9) and references therein.

The second step in the estimation procedure is to estimate the factors given $B$ and $\Sigma$. Since the factors are derived from the covariance structure, the means are not specified in (6.4.1). Without loss of generality, we can restrict the factors to have zero means and express the factor model in terms of deviations about the means,

$$\begin{align*}
(R_t - \mu) &= B\{f_t + \epsilon_t\}. 
\end{align*}$$

(6.4.5)

Given (6.4.5), a candidate to proxy for the factor realizations for time period $t$ is the cross-sectional generalized least squares (GLS) regression estimator. Using the maximum likelihood estimators of $B$ and $D$ we have for each $t$

$$\hat{f}_t = (\hat{B}'\hat{D}^{-1}\hat{B})^{-1}\hat{B}'\hat{D}^{-1}(R_t - \hat{\mu}).$$

(6.4.6)

Here we are estimating $f_t$ by regressing $(R_t - \hat{\mu})$ onto $\hat{B}$. The factor realization series, $\hat{f}_t$, $t = 1, \ldots, T$, can be employed to test the model using the approach in Section 6.2.3.

Since the factors are linear combinations of returns we can construct portfolios which are perfectly correlated with the factors. Denoting $\hat{R}_{Kt}$ as the $(K \times 1)$ vector of factor portfolio returns for time period $t$, we have

$$\hat{R}_{Kt} = AWR_t,$$

(6.4.7)

where

$$W = (\hat{B}'\hat{D}^{-1}\hat{B})^{-1}\hat{B}'\hat{D}^{-1},$$

and $A$ is defined as a diagonal matrix with $1/W_j$ as the $j$th diagonal element, where $W_j$ is the $j$th element of $W_t$.

The factor portfolio weights obtained for the $j$th factor from this procedure are equivalent to the weights that would result from solving the following optimization problem and then normalizing the weights to sum to one:

$$\begin{align*}
\text{Min} & \quad \omega_j'\hat{D}W, \\
\text{subject to} & \quad \omega_j'\hat{b}_k = 0 \quad \forall k \neq j, \\
& \quad \omega_j'\hat{b}_k = 1 \quad \forall k = j.
\end{align*}$$

(6.4.8, 6.4.9, 6.4.10)
That is, the factor portfolio weights minimize the residual variance subject to the constraints that each factor portfolio has a unit loading on its own factor and zero loadings on other factors. The resulting factor portfolio returns can be used in all the approaches discussed in Section 6.2.

If B and D are known, then the factor estimators based on GLS with the population values of B and D will have the maximum correlation with the population factors. This follows from the minimum-variance unbiased estimator property of generalized least squares given the assumed normality of the disturbance vector. But in practice the factors in (6.4.6) and (6.4.7) need not have the maximum correlation with the population common factors since they are based on estimates of B and D. Lehmann and Modest (1988) present an alternative to GLS. In the presence of measurement error, they find this alternative can produce factor portfolios with a higher population correlation with the common factors. They suggest for the jth factor to use \( \hat{\omega}_j^* \), where the \( N \times 1 \) vector \( \hat{\omega}_j \) is the solution to the following problem:

\[
\min_{\omega_j} \omega_j^* D \omega_j^* 
\]

subject to

\[
\omega_j^* b_k = 0 \quad \forall k \neq j 
\]

\[
\omega_j^* e = 1. 
\]

This approach finds the portfolio which has the minimum residual variance of all portfolios orthogonal to the other \( (K-1) \) factors. Unlike the GLS procedure, this procedure ignores the information in the factor loadings of the jth factor. It is possible that this is beneficial because of the measurement error in the loadings. Indeed, Lehmann and Modest find that this method of forming factor portfolios results in factors with less extreme weightings on the assets and a resulting higher correlation with the underlying common factors.

Principal Components

Factor analysis represents only one statistical method of forming factor portfolios. An alternative approach is principal components analysis. Principal components is a technique to reduce the number of variables being studied without losing too much information in the covariance matrix. In the present application, the objective is to reduce the dimension from N asset returns to K factors. The principal components serve as the factors. The first principal component is the (normalized) linear combination of asset returns with maximum variance. The second principal component is the (normalized) linear combination of asset returns with maximum variance of all combinations orthogonal to the first principal component. And so on.
6.4. Selection of Factors

The first sample principal component is \( x_1^* \mathbf{R} \), where the \((N \times 1)\) vector \( x_1^* \) is the solution to the following problem:

\[
\text{Max} \ x_1^* \hat{\Omega} x_1
\]

subject to

\[
x_1^* x_1 = 1.
\]

\( \hat{\Omega} \) is the sample covariance matrix of returns. The solution \( x_1^* \) is the eigenvector associated with the largest eigenvalue of \( \hat{\Omega} \). To facilitate the portfolio interpretation of the factors we can define the first factor as \( \omega_1^* \mathbf{R} \), where \( \omega_1 \) is \( x_1^* \) scaled by the reciprocal of \( \ell^* x_1^* \) so that its elements sum to one. The second sample principal component solves the above problem for \( x_2 \) in the place of \( x_1 \) with the additional restriction \( x_1^* x_2 = 0 \). The solution \( x_2^* \) is the eigenvector associated with the second largest eigenvalue of \( \hat{\Omega} \). \( x_2^* \) can be scaled by the reciprocal of \( \ell^* x_2^* \) giving \( \omega_2 \), and then the second factor portfolio will be \( \omega_2^* \mathbf{R} \). In general the jth factor will be \( \omega_j^* \mathbf{R} \), where \( \omega_j \) is the rescaled eigenvector associated with the jth largest eigenvalue of \( \hat{\Omega} \). The factor portfolios derived from the first K principal components analysis can then be employed as factors for all the tests outlined in Section 6.2.

Another principal components approach has been developed by Connor and Korajczyk (1986, 1988). They propose using the eigenvectors associated with the \( K \) largest eigenvalues of the \((T \times T)\) centered returns cross-product matrix rather than the standard approach which uses the principal components of the \((N \times N)\) sample covariance matrix. They show that as the cross section becomes large the \((K \times T)\) matrix with the rows consisting of the \( K \) eigenvectors of the cross-product matrix will converge to the matrix of factor realizations (up to a nonsingular linear transformation reflecting the rotational indeterminacy of factor models). The potential advantages of this approach are that it allows for time-varying factor risk premia and that it is computationally convenient. Because it is typical to have a cross section of assets much larger than the number of time-series observations, analyzing a \((T \times T)\) matrix can be less burdensome than working with an \((N \times N)\) sample covariance matrix.

Factor Analysis or Principal Components?

We have discussed two statistical primary approaches for constructing the model factors—factor analysis and principal components. Within each approach there are possible variations in the process of estimating the factors. A question arises as to which technique is optimal in the sense of providing the most precise measures of the population factors given a fixed sample of returns. Unfortunately, the answer in finite samples is not clear although all procedures can be justified in large samples.

\(^4\text{See also Mei (1993)}\)
Chamberlain and Rothschild (1933) show that consistent estimates of the factor loading matrix $B$ can be obtained from the eigenvectors associated with the largest eigenvalues of $\mathbf{\Psi}^{-1}\mathbf{\Omega}$, where $\mathbf{\Psi}$ is any arbitrary positive definite matrix with eigenvalues bounded away from zero and infinity. Both standard factor analysis and principal components fit into this category, for factor analysis $\mathbf{\Psi} = \mathbf{D}$ and for principal components $\mathbf{\Psi} = \mathbf{I}$. However, the finite-sample applicability of the result is unclear since it is required that both the number of assets $N$ and the number of time periods $T$ go to infinity.

The Connor and Korajczyk principal components approach is also consistent as $N$ increases. It has the further potential advantage that it only requires $T \geq K$ and does not require $T$ to increase to infinity. However, whether in finite samples it dominates factor analysis or standard principal components is an open question.

### 6.4.2 Number of Factors

The underlying theory of the multifactor models does not specify the number of factors that are required, that is, the value of $K$. While, for the theory to be useful, $K$ should be reasonably small, the researcher still has significant latitude in the choice. In empirical work this lack of specification has been handled in several ways. One approach is to repeat the estimation and testing of the model for a variety of values of $K$ and observe if the tests are sensitive to increasing the number of factors. For example Lehmann and Modest (1988) present empirical results for five, ten, and fifteen factors. Their results display minimal sensitivity when the number of factors increases from five to ten to fifteen. Similarly Connor and Korajczyk (1988) consider five and ten factors with little sensitivity to the additional five factors. These results suggest that five factors are adequate.

A second approach is to test explicitly for the adequacy of $K$ factors. An asymptotic likelihood ratio test of the adequacy of $K$ factors can be constructed using $-2$ times the difference of the value of the log-likelihood function of the covariance matrix evaluated at the constrained and unconstrained estimators. Morrison (1990, p. 362) presents this test. The likelihood ratio test statistic is

$$J_K = - \left( T - 1 - \frac{1}{6}(2N + 5) - \frac{2}{3}K \right) \left[ \log |\hat{\Omega}| - \log |\hat{B}\hat{B}^\prime + \hat{D}| \right], \quad (6.4.16)$$

where $\hat{\Omega}$ is the maximum likelihood estimator of $\Omega$ and $B$ and $D$ are the maximum likelihood estimators of $B$ and $D$, respectively. The leading term is an adjustment to improve the convergence of the finite-sample null distribution to the large-sample distribution. Under the null hypothesis that $K$ factors are adequate, $J_K$ will be asymptotically distributed ($T \to \infty$) as a chi-square variate with $\frac{1}{2}[((N-K)^2 - N - K)$ degrees of freedom. Roll and
Ross (1980) use this approach and conclude that three or four factors are adequate.

A potential drawback of using the test from maximum likelihood factor analysis is that the constrained model assumes a strict factor structure—an assumption which is not theoretically necessary. Connor and Korajczyk (1993) develop an asymptotic test \((N \to \infty)\) for the adequacy of \(K\) factors under the assumption of an approximate factor structure. Their test uses the result that with an approximate factor structure the average cross-sectional variation explained by the \(K+1\)'st factor approaches zero as \(N\) increases,

\[
\lim_{N \to \infty} \frac{1}{N} b_{K+1}' b_{K+1} = 0, \tag{6.4.17}
\]

where the dependence of \(b_{K+1}\) on \(N\) is implicit. This implies that in a large cross section generated by a \(K\)-factor model, the average residual variance in a linear factor model estimated with \(K+1\) factors should converge to the average residual variance with \(K\) factors. This is the implication Connor and Korajczyk test. Examining returns from stocks listed on the New York Stock Exchange and the American Stock Exchange they conclude that there are up to six pervasive factors.

### 6.4.3 Theoretical Approaches

Theoretically based approaches for selecting factors fall into two main categories. One approach is to specify macroeconomic and financial market variables that are thought to capture the systematic risks of the economy. A second approach is to specify characteristics of firms which are likely to explain differential sensitivity to the systematic risks and then form portfolios of stocks based on the characteristics.

Chen, Roll, and Ross (1986) is a good example of the first approach. The authors argue that in selecting factors we should consider forces which will explain changes in the discount rate used to discount future expected cash flows and forces which influence expected cash flows themselves. Based on intuitive analysis and empirical investigation a five-factor model is proposed. The factors include the yield spread between long and short interest rates for US government bonds (maturity premium), expected inflation, unexpected inflation, industrial production growth, and the yield spread between corporate high- and low-grade bonds (default premium). Aggregate consumption growth and oil prices are found not to have incremental effects beyond the five factors.\(^3\)

\(^3\)An alternative implementation of the first approach is given by Campbell (1996a) and is discussed in Chapter 8.
The second approach of creating factor portfolios based on firm characteristics has been used in a number of studies. These characteristics have mostly surfaced from the literature of CAPM violations discussed in Chapter 5. Characteristics which have been found to be empirically important include market value of equity, price-to-earnings ratio, and ratio of book value of equity to market value of equity. The general finding is that factor models which include a broad based market portfolio (such as an equal-weighted index) and factor portfolios created using these characteristics do a good job in explaining the cross section of returns. However, because the important characteristics have been identified largely through empirical analysis, their importance may be overstated because of data-snooping biases. We will discuss this issue in Section 6.6.

6.5 Empirical Results

Many empirical studies of multifactor models exist. We will review four of the studies which nicely illustrate the estimation and testing methodology we have discussed. Two comprehensive studies using statistical approaches to select the factors are Lehmann and Modest (1988) and Connor and Korajczyk (1988). Lehmann and Modest [LM] use factor analysis and Connor and Korajczyk [CK] use (T x T) principal components. Two studies using the theoretical approach to factor identification are Fama and French (1993) and Chen, Roll, and Ross (1986). Fama and French [FF] use firm characteristics to form factor portfolios and Chen, Roll, and Ross [CRR] specify macroeconomic variables as factors. The first three studies include tests of the implications of exact factor pricing, while Chen, Roll, and Ross focus on whether or not the factors are priced. The evidence supporting exact factor pricing is mixed. Table 6.1 summarizes the main results from LM, CK, and FF.

A number of general points emerge from this table. The strongest evidence against exact factor pricing comes from tests using dependent portfolios based on market value of equity and book-to-market ratios. Even multifactor models have difficulty explaining the "size" effect and "book to market" effect. Portfolios which are formed based on dividend yield and based on own variance provide little evidence against exact factor pricing. The CK results for January and non-January months suggest that the evidence against exact factor pricing does not arise from the January effect.

Using the statistical approaches, CK and LM find little sensitivity to increasing the number of factors beyond five. On the other hand FF find some improvement going from two factors to five factors. In results not included, FF find that with stocks only three factors are necessary and that when bond portfolios are included then five factors are needed. These
Table 6.1. Summary of results for tests of exact factor pricing using zero-intercept F-test.

<table>
<thead>
<tr>
<th>Study</th>
<th>Time period</th>
<th>Portfolio characteristic</th>
<th>N</th>
<th>K</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CK</td>
<td>64:01-83:12</td>
<td>market value of equity</td>
<td>10</td>
<td>5</td>
<td>0.002</td>
</tr>
<tr>
<td>CK</td>
<td>64:01-83:12</td>
<td>market value of equity</td>
<td>10</td>
<td>10</td>
<td>0.002</td>
</tr>
<tr>
<td>CK</td>
<td>64:01-83:12</td>
<td>market value of equity</td>
<td>10</td>
<td>5</td>
<td>0.236</td>
</tr>
<tr>
<td>CK</td>
<td>64:01-83:12</td>
<td>market value of equity</td>
<td>10</td>
<td>10</td>
<td>0.171</td>
</tr>
<tr>
<td>CK</td>
<td>64:01-83:12</td>
<td>market value of equity</td>
<td>10</td>
<td>5</td>
<td>0.011</td>
</tr>
<tr>
<td>CK</td>
<td>64:01-83:12</td>
<td>market value of equity</td>
<td>10</td>
<td>10</td>
<td>0.019</td>
</tr>
</tbody>
</table>

| LM    | 63:01-82:12 | market value of equity   | 5 | 5 | **      |
| LM    | 63:01-82:12 | market value of equity   | 5 | 10| **      |
| LM    | 63:01-82:12 | market value of equity   | 5 | 15| **      |
| LM    | 63:01-82:12 | market value of equity   | 20| 5 | 0.11    |
| LM    | 63:01-82:12 | market value of equity   | 20| 10| 0.14    |
| LM    | 63:01-82:12 | market value of equity   | 20| 15| 0.42    |

| LM    | 63:01-82:12 | dividend yield           | 5 | 5 | 0.17    |
| LM    | 63:01-82:12 | dividend yield           | 5 | 10| 0.18    |
| LM    | 63:01-82:12 | dividend yield           | 5 | 15| 0.17    |
| LM    | 63:01-82:12 | dividend yield           | 20| 5 | 0.94    |
| LM    | 63:01-82:12 | dividend yield           | 20| 10| 0.97    |
| LM    | 63:01-82:12 | dividend yield           | 20| 15| 0.98    |

| LM    | 63:01-82:12 | own variance             | 5 | 5 | 0.29    |
| LM    | 63:01-82:12 | own variance             | 5 | 10| 0.57    |
| LM    | 63:01-82:12 | own variance             | 5 | 15| 0.55    |
| LM    | 63:01-82:12 | own variance             | 20| 5 | 0.83    |
| LM    | 63:01-82:12 | own variance             | 20| 10| 0.97    |
| LM    | 63:01-82:12 | own variance             | 20| 15| 0.98    |

| FF    | 63:07-91:12 | stocks and bonds         | 32| 2 | 0.010   |
| FF    | 63:07-91:12 | stocks and bonds         | 32| 3 | 0.039   |
| FF    | 63:07-91:12 | stocks and bonds         | 32| 5 | 0.025   |

**Less than 0.001.

CK refers to Connor and Korajczyk (1988), LM refers to Lehmann and Modest (1988), and FF refers to Fama and French (1993). The CK factors are derived using (TXT) principal components, the LM factors are derived using maximum likelihood factor analysis, and the FF factors are prespecified factor portfolios. For the FF two-factor case the factors are the return on a portfolio of low market value of equity firms minus a portfolio of high market value of equity firms and the return on a portfolio of high book-to-market value firms minus a portfolio of low book-to-market value firms. For the three-factor case the factors are those in the two-factor case plus the return on the CRSP value-weighted stock index. For the five-factor case the returns on a term structure factor and a default risk factor are added. CK include tests separating the intercept for January from the intercept for other months. CKJ are results of tests of the hypothesis that the January intercept is zero and CKNJ are results of tests of the hypothesis that the non-January intercept is zero. CK and FF work with a monthly sampling interval. LM use a daily interval to estimate the factors and a weekly interval for testing. The test results from CK and LM are based on tests from four five-year periods aggregated together. The portfolio characteristic represents the firm characteristic used to allocate stocks into the dependent portfolios. FF use 25 stock portfolios and 7 bond portfolios. The stock portfolios are created using a two way sort based on market value of equity and book-value-to-market-value ratios. The bond portfolios include five US government bond portfolios and two corporate bond portfolios. The government bond portfolios are created based on maturity and the corporate bond portfolios are created based on the level of default risk. N is the number of dependent portfolios and K is the number of factors. The p-values are reported for the zero-intercept F-test.
results are generally consistent with direct tests for the number of factors discussed in Section 6.4.2. The LM results display considerable sensitivity to the number of dependent portfolios included. The \( p \)-values are considerably lower with fewer portfolios. This is most likely an issue of the power of the test. For these tests with an unspecified alternative hypothesis, reducing the number of portfolios without eliminating the deviations from the null hypothesis can lead to substantial increases in power, because fewer restrictions must be tested.

The CRR paper focuses on the pricing of the factors. They use a cross-sectional regression methodology which is similar to the approach presented in Section 6.3. As previously noted they find evidence of five priced factors. The factors include the yield spread between long and short interest rates for US government bonds (maturity premium), expected inflation, unexpected inflation, industrial production growth, and the yield spread between corporate high- and low-grade bonds (default premium).

6.6 Interpreting Deviations from Exact Factor Pricing

We have just reviewed empirical evidence which suggests that, while multifactor models do a reasonable job of describing the cross section of returns, deviations from the models do exist. Given this, it is important to consider the possible sources of deviations from exact factor pricing. This issue is important because in a given finite sample it is always possible to find an additional factor that will make the deviations vanish. However the procedure of adding an extra factor implicitly assumes that the source of the deviations is a missing risk factor and does not consider other possible explanations.

In this section we analyze the deviations from exact factor pricing for a given model with the objective of exploring the source of the deviations. For the analysis the potential sources of deviations are categorized into two groups—risk-based and nonrisk-based. The objective is to evaluate the plausibility of the argument that the deviations from the given factor model can be explained by additional risk factors.

The analysis relies on an important distinction between the two categories, namely, a difference in the behavior of the maximum squared Sharpe ratio as the cross section of securities is increased. (Recall that the Sharpe ratio is the ratio of the mean excess return to the standard deviation of the excess return.) For the risk-based alternatives the maximum squared Sharpe ratio is bounded and for the nonrisk-based alternatives the maximum squared Sharpe ratio is a less useful construct and can, in principle, be unbounded.
6.6. Interpreting Deviations from Exact Factor Pricing

6.6.1 Exact Factor Pricing Models, Mean-Variance Analysis, and the Optimal Orthogonal Portfolio

For the initial analysis we drop back to the level of the primary assets in the economy. Let N be the number of primary assets. Assume that a riskfree asset exists. Let \( \mathbf{Z}_t \) represent the (N x 1) vector of excess returns for period \( t \). Assume \( \mathbf{Z}_t \) is stationary and ergodic with mean \( \mathbf{\mu} \) and covariance matrix \( \mathbf{\Omega} \) that is full rank. We also take as given a set of K factor portfolios and analyze the deviations from exact factor pricing. For the factor model, as in (6.2.2), we have

\[
\mathbf{Z}_t = \mathbf{a} + \mathbf{BZ}_K + \mathbf{\epsilon}_t.
\]  

(6.6.1)

Here \( \mathbf{B} \) is the (N x K) matrix of factor loadings, \( \mathbf{Z}_K \) is the (K x 1) vector of time-\( t \) factor portfolio excess returns, and \( \mathbf{a} \) and \( \mathbf{\epsilon}_t \) are (N x 1) vectors of asset return intercepts and disturbances, respectively. The variance-covariance matrix of the disturbances is \( \mathbf{\Sigma} \) and the variance-covariance matrix of the factors is \( \mathbf{\Omega}_K \), as in (6.2.3)–(6.2.6). The values of \( \mathbf{a} \), \( \mathbf{B} \), and \( \mathbf{\Sigma} \) will depend on the factor portfolios, but this dependence is suppressed for notational convenience.

If we have exact factor pricing relative to the K factors, all the elements of the vector \( \mathbf{a} \) will be zero; equivalently, a linear combination of the factor portfolios forms the tangency portfolio (the mean-variance efficient portfolio of risky assets given the presence of a riskfree asset). Let \( \mathbf{Z}_\tau \) be the excess return of the (ex ante) tangency portfolio and let \( \mathbf{\omega}_q \) be the (N x 1) vector of portfolio weights. From mean-variance analysis (see Chapter 5),

\[
\mathbf{\omega}_q = (\mathbf{\mu}'\mathbf{\Omega}^{-1}\mathbf{\mu})^{-1}\mathbf{\Omega}^{-1}\mathbf{\mu}.
\]  

(6.6.2)

In the context of the K-factor model in (6.6.1), we have exact factor pricing when the tangency portfolio in (6.6.2) can be formed from a linear combination of the K factor portfolios.

Now consider the case where we do not have exact factor pricing, so the tangency portfolio cannot be formed from a linear combination of the factor portfolios. Our interest is in developing the relation between the deviations from the asset pricing model, \( \mathbf{a} \), and the residual covariance matrix, \( \mathbf{\Sigma} \). To facilitate this, we define the optimal orthogonal portfolio,\(^6\) which is the unique portfolio that can be combined with the K factor portfolios to form the tangency portfolio and is orthogonal to the factor portfolios.

**Definition (optimal orthogonal portfolio).** Take as given K factor portfolios which cannot be combined to form the tangency portfolio or the global minimum-variance portfolio. A portfolio \( \mathbf{h} \) will be defined as the optimal orthogonal portfolio with respect to these K factor portfolios if

\[
\mathbf{\omega}_q = \mathbf{W}_p \mathbf{\omega} + \mathbf{\omega}_h (1 - \mathbf{\omega}' \mathbf{\omega})
\]  

(6.6.3)

\(^6\)See Roll (1980) for general properties of orthogonal portfolios.
and
\[ \omega_h' \Omega W_p = 0 \]  
(6.6.4)

for a \((K \times 1)\) vector \(w\) where \(W_p\) is the \((N \times K)\) matrix of asset weights for the factor portfolios, \(w_h\) is the \((N \times 1)\) vector of asset weights for the optimal orthogonal portfolio, and \(w_q\) is the \((N \times 1)\) vector of asset weights for the tangency portfolio. If one considers a model without any factor portfolios \((K = 0)\) then the optimal orthogonal portfolio will be the tangency portfolio.

The weights of portfolio \(h\) can be expressed in terms of the parameters of the \(K\)-factor model. The vector of weights is
\[
\omega_h = (\tau'I^{-1}a)^{-1}a = (\tau'\Sigma^+a)^{-1}\Sigma^+a, \tag{6.6.5}
\]
where the \(\dagger\) superscript indicates the generalized inverse. The usefulness of this portfolio comes from the fact that when added to (6.6.1) the intercept will vanish and the factor loading matrix \(B\) will not be altered. The optimality restriction in (6.6.3) leads to the intercept vanishing, and the orthogonality condition in (6.6.4) leads to \(B\) being unchanged. Adding in \(Z_{ht}\):
\[
Z_t = BZ_{Kt} + \beta_h Z_{ht} + u_t \tag{6.6.6}
\]

We can relate the optimal orthogonal portfolio parameters to the factor model deviations by comparing (6.6.1) and (6.6.6). Taking the unconditional expectations of both sides,
\[
E[u_t] = 0 \tag{6.6.7}
\]
\[
E[u_t u_t'] = \Phi \tag{6.6.8}
\]
\[
E[Z_{ht}] = \mu_h, \quad E[(Z_{ht} - \mu_h)^2] = \sigma_h^2 \tag{6.6.9}
\]
\[
\text{Cov}[Z_{Kt}, u_t] = 0 \tag{6.6.10}
\]
\[
\text{Cov}[Z_{ht}, u_t] = 0 \tag{6.6.11}
\]

We can relate the optimal orthogonal portfolio parameters to the factor model deviations by comparing (6.6.1) and (6.6.6). Taking the unconditional expectations of both sides,
\[
a = \beta_h \mu_h, \tag{6.6.12}
\]

and by equating the variance of \(\epsilon_t\) with the variance of \(\beta_h Z_{ht} + u_t\),
\[
\Sigma = \beta_h \sigma_h^2 + \Phi = \alpha \sigma_h^2 + \Phi. \tag{6.6.13}
\]

The key link between the model deviations and the residual variances and covariances emerges from (6.6.13). The intuition for the link is straightforward. Deviations from the model must be accompanied by a common
component in the residual variance to prevent the formation of a portfolio with a positive deviation and a residual variance that decreases to zero as the number of securities in the portfolio grows, that is, an asymptotic arbitrage opportunin.

6.6.2 Squared Sharpe Ratios

The squared Sharpe ratio is a useful construct for interpreting much of the ensuing analysis. The tangency portfolio $q$ has the maximum squared Sharpe measure of all portfolios. The squared Sharpe ratio of $q$, $s_q^2$, is

$$s_q^2 = \mu' \Omega^{-1} \mu. \quad (6.6.14)$$

Given that the $K$ factor portfolios and the optimal orthogonal portfolio $h$ can be combined to form the tangency portfolio, the maximum squared Sharpe ratio of these $K+1$ portfolios will be $s_q^2$. Since $h$ is orthogonal to the portfolios $K$, MacKinlay (1995) shows that one can express $s_q^2$ as the sum of the squared Sharpe ratio of the orthogonal portfolio and the squared maximum Sharpe ratio of the factor portfolios,

$$s_q^2 = s_h^2 + s_K^2, \quad (6.6.15)$$

where $s_h^2 = \mu_h^2 / \sigma_h^2$ and $s_K^2 = \mu_K' \Omega_K^{-1} \mu_K.$

Empirical tests of multifactor models employ subsets of the $N$ assets. The factor portfolios need not be linear combinations of the subset of assets. Results similar to those above will hold within a subset of $N$ assets. For subset analysis when considering the tangency portfolio (of the subset), the maximum squared Sharpe ratio of the assets and factor portfolios, and the optimal orthogonal portfolio for the subset, it is necessary to augment the $N$ assets with the factor portfolios $K$. Defining $Z_s'$ as the $(N+K \times 1)$ vector $[Z', Z_{Ks}]'$ with mean $\mu^*_s$ and covariance matrix $\Omega^*_s$, for the tangency portfolio of these $N+K$ assets we have

$$s_q^2 = \mu^*_s' \Omega^*_s^{-1} \mu^*_s. \quad (6.6.16)$$

The subscript $s$ indicates that a subset of the assets is being considered. If any of the factor portfolios is a linear combination of the $N$ assets, it will be necessary to use the generalized inverse in (6.6.16).

$^7$This result is related to the work of Gibbons, Ross, and Shanken (1989).
6. Multifactor Pricing Models

The analysis (with a subset of assets) involves the quadratic $\mathbf{a}^T \Sigma^{-1} \mathbf{a}$ computed using the parameters for the $N$ assets. Gibbons, Ross, and Shanken (1989) and Lehmann (1987, 1992) provide interpretations of this quadratic term using Sharpe ratios. Assuming $\Sigma$ is of full rank, they show

$$\mathbf{a}_i^T \Sigma^{-1} \mathbf{a}_i = s_{h_i}^2 - s_K^2.$$  \hspace{1cm} (6.6.17)

Consistent with (6.6.15), for the subset of assets $\mathbf{a}^T \Sigma^{-1} \mathbf{a}$ is the squared Sharpe ratio of the subset's optimal orthogonal portfolio $h_i$. Therefore, for a given subset of assets:

$$s_{h_i}^2 = \mathbf{a}_i^T \Sigma^{-1} \mathbf{a}_i,$$  \hspace{1cm} (6.6.18)

and

$$s_{h_i}^2 = s_{h_i}^2 + s_K^2.$$  \hspace{1cm} (6.6.19)

Note that the squared Sharpe ratio of the subset's optimal orthogonal portfolio is less than or equal to that of the population optimal orthogonal portfolio, that is,

$$s_{h_i}^2 \leq s_{h^*}^2.$$  \hspace{1cm} (6.6.20)

Next we use the optimal orthogonal portfolio and the Sharpe ratios results together with the model deviation residual variance link to develop implications for distinguishing among asset pricing models. Hereafter the $s$ subscript is suppressed. No ambiguity will result since, in the subsequent analysis, we will be working only with subsets of the assets.

6.6.3 Implications for Separating Alternative Theories

If a given factor model is rejected a common interpretation is that more (or different) risk factors are required to explain the risk-return relation. This interpretation suggests that one should include additional factors so that the null hypothesis will be accepted. A shortcoming of this popular approach is that there are multiple potential interpretations of why the hypothesis is accepted. One view is that genuine progress in terms of identifying the "right" asset pricing model has been made. But it could also be the case that the apparent success in identifying a better model has come from finding a good within-sample fit through data-snooping. The likelihood of this possibility is increased by the fact that the additional factors lack theoretical motivation.

This section attempts to discriminate between the two interpretations. To do this, we compare the distribution of the test statistic under the null hypothesis with the distribution under each of the alternatives.

We reconsider the zero-intercept F-test of the null hypothesis that the intercept vector $\mathbf{a}$ from (6.6.1) is $0$. Let $H_0$ be the null hypothesis and $H_A$
be the alternative:

\[
H_0: \ a = 0 \\
H_A: \ a \neq 0.
\]

\(H_0\) can be tested using the test statistic \(J_1\) from (6.2.12):

\[
J_1 = \frac{(T - N - K)}{N} [1 + \mu_k' \hat{\Omega}_K^{-1} \hat{\mu}_k]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha},
\]

(6.6.21)

where \(T\) is the number of time-series observations, \(N\) is the number of assets or portfolios of assets included, and \(K\) is the number of factor portfolios. The hat superscripts indicate the maximum likelihood estimators. Under the null hypothesis, \(J_1\) is unconditionally distributed central \(F\) with \(N\) degrees of freedom in the numerator and \((T - N - K)\) degrees of freedom in the denominator.

To interpret deviations from the null hypothesis, we require a general representation for the distribution of \(J_1\). Conditional on the factor portfolio returns the distribution of \(J_1\) is

\[
J_1 \sim F_{N, T - N - K}(\delta),
\]

(6.6.22)

\[
\delta = \frac{T [1 + \hat{\mu}_K' \hat{\Omega}_K^{-1} \hat{\mu}_k]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{N},
\]

(6.6.23)

where \(\delta\) is the noncentrality parameter of the \(F\) distribution. If \(K = 0\) then the term \([1 + \hat{\mu}_K' \hat{\Omega}_K^{-1} \hat{\mu}_k]^{-1}\) will not appear in (6.6.21) or in (6.6.23), and \(J_1\) will be unconditionally distributed non-central \(F\).

We consider the distribution of \(J_1\) under two different alternatives, which are separated by their implications for the maximum value of the squared Sharpe ratio. With the risk-based multifactor alternative there will be an upper bound on the squared Sharpe ratio, whereas with the nonrisk-based alternatives the maximum squared Sharpe ratio is unbounded as the number of assets increases.

First consider the distribution of \(J_1\) under the alternative hypothesis that deviations are due to missing factors. Drawing on the results for the squared Sharpe ratios, the noncentrality parameter of the \(F\) distribution is

\[
\delta = T [1 + \hat{\mu}_K' \hat{\Omega}_K^{-1} \hat{\mu}_k]^{-1} s_h^2.
\]

(6.6.24)

From (6.6.20), the third term in (6.6.24) is bounded above by \(s_h^2\) and positive. The second term is bounded between zero and one. Thus there is an upper bound for \(\delta\).

\[
\delta < Ts_h^2 \leq Ts_s^2.
\]

(6.6.25)

The second inequality follows from the fact that the tangency portfolio \(q\) has the maximum Sharpe ratio of any asset or portfolio.
Given a maximum value for the squared Sharpe ratio, the upper bound on the noncentrality parameter can be important. With this bound, independent of how one arranges the assets to be included as dependent variables in the pricing model regression and for any value of $N$, there is a limit on the distance between the null distribution and the distribution of the test statistic under the missing-factor alternative. All the assets can be mispriced and yet the bound will still apply.

In contrast, when the alternative one has in mind is that the source of deviations is nonrisk-based, such as data snooping, market frictions, or market irrationalities, the notion of a maximum squared Sharpe ratio is not useful. The squared Sharpe ratio (and the noncentrality parameter) are in principle unbounded because the theory linking the deviations and the residual variances and covariances does not apply. When comparing alternatives with the intercepts of about the same magnitude, in general, one would expect to see larger test statistics in this nonrisk-based case.

We examine the informativeness of the above analysis by considering alternatives with realistic parameter values. We consider the distribution of the test statistic for three cases: the null hypothesis, the missing risk factors alternative, and the nonrisk-based alternative. For the risk-based alternative, the framework is designed to be similar to that in Fama and French (1993). For the nonrisk-based alternative we use a setup that is consistent with the analysis of Lo and MacKinlay (1990b) and the work of Lakonishok, Shleifer, and Vishny (1994).

Consider a one-factor asset pricing model using a time series of the excess returns for 32 portfolios as the dependent variable. The one factor (independent variable) is the excess return of the market so that the zero-intercept null hypothesis is the CAPM. The length of the time series is 342 months. This setup corresponds to that of Fama and French (1993, Table 9, regression (ii)). The null distribution of the test statistic $J_1$ is

$$J_1 \sim F_{32,309}(0)$$

(6.6.26)

To define the distribution of $J_1$ under the alternatives of interest one needs to specify the parameters necessary to calculate the noncentrality parameter. For the risk-based alternative, given a value for the squared Sharpe ratio of the optimal orthogonal portfolio, the distribution corresponding to the upper bound of the noncentrality parameter from (6.6.25) can be considered. The Sharpe ratio of the optimal orthogonal portfolio can be obtained using (6.6.15) given the squared Sharpe ratios of the tangency portfolio and of the included factor portfolio.

---

8In practice when using the F-test it will be necessary for $N$ to be less than $T-K$ so that $\Sigma$ will be of full rank.
MacKinlay (1995) argues that in a perfect capital markets setting, a reasonable value for the Sharpe ratio squared of the tangency portfolio for an observation interval of one month is 0.031 (or approximately 0.6 for the Sharpe ratio on an annualized basis). This value, for example, corresponds to a portfolio with an annual expected excess return of 10% and a standard deviation of 16%. If the maximum squared Sharpe ratio of the included factor portfolios is the \textit{ex post} squared Sharpe ratio of the CRSP value-weighted index, the implied maximum squared Sharpe ratio for the optimal orthogonal portfolio is 0.021. This monthly value of 0.021 would be consistent with a portfolio which has an annualized mean excess return of 8% and annualized standard deviation of 16%. We work through the analysis using this value.

Using this squared Sharpe ratio for the optimal orthogonal portfolio to calculate $\theta$, the distribution of $J_1$ from equation (6.2.1) is

$$J_1 \sim F_{32.309}(7.1).$$

(6.6.27)

This distribution will be used to characterize the risk-based alternative.

One can specify the distribution for two nonrisk-based alternatives by specifying values of $a$, $\Sigma$, and $\hat{\mu}_K \hat{\Omega}_K^{-1} \hat{\mu}_K$, and then calculating $\theta$ from (6.6.23). To specify the intercepts we assume that the elements of $a$ are normally distributed with a mean of zero. We consider two values for the standard deviation, 0.0007 and 0.001. When the standard deviation of the elements of $a$ is 0.001 about 95% of deviations will lie between $-0.002$ and $+0.002$, an annualized spread of about 4.8%. A standard deviation of 0.0007 for the deviations would correspond to an annual spread of about 3.4%. These spreads are consistent with spreads that could arise from data-snooping. They are plausible and even somewhat conservative given the contrarian strategy returns presented in papers such as Lakonishok, Shleifer, and Vishny (1993). For $\Sigma$ we use a sample estimate based on portfolios sorted by market capitalization for the Fama and French (1993) sample period 1963 to 1991. The effect of $\hat{\mu}_K \hat{\Omega}_K^{-1} \hat{\mu}_K$ on $\theta$ will typically be small, so it is set to zero. To get an idea of a reasonable value for the noncentrality parameter given this alternative, the expected value of $\theta$ given the distributional assumption for the elements of a conditional upon $\Sigma = \hat{\Sigma}$ is considered. The expected value of the noncentrality parameter is 39.4 for a standard deviation of 0.0007 and 80.3 for a standard deviation of 0.001. Using these values for the noncentrality parameter, the distribution of $J_1$ is

$$J_1 \sim F_{32.309}(39.4)$$

(6.6.28)

\footnote{With data-snooping the distribution of $J_1$ is not exactly a noncentral F (see Lo and MacKinlay [1990b]). However, for the purposes of this analysis, the noncentral F will be a good approximation.}
when $a_0 = 0.0007$ and

$$J_1 \sim F_{32,309}(80.3)$$

(6.6.29)

when $a_0 = 0.001$.

A plot of the four distributions from (6.6.26), (6.6.27), (6.6.28), and (6.6.29) is in Figure 6.1. The vertical bar on the plot represents the value 1.91 which Fama and French calculate for the test statistic. From this figure, notice that the distributions under the null hypothesis and the risk-based alternative hypothesis are quite close together.\(^{10}\) This reflects the impact of the upper bound on the noncentrality parameter. In contrast, the nonrisk-based alternatives' distributions are far to the right of the other two distributions, consistent with the unboundedness of the noncentrality parameter for these alternatives.

Given that Fama and French find a test statistic of 1.91, these results suggest that the missing-risk-factors argument is not the whole story. From Figure 6.1 one can see that 1.91 is still in the upper tail when the distribution of $J_1$ in the presence of missing risk factors is tabulated. The $p$-value using this distribution is 0.03 for the monthly data. Hence it seems unlikely that missing factors completely explain the deviations.

The data offer some support for the nonrisk-based alternative views. The test statistic falls almost in the middle of the nonrisk-based alterna-

\(^{10}\) See MacKinlay (1987) for detailed analysis of the risk-based alternative.
tive with the lower standard deviation of the elements of $a$. Several of the nonrisk-based alternatives could equally well explain the results. Different nonrisk-based views can give the same noncentrality parameter and test-statistic distribution. The results are consistent with the data-snooping alternative of Lo and MacKinlay (1990b), with the related sample selection biases discussed by Breen and Korajczyk (1993) and Kothari, Shanken, and Sloan (1995), and with the presence of market inefficiencies.

### 6.7 Conclusion

In this chapter we have developed the econometrics for estimating and testing multifactor pricing models. These models provide an attractive alternative to the single-factor CAPM, but users of such models should be aware of two serious dangers that arise when factors are chosen to fit existing data without regard to economic theory. First, the models may overfit the data because of data-snooping bias; in this case they will not be able to predict asset returns in the future. Second, the models may capture empirical regularities that are due to market inefficiency or investor irrationality; in this case they may continue to fit the data but they will imply Sharpe ratios for factor portfolios that are too high to be consistent with a reasonable underlying model of market equilibrium. Both these problems can be mitigated if one derives a factor structure from an equilibrium model, along the lines discussed in Chapter 8. In the end, however, the usefulness of multifactor models will not be fully known until sufficient new data become available to provide a true out-of-sample check on their performance.

### Problems—Chapter 6

**6.1** Consider a multiple regression of the return on any asset or portfolio $R_a$ on the returns of any set of portfolios from which the entire minimum-variance boundary can be generated. Show that the intercept of this regression will be zero and that the factor regression coefficients for any asset will sum to unity.

**6.2** Consider two economies, economy A and economy B. The mean excess-return vector and the covariance matrix is specified below for each of the economies. Assume there exist a riskfree asset, $N$ risky assets with mean excess return $\mu$ and nonsingular covariance matrix $\Omega$, and a risky factor portfolio with mean excess return $\mu_p$ and variance $\sigma_p^2$. The factor portfolio is not a linear combination of the $N$ assets. (This criterion can be met by eliminating one of the assets which is included in the factor portfolio
if necessary.) For both economies A and B:

\[ \mu = a + \beta \mu_p \]  

(6.7.1)

\[ \Omega = \beta \beta' \sigma^2_p + \delta \delta' \sigma^2_h + I \sigma^2_e . \]  

(6.7.2)

Given the above mean and covariance matrix and the assumption that the factor portfolio \( \mu_p \) is a traded asset, what is the maximum squared Sharpe ratio for the given economies?

6.3 Returning to the above problem, the economies are further specified. Assume the elements of \( a \) are cross-sectionally independent and identically distributed,

\[ a_i \sim \text{IID}(0, \sigma^2_a) \quad i = 1, \ldots, N. \]  

(6.7.3)

The specification of the distribution of the elements of \( \delta \) conditional on \( a \) differentiates economies A and B. For economy A:

\[ \delta_i \mid a \sim \text{IID}(a_i, 0) \quad i = 1, \ldots, N, \]  

(6.7.4)

and for economy B:

\[ \delta_i \mid a \sim \text{IID}(0, \sigma^2_\delta) \quad i = 1, \ldots, N. \]  

(6.7.5)

Unconditionally the cross-sectional distribution of the elements of \( \delta \) will be the same for both economies, but for economy A conditional on \( a \), \( \delta \) is fixed. What is the maximum squared Sharpe ratio for each economy? What is the maximum squared Sharpe ratio for each economy as the \( N \) increases to infinity?