Testing the APT with the Maximum Sharpe Ratio of Extracted Factors

Chu Zhang
Department of Finance, Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong, czhang@ust.hk

This paper develops a test of the asymptotic arbitrage pricing theory (APT) via the maximum squared Sharpe ratio of the factors extracted from individual stocks using the Connor-Korajczyk method. The test treats the beta pricing relation as approximate without predetermining the systematic factors, unlike the existing tests that take the relationship as exact and systematic factors as given. This paper also examines the magnitude of pricing errors bounded partly by the maximum squared Sharpe ratio. For most 60-month subperiods of the sample, the hypothesis that the maximum squared Sharpe ratio for monthly returns is greater than 0.25 can be rejected. Simulation indicates that the average pricing error in monthly returns is less than 0.001. These results support the asymptotic APT.

Key words: asymptotic APT; beta pricing errors; Sharpe ratio; Connor-Korajczyk method; eigenvalue; eigenvector

1. Introduction

The mainstream asset pricing models, i.e., the capital asset pricing model (CAPM) developed by Sharpe (1964), the intertemporal capital asset pricing models by Merton (1973), and the arbitrage pricing theory (APT) by Ross (1976), are based on the notion of systematic risks that are represented by marketwide factors. Expected returns on individual securities are linear functions of their standardized covariances, or betas, with the marketwide factors. Despite their vintage, these models remain the most commonly used in theoretical and empirical analyses by academics and practitioners. The most popular method to test a beta pricing model has been the method of Gibbons et al. (1989; GRS hereafter). Although the GRS test does a perfect job for what it is meant to do, it has two shortcomings as a test of general beta pricing theories. First, as the title of the GRS paper makes clear, it examines the pricing error from the beta pricing model with respect to a set of given portfolio returns as systematic factors. If the test rejects a particular model, it is always possible that the factors in the test are misspecified, although the beta pricing principle is still correct. Second, because the test requires that the number of testing assets be much smaller than the number of time-series observations, a beta pricing model is typically tested on a small number of portfolios as the testing assets. If the test does not reject a model, it does not imply that the model works for all the assets.1

The purpose of the current paper is to develop a test of the beta pricing theory using factors extracted from all available individual stock returns. The method of extracting factors was developed by Connor and Korajczyk (1986, 1988) and has been widely used in testing asset pricing anomalies.2 The factors extracted by the Connor-Korajczyk (CK) method guarantee that they are systematic factors, but there is no guarantee that the pricing errors associated with these factors are bounded as the theory claims. Surprisingly, however, the implication of the asymptotic APT that the error from the beta pricing relationship should be contained has never been tested in this context, although the method of extracting factors has been well developed. The test developed in this paper examines the sum of squared pricing errors of all individual assets with respect to extracted factors that are

1 Another popular methodology, the two-pass methodology, also takes the factors as given. It examines whether the factors are priced (i.e., the factor premiums are nonzero), typically without examining whether the pricing errors are zero.

2 For example, Connor and Korajczyk (1986) apply the method to mutual fund performance evaluation, McCulloch and Rossi (1991) use the CK factors to test the firm size anomaly, and Brennan et al. (1998) extract the CK factors to examine the size and book-to-market effects.
There has been a plethora of evidence in the recent literature that challenges various beta pricing models with prespecified factors. The new evidence provided here points to the possibility that the evidence more likely challenges the completeness of the prespecified factors, rather than the beta pricing theory per se. This second contribution, likewise, is limited to the case of unconditional models.

The rest of this paper is organized as follows. Section 2 presents theoretical results on the implication of the asymptotic APT regarding the number of unbounded eigenvalues of the second-moment and variance matrices. The connection between extracted factors and the Sharpe ratio is given as well, followed by the test statistic based on the minimum eigenvalue of the variance matrix of the extracted eigenvectors. Section 3 reports empirical results on the maximum squared Sharpe ratio using individual stock returns from the United States in eight 60-month sample periods during 1965–2004. Section 4 conducts simulations about the magnitude of the average pricing error. The last section concludes the paper, and the appendix presents the proof of the analytical results.

2. Methodology

2.1. Theoretical Foundation

Suppose that \( r_t \) is the vector of returns in excess of the risk-free rate on \( n \) assets in month \( t \). In a factor model, the excess returns are driven by

\[
   r_t = a + B f_t + e_t, \tag{1}
\]

where \( a \) is an \( n \)-vector of constants, known as Jensen’s (1968) alpha, \( B \) is an \( n \times k \) matrix of betas of the \( k \)-vector of factors, \( f_t \), and \( e_t \) is the idiosyncratic risk of the \( n \) assets, satisfying \( E(e_t | f_t) = 0 \). Without loss of generality, we can assume that \( a \) is orthogonal to \( B \); i.e., \( B a = 0 \).\(^4\) The factors may be unobserved with mean \( \mu_f \) and variance \( \Sigma_f \). Without loss of generality, \( \Sigma_f \) is assumed to be positive definite, which means that \( f_t \) contains no redundancy. The idiosyncratic risk has a variance matrix denoted as \( \Sigma_e \). The variance matrix of \( r_t \) is \( \Sigma_r = B \Sigma_f B' + \Sigma_e \).

When \( f_t \) itself is the excess returns on certain portfolios, \( a \) represents pricing errors of the beta pricing theory: \( E(r_t) = B E(f_t) \). Huberman (1982) assumes that idiosyncratic risks are uncorrelated and establishes the asymptotic APT result that the sum of the squared pricing errors, \( a' a \), is bounded. Chamberlain

\(^4\) For a given \( u \) and the \( f_t \) constructed as the returns on portfolios of the assets on the left-hand side, \( f_t = \Gamma r_t \); multiplying \( \Gamma \) to (1) gives \( \Gamma a = 0 \) and \( \Gamma B = \Lambda_f \), which implies \( B a = 0 \). As \( u \) increases, the identity of \( f_t \) changes with \( r_t \), and \( B a = 0 \) remains true. For nonreturn factors, we can let \( \tilde{a} = a - B(\tilde{B}' \tilde{B})^{-1} \tilde{B} a, \tilde{f}_t = f_t + (\tilde{B}' \tilde{B})^{-1} \tilde{B} a, \) and rewrite (1) as \( r_t = \tilde{a} + B \tilde{f}_t + e_t \) with \( B \tilde{a} = 0 \).

---

\(^3\) One exception is Stambaugh (1983), who provides a conditional version of the APT.
and Rothschild (1983) relax the assumption of the uncorrelated idiosyncratic risk in Huberman (1982) and propose an approximate $k$-factor model, in which $k$ largest eigenvalues of $\Sigma$ increase without bound, whereas the other eigenvalues remain bounded as the number of securities, $n$, tends to infinity. In the approximate factor model, the idiosyncratic risk can be cross-sectionally correlated. If the asymptotic APT holds, the returns in an approximate factor model can also be written as (1), where the factors can be taken as the $k$ linear combinations of the returns where the coefficients are the $k$ eigenvectors of $\Sigma$, corresponding to the $k$ largest eigenvalues and the idiosyncratic risks are the linear combinations of returns associated with the remaining eigenvectors. The implication of the APT fails if, in (1), $a'$ is unbounded as $n$ goes to infinity.

The Connor-Korajczyk method of extracting factors is based on the second-moment matrix of the returns, rather than on the variance matrix, to deal with the situation in which the number of assets is much greater than the number of time-series return observations. This introduces a potential problem unnoticed by previous researchers who used the CK factors in applications regarding the predictive power of firm-specific variables. The problem is that the extracted “factors” may be contaminated by pricing errors. This can be seen more easily in a one-factor model, $r_t = a + b f_t + e_t$, with $a'b = 0$. The variance matrix of the returns has one unbounded eigenvalue equal to $(b'b)\sigma_f^2$, where $\sigma_f^2$ is the variance of $f_t$. If $a'a$ is also unbounded, the second-moment matrix of the returns has two unbounded eigenvalues, $a'a$ and $(b'b)\sigma_f^2$, where $\sigma_f^2$ is the second-moment of $f_t$. Now let $x$ be a constant $n$-vector of a firm-specific variable that is highly, cross-sectionally correlated with $a$. Naturally, the expected excess returns are found to be related to the market beta, $b$, and the firm-specific variable, $x$, because $x$ is highly correlated with $a$. To investigate whether $x$ is a proxy for the beta of an unobserved factor, two “factors” are extracted from the second-moment matrix of the returns because the matrix has two unbounded eigenvalues. Ignoring finite-sample errors, the “factors” can be written as $g_t = C(1 \ f_t)'$ for a $2 \times 2$ nonsingular matrix, $C$.\footnote{Suppose $f_t$ is standardized such that $s^2_f = 1$. As $n \to \infty$, the second-moment matrix of $r_t$ is asymptotically equivalent to $(a'a + b'b)'$, which has two nonzero eigenvalues, $a'a$ and $b'b$, with corresponding eigenvectors, $a$ and $b$. The (unscaled) mimicking portfolios are $g_t = (a, b)' r_t = (a, b)' [(a, b)' (1)] + e_t$, which, as $n \to \infty$, approaches 
\[
\begin{pmatrix}
a' a \\
b' b
\end{pmatrix}
\begin{pmatrix}
1 \\
f_t
\end{pmatrix}
= C(1 \ f_t).\]
The problem remains if the extracted “factors” contain the idiosyncratic component of the excess returns.}

The $n \times 2$ beta matrix of the extracted “factors” are
\[
B_g = E(r_t g_t' ) [E(g_t g_t') ]^{-1} = (a \ b) C^{-1}.
\]

Therefore, $B_g$ spans the same subspace as $(a \ b)$. Given $B_g$, $x$ will have no additional explanatory power for the expected excess returns. The puzzle of the dependence of expected returns on $x$ is then declared as being solved, and a two-factor beta pricing model holds. This inference, of course, is erroneous because we began with a model for which the beta pricing theory fails. This highly simplified example illustrates the danger of using extracted “factors” without verifying their validity.\footnote{Ferson et al. (1999) make a similar point for factors constructed from the firm-specific variables, rather than extracted statistically.}

The features of pricing errors seen in the one-factor example can be generalized to the case of an arbitrary $k$. Let $S_t = E(r_t r_t')$ be the second-moment matrix of $r_t$. From (1),
\[
S_t = a a' + B_S B' + \Sigma_{r_t}, \tag{2}
\]
where $S_t = E(f_t f_t')$ is the second-moment matrix of $f_t$. A test of the existence of unbounded pricing errors can be based on the following proposition.

**Proposition 1.** Suppose that the returns follow the approximate $k$-factor model (1). Let $k^*$ be the number of unbounded eigenvalues of the second-moment matrix of the excess returns, $S_t$. Then, as $n$ goes to infinity, $k^* = k + 1$ if $a'a$ is unbounded, or $k^* = k$ if $a'a$ is bounded.

By writing $a + B f_t = (a \ b)(1 \ f_t)'$, we see that the difference between a $k$-factor model with a pricing error and a $(k+1)$-factor model without a pricing error is that, in the model with a pricing error, the “factors” contain a constant. The pricing error, $a$, is contained in the second-moment matrix of the returns, but it is absent from the variance matrix because the variance is invariant of an additive constant. Suppose that $g_t$ is a $k^*$-vector of extracted “factors” from the second-moment matrix of returns. From Proposition 1, we know that, as the number of stocks tends to infinity, $g_t$ is a linear transformation of $f_t$ if the pricing error is bounded, or a linear transformation of $(1 \ f_t')$ if the pricing error is unbounded. We can write $g_t = C f_t$ where the dimension of $g_t$ is $k^*$. Either $f_t^* = f_t$ or $f_t^* = (1 \ f_t')$ depending on whether there is an unbounded pricing error. The matrix $C$ is nonsingular because $\Sigma_f$ is assumed to be positive definite. In terms of extracted “factors,” the difference between a $k$-factor model with a pricing error and a $(k+1)$-factor model without a pricing error is that, in the former situation, the variance matrix of $g_t$ is degenerate, whereas in the latter, the variance matrix is positive definite.
The proposition gives a simple criterion to test the existence of unbounded pricing errors if the number of unbounded eigenvalues can be easily determined from the variance matrix and the second-moment matrix of the excess returns. Unfortunately, the econometric issue of determining the number of unbounded eigenvalues from the sample version of $\Sigma$ and $S$ turns out to be difficult to resolve.\(^7\) In short, there has been no consensus in the literature on what is the best way to determine the number of systematic factors, and there has been no consensus on the actual number of factors found in the U.S. stock market.

Fortunately, we find that the inferences using our new test appear relatively insensitive to the number of factors, given $k > 1$. The new test is related to the maximum squared Sharpe ratio of all excess returns, defined as

$$s = \max_{u, u \neq 0} u \Sigma u / \text{Var}(u' r).$$

The Sharpe ratio plays an important role in modern finance, representing the trade-off between the expected return and the standard deviation of the return among efficient portfolios. The mean variance portfolio analysis can be recast as maximizing the Sharpe ratio of the portfolio. The CAPM implies that the market portfolio has the maximum Sharpe ratio among all portfolios of risky assets. MacKinlay (1995) uses the squared Sharpe ratio to examine the plausibility for multifactor models to explain anomalies to the CAPM. In more recent work, Cochrane and Saa-Requejo (2000) derive good-deal asset pricing bounds, using an idea similar to that of restricting the Sharpe ratio, for pricing derivatives in the discrete-time framework where markets are incomplete.\(^8\)

Chamberlain and Rothschild (1983) make formal connections between mean variance analysis and the arbitrage pricing theory. No-arbitrage conditions require that the maximum Sharpe ratio be finite. With an approximate factor structure, no arbitrage implies that the pricing error is partly bounded by the squared Sharpe ratio as in the inequality

$$a' a \leq s \hat{\sigma}_e^2,$$

\(^3\)Trzcinka (1986) finds that the largest eigenvalue of the variance matrix dominates the rest of the eigenvalues. Brown (1989), however, argues convincingly that one dominant eigenvalue does not mean that there is only one systematic factor in the returns. Connor and Korajczyk (1993) propose a method to determine the number of factors that measures the marginal contribution of an additional factor and report findings of one to six factors for various subperiods. Geweke and Zhou (1996) apply a Bayesian approach and find little improvement in reducing pricing errors by having additional factors beyond the first one. Bai and Ng (2002) design various test statistics under general factor structures and draw the conclusion that two factors are adequate for U.S. stock returns.

\(^4\)A good-deal bound is imposed on the volatility of all stochastic discount factors that price primitive assets. Because of the relationship between the volatility of stochastic discount factors and the Sharpe ratio, the volatility bound is also an upper bound for the squared Sharpe ratio.

where $\hat{\sigma}_e^2 = \max_{u, u \neq 0} u \Sigma u / u' u$ is an upper bound of the idiosyncratic risk. The test developed in this paper is based on the maximum squared Sharpe ratio of the CK factors, which are orthonormal eigenvectors of the sample second-moment matrix of the excess returns satisfying $E g_i g_i' = I_r$. Let the mean of $g_i$ be denoted as $\mu_g$. The proposition below characterizes the theoretical properties of the CK factors.

**Proposition 2.** Let $g_i$ be the $k$-vector of “factors” extracted from the second-moment matrix of the returns, normalized to have $E g_i g_i' = I_r$. Denote $\mu_g = E g_i$ and $\gamma = \mu_g' \mu_g$. Then, (i) $\gamma = \gamma(1 - \gamma)$ if and only if $\gamma = 1$ as $n \to \infty$; and (ii) the maximum squared Sharpe ratio equals $s = \gamma/(1 - \gamma)$, or infinity if $\gamma = 1$.\(^5\)

Proposition 2 suggests that we can test APT through $s$ or $\gamma$. If the implication of the APT is confined to the inequality (3) with $s$ unspecified, then, as Shanken (1992) correctly points out, testing the APT is not very meaningful because that inequality constraint is more or less a tautology. For the asymptotic APT to be useful, the expected returns to be well approximated by linear functions of factor betas, $s$ has to be small despite the fact that the notion of no asymptotic arbitrage requires only that $s$ be finite.

Because the standard APT does not make assumptions on investors’ preferences, the theory does not involve the determination of the maximum squared Sharpe ratio, $s$. What data say about the existence of asymptotic arbitrage opportunities and, more generally, about the value of the maximum squared Sharpe ratio, is certainly a meaningful question to ask. For practical purposes, a large $s$ can be regarded as the failure of the asymptotic APT. In the empirical tests below, we estimate maximum Sharpe ratios using individual stocks and examine a range of potential values for the limiting maximum Sharpe ratio.

### 2.2. Econometric Method

Let $R = (r_1, \ldots, r_\tau)$ be the observations of $r_t$ for $t = 1, \ldots, \tau$, $F = (f_1, \ldots, f_\tau)$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_\tau)$. In matrix form,

$$R = aI_\tau + BF + \varepsilon,$$

\(^6\)where $1_\tau$ is the $\tau$-vector of ones. Let

$$\Omega = \frac{1}{n} R' R.$$

For a given number $k^*$, as a candidate of the number of unbounded factors, the Connor and Korajczyk (1986, 1988) method is to extract (transformed) factors as the $k^* \times \tau$ matrix of eigenvectors corresponding to the largest $k^*$ eigenvalues of $\Omega$. It is shown that, in the case of $a = 0$, the eigenvectors converge to a nonsingular linear transformation of the factors, $F$, as $n$ goes to infinity. They do not consider the pricing
errors, however. With an unbounded pricing error, \( a \), the extracted factors are a transform of \((1, F')\) in the limit as \( n \) goes to infinity, rather than just \( F \), as we can see formally that \( R = (a \ B)(1, F') + e \).

The existence of a systematic pricing error will show up in the difference between the second-moment and variance matrices of the eigenvectors. Let \( \tau^{-1/2} G \) be the \( k^* \times \tau \) orthonormal matrix of the \( k^* \) eigenvectors of \( \Omega \) corresponding to the \( k^* \) largest eigenvalues of \( \Omega \), arranged as row vectors. Let \( \tilde{g} = (1/\tau)G_{1*} \). By construction, \( \tau^{-1/2}GG' \) has \( k^* \) eigenvalues all equal to one. To see whether all the \( k^* \) eigenvectors are true factors or one of them is a pricing error after transformation, we can examine the smallest eigenvalue of \( \tau^{-1}(G - \tilde{g}1_\tau)'(G - \tilde{g}1_\tau) \), denoted as \( \lambda_{k^*} \). If the pricing error is unbounded, then in the limit as \( n \) goes to infinity, \( \tau^{-1/2}G \) is transformed from \((1, F')\), which contains a constant row. As a result, \( \tau^{-1}(G - \tilde{g}1_\tau)'(G - \tilde{g}1_\tau) \) will be degenerate. This can be found by testing \( \lambda_{k^*} = 0 \). The conditions and the distribution of the formal test are stated in the following proposition.

**Proposition 3.** Suppose that \( \Omega = (1/n)R'R \) is a \( \tau \times \tau \) positive definite random matrix, the \( k^* \times \tau \) matrix \( \tau^{-1/2}G \) is the \( k^* \) orthonormal eigenvectors of \( \Omega \) corresponding to the largest \( k^* \) eigenvalues, \( \tilde{g} = (1/\tau)G_{1*} \), and \( \tilde{g} = \tilde{g} \tilde{g}' \).

Then,

(i) \( \tau^{-1}(G - \tilde{g}1_\tau)'(G - \tilde{g}1_\tau) \) has \( k^* - 1 \) eigenvalues equal to one and one eigenvalue, \( \lambda_{k^*} = 1 - \gamma \), between zero and one, both inclusive.

(ii) Suppose that \( F \) is normally distributed. Then, the asymptotic distribution of \( \tilde{g} \) is a noncentral beta with density function

\[
p(x) = \frac{\Gamma(\tau/2)}{\Gamma((\tau - k^*)/2)\Gamma(\tau/2)}x^{k^*/2 - 1}(1 - x)^{(\tau - k^*)/2 - 1}e^{-\gamma s/2}F_1\left(\frac{\tau}{2}, \frac{\tau s}{2}, \frac{x}{2}, \frac{1}{2} \right), \quad 0 \leq x \leq 1,
\]

where \( s = \gamma/(1 - \gamma) \) is the maximal squared Sharpe ratio, \( \gamma = p'\mu_p' \mu_p \), \( \mu_p = E\mu_p \), and \( F_1(\cdot, \cdot, \cdot) \) is the confluent hypergeometric function.

When \( s = 0 \), the distribution is the familiar central beta distribution with \((k^*/2, (\tau - k^*)/2)\) degrees of freedom. In typical applications, \( \tau \gg k^* \), so the majority of the mass of the distribution leans toward zero. A noncentral beta distribution with a noncentrality parameter, \( \tau s/2 \), shifts the mass of the distribution to the right as \( s \) becomes greater. In the extreme case where \( s \rightarrow \infty \), the distribution becomes degenerate and concentrates on one. Figure 1 depicts three such noncentral beta density functions corresponding to \( s = 0.25, s = 1, \) and \( s = 9 \) (or \( \gamma = 0.2, 0.5, \) and 0.9) with \( \tau = 60 \) and \( k^* = 1, 3, 5, 10 \). One observation from the figure is that, as \( k^* \) becomes greater, the distribution shifts more to the right. For \( k^* = 1 \), the modal point of the distribution is close to the parameter, \( \gamma \). For \( k^* = 10 \), however, the modal point of the distribution is greater than the parameter, \( \gamma \). The distribution is not extremely sensitive to \( k^* \), however, especially for large \( \gamma \).

Proposition 3 suggests the following way of testing the asymptotic APT. For a plausible value of \( k^* \), we take \( k^* \) orthonormal eigenvectors of \( \Omega \), \( \tau^{-1/2}G \). We then calculate \( \tilde{g} \) from \( \tilde{g} \). We reject the asymptotic APT if \( \tilde{g} \) is large. The rejection decision can be based on the asymptotic distribution in (6). In particular, suppose that we set the null hypothesis to \( s = 0.25 \), we can test the hypothesis by calculating the left-tail \( p \)-value of the test statistic, \( \gamma \), using the distribution corresponding to \( \gamma = 0.2 \). If the hypothesis \( \gamma \geq 0.2 \) is rejected, the evidence is then in favor of the asymptotic APT.

The test in Proposition 3 still involves \( k \). To partially circumvent the difficulty in determining the number of factors, the following scheme is adopted. Suppose, from the existing work on the number of factors, we know that the actual number of factors, \( k \), is below a certain number, \( K \), although we are not sure what exactly \( k \) is. For example, for the case of the U.S. stock returns, the literature cited above leads to the belief that \( k \leq K = 20 \). For any \( k^* \leq K \), we use Proposition 3 to test the asymptotic APT. If for all \( k^* \leq K \), the test assuming \( k^* \) rejects \( \gamma \geq 0.2 \), then the overall test rejects \( \gamma \geq 0.2 \). However, if one test assuming a particular \( k^* \leq K \) does not reject \( \gamma \geq 0.2 \) at a given test size,
then the overall test does not reject $\gamma \geq 0.2$. The size of the overall test will be smaller than that of individual tests assuming the value of $k^*$. As such, the test scheme partially circumvents the problem of not knowing $k^*$. The test in Proposition 3 takes $\tau$ as given. It can be shown that a larger $\tau$ with additional time-series observations results in a tighter distribution (i.e., a smaller variance) of the test statistic, $\tilde{\gamma}$, and therefore, a more powerful test. In practice, however, it is not always preferable to use the largest possible $\tau$. As $\tau$ becomes larger, the number of stocks that have $\tau$ consecutive observations becomes smaller. A balance between $\tau$ and $n$ needs to be struck. In addition, over a longer period, the potential problems of heteroskedasticity and time-varying betas in the returns become more serious. The latter concern gives rise to a common practice in the empirical asset pricing studies that use 60 months as a standard choice of the number of time-series observations for monthly data. We follow this practice in this paper.

The discussion so far is based on the returns in excess of the risk-free rate. When the risk-free asset does not exist, the Sharpe ratio can still be defined for all zero-cost portfolios as in Chamberlain and Rothschild (1983). In that context, the approximate beta pricing still holds where the excess returns are returns in excess of a zero-beta rate, i.e., the expected return on a systematic factor whose covariance with other factors is zero. If the zero-beta rate is known, all the analysis remains the same. However, if the zero-beta rate is unknown, it needs to be estimated.

3. Empirical Results

The data used for the empirical tests are the monthly stock returns at the firm level in the United States between 1965 and 2004. The data are from the Center for Research in Security Prices at the University of Chicago. All the common stock traded on the NYSE/AMEX/NASDAQ are included except for American Deposit Receipts. Preferred stocks are excluded, but common stocks of utility companies and financial companies are not excluded. The one-month $T$-bill rate is used as the risk-free rate to calculate the excess returns.

Following the convention in the empirical asset pricing literature, the entire sample period is broken into eight 60-month subperiods, and tests of the asymptotic APT are conducted within each subperiod. For each subperiod, stocks with nonmissing monthly returns are collected, and the second-moment matrix, \( \bar{\Omega} = (1/\tau)R'R \), is calculated where $R$ is the $n \times \tau$ excess returns with $n$ being the number of stocks with nonmissing returns and $\tau = 60$ being the number of observations for each stock. Note that $\bar{\Omega}$ differs from $\Omega$ by a factor of $n/\tau$, and $\bar{\Omega}$ has the same eigenvalues as the second-moment matrix of the excess returns, $(1/n)RR'$. From $\bar{\Omega}$, 60 eigenvalues are obtained and are arranged in descending order. Table 1 reports for each subperiod the first 10 eigenvalues, $\bar{l}_j$, $j = 1, \ldots, 10$, the proportion of each eigenvalue in the total, $\bar{l}_j/L$, where $L = \sum_{j=1}^{60} \bar{l}_j$, the first $j$ cumulative eigenvalues, $\sum_{j=1}^{\infty} \bar{l}_j$, and the proportional first $j$ cumulative eigenvalues, $\sum_{j=1}^{\infty} \bar{l}_j/L$. The results for $k = 11, \ldots, 20$ are not reported to save space.

The number of stocks in each period ranges from 1,487 to 3,858, and the sum of the eigenvalues, $L$, ranges from 21 to 139. Early periods have relatively smaller numbers of stocks with nonmissing values. The largest sum of the eigenvalues occurs in the last subperiod in which stock prices are very volatile. The big differences across subperiods is one reason for testing within each subperiod. The first eigenvalue accounts for 12% to 34% of the total sum of all the eigenvalues. The second eigenvalue is considerably smaller than the first one. The rest of them decline gradually without showing a clear point on the number of dominant eigenvalues. Though not reported here, the pattern continues until $j = 60$. This phenomenon has been documented in the literature and causes difficulty in determining the number of factors in the stock returns.

The Connor-Korajczyk method is then used to extract factors from the second-moment matrix of the excess returns, $\bar{\Omega}$. For an integer $k$, let $\tau^{-1/2}G = (g_1, \ldots, g_\tau)$ be the $k$ eigenvectors of $\bar{\Omega}$ corresponding to the largest $k$ eigenvalues of $\bar{\Omega}$. As explained earlier, the smallest eigenvalue of $\tau^{-1}(G - \bar{\Omega})G$ is $1 - \bar{\gamma}$ where $\bar{\gamma} = \hat{\gamma}/\tilde{\gamma}$ and $\hat{\gamma}$ is the $k$-vector mean of $G$. Proposition 3 establishes that $\bar{\gamma}$ has a noncentral beta distribution whose noncentrality parameter is a multiple of $s$, where $s$ is the asymptotic maximum squared Sharpe ratio. Table 2 reports the statistic $\bar{\gamma}$ for $k = 1, 2, \ldots, 10$. Beside $\bar{\gamma}$ is $\bar{\gamma}/(1 - \bar{\gamma})$. The left-tail $p$-values of $\bar{\gamma}$ under the hypotheses $\gamma = 0.5$, $\gamma = 0.2$, and $\gamma = 0.1$, which correspond to $s = 1, s = 0.25$, and $s = 0.11$, are reported. To put these hypothesized $s$ values in perspective, we note that the monthly based sample squared Sharpe ratio of the value-weighted market portfolio in the U.S. market during the 1965-2004 period is less than 0.01.

The test statistic, $\bar{\gamma}$, ranges from 0.016 to 0.188, with corresponding $\bar{\gamma}/(1 - \bar{\gamma})$ from 0.016 to 0.231, depending on the subperiod and $k$. The left-tail $p$-values at $\gamma = 0.5$ are virtually zero, and the left-tail $p$-value at $\gamma = 0.2$ is
Table 1: Eigenvalue Decomposition of the Second-Moment Matrix of Individual Stocks

<table>
<thead>
<tr>
<th>Period</th>
<th>n</th>
<th>L</th>
<th>j</th>
<th>( \sum_{i=1}^{j} l_i )</th>
<th>( l_i/L )</th>
<th>( \sum_{i=1}^{j} l_i/L )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965–1969</td>
<td>1.487</td>
<td>21.0684</td>
<td>1</td>
<td>6.077</td>
<td>0.288</td>
<td>0.288</td>
</tr>
<tr>
<td>1970–1974</td>
<td>1.824</td>
<td>29.4620</td>
<td>1</td>
<td>10.027</td>
<td>0.340</td>
<td>0.340</td>
</tr>
<tr>
<td>1975–1979</td>
<td>3.045</td>
<td>57.8047</td>
<td>1</td>
<td>17.380</td>
<td>0.301</td>
<td>0.301</td>
</tr>
<tr>
<td>1980–1984</td>
<td>3.063</td>
<td>55.1144</td>
<td>1</td>
<td>11.400</td>
<td>0.207</td>
<td>0.207</td>
</tr>
<tr>
<td>1985–1989</td>
<td>3.352</td>
<td>73.6322</td>
<td>1</td>
<td>12.176</td>
<td>0.165</td>
<td>0.165</td>
</tr>
<tr>
<td>1990–1994</td>
<td>3.812</td>
<td>115.4322</td>
<td>1</td>
<td>18.010</td>
<td>0.156</td>
<td>0.156</td>
</tr>
</tbody>
</table>

Table 1 (Continued)

<table>
<thead>
<tr>
<th>Period</th>
<th>n</th>
<th>L</th>
<th>j</th>
<th>( \sum_{i=1}^{j} l_i )</th>
<th>( l_i/L )</th>
<th>( \sum_{i=1}^{j} l_i/L )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995–1999</td>
<td>3.858</td>
<td>116.1262</td>
<td>1</td>
<td>14.115</td>
<td>0.122</td>
<td>0.122</td>
</tr>
<tr>
<td>2000–2004</td>
<td>3.708</td>
<td>139.2781</td>
<td>1</td>
<td>30.214</td>
<td>0.217</td>
<td>0.217</td>
</tr>
</tbody>
</table>

Notes. This table reports the second-moment decomposition of individual stocks in eight 60-month periods. \( n \) is the number of stocks. \( l_j \) is the \( j \)-th largest eigenvalue of the second-moment matrix of excess returns. \( L = \sum_{j=1}^{J} l_j \).

Zhang: Testing the APT with the Maximum Sharpe Ratio of Extracted Factors

Management Science 55(7), pp. 1255–1266, © 2009 INFORMS

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s).

Additional information, including rights and permission policies, is available at http://journals.informs.org.

also very low and only occasionally goes above 0.10.

The hypothesis \( \gamma = 0.1 \) cannot be strongly rejected because \( \bar{y} \) itself sometimes is greater than 0.1. The rejection of \( \gamma \geq 0.2 \) (or \( s \geq 0.25 \)) means that there does not appear to be any unconditional portfolio strategy that can generate a Sharpe ratio to be termed as an asymptotic arbitrage opportunity under a reasonable subjective definition of an arbitrage opportunity. As far as the number of possible factors considered here, evidence is strong to uphold the implication of the asymptotic APT. For \( k \) greater than 10, although \( \bar{y} \) and \( s \) increase, the left-tail p-values do not necessarily increase, as we see from the pattern in \( p_{0.2} \) for \( k < 10 \) in Table 2. This is because the asymptotic distribution of \( \bar{y} \) becomes more skewed as \( k \) becomes larger, as we see in Figure 1.

The results reported here can be compared with those in MacKinlay (1995), who uses a variant of the GRS test on a small number of portfolios. Based on time-series estimates of the Sharpe ratio of several stock index returns, MacKinlay (1995) makes a choice of \( s \leq 0.031 \) and interprets the rejection of that hypothesis as evidence against multifactor models in favor of non-risk-based explanations of CAPM anomalies. MacKinlay’s test rejects \( s \leq 0.031 \) and the results in Table 2 reject \( s \geq 0.25 \), so the difference is not in the evidence, but rather in the interpretation. Because the CAPM is rejected simply because the index returns have too low a squared Sharpe ratio, using the squared Sharpe ratio of indexes as
Table 2 Test of Asymptotic APT with CK Factors

<table>
<thead>
<tr>
<th>Period</th>
<th>(k^*)</th>
<th>(\bar{y})</th>
<th>(\bar{s})</th>
<th>(p_{0.05})</th>
<th>(p_{0.02})</th>
<th>(p_{0.01})</th>
<th>(p_{0.00})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965–1969</td>
<td>1</td>
<td>0.052</td>
<td>0.055</td>
<td>0.000</td>
<td>0.020</td>
<td>0.218</td>
<td>0.510</td>
</tr>
<tr>
<td>1970–1974</td>
<td>1</td>
<td>0.026</td>
<td>0.027</td>
<td>0.000</td>
<td>0.005</td>
<td>0.039</td>
<td>0.539</td>
</tr>
<tr>
<td>1975–1979</td>
<td>1</td>
<td>0.109</td>
<td>0.123</td>
<td>0.000</td>
<td>0.122</td>
<td>0.536</td>
<td>0.518</td>
</tr>
<tr>
<td>1980–1984</td>
<td>1</td>
<td>0.021</td>
<td>0.022</td>
<td>0.000</td>
<td>0.003</td>
<td>0.073</td>
<td>0.458</td>
</tr>
<tr>
<td>1985–1989</td>
<td>1</td>
<td>0.016</td>
<td>0.016</td>
<td>0.000</td>
<td>0.002</td>
<td>0.217</td>
<td>0.422</td>
</tr>
<tr>
<td>1990–1994</td>
<td>1</td>
<td>0.032</td>
<td>0.033</td>
<td>0.000</td>
<td>0.002</td>
<td>0.160</td>
<td>0.515</td>
</tr>
</tbody>
</table>

Table 2 (Continued)

<table>
<thead>
<tr>
<th>Period</th>
<th>(k^*)</th>
<th>(\bar{y})</th>
<th>(\bar{s})</th>
<th>(p_{0.05})</th>
<th>(p_{0.02})</th>
<th>(p_{0.01})</th>
<th>(p_{0.00})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995–1999</td>
<td>1</td>
<td>0.096</td>
<td>0.107</td>
<td>0.000</td>
<td>0.008</td>
<td>0.465</td>
<td>0.627</td>
</tr>
<tr>
<td>2000–2004</td>
<td>1</td>
<td>0.029</td>
<td>0.030</td>
<td>0.000</td>
<td>0.004</td>
<td>0.075</td>
<td>0.352</td>
</tr>
<tr>
<td>1965–1969</td>
<td>2</td>
<td>0.052</td>
<td>0.055</td>
<td>0.000</td>
<td>0.012</td>
<td>0.151</td>
<td>0.360</td>
</tr>
<tr>
<td>1970–1974</td>
<td>2</td>
<td>0.087</td>
<td>0.096</td>
<td>0.000</td>
<td>0.033</td>
<td>0.253</td>
<td>0.480</td>
</tr>
<tr>
<td>1975–1979</td>
<td>2</td>
<td>0.131</td>
<td>0.151</td>
<td>0.000</td>
<td>0.114</td>
<td>0.486</td>
<td>0.419</td>
</tr>
<tr>
<td>1980–1984</td>
<td>2</td>
<td>0.090</td>
<td>0.099</td>
<td>0.000</td>
<td>0.024</td>
<td>0.200</td>
<td>0.366</td>
</tr>
<tr>
<td>1985–1989</td>
<td>2</td>
<td>0.091</td>
<td>0.101</td>
<td>0.000</td>
<td>0.016</td>
<td>0.147</td>
<td>0.282</td>
</tr>
<tr>
<td>1990–1994</td>
<td>2</td>
<td>0.096</td>
<td>0.106</td>
<td>0.000</td>
<td>0.012</td>
<td>0.118</td>
<td>0.234</td>
</tr>
<tr>
<td>1965–1969</td>
<td>3</td>
<td>0.138</td>
<td>0.160</td>
<td>0.000</td>
<td>0.073</td>
<td>0.363</td>
<td>0.282</td>
</tr>
<tr>
<td>1970–1974</td>
<td>3</td>
<td>0.149</td>
<td>0.175</td>
<td>0.000</td>
<td>0.071</td>
<td>0.343</td>
<td>0.277</td>
</tr>
<tr>
<td>1975–1979</td>
<td>3</td>
<td>0.153</td>
<td>0.181</td>
<td>0.000</td>
<td>0.057</td>
<td>0.292</td>
<td>0.243</td>
</tr>
<tr>
<td>1980–1984</td>
<td>3</td>
<td>0.153</td>
<td>0.181</td>
<td>0.000</td>
<td>0.040</td>
<td>0.228</td>
<td>0.188</td>
</tr>
<tr>
<td>1985–1989</td>
<td>3</td>
<td>0.154</td>
<td>0.182</td>
<td>0.000</td>
<td>0.028</td>
<td>0.176</td>
<td>0.139</td>
</tr>
<tr>
<td>1990–1994</td>
<td>3</td>
<td>0.166</td>
<td>0.199</td>
<td>0.000</td>
<td>0.028</td>
<td>0.169</td>
<td>0.126</td>
</tr>
<tr>
<td>1965–1969</td>
<td>4</td>
<td>0.084</td>
<td>0.092</td>
<td>0.000</td>
<td>0.012</td>
<td>0.122</td>
<td>0.458</td>
</tr>
<tr>
<td>1970–1974</td>
<td>4</td>
<td>0.084</td>
<td>0.092</td>
<td>0.000</td>
<td>0.012</td>
<td>0.122</td>
<td>0.458</td>
</tr>
<tr>
<td>1975–1979</td>
<td>4</td>
<td>0.105</td>
<td>0.117</td>
<td>0.000</td>
<td>0.004</td>
<td>0.047</td>
<td>0.148</td>
</tr>
<tr>
<td>1980–1984</td>
<td>4</td>
<td>0.105</td>
<td>0.117</td>
<td>0.000</td>
<td>0.002</td>
<td>0.029</td>
<td>0.101</td>
</tr>
<tr>
<td>1985–1989</td>
<td>4</td>
<td>0.114</td>
<td>0.128</td>
<td>0.000</td>
<td>0.006</td>
<td>0.063</td>
<td>0.150</td>
</tr>
<tr>
<td>1990–1994</td>
<td>4</td>
<td>0.114</td>
<td>0.129</td>
<td>0.000</td>
<td>0.006</td>
<td>0.064</td>
<td>0.111</td>
</tr>
</tbody>
</table>

Notes. This table reports the test statistics, \(\bar{y}\) and \(\bar{s} = \bar{y}/(1 − \bar{y})\), of the hypotheses \(\gamma = 0.5, \gamma = 0.2,\) and \(\gamma = 0.1\), which correspond to \(s = 0.25,\) and \(s = 0.11,\) respectively. \(p_{0.05}\), \(p_{0.02}\), and \(p_{0.01}\) are the left-tail \(p\)-values of \(\bar{y}\) under the theoretical/asymptotic distribution. \(p_{0.00}\) is the \(p\)-value of \(\bar{y}\) under the empirical distribution from simulations with \(a = 0\), described in §4.

Theoretical Empirical

4. Simulation Results

The estimates and tests conducted in the last section provide information about the maximum squared Sharpe ratio inappropriate. The results presented in Table 2, based on factors extracted from all individual stocks, provide evidence that the maximum squared Sharpe ratio is much greater than what MacKinlay (1995) specifies, but is still well contained to be reasonably interpreted as being consistent with the asymptotic APT. One of the contributions of this paper is to provide a sensible estimate of the maximum squared Sharpe ratio using individual stock returns, rather than relying on a few stock indexes with time-series estimates.

The Connor-Korajczyk method to extract factors is based on the homoskedasticity assumption that the aggregate idiosyncratic volatilities, \((1/n)e_i'\sigma_{e_i}\), where \(e_i\) is the \(n\)-vector of idiosyncratic risk, have the same limit for all \(t\) within each subperiod. Jones (2001) argues that such an assumption might be violated by more recent data from the United States and develops a methodology that relaxes the homoskedasticity assumption. The results based on Jones‘ (2001) method (not reported here to save space) are qualitatively the same as those in Table 2.

4. Simulation Results

The estimates and tests conducted in the last section provide information about the maximum squared
Sharpe ratio. To relate the results to traditional tests such as the GRS test, we use simulations to discuss the issue of the pricing error. Along with the discussion of the pricing error, we also address a few robustness issues about the tests in the last section.

The simulation procedure is explained as follows. Because \( s \) or \( \gamma \) does not enter the return-generating process (1) explicitly, we generate simulated returns using the pricing error, \( a \), directly. The parameters used for the simulation are taken from actual data for their relevance. We take the return data, \( R \), from a subperiod with \( \tau = 60 \) months and all stocks with nonmissing values. The variance matrix of the returns is calculated as \( (R - \bar{R}_1)(R - \bar{R}_1)'/\tau \), where \( \bar{R} \) is the \( n \)-vector of mean returns. To generate returns, we take the five eigenvectors of the \( \Omega = (R - \bar{R}_1)'(R - \bar{R}_1)/n \) corresponding to the largest five eigenvalues as realizations of five demeaned factors, \( F \). We then regress \( R \) on \([1, \ldots, F]\) to obtain the pricing error and the beta matrix, \([\alpha_1, \ldots, \alpha_F]\). The residuals of the regression are used to calculate its variance matrix, \( \Sigma_\tau \). The sample variance matrix of \( F \) is denoted as \( \Sigma_\tau \) and \( \mu_\tau \) is defined as \( \mu_\tau = (B_1'\Sigma_\tau^{-1}B_1)^{-1}B_1'\bar{R} \). The vector \( \alpha_\tau \) is discarded. Instead, a normally distributed \( n \)-vector is generated and regressed on \( B_\tau \) to obtain the residual. The pricing error, \( \epsilon_i \), is the residual multiplied by a constant to achieve a certain magnitude of mispricing. The choice of the constant will be explained below. In ith simulation, an \( n \times \tau \) return matrix is generated by

\[
R(t) = a_\tau + B_\tau(\mu_\tau 1_\tau + \Sigma_\tau^{1/2} \epsilon(t) + \Sigma_\tau^{1/2} \xi(t)),
\]

where \( \eta(t) \) is a \( 5 \times \tau \) matrix of simulated independent random variables with zero mean and unit variance, \( \xi(t) \) is an \( n \times \tau \) matrix of simulated independent random variables with zero mean and unit variance, and \( \eta(t) \) and \( \xi(t) \) are independent of each other. The reason that the factor structure is obtained from the sample variance matrix of the actual returns, rather than from the second-moment matrix, is to ensure that the potential pricing error does not enter the factor construction. It should be obvious that the choice of five factors does not play any important role in generating returns. The variance matrix of the simulated returns is the same as the sample matrix of the actual returns in the chosen subperiod. The only purpose of having a \( B_\tau \) matrix is to construct a pricing error, \( a_\tau \), which is orthogonal to \( B_\tau \), so that the magnitude of the pricing error can be more meaningful.

Following this simulation procedure, we conduct three sets of simulations. In the first set, \( a_\tau \) is set to zero. For each of the 60-month subperiods, one thousand replications of returns are generated according to the procedure with \( \eta(t) \) and \( \xi(t) \) being normally distributed. For each replication, \( i \), the test statistic, \( \bar{\gamma}(t) \), is calculated for \( k^* = 1, 2, \ldots, 10 \). An empirical distribution of \( \bar{\gamma} \) is then generated from the one thousand replications for each \( k^* = 1, 2, \ldots, 10 \). The left-tail \( p \)-value of the statistic, \( \bar{\gamma} \), from the actual returns is calculated against this empirical distribution and is reported in the last column of Table 2. These \( p \)-values, ranging from 0.096 to 0.721, differ quite a lot across subperiods and are typically large for smaller \( k^* \)'s. This produces no strong evidence against the hypothesis that \( a = 0 \).

In the second set of simulations, we answer the following questions. What is the most plausible value of the average pricing error per stock that generates the test statistics calculated from the actual returns? Is there an impact of the number assets used to calculate the test statistics? We consider three choices of the magnitude of the pricing error. More specifically, we choose the pricing error, \( a \), such that the pricing error per stock, \( \sqrt{a/n} \), equals 0.01, 0.005, or 0.001. As a very crude estimate, the cross-sectional standard deviation of average monthly excess returns is about 0.02. The choice of \( \sqrt{a/n} = 0.01 \) is interpreted as a relatively large pricing error, the choice of \( \sqrt{a/n} = 0.005 \) can be regarded as a medium-size pricing error, and the choice of \( \sqrt{a/n} = 0.001 \) corresponds to a small pricing error. We conduct simulations for the first subperiod in which \( n = 1,487 \) and the last subperiod in which \( n = 3,708 \). For each of the two subperiods and each choice of \( a \), ten thousand replications are generated. In each replication, \( i \), an \( n \times \tau \) matrix of excess returns is generated according to (7), and the statistic \( \bar{\gamma}(t) \) is calculated for a given \( k = 1, 2, \ldots, 10 \), using the CK method. Correspondingly, the \( \bar{\gamma}(t) \) and the left-tail \( p \)-values, \( p_{0.01}^{(t)}, p_{0.05}^{(t)}, \) and \( p_{0.1}^{(t)} \), are calculated. The average values of these statistics are reported in Table 3.

Two points can be made about the results in Table 3. First, the average \( p \)-values for \( \gamma = 0.5 \) are virtually zero in all cases, whereas the average \( p \)-values for \( \gamma = 0.1 \) are considerably high. On the other hand, the average \( p \)-values in panel A are higher than those in panel B, which in turn are slightly higher than those in panel C, especially for large \( k^* \). This indicates that the test statistic, \( \bar{\gamma} \), has the desired property of distinguishing alternative hypotheses regarding \( \gamma \), but not so about the average pricing error when it is small. Second, the average \( p \)-values in Table 3, even those for \( \sqrt{a/n} = 0.001 \), are higher than the corresponding \( p \)-values in Table 2. This means that the pricing error in the actual data may indeed be small, probably smaller than 0.001 per stock per month on average, consistent with the empirical \( p \)-values reported in Table 2.

In the third set of simulations, we address the issue related to return distribution. The theoretical distribution of \( \bar{\gamma} \) is derived under the normality assumption. Although normality is a reasonable assumption for monthly returns, potential deviations from normality by the actual return data may impact the test results.
To investigate the impact, we simulate returns as in the second set of simulations, except that \( \xi^{(0)} \) and \( \eta^{(0)} \) are \( t \)-distributed with a degree of freedom equal to 10 and with variance adjusted to remain one. The resultant average \( p \)-values are reported in Table 4, which can be correspondingly compared with those in Table 3. The results indicate that \( p \)-values calculated with \( t \)-distributed factors and idiosyncratic returns tend to be greater. Therefore, if the distribution of returns has a fatter tail than the normal distribution, the test would be less likely to reject the hypothesis of the form \( \gamma \geq \gamma_0 \) or \( s \geq s_0 \). Because the results in Table 2 reject the hypothesis \( \gamma \geq 0.2 \) or \( s \geq 0.25 \) in most cases, the rejection cannot be attributed to nonnormality of the returns.

### 5. Conclusions

The existing studies on beta pricing models test the beta pricing restriction with prespecified systematic factors on small sets of testing assets. Tests with prespecified factors can be, and in many cases are, inconclusive, for example, because the factors are measured with error. This paper develops a test of the beta pricing restriction using factors extracted from returns, based on the asymptotic arbitrage pricing theory advanced by Ross (1976),
Huberman (1982), and Chamberlain and Rothschild (1983), and the factor-extraction method by Connor and Korajczyk (1986, 1988). The developed test statistic is related to the maximum squared Sharpe ratio of the extracted factors. The test addresses Shanken’s (1992) criticism about the testability of the APT. Unlike the Gibbons-Ross-Shanken F-test and its variants with asymptotic Chi-square distributions, the test developed in this paper follows a noncentral beta distribution asymptotically, with the noncentrality parameter proportional to the theoretical maximum squared Sharpe ratio of all portfolios of the returns.

The testing method is applied to individual stock returns in the U.S. market during eight 60-month sub-periods from 1965 to 2004. The empirical evidence shows that the sample maximum squared Sharpe ratio of the extracted factors can be much greater than those calculated from stock indexes. However, the hypothesis that the maximum squared Sharpe ratio is greater than 0.25 can be comfortably rejected in most cases, so the risk-return trade-off is well contained in the range that is consistent with the asymptotic APT. Simulation experiments show that the test has the ability to differentiate the hypotheses about the maximum squared Sharpe ratio and to detect large pricing errors, so the results based on actual U.S. return data are indeed evidence of low squared Sharpe ratios and small unconditional deviations in expected returns from the beta pricing theory, very likely less than 0.001 on the monthly basis.

Acknowledgments
The author thanks Chuanyang Hwang, Raymond Kan, Qiao Liu, Keshab Shrestha, Mungo Wilson, Guofu Zhou, seminar participants at Nanyang Technology University and the Chinese International Conference in Finance 2006 and, especially, two anonymous referees for their helpful comments on earlier versions of this paper. Financial support from the RGC Competitive Earmarked Research Grant HKUST6268/04H is gratefully acknowledged. All remaining errors are the author’s responsibility.

Appendix

Proof of Proposition 1. Under the assumption of an approximate k-factor structure, \( r_t = a + B_f f_t + e_t \), where \( B_f \) to emphasize its relation with \( f_t \), and the variance matrix of \( B_f f_t, B_f \Sigma f f_t \), has \( k \) unbounded eigenvalues, whereas the variance matrix of \( e_t, \Sigma e \) has bounded eigenvalues. Let \( g_t = S^{-1/2} f_t \) and \( B_g = S f f_t \) where \( S = E f f_t \) is the second-moment matrix of \( f_t \). By construction, \( E g g_t = I_k \).

Consider the second moment of \( a + B_g g_t \), denoted \( S \):

\[
S = E[(a + B_g g_t)(a + B_g g_t)] = aa' + B_g B_g' = (a B_g)(a B_g)' .
\]

The positive eigenvalues of \( S \) are the same as those of \( \Omega \equiv (a B_g)'(a B_g) \). From \( a B_g = a B_g^{1/2} B_g'^{1/2} = 0' \), it follows that

\[
Q = \begin{pmatrix}
aa' & 0_k \\
0_k & B_g B_g'
\end{pmatrix} .
\]

The characteristic function of \( Q \) equals \( |Q - I_k| = |a' a - \lambda| \). The eigenvalues of \( Q \) are therefore \( a' a \) and the eigenvalues of \( B_g B_g' \). The eigenvalues of \( B_g B_g' \) are the same as the positive eigenvalues of \( B_g B_g' = B_g S_f f_t' > B_f \Sigma f f_t' \).

Therefore, \( S \) has \( k \) unbounded eigenvalues if \( a' a \) is bounded and \( S \) has \( k + 1 \) unbounded eigenvalues if \( a' a \) is unbounded. The same is true to the second-moment matrix of \( r, \Sigma_r \).

Q.E.D.

Proof of Proposition 2. (i) It is obvious that \( \mu_g \mu_g' \geq 0 \). Expanding the determinant below in two ways gives

\[
\begin{vmatrix}
I_k & \mu_g \\
\mu_g' & 1
\end{vmatrix} = [I_k - \mu_g \mu_g'] = 1 - \mu_g' \mu_g .
\]

Because \( I_k - \mu_g \mu_g' \) is the variance matrix of \( g_t \), its determinant is nonnegative. This establishes that \( \mu_g \mu_g' \leq 1 \).

(ii) \((\Rightarrow)\) When there is a systematic pricing error, \( g_t = C(1 f_t)' \). Let \( c' \) be the first row of \( C^{-1} \). Then \( c' c = 1 \). Because \( E g_t g_t' = I_k \), it follows that \( c' c = 1 \). On the other hand, \( c' = I_k c = E g_t g_t' = E g_t g_t' = \mu_g \). Therefore, \( \mu_g \mu_g' = c' c = 1 \). \((\Leftarrow)\) If \( \mu_g \mu_g' = 1 \), then obviously \( \mu_g \neq 0 \). Because \( \text{Var}(\mu g t) = \mu_g (I_k - \mu_g \mu_g') \mu_g = 0 \), \( \mu_g \mu_g' \) is a constant, and, therefore, the original \( f_t \) contains a constant, which corresponds to a systematic pricing error.

(iii) Only the case in which \( n \) is finite needs to be considered.\(^{11}\) Because \( r_t = a + B_g g_t + e_t \), we have \( \mu_g \equiv E r_t = a + B_g \mu_g \) and \( \Sigma_r \equiv \text{Var} r_t = B_g \Sigma B_g' + \Sigma_e \) with \( a' a \) and the maximum eigenvalue of \( \Sigma_e \) bounded but \( B_g \Sigma B_g' \) unbounded. It follows that, in the limit when \( n \) goes to infinity, the maximum squared Sharpe ratio is

\[
s = \lim_{n \to \infty} \mu_g \Sigma_e^{-1} \mu_g = \mu_g (B_g \Sigma B_g')^{-1} (a + B_g \mu_g)
\]

\[
= \mu_g (I_k - \mu_g \mu_g')^{-1} \mu_g
\]

\[
= \mu_g [I_k + \mu_g \mu_g' (1 + \mu_g \mu_g')] \mu_g = \gamma / (1 - \gamma),
\]

where \( A^+ \) is the generalized inverse of \( A \), and the last line invokes a formula for matrix inverse. Q.E.D.

Proof of Proposition 3. (i) From the factor structure with \( \Omega = (a B) \) and \( \bar{F} = (1, F)' \),

\[
\Omega = \bar{F} \left( \frac{1 - \mu B}{n} \right) \bar{F} + \frac{1}{n} \bar{e} e' + \frac{1}{n} \bar{F} B e + \frac{1}{n} \bar{F} B \bar{F} \Delta = H' H + D,
\]

where \( H = (\text{plim}_{n \to \infty} (1/n) \mu B) \) is a \((k + 1) \times \tau \) matrix if \( a' a \) is unbounded or \( H = (\text{plim}_{n \to \infty} (1/n) \mu B) \) is a \( k \times \tau \) matrix if \( a' a \) is bounded, \( D = \text{plim}_{n \to \infty} (1/n) \mu B \) is a scalar matrix, and \( \Delta \) means asymptotic equivalence. The last two terms in the first line above are zero because \( e \) is uncorrelated with \( \bar{F} \), \( H = (h_1, h_2, \ldots, h_\tau) \) is unique up to a premultiplier of an orthogonal matrix. Let \( \tau^{-1/2} C \) be the \( k^* \) orthonormal eigenvectors of \( \Omega \) corresponding to

11\(^{10}\) For two \( n \times n \) symmetric matrices, \( A_1 \) and \( A_2 \), \( A_1 > A_2 \) means that \( A_1 - A_2 \) is positive definite. This implies that the \( i \)th largest eigenvalue of \( A_1 \) is greater than that of \( A_2 \), for \( i = 1, \ldots, n \).

11 I thank a referee who provided the proof in this part, which is simpler than the original one.
the $k^*$ largest eigenvalues of $\Omega$, which are also the eigenvectors of $H'H$. The components of $G = (g_1, \ldots, g_n)$ have their sample second-moment matrix equal to $I_n$ because, by construction, $\tau^{-1}G'G = I_n$. Without loss of generality, $\tau^{-1/2}G = (HH')^{-1/2}H$. Note that

$$\tau^{-1}(G - \bar{G}I_n)(G - \bar{G}I_n)' = (\tau^{-1/2}G)(I_n - 1_n(1', 1')^{-1}1_n')(\tau^{-1/2}G)'$$

where $I_n - 1_n(1', 1')^{-1}1_n'$ is a symmetric idempotent matrix with $\sigma - 1$ eigenvalues equal to one and one eigenvalue equal to zero. Because $\tau^{-1/2}G$ comprises orthonormal vectors, it follows from the Poincare separation (interlacing) theorem that $\tau^{-1}(G - \bar{G}I_n)(G - \bar{G}I_n)'$ has $k^*$ eigenvalues equal to one and one eigenvalue between zero and one. The smallest eigenvalue equals

$$\min_{u' = 1} \tau^{-1}u'(G - \bar{G}I_n)(G - \bar{G}I_n)'u = 1 - \max_{u' = 1} \bar{g}u'\bar{g}u = 1 - (\bar{g}'\bar{g})$$

which is obtained at $u = \bar{g}/\sqrt{\bar{g}'\bar{g}}$.

(ii) Write

$$\bar{\eta} = (\bar{g}'\bar{g})^{-1/2}G'1_n = \tau^{-1/2}H(\bar{H}'H)^{-1}H'1_n$$

$$= \bar{h}'\left(\frac{1}{\tau} \sum h_i h_i'\right)^{-1} \bar{h}$$

$$= \bar{h}' \left(\tilde{S}^{-1} + \bar{h} \bar{h}'\right)^{-1} \bar{h}$$

where $h_i = (1/\tau) \sum h_i h_i'$, $\tilde{S}_n = \sum (h_i - \bar{h})(h_i - \bar{h})'$ and the last equality is obtained using a matrix inversion formula. It is well known that, under the assumption of normality, $\tilde{S}^{-1} + \bar{h} \bar{h}'$ is the so-called noncentral Hotelling’s $T^2$ variate, and $\eta = ((\tau - k^*)/k^*)\bar{h}' \tilde{S}_n^{-1} \bar{h}$ has a noncentral $F(\tau s)$ distribution where $s = \mu_1^2 \sum h_i^2 - \mu_2 \sum h_i$, because $h$ is a nonsingular transformation of $f^c$. See, for example, Muirhead [1982, p. 98]. The density function of $\eta$ is

$$p_{\eta}(\eta) = \frac{\Gamma(\tau/2)}{\Gamma((\tau - k)/2)\Gamma(k^*/2)} (1 + (k^*/(\tau - k^*)\eta))^{\tau - k^*}$$

$$\cdot \cdot F_1\left(\frac{\tau}{2}, \frac{k^*}{2}, \frac{k^* \epsilon}{2(\tau - k^* + k^* \eta)}\right), \quad \eta \geq 0.$$ 

The change of variable $y = (\tau - k^*)\eta/k^*(1 - x)$ with the Jacobian equal to $dy/dx = (\tau - k^*)/k^*(1 - x)^2$ leads to the density function of $\bar{\eta}$ as

$$p_{\bar{\eta}}(x) = \frac{\Gamma(\tau/2)}{\Gamma((\tau - k)/2)\Gamma(k^*/2)} \left(1 - (1 - x)^{(\tau - k^*)/2 - 1}\epsilon^{\tau - k^* /2}ight)$$

$$\cdot \cdot F_1\left(\frac{\tau}{2}, \frac{k^*}{2}, \frac{\tau \epsilon x}{2}\frac{\tau \epsilon x}{2}\right), \quad 0 \leq x \leq 1. \quad \text{Q.E.D.}$$

References


