ELSEVIER

Journal of Empirical Finance 6 (1999) 219–241

Journal of EMPIRICAL FINANCE

www.elsevier.com/locate/econbase

Testing multi-beta asset pricing models

Raja Velu [a], Guofu Zhou [b,*]

[a] *School of Management, Syracuse University, Syracuse, NY 13244, USA*
[b] *Olin School of Business, Washington University, Campus Box 1133, St. Louis, MO 63130, USA*

Received 1 September 1997; accepted 1 March 1999

Abstract

This paper presents a complete solution to the estimation and testing of multi-beta models by providing a small sample likelihood ratio test when the usual normality assumption is imposed and an almost analytical GMM test when the normality assumption is relaxed. Using 10 size portfolios from January 1926 to December 1994, we reject the joint efficiency of the CRSP value-weighted and equal-weighted indices. We also apply the tests to analyze a new version of Fama and French's [Fama, E.F., French, K.R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics. 33, 3–56.] three-factor model in addition to two standard ones, and find that the new version performs the best. © 1999 Elsevier Science B.V. All rights reserved.

JEL classification: G12; G11; C11; C31

Keywords: Asset pricing; Multi-beta; Small sample test; Factor model; APT

1. Introduction

A fundamental problem in finance is to examine the tradeoff between risk and return. Sharpe (1964), Lintner (1965), Black (1972), Ross (1976), Merton (1973) and Breeden (1979), among others, develop single and multi-beta asset pricing models which imply that the expected return on a security is a linear function of factor risk premiums and their associated betas. There has been an enormous amount of empirical research on the validity of such a linear relationship. Kandel

* Corresponding author. Tel.: +1-314-935-6384; fax: +1-314-935-6359; e-mail: zhou@zhoufin.wustl.edu or zhou@mail.olin.wustl.edu

0927-5398/99/$ - see front matter
PII: S0927-5398(99)00002-X

and Stambaugh (1989) provide an excellent survey of the earlier literature, and Shanken (1996) provides such a survey of the most recent developments.

When there exists a riskless asset, Gibbons et al. (1989) provides an exact test of a portfolio's efficiency (single beta model) as well as an exact test of the multi-beta linear pricing relationship. In this case, the zero-beta rate is usually taken as the U.S. Treasury bill rate. As the T-bills are only nominally riskless, the assumption that the zero-beta rate is the bill rate may not be appropriate if there is concern about the use of the real returns in the test. Even in the nominal case, the zero-beta rate may not be equal to the T-bill rate due to the restrictions on borrowing. In an extension where the zero-beta rate deviates from the T-bill rate by some constant, the efficiency test will be econometrically the same as testing efficiency in the standard zero-beta model where no riskless asset exists (see Section 2.2). Hence, even when a nominally riskless asset exists, the methodology of testing the standard zero-beta models is useful for assessing the impact of restrictions on borrowing.

The standard zero-beta model is often implemented empirically by assuming a constant zero-beta rate over the testing period. In this case, no information on the bill rates are used explicitly, and the zero-beta rate is estimated from security return data. Since the zero-beta rate is unobservable, it is more difficult to develop a small sample test than in the riskless asset case. With the multivariate framework initiated by Gibbons (1982), Kandel (1984; 1986) obtains an explicit solution for the maximum likelihood estimator of the zero-beta rate. Shanken (1986) extends this result to market model parameterization and to the multi-beta case. Although the estimation problem is elegantly solved, the testing problem remains. Zhou (1991) provides a small sample test, but it applies only to the single beta case, and the approach does not extend to the multi-beta case in which only bounds on the exact distribution of the likelihood ratio test (LRT) are obtained by Shanken (1985).

In this paper, we derive, taking a new approach, the exact distribution of the LRT for the multi-beta pricing models. Unlike the riskless asset case, the exact distribution is unfortunately far more complex, and dependent monotonically on a nuisance parameter. As a result, the best one can do is to evaluate the exact distribution at an estimate of this parameter. This, however, does not pose any serious problem for assessing the exact P-value, because the maximum likelihood estimator of the nuisance parameter is usually quite large and the P-values computed from sizable perturbations of the estimate are close to one another. [1] To help make easy inference, we also provide *optimal* upper and lower bounds, which are independent of the nuisance parameter, on the exact P-value. The upper

[1] For example, the maximum likelihood estimator of the nuisance parameter for the first period in Table 1 is 58.53, and the P-value evaluated at half of this value is 0.00705, almost the same as 0.00718 evaluated at 58.53.

bound, which is tighter than that of Shanken (1986) and follows a simple F-distribution, is particularly useful when the hypothesis is close to being rejected. For example, if we find an upper bound of 5%, the true significance level must be less than or equal to 5%, so we can reject the model at the usual 5% level. On the other hand, if a lower bound of 6% is found, the exact P-value must be higher than 6% and we cannot reject the null regardless of what value the nuisance parameter is.

The above small sample test tells precisely to what probability extent we reject or do not reject an asset pricing model if the data is perfectly normally distributed. Since the normality assumption may be violated in certain applications, Shanken (1992) initiates multi-beta tests that are valid as long as the data is identically and independently distributed (iid). The second contribution of this paper is to provide a test based on the generalized method of moments (GMM) which allows a very general form of conditional heteroskedasticity, including as special cases the iid and normal distributions.[2] Our test can be computed almost analytically, which requires minimizing the objective function over only a single variable (the zero-beta rate), and hence the test is straightforward to apply in practice. In addition, Monte Carlo experiments become much easier with our almost analytical GMM test. In contrast to the LRT, the advantage of the GMM test is that it is valid even if the data has a very general form of conditional heteroskedasticity, but the price paid is that the test no longer has a tractable exact distribution, but only an asymptotic one.

The paper is organized as follows. In Section 2, we examine the zero-beta asset pricing restrictions and show that an extension of the Sharpe (1964) and Lintner (1965) capital asset pricing model (CAPM) also gives rise to the same zero-beta restrictions. Then, we derive both a small sample test and a GMM test for these restrictions. In Section 3, we extend the tests to a more general multi-beta asset pricing model studied by Shanken (1992). We also show that how the pricing restrictions of the arbitrage pricing theory (APT) of Ross (1976) can be transformed into restrictions similar to those of the multi-beta asset pricing model. Hence, the proposed tests can also be applied to test the APT without having to estimate the factors. In Section 4, we examine empirically whether a portfolio of the CRSP value-weighted and equal-weighted indices is efficient relative to the asset space spanned by them and the 10 standard size portfolios, with monthly data readily available from the Center for Research in Security Prices (CRSP) at the University of Chicago. In Section 5, we apply the tests to analyze three versions of the Fama and French (1993) multi-beta model. Section 6 concludes the paper.

[2] In the riskless asset case differing from the no riskless asset case here, MacKinlay and Richardson (1991) and Harvey and Zhou (1993) provide GMM tests of a portfolio's efficiency.

2. The model and tests

In this section, we discuss first the econometric model and the standard zero-beta asset pricing restrictions. Then, in Section 2.2, we introduce a particular form of time-varying zero-beta rate model which is an extension of the Sharpe (1964) and Lintner (1965) capital asset pricing model (CAPM), and show that a test of this model is econometrically the same as testing the standard zero-beta model. In Section 2.3, we provide the exact distribution of the likelihood ratio test (LRT) as well as its optimal bounds under the normality assumption. Finally, in Section 2.4, we provide the GMM test that is analytical up to one parameter and valid under general conditional heteroskedasticity conditions.

2.1. Asset pricing restrictions: Constant zero-beta rate

Our objective is to examine whether an efficient portfolio can be constructed from a given group of K reference portfolios, which is a subset of a larger universe of assets that contains N additional assets. In other words, we are going to test a K-beta model by using N assets and K reference portfolios. As emphasized by Fama (1991), we cannot test any asset pricing theory without specifying the law of motion for the asset returns. Given statistical assumptions on the stochastic behavior of the returns, the efficiency problem has testable restrictions on the parameters of the statistical model.

Following most studies, we assume that the asset returns are governed by the standard multi-beta market model:

$$r_{it} = \alpha_i + \beta_{i1} R_{\mathrm{p}1t} + \cdots + \beta_{iK} R_{\mathrm{p}Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where r_{it} = the return on asset i in period t $(1 \le i \le N)$, $R_{\mathrm{p}jt}$ = the return on the j-th reference portfolio in period t $(1 \le j \le K)$, ϵ_{it} = the disturbances or random errors, and T is the number of observations. Like most studies, we maintain the normality assumption, which will be relaxed later in Section 2.4, that the disturbances are independent over time, and jointly normally distributed in each period with a zero mean and a nonsingular covariance matrix $\boldsymbol{\Sigma}$.

The null hypothesis is that some portfolio of the K reference portfolios is efficient with respect to the total set of $N+K$ assets, which translates into the standard nonlinear restriction on the parameters,

$$H_0\colon \alpha_i = \gamma(1 - \beta_{i1} - \cdots - \beta_{iK}), \quad i = 1, \ldots, N, \tag{2}$$

where γ is the expected return of the zero-beta asset. In empirical studies, γ is usually assumed to be a constant over the T period, and hence it becomes a parameter that has to be estimated from the data. In the special case of $K = 1$, a test of Eq. (2) is a test of efficiency for a given portfolio. When this portfolio is identified as the market portfolio, Eq. (2) is the well-known zero-beta CAPM restriction. The primary objective of this paper is to develop both a small sample likelihood ratio test and an asymptotic GMM test for the general asset pricing restriction Eq. (2).

2.2. Asset pricing restrictions: Time-varying zero-beta rate

The constant zero-beta rate assumption may not be appropriate if one takes the view that the return on a T-bill should be a close approximation of the zero-beta rate. Because the T-bill return fluctuates over time, so does the zero-beta rate. In the special case where the zero-beta rate is equal to the T-bill rate every period, i.e., $\gamma_t = r_{ft}$ for $t = 1, \ldots, T$, the empirical tests are often carried out in the following multi-beta market model:

$$r_{it} - r_{ft} = \alpha_i + \beta_{i1}\left(R_{p1t} - r_{ft}\right) + \cdots + \beta_{iK}\left(R_{pKt} - r_{ft}\right) + \epsilon_{it}, \tag{3}$$

where the regressions are run, in contrast to Eq. (1), on excess returns, i.e., on returns in excess of the T-bill rate r_{ft}. The asset pricing restrictions are the standard zero intercepts hypothesis,

$$H_0^*: \alpha_i = 0, \quad i = 1, \ldots, N. \tag{4}$$

Hypothesis (4) is the restrictions tested in the riskless asset case. The well-known Gibbons et al. (1989) test (GRS) is often used for this purpose. Under the earlier normality assumption on the residuals, the GRS test has a simple finite sample F distribution, in addition to its interesting economic interpretations.

As the T-bills are only nominally riskless, and there are restrictions on borrowing even in the nominal case, the assumption that $\gamma_t = r_{ft}$, for $t = 1, \ldots, T$, may not be appropriate. As a simple extension, we assume that

$$\gamma_t = r_{ft} + c_0, \quad t = 1, \ldots, T, \tag{5}$$

where c_0 is a fixed constant over the testing period which reflects perhaps the difference of borrowing and lending rates. Then, the asset pricing restrictions become

$$H_0^{**}: \alpha_i = c_0(1 - \beta_{i1} - \cdots - \beta_{iK}), \quad i = 1, \ldots, N. \tag{6}$$

In the special case where $c_0 = 0$, Eq. (6) collapses into the standard restrictions (Eq. (4)).

It should be acknowledged that the concept of the time-varying zero-beta rate γ_t is not easily justified in a conditional asset pricing model. This is because a conditional model often imply time-varying betas while the betas here are assumed constant. Hence, Eq. (5) can only be understood as a simple extension of the usual practice, say Gibbons et al. (1989), that take the riskfree rate to be time-varying while holding other parameters constant.

In particular, if there is only one beta and the benchmark portfolio is the return on the market portfolio, the multi-beta market model (3) with restrictions (4) is the CAPM of Sharpe (1964) and Lintner (1965),

$$r_{it} - r_{ft} = \beta_{i1}(R_{Mt} - r_{ft}) + \epsilon_{it}, \quad i = 1, \dots, N. \tag{7}$$

The implicit assumption is that the zero-beta rate is equal to the T-bill rate every period. In contrast, with assumption (5) on the zero-beta rate and the asset pricing restrictions (6), the multi-beta market model (3) becomes (when $K = 1$)

$$r_{it} - r_{ft} - c_0 = \beta_{i1}(R_{Mt} - r_{ft} - c_0) + \epsilon_{it}, \quad i = 1, \dots, N. \tag{8}$$

As seen from Eqs. (7) and (8), with assumption (5) on the zero-beta rate, we obtain a simple extension of the CAPM.

Comparing Eq. (6) with Eq. (2), despite of different economic interpretations one may have on the regressions (1) and (3), it is clear that testing (6) is econometrically equivalent to testing (2). In other words, the test for (6) is obtained exactly the same way as the test for (2), except that the returns are replaced by the excess returns in the latter case. Therefore, we will focus on testing (2) in what follows.

2.3. Likelihood ratio test

To construct the likelihood ratio test (LRT) for (2), both parameter estimates, without the restriction and with the restriction, are needed. The unconstrained estimates of the alphas and betas are straightforwardly obtained by using the ordinary least squares (OLS) procedure. The unconstrained maximum likelihood estimator of $\boldsymbol{\Sigma}$ is then the cross-products of the OLS residuals divided by T.

With restriction (2), the parameter estimates are complex. But the estimation problem has been solved elegantly by Shanken (1986), who extended the single beta case of Kandel (1984) to a multi-beta one. A key that leads to the solution is to note that, conditional on γ, the multi-beta market model regression (1) can be written as

$$\mathbf{R}_t - \gamma 1_N = \boldsymbol{\alpha}(\gamma) + \boldsymbol{\beta}(\mathbf{R}_{pt} - \gamma \mathbf{1}_K) + \epsilon_t, \tag{9}$$

where $\mathbf{R}_t$ is an N-vector of the asset returns for period t, $\mathbf{R}_{pt}$ is a K-vector of the returns on the reference portfolios, $\boldsymbol{\alpha}(\gamma)$ is an N-vector of the regression intercepts, and $\boldsymbol{\beta}$ is an $N \times K$ matrix of the betas. Under the null that (2) is valid, there must exist a scalar γ such that $\boldsymbol{\alpha}(\gamma)$ is zero. Shanken (1986) shows that the constrained estimates of both $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are the same as the unconstrained estimates (they do not depend on γ!), while the constrained estimator of $\boldsymbol{\alpha}(\gamma)$ is, conditional on γ,

$$\tilde{\boldsymbol{\alpha}}(\gamma) = \hat{\boldsymbol{\alpha}} - \gamma(\mathbf{1}_N - \hat{\boldsymbol{\beta}} \mathbf{1}_K), \tag{10}$$

where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are the unconstrained estimate of the alphas and betas. Clearly, both $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are independent of γ. Hence, we need only to find the constrained estimate of γ, say $\tilde{\gamma}$.

Let $\overline{\mathbf{R}}_{\mathrm{p}}$, $K \times 1$, be a K-vector of sample means and $\hat{\boldsymbol{\Delta}}$ be the sample covariance matrix (the cross-products of deviations from the sample means divided by T) of the returns on the K reference portfolios. Consider the following function of γ: [3]

$$Q(\gamma) \equiv \frac{\tilde{\boldsymbol{\alpha}}(\gamma)'\hat{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\alpha}}(\gamma)}{1 + \left(\overline{\mathbf{R}}_{\mathrm{p}} - \gamma\mathbf{1}_K\right)'\hat{\boldsymbol{\Delta}}^{-1}\left(\overline{\mathbf{R}}_{\mathrm{p}} - \gamma\mathbf{1}_K\right)}. \tag{11}$$

As shown by Shanken (1985) and Gibbons et al. (1989), the $Q(\gamma)$ function not only enjoys interesting economic interpretations, but also plays an important role in obtaining both the constrained estimates of the parameters and the distribution of the likelihood ratio test (LRT). Indeed, $\tilde{\gamma}$ is the numerical value that minimizes $Q(\gamma)$. Based on this, Shanken (1986) provides an explicit solution for $\tilde{\gamma}$. Moreover, Shanken (1985; 1986) shows the following relationship between the LRT and $Q(\gamma)$:

$$\mathrm{LRT} = T\log\left[1 + Q(\tilde{\gamma})\right]. \tag{12}$$

Based on Eq. (12), Shanken (1986) is able to obtain bounds on the distribution of the LRT. In what follows, we also rely on equality (12) to obtain the exact distribution and optimal bounds for the distribution of the LRT (as a byproduct, our procedure also yields an alternative and seemingly simpler analytical expression for $\tilde{\gamma}$).

The idea is to link $Q(\gamma)$ to eigenvalues of some matrices, because there are well developed distribution theory on eigenvalues (Muirhead, 1982). Let $\boldsymbol{\xi} = \mathbf{1}_N - \boldsymbol{\beta}\mathbf{1}_K$ and $\hat{\boldsymbol{\xi}} = \mathbf{1}_N - \hat{\boldsymbol{\beta}}\mathbf{1}_K$. It is observed that $Q(\gamma)$ is a ratio of two quadratic functions of γ, which can be written as

$$Q(\gamma) = \frac{\mathbf{w}'\mathbf{A}\mathbf{w}}{\mathbf{w}'\mathbf{B}\mathbf{w}}, \quad \mathbf{w} \equiv \begin{pmatrix} 1 \\ \gamma \end{pmatrix}, \tag{13}$$

where

$$\mathbf{A} = \begin{pmatrix} \hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}} & -\hat{\boldsymbol{\xi}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}} \\ -\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\xi}} & \hat{\boldsymbol{\xi}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\xi}} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 + \overline{\mathbf{R}}_{\mathrm{p}}'\hat{\boldsymbol{\Delta}}^{-1}\overline{\mathbf{R}}_{\mathrm{p}} & -1_K'\hat{\boldsymbol{\Delta}}^{-1}\overline{\mathbf{R}}_{\mathrm{p}} \\ -\overline{\mathbf{R}}_{\mathrm{p}}'\hat{\boldsymbol{\Delta}}^{-1}1_K & 1_K'\hat{\boldsymbol{\Delta}}^{-1}1_K \end{pmatrix}.$$

Expressing $Q(\gamma)$ in form of Eq. (13) makes it possible to obtain the minimum and

[3] For notational brevity in our later derivations, $Q(\gamma)$ used here is different from that of Shanken (1986) by a scale $(T-K-1)$. (The T term is omitted and that $\hat{\Sigma}$ is the ML estimator instead of the unbiased estimator.).

hence its distribution. More specifically, let λ_2 be the smallest eigenvalue of the following determinantal equation [4]

$$|\mathbf{A} - \lambda(\mathbf{A} + \mathbf{B})| = 0. \tag{14}$$

Then, we have

Theorem 1. *If $N \geq 2$ and $T \geq N + K + 1$, then the minimized $Q(\gamma)$ is a scale function of λ_2, i.e.,*

$$Q(\tilde{\gamma}) = \frac{\lambda_2}{1 - \lambda_2}, \tag{15}$$

and λ_2 is distributed the same as the smallest eigenvalue of

$$|\mathbf{U}\mathbf{U}' - \lambda(\mathbf{U}\mathbf{U}' + \mathbf{V}\mathbf{V}')| = 0, \tag{16}$$

with the random matrices $\mathbf{U}$ *and* $\mathbf{V}$ *are defined as*

$$\mathbf{U} = \begin{pmatrix} \sqrt{\chi^2_N(\omega_1)} & \mathrm{N}(0,1) \\ 0 & \sqrt{\chi^2_{N-1}} \end{pmatrix}, \text{ and } \mathbf{V} = \begin{pmatrix} \sqrt{\chi^2_{T+1-N-K}} & \mathrm{N}(0,1) \\ 0 & \sqrt{\chi^2_{T-N-K}} \end{pmatrix},$$

where $\chi^2_N(\omega_1)$ denotes a non-central χ^2 random variate with noncentrality parameter $\omega_1 = T(\boldsymbol{\xi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi})\ (\gamma, -1)\mathbf{B}^{-1}\ (\gamma, -1)'$, χ^2_{N-1} a central χ^2 variate and N(0, 1) *a standard normal variate (all the random variates are independent).*

Proof. See Appendix A.

Since λ_2, $Q(\tilde{\gamma})$ and the LRT are monotonic transformations of one another, they are equivalent tests (the P-values are the same). Hence, to test Eq. (2), we need only to determine the P-value from the λ_2 statistic. Theorem 1 shows that the P-value, Prob$(\lambda_2 > x | \omega_1)$, depends on the unknown nuisance parameter ω_1. However, at any given value of ω_1, such as its maximum likelihood estimate computed by replacing $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{\xi}}$ and $\hat{\Sigma}$ in its formula, Prob$(\lambda_2 > x | \omega_1)$ can be straightforwardly computed. Samples of $\mathbf{U}$ and $\mathbf{V}$ are easily generated from their probability distributions, and λ_2 can be computed from Eq. (14). If we repeat this process 10,000 times, for example, the percentage of times that λ_2 is greater than x is readily computed, which is a numerical approximation to Prob$(\lambda_2 > x)$ at the given ω_1.

The above procedure is essentially an application of the Monte Carlo integration approach (see, e.g., Geweke, 1989). With the number of simulation being 10,000, it provides results that are generally accurate to 2 or 3 decimal points,

[4] In the single beta case, this λ_2 reduces to the λ_2 in Zhou (1991).

which are sufficient for our purposes. An alternative approach is to integrate either analytically or numerically the exact density function of λ_2, an explicit (but complex) expression of which is provided in Appendix B. The Monte Carlo integration approach is recommended to apply in practice because of its simplicity.

The effect of using estimated ω_1 can be easily examined because the P-value is an increasing function of ω_1 (by Theorem 1 and Perlman and Oklin (1980)). In other words, if the P-value is evaluated at a value of ω_1 lower than the true one, we will underestimate the true probability; and if evaluated at a value higher than the true one, we will overestimate the true probability. Nevertheless, numerical experiments show that there are no substantial changes in the P-value when small perturbations in ω_1 are allowed.

The monotonicity implies that if $\lim_{\omega_1 \to \infty} \mathrm{Prob}(\lambda_2 > x | \omega_1)$ exists, it must be an upper bound on $\mathrm{Prob}(\lambda_2 > x | \omega_1)$ for all possible values of ω_1. Conversely, $\mathrm{Prob}(\lambda_2 > x | 0)$ must be a lower bound. Based on Nanda (1947) and Schott (1984), we can obtain the bounds explicitly as follows.

Theorem 2. *Under the null hypothesis that the constraints* (2) *are true, if* $N \geq 2$ *and* $T \geq N + K + 1$, *then we have the following upper and lower bounds:*

$$P(\lambda_2 \geq x | \omega_1) \leq P\left[F_{N-1, T+1-N-K} \geq x(T+1-N-K) / (1-x)(N-1) \right], \quad \forall \omega_1 \geq 0,$$

and

$$P(\lambda_2 \leq x | \omega_1) \geq k_2 \left(2 \int_0^{1-x} t^{2n_2+1} (1-t)^{2n_1+1} \mathrm{d}t - x^{n_1+1} (1-x)^{n_2+1} \times \int_0^{1-x} t^{n_2} (1-t)^{n_1} \mathrm{d}t \right),$$

where $n_1 = (N-3)/2$, $n_2 = (T-K-N-2)/2$, $k_1 = \Gamma(T-K)/4\Gamma(N-1)\Gamma(T-K-N)$, $k_2 = k_1/(n_1 + n_2 + 2)$, *and* $\Gamma(a)$ *is the gamma function* (*evaluated at* a).

The bounds are independent of any unknown parameters, and are optimal in the sense that they are the tightest possible among all distributions that are not dependent on ω_1. In light of Theorem 1, the importance of the bounds is limited. However, the bounds seem useful in understanding how the exact distribution varies as the nuisance parameter takes extreme values. In addition, the bounds are much easier to compute compared with the previous exact procedure, and hence may be used by a broader spectrum of researchers to make quick inference on the validity of the null hypothesis. An upper bound on the P-value, if it is small, is useful for rejecting the null hypothesis. On the other hand, a lower bound on the P-value, if it is large, is useful for accepting the hypothesis. For example, if a 4% upper bound is found, we can surely reject the null at the usual 5% level.

2.4. GMM test

The generalized method of moments (GMM) approach provides an over-identification test of Eq. (2) by examining the moment conditions imposed by Eqs. (1) and (2). Let $\boldsymbol{\theta} = \mathrm{vec}(\gamma, \boldsymbol{\beta})$ be an $(NK+1)$-vector of the parameters, and $\mathbf{Z}_t$ be an L-vector of instrumental variables $(L \geq (K+1))$. Following Hansen (1982), we obtain the GMM estimator by minimizing a weighted quadratic form of the sample moments:

$$\min \boldsymbol{g}'_T(\boldsymbol{\theta}) \mathbf{W}_T \mathbf{g}_T(\boldsymbol{\theta}), \tag{17}$$

over the parameter space $\boldsymbol{\theta}$, where $\mathbf{g}_T(\boldsymbol{\theta}) = 1/T \sum_{t=1}^{T} \mathscr{E}_t(\boldsymbol{\theta}) \otimes \mathbf{Z}_t$, $\mathscr{E}_t(\boldsymbol{\theta}) = (\mathbf{R}_t - \gamma 1_N) - \boldsymbol{\beta}(\mathbf{R}_{\mathrm{p}t} - \gamma 1_K)$, and $\mathbf{W}_T$, $NL \times NL$, is the weighting matrix. In the multi-beta model (1) or (9), the first order conditions are $E(\mathscr{E}_t | \mathbf{R}_{\mathrm{p}t}) = \mathbf{0}$. In the GMM framework, the model residuals $\mathscr{E}_t$ can have general heteroskedasticity that includes as special cases the independently and identically normally distributed assumption and the normality assumption. To specify an estimator for the covariance matrix of the residuals, it is natural to assume that the first order conditions continue to hold conditional on past information of $\mathbf{R}_{\mathrm{p}t}$, then [5] a consistent estimator, $\mathbf{S}_T$, of the covariance matrix is, as given by MacKinlay and Richardson (1991),

$$\mathbf{S}_T = \frac{1}{T} \sum_{t=1}^{T} (\mathscr{E}_t \mathscr{E}'_t \otimes \mathbf{Z}_t \mathbf{Z}'_t), \tag{18}$$

where $\mathbf{Z}_t = (1, \mathbf{R}'_{\mathrm{p}t})'$ is an $L = K+1$ vector.

Because of Hansen (1982), it is well-known that the optimal GMM estimator is obtained by solving (17) with $\mathbf{W}_T = \mathbf{S}_T^{-1}$, and the resulted quadratic form multiplied by the same size provides the standard GMM test for asset pricing restriction (2). However, as the hypothesis is complex and non-linear, there does seem any analytical solutions, and numerical optimization techniques have to be used to solve (17). In our later applications, there are up to 76 parameters. Clearly, numerical optimization is not an easy task to accomplish. In addition, a well-known problem with numerical optimization is that the solution may not converge to the global optimum or not converge at all. Hence, it is important to reduce the problem analytically so that the solution requires only one parameter to find numerically. There are two approaches for obtaining such almost analytical GMM

[5] Other assumptions may also be assumed, then $\mathbf{S}_T$ may no longer have an expression like (18), and an estimator, such as that of Newey and West (1987), may have to be used instead. However, this presents no difficulties at all because $\mathbf{S}_T$ enters into the GMM test only through Eq. (23) which, for an arbitrary $\mathbf{S}_T$, can be computed as easily as for a $\mathbf{S}_T$ of the form (18).

estimators. The first approach is to solve the GMM optimization problem by using not the optimal weighting matrix, but one with a structure form of

$$\mathbf{W}_T \equiv \mathbf{W}_1 \otimes \mathbf{W}_2, \quad \mathbf{W}_1{:}N \times N, \quad \mathbf{W}_2{:}(K+1) \times (K+1). \tag{19}$$

Then, we can easily solve the associated GMM optimization analytically in terms of one unknown parameter. Indeed, with the weighting matrix given by Eq. (19), the GMM estimator of $\boldsymbol{\beta}$ conditional on γ can be explicitly obtained as

$$\tilde{\boldsymbol{\beta}}' = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{Y}, \tag{20}$$

where $\mathbf{P} = \mathbf{Z}\mathbf{W}_2\mathbf{Z}'$ is $T \times T$ matrix, and $\mathbf{Y}$ and $\mathbf{X}$ are $T \times N$ and $T \times K$ matrices formed by stacking $\mathbf{R}_t - \gamma 1_N$ and $\mathbf{R}_{pt} - \gamma 1_K$, respectively. Furthermore, it is easy to show that the objective function becomes $\mathrm{tr}(\mathbf{W}_1\mathbf{Y}'[\mathbf{P} - \mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}\mathbf{P}]\mathbf{Y})$ divided by the squared value of T, where tr is the trace operator for a matrix. Hence, the GMM estimator of γ can be obtained by minimizing this function. Since there is only one variable, γ, a numerical subroutine, such as the golden section search (see, e.g, Press et al., 1989), may be used to accomplish the minimization easily in practice. It is observed that, conditional on $\boldsymbol{\beta}$, the estimator of γ can be explicitly written as

$$\gamma = \mathrm{tr}\big[\mathbf{W}_1\boldsymbol{\Phi}'\mathbf{P}(\mathbf{R} - \boldsymbol{\beta}\mathbf{R}_{\mathrm{p}})\big] / \mathrm{tr}(\mathbf{W}_1\boldsymbol{\Phi}'\mathbf{P}\boldsymbol{\Phi}), \tag{21}$$

where $\boldsymbol{\Phi} = 1_T \otimes (1_N - \boldsymbol{\beta}1_K)'$ is a $T \times N$ matrix. Therefore, an alternative and simpler approach is to start from an arbitrary initial value for γ, Eq. (20) obtains an estimate for $\boldsymbol{\beta}$. Then, based on this, Eq. (21) provides a new estimate for γ which can be used again in Eq. (20) to obtain another estimate for $\boldsymbol{\beta}$. Iterating back and forth between Eqs. (20) and (21), a convergent solution is easily computed in practice. This solution is the GMM estimator when the weighting matrix is chosen of form (19).

Because of the use of $\mathbf{W}_T$ in the particular form of Eq. (19), the standard asymptotic chi-squared distribution does not apply to the resulted quadratic form. Fortunately, based on Zhou (1994), a simple asymptotic chi-squared test can still be easily constructed. First, choosing the weighting matrix as the Kronecker product of an N dimensional identity matrix with $(\sum_{t=1}^{T} \mathbf{Z}_t\mathbf{Z}_t'/T)^{-1}$, the above procedure yields easily a consistent estimate of $\boldsymbol{\theta}$. Then, a weighting matrix of $\mathbf{W}_T = (\sum_{t=1}^{T}\mathscr{E}_t\mathscr{E}_t'/T)^{-1} \otimes (\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{Z}_t'/T)^{-1}$ can be chosen to get a new consistent estimate [6] of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}$. Let $\mathbf{g}_T = \mathbf{g}_T(\tilde{\boldsymbol{\theta}})$ and $\mathbf{D}_T = \mathbf{D}_T(\tilde{\boldsymbol{\theta}})$ be an $NL \times q$, $q = NK + 1$, matrix of the first order derivatives of $\mathbf{g}_T$ with respect to the free parameters.

[6] As pointed out by Zhou (1994), an efficient analytical estimator can be constructed by the score algorithm, as demonstrated in Newey (1985), if such an estimator is desired.

Then, the GMM test is given by

$$H_z \equiv T(\boldsymbol{\Psi}_T \mathbf{g}_T)' \mathbf{V}_T (\boldsymbol{\Psi}_T \mathbf{g}_T), \tag{22}$$

where $\mathbf{V}_T$ is an $NL \times NL$ diagonal matrix, $\mathbf{V}_T = \mathrm{Diag}(1/v_1, \dots, 1/v_d, 0, \dots, 0)$, formed by $v_1 > \dots > v_d > 0$ $(d = N-1)$, the positive eigenvalues of the following $NL \times NL$ semi-definite matrix:

$$\boldsymbol{\Omega}_T \equiv \left[\mathbf{I} - D_T(\mathbf{D}'_T \mathbf{W}_T \mathbf{D}_T)^{-1} D'_T \mathbf{W}_T\right] \mathbf{S}_T \left[\mathbf{I} - D_T(\mathbf{D}'_T \mathbf{W}_T \mathbf{D}_T)^{-1} \mathbf{D}'_T \mathbf{W}_T\right]'; \tag{23}$$

$\boldsymbol{\Psi}_T$ is an $NL \times NL$ matrix, of which the i-th row is the standardized eigenvector corresponding to the i-th largest eigenvalue of $\boldsymbol{\Omega}_T$ for $i = 1, \dots, NL$; and $\mathbf{S}_T$ is the estimator of the underlying covariance structure of the model residuals given by Eq. (18). As shown by Zhou (1994), assuming the very general conditions of Hansen (1982), H_z is asymptotically χ^2 distributed with degrees of freedom $N-1$.

An alternative approach for obtaining the almost analytical GMM estimator is to solve Eq. (17) directly in terms of γ by using the optimal weighting matrix. It is easy to show that vec $\boldsymbol{\beta}' = (\mathbf{x}' \mathbf{W}_T \mathbf{x})^{-1} \mathbf{x}' \mathbf{W}_T \mathbf{y}$, where $\mathbf{W}_T = \mathbf{S}_T^{-1}$, $\mathbf{y} = \mathrm{vec}(\mathbf{Z}\mathbf{Y}')/T$ and $\mathbf{x} = \mathbf{I}_N \otimes (\mathbf{Z}'\mathbf{X})/T$. Then, the GMM estimator is obtained similarly as before by solving a one-dimensional numerical optimization problem. An advantage of this approach is that the χ^2 test is simply given by the minimized quadratic form multiplied by T without the use of Eq. (22). However, as the inversion of $\mathbf{S}_T$ is used explicitly, the estimation and test are numerically less stable than the first approach in empirical applications. For simplicity, we will use only the first approach in what follows.

3. Extensions

In this section, we extend the GMM test to a more general multi-beta asset pricing model which has a vector of general economic factors in addition to the vector of the reference portfolio returns on the right hand side of Eq. (1). Then, in Section 3.2, we show how the pricing restrictions of arbitrage pricing theory (APT) by Ross (1976) can be transformed into restrictions similar to those in Eq. (2). Hence, the proposed tests can also be applied to test for the APT.

3.1. Additional non-portfolio factors

Shanken (1992) examines the following general multi-beta asset pricing model:

$$\mathbf{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 \mathbf{F}_{1t} + \boldsymbol{\beta}_2 \mathbf{F}_{2t} + \boldsymbol{\epsilon}_t, \tag{24}$$

where $\mathbf{F}_{1t}$ is a vector of general economic factors with $E[\mathbf{F}_{1t}] = 0$ and $\mathbf{F}_{2t}$ is a K-vector of asset returns on reference portfolios. For example, $\mathbf{F}_{1t}$ might contain innovations in certain macroeconomic variables, whereas $\mathbf{F}_{2t}$ could include stock and bond index returns. The asset pricing restrictions impose the following constraints on the parameters in Eq. (24):

$$\boldsymbol{\alpha} = \boldsymbol{\beta}_1 \boldsymbol{\gamma}_1 + \gamma(\mathbf{1}_N - \boldsymbol{\beta}_2 \mathbf{1}_K), \tag{25}$$

where γ is the return on the zero-beta asset and $\boldsymbol{\gamma}_1$ is a vector of risk premiums on the general factor $\mathbf{F}_1$. In the absence of $\mathbf{F}_1$, Eq. (24) clearly reduces to Eq. (1), and Eq. (25) collapses to Eq. (2). In his extensive study, Shanken (1992) was the first to develop methods for testing Eq. (25) by relaxing the normality assumption that had been usually assumed. Following Shanken (1992), we generalize the GMM test provided in the previous section to test Eq. (25) under a very general form of conditional heteroskedasticity, which contains as special cases the usual normality assumption and the iid assumption used in Shanken (1992).

First, the GMM estimator of the parameters can be obtained as follows. Conditional on γ and $\boldsymbol{\gamma}_1$, the GMM estimator of $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is $\tilde{\boldsymbol{\beta}}' = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{Y}$, where $\mathbf{P}$ and $\mathbf{Y}$ are defined the same way as before, but $\mathbf{X}$ is enlarged to include $\mathbf{F}_1$ plus its risk premiums as its first submatrix whereas its second submatrix is unchanged. Then, the objective function will still have the earlier expression, and hence the GMM estimator of γ, $\boldsymbol{\gamma}_1$, and that of $\boldsymbol{\beta}$ can be obtained by minimizing the objective function over a few parameters, γ and $\boldsymbol{\gamma}_1$. Although this is more complex than the one variable case, a numerical subroutine can clearly be used for this task which seem easy to accomplish in practice. Now, with the parameter estimates, the GMM test is straightforwardly computed by adjusting the components in Eq. (23) accordingly.

3.2. APT as multi-beta restrictions

As an important extension of the asset pricing model of Sharpe (1964) and Lintner (1965), Ross (1976) derives the arbitrage pricing theory (APT). In contrast to the multi-beta pricing model where asset returns are linear combinations of returns on observable reference portfolios plus noises, the APT assumes that the returns on a vector of N assets are related to K pervasive and unknown factors by a K-factor model:

$$r_{it} = E[r_{it}] + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{26}$$

where f_{kt} is the k-th pervasive factor, ϵ_{it} is the idiosyncratic factor of asset i, β_{ik} is the beta or factor loading of the k-th factor for asset i, N is the number of assets, and T is the number of periods.

The APT implies that the expected return of an asset is approximately a linear function of the risk premiums on systematic factors in the economy. There have been both a large theoretical literature extending the APT and a large empirical literature testing its implications. [7] The often tested restrictions of the APT is

$$E[\mathbf{r}_t] = r_{\mathrm{f}t}\mathbf{1}_N + \boldsymbol{\beta}\boldsymbol{\lambda}_t, \tag{27}$$

where $r_{\mathrm{f}t}$ is the risk-free rate of return, $\mathbf{1}_N$ is an N-vector of ones and $\boldsymbol{\lambda}_t$ is an N-vector of the risk premiums. Eq. (27) is the implication, for example, of the equilibrium version APT (Connor, 1984). In the case where a riskless asset does not exist, the restrictions become

$$E[\mathbf{r}_t] = \lambda_{0t}\mathbf{1}_N + \boldsymbol{\beta}\boldsymbol{\lambda}_t, \tag{28}$$

where λ_0 is the return on the zero-beta asset.

The most widely used approach for testing the APT is a two-pass procedure. In the first pass, either the factor loadings or the factors are estimated. Then, in the second pass, the regression of the returns on the estimated loadings or the factors is estimated. Treating the estimates as the true variables, the APT restrictions become linear constraints (implying zero-intercepts) on the regression coefficients in a multivariate regression and hence can be tested by using standard methods such as the Gibbons et al. (1989) test. However, this procedure suffers an errors-in-variables problem, because the estimated rather than the actual factor loadings or factors are used in the second pass tests. As known in the errors-in-variables literature, ignoring the uncertainty of the estimates can potentially lead to incorrect inference.

The errors-in-variables problem arises from using estimated rather than the actual factor loadings or factor. In fact, the APT restrictions can be tested without using the factors at all. Motivated by Gibbons and Ferson (1985), we assume, without loss of generality, that the first $K \times K$ submatrix of $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)$ is nonsingular. In the riskless asset case, solving $\boldsymbol{\lambda}$ from the first K equations of (27), we obtain $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1^{-1}(E[\mathbf{r}_{1t}] - r_{\mathrm{f}t}\mathbf{1}_K)$, where $\mathbf{r}_{1t}$ is the first K returns of $\mathbf{r}_t = (\mathbf{r}'_{1t}, \mathbf{r}'_{2t})$. Then, from the remaining $N-K$ equations of (27), we have

$$E[\mathbf{r}_{2t}] = r_{\mathrm{f}t}\mathbf{1}_{N-K} + \boldsymbol{\beta}_2\boldsymbol{\beta}_1^{-1}\big(E[\mathbf{r}_{1t}] - r_{\mathrm{f}t}\mathbf{1}_K\big). \tag{29}$$

Hence, the APT restriction implies zero intercepts, i.e.,

$$H_0 \colon \boldsymbol{\alpha} = 0 \tag{30}$$

in the following regression of excess returns $\mathbf{R}_{2t} = \mathbf{r}_{2t} - r_{\mathrm{f}t}\mathbf{1}_{N-K}$ on $\mathbf{R}_{1t} = \mathbf{r}_{1t} - r_{\mathrm{f}t}\mathbf{1}_K$:

$$\mathbf{R}_{2t} = \boldsymbol{\alpha} + \boldsymbol{\theta}\mathbf{R}_{1t} + \mathbf{u}_t. \tag{31}$$

It is easy to verify that H_0 is not only necessary, but also sufficient for the validity of the APT. In other words, if Eqs. (30) and (31) are true, so are the APT

[7] Connor and Korajczyk (1995) provide an excellent survey of the literature.

restrictions. With the standard multivariate normality assumption on returns (which is also consistent with the factor model (26)), Gibbons et al. (1989) applies to Eq. (31) straightforwardly to yield an exact test of the APT. When there is a concern about the normality assumption, a GMM test can be used. Given any weighting matrix of the form (19) with N being replaced by $(N-K)$, the GMM estimator of α under the null is $\tilde{\boldsymbol{\alpha}} = (\mathbf{R}_1'\mathbf{P}\mathbf{R}_1)^{-1}\mathbf{R}_1'\mathbf{P}\mathbf{R}_2$, where $\mathbf{P}$ is defined accordingly. Then, a GMM test is similarly defined as before, and it is asymptotically χ^2 distributed with degrees of freedom $N-K$. Alternatively, the GMM tests of MacKinlay and Richardson (1991) and Harvey and Zhou (1993), which are developed in the CAPM context, can be extended to test Eq. (31).

In the no riskless asset case, we assume that the risk premiums $\boldsymbol{\lambda}_t$ are no longer time-varying, but $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}$, a constant vector over the testing period. In addition, assume $\lambda_{0t} = \lambda_0$. Then, we solve $\boldsymbol{\lambda}$ from Eq. (28) to obtain $\boldsymbol{\lambda} = \boldsymbol{\beta}_1^{-1}E[\mathbf{r}_{1t}] - \lambda_0\boldsymbol{\beta}_1^{-1}\mathbf{1}_K$. Hence,

$$E[\mathbf{r}_{2t}] = \lambda_0\left(\mathbf{1}_{N-K} - \boldsymbol{\beta}_2\boldsymbol{\beta}_1^{-1}\mathbf{1}_K\right) + \boldsymbol{\beta}_2\boldsymbol{\beta}_1^{-1}E[\mathbf{r}_{1t}]. \tag{32}$$

This implies that a test of the K-factor APT is a test of the hypothesis,

$$H_0\colon \boldsymbol{\alpha} = \lambda_0(\mathbf{1}_{N-K} - \boldsymbol{\theta}\mathbf{1}_K) \tag{33}$$

in the multivariate regression

$$\mathbf{r}_{2t} = \boldsymbol{\alpha} + \boldsymbol{\theta}\mathbf{r}_{1t} + \mathbf{u}_t. \tag{34}$$

Since Eq. (33) has exactly the same form as the multi-beta pricing restrictions (2), both the small sample and GMM tests developed in the previous section are straightforward to apply. In fact, only a slight change of the notations is needed. For example, $N-K$ here plays the role of previous N and $\mathbf{r}_{1t}$ plays the role of previous $\mathbf{R}_{\mathrm{p}t}$.

4. Efficiency test

In this section, we examine the efficiency of a portfolio of two stock indices, the value-weighted index and equal-weighted index available from the Center for Research in Security Prices (CRSP) at the University of Chicago. The data is monthly returns on market-value-sorted New York Stock Exchange (NYSE) portfolio deciles varying from size 1 to size 10, also available from the CRSP. The returns are from January 1926 to December 1994, a total of 69 years data. As shown by Shanken (1985), MacKinlay and Richardson (1991) and Zhou (1991), the efficiency of either of the two indices is usually rejected by the data (a further test is provided below). Hence, it is of interest to examine whether the efficiency of a portfolio of the two indices can be rejected. In this case, $N = 10$, $K = 2$, $\mathbf{R}_t$ is a vector of returns on the 10 size portfolios, and $\mathbf{R}_{pt}$ is a vector of returns on the two indices. Concerning about the parameter stability, we conduct our study for both 10-year subperiods and for the whole period, implying $T = 120$ and $T = 828$, respectively. The LRT estimation is straightforward to carry out, and the GMM estimation is implemented by using both the golden search and iteration methods which yield, as expected, the same results. Then, the test statistics and the associated P-values are readily evaluated.

Panel A of Table 1 provides the results. The second, third and the fourth columns of the table provide the exact P-value of the likelihood ratio test (LRT), its upper and lower bounds. Consider first the results in the subperiods. The LRT P-values are successively 0.7%, 6% and greater than 14% for all of the subperiods. Hence, it is interesting that, although we often reject the efficiency of either of the two indices at the usual 5% level, we can no longer reject efficiency of a portfolio of the two indices in all the 10-year subperiods except the first one. The rejection over the first 10-year subperiod, February 1926 to January 1935, might bear some relation to the enormous volatility of the stock returns during the Great Depression period. In the fifth column of the table, the P-value of the GMM test is

Table 1
Joint efficiency of stock indices
Panel A of the table reports the exact P-value from the likelihood ratio test, its upper and lower bounds, and the P-value based on the GMM test. Panel B of the table reports the P-value of the GMM test, H_V (H_E), for whether or not it is sufficient to use only the value-weighted index (the equal-weighted index) vs. using both of them in the efficiency restrictions.

Panel A: Joint efficiency

Period	LRT	U	L	GMM
2/1926–1/1935	0.007	0.007	0.000	0.030
2/1936–1/1945	0.060	0.061	0.002	0.022
2/1946–1/1955	0.145	0.147	0.014	0.166
2/1956–1/1965	0.341	0.343	0.085	0.215
2/1966–1/1975	0.473	0.474	0.175	0.461
2/1976–1/1985	0.782	0.784	0.325	0.591
2/1926–12/1994	0.153	0.153	0.016	0.055

Panel B: Single index efficiency

Period	H_V	H_E
2/1926–1/1935	<0.001	<0.001
2/1936–1/1945	<0.001	<0.001
2/1946–1/1955	<0.001	<0.001
2/1956–1/1965	<0.001	<0.001
2/1966–1/1975	<0.001	<0.001
2/1976–1/1985	<0.001	<0.001
2/1926–12/1994	<0.001	<0.001

reported. The GMM allows a much more general form of distributional assumptions than the normality one which the LRT relies on. Interestingly, the GMM test yields very similar conclusions, although the P-values are generally smaller than those of the LRT. Over the whole period, the P-values from both the LRT and the GMM tests are 15.30% and 5.51%, respectively. This suggests again that we cannot reject the efficiency at the usual 5% level. Zhou (1994) shows that the above GMM test has good finite sample properties with sample size as small as 120. Hence, one may expect the GMM test is reliable in the whole sampling period with $T = 828$. And yet, the P-value of the GMM test is 5.51%, about three times smaller than that from the exact LRT test. There are two explanations for this apparent paradox. The first one is due to the fact that the GMM test has lower P-values over the subperiods, resulting in a lower whole period P-value as the whole period P-value is in some way an aggregate of the subperiod P-values. A second explanation is that the paradox may be caused by the underlying distributional assumptions. In fact, if the normality assumption is imposed, the consistent estimator of the covariance matrix is then $(\sum_{t=1}^{T} \mathscr{E}_t \mathscr{E}_t'/T) \otimes (\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t'/T)$, and the resulted GMM test yields a P-value of 15.31%, which is virtually the same as the P-value from the exact LRT test. So, the earlier P-value of 5.51% might simply reveal that there are some evidence of conditional heteroskedasticity in the data.

Another interesting fact from the table is that the upper bounds [8] (the P-values evaluated at an infinity value of the nuisance parameter) are very close to the P-values evaluated at the maximum likelihood estimates. This seems due to the fact that the nuisance parameter is usually large and the P-values computed from sizable perturbations of the nuisance parameter are still close to one another (see footnote 1). From this and good finite sample properties of the asymptotic distribution of the GMM test, it appears that the P-values evaluated at the maximum likelihood estimates should be close to the true but unknown P-values (since the true nuisance parameter will never be known exactly). In contrast to the upper bounds, the lower bounds are generally far away from the P-values evaluated at the maximum likelihood estimates, suggesting that the exact distribution does shift substantially when the nuisance parameter approaches zero. Nevertheless, even if one is extremely conservative by assigning a zero value to the nuisance parameter, he still cannot reject the efficiency at the usual 5% level for the last three subperiods.

Finally, it is of interest to see how the efficiency test results change with only the CRSP value-weighted index or only the equal-weighted index. There are two

[8] Although not reported in the table, we have done experiments with the bounds of Shanken (1985), and find that, like the single beta case reported in Zhou (1991), they are fairly close to the optimal bounds in many cases.

kinds of tests for the efficiency of a single index. The first is to take $K=1$ as the true model, and the second is to take $K=2$ as the true model and nest the one factor within to test whether all the betas of the excluded factor are zero in addition to the single zero-beta asset pricing restrictions. As Shanken (1985), MacKinlay and Richardson (1991) and Zhou (1991) are exactly those that carry out the first kind, we will not repeat them here. Hence, we will provide only the second kind which is easily implemented by simply adjusting the previous formulas for the GMM test. Panel B of Table 1 provides the results. The P-values of the GMM test are less than 0.1% for all the periods, subperiods and entire period, and for either of the indices, and so we reject the efficiency of either of the indices. Comparatively speaking, the P-values are generally much smaller than those of the joint efficiency test and those of the first kind test. Econometrically, if we are to test for the univariate efficiency of the value-weighted index, we have N more moment conditions resulted from the equal-weighted index than the first kind of test which uses only the moment conditions resulted from the value-weighted index. As the first kind of test already rejects the model, it is not surprising that the nested second kind test rejects even much stronger, as evidenced here by the very small P-values. From an economic standpoint, the first kind test compares the efficiency of the value-weighted index with the group formed by this index and other assets. In contrast, the second kind test compares it with the group formed by both of the indices and other assets. Hence, the second kind efficiency test should be more powerful and have in general smaller P-values than the earlier one.

5. Testing Fama–French model

In this section, we test three versions of the Fama and French (1993) factor model. They provide three common risk factors to explain the average returns on 25 stock portfolios formed on size and book-to-market (B/M). The factors are an overall market factor, M, as represented by the returns on the weighted market portfolio; a size factor, SMB, as represented by the return on small stocks minus those on large stocks and a book-to-market factor, HML, as represented by the return on high B/M stocks minus those on low B/M stocks. The data are monthly from July 1964 to December 1992. So there are 341 observations. [9]

[9] See Fama and French (1993), Harvey and Kirby (1995), and He et al. (1996) for a detailed description of the data. The authors are grateful to C. Harvey and R. Kan for permission and forward of their data.

We test first the standard zero-beta version of the three-factor model of Fama and French (1993). In this case, we have the following linear regression:

$$r_{it} = \alpha_i + \beta_{i1} f_{M,t} + \beta_{i2} f_{\mathrm{SMB},t} + \beta_{i3} f_{\mathrm{HML},t} + \epsilon_{it}, \tag{35}$$

where r_{it}'s are returns on the 25 stock portfolios, and f_M, f_{SMB} and f_{HML} are returns on the factors. The pricing restriction is

$$\alpha_i = \gamma(1 - \beta_{i1}), \quad i = 1, \dots, N. \tag{36}$$

Notice that this restriction is similar but different from Eq. (2). However, the GMM objective function still has the same form except that f_M, f_{SMB} and f_{HML} enter the formula without subtracting the zero-beta rate. Hence, both the estimation and test are computed in the same way with a slight change of the notations. The first row of Table 2 provides the results. The first column $\gamma_t = \gamma$ indicates the case where the gamma rate is assumed to be constant. The second and the third columns report the GMM test statistic, 51.8395, and the associated P-value, 0.0008, respectively. As the P-value is too small, the test suggests a strong rejection of the model.

Now we test a CAPM version of the three-factor model of Fama and French (1993) and its extension. The market model regression is

$$r_{it} - r_{\mathrm{f}t} = \alpha_i + \beta_{i1}(f_{M,t} - r_{\mathrm{f}t}) + \beta_{i2} f_{\mathrm{SMB},t} + \beta_{i3} f_{\mathrm{HML},t} + \epsilon_{it}, \tag{37}$$

where r_{it}'s and the f's are the same as before, and $r_{\mathrm{f}t}$ is the 30-day T-bill rate. In the case where the zero-beta rate equals to the T-bill rate, the pricing restriction is the standard CAPM type restriction

$$\alpha_i = 0, \quad i = 1, \dots, N. \tag{38}$$

The second row of Table 2 provides the GMM statistic, 55.1252 and the associated P-value 0.0005, suggesting again a rejection. In the extension of the CAPM type model where the zero-beta rate is the T-bill rate plus a constant c_0, the pricing restriction is

$$\alpha_i = c_0(1 - \beta_{i1}), \quad i = 1, \dots, N. \tag{39}$$

The last row of Table 2 provides the GMM statistic, 50.8437, with a P-value of 0.0011. In contrast with testing the CAPM type restriction (38), the P-value is

Table 2
Testing Fama–French model

Assumptions on zero-beta rate	GMM test statistic	P-value
$\gamma_t = r_{\mathrm{f}t}$	55.1252	0.0005
$\gamma_t = r_{\mathrm{f}t} + c_0$	50.8437	0.0011
$\gamma_t = \gamma_0$	51.8395	0.0008

The table provides a GMM test and the associated P-value for the asset pricing restrictions of the three different versions of the Fama–French model.

twice as large. However, it is still within the usual rejection range. But it is clear that the extended model performs the best among the three models. However, the better performance does not seem overwhelming.

The reason why the models are rejected seems related to the recent CAPM debate, in which Fama and French (1992), among others, argue against the CAPM whereas Jagannathan and Wang (1996), among others, argue for it. It is of importance to examine whether the rejection or acceptance of the CAPM is due to problems in the statistical procedures used or the fundamental behavior of the data. As this requires substantial amount of study and analysis, We leave it as part of our future research.

6. Conclusions

Multi-beta asset pricing models are an important class of models in finance, and hence their empirical validity is of interest to both finance researchers and practitioners. Empirical tests are usually carried out with the normality assumption and without the normality assumption. Under the normality assumption, we obtain the exact distribution of the likelihood ratio test, based on which a small sample test can be used to test the model. Without the normality assumption, extending Shanken (1992), we propose a GMM (generalized method of moments) test that allows a very general form of conditional heteroskedasticity. Unlike a standard GMM application, our GMM test can be computed almost analytically (requiring the minimization over only a single variable), and hence it is straightforward to apply in practice. It appears that we give a complete solution to testing multi-beta asset pricing models: a small sample likelihood ratio test if normality is assumed, and a GMM test if the normality assumption is relaxed. These tests are also shown to be applicable to Ross's (1976) APT.

Based on the proposed tests, we examine the efficiency of a portfolio of two stock indices, the value-weighted index and equal-weighted index available from the Center for Research in Security Prices (CRSP) at the University of Chicago. Using monthly size portfolio returns from January 1926 to December 1994, we find that, while most existing studies reject the efficiency of a single index, the efficiency of a portfolio of the CRSP value-weighted index and equal-weighted index cannot be rejected for all of the 10-year subperiods except for the first one. We also apply the tests to analyze a new version of the three-beta model of Fama and French (1993) in addition to two standard ones. Although all three versions are rejected by the data at the usual 5% significance level, our results suggest that the new version performs the best among the three.

Acknowledgements

We are grateful to Philip Dybvig, Campbell Harvey, Steve Heston, Raymond Kan, Lilian Ng, Carmela Quintos, and especially John Thatcher, Christian Wolff (the Editor), an anonymous referee and an Associate Editor for many helpful comments that substantially improved the paper.

Appendix A. Proof of Theorem 1

To show Eq. (15), consider minimizing $Q(\gamma)$ for all $\mathbf{w}=(w_1,w_2)'$. By a result of matrix theory (Anderson, 1984, p. 590), we know that the minimum is equal to the smallest eigenvalue of $|\mathbf{A}-f\mathbf{B}|=0$ and the minimum is achieved if $\mathbf{w}$ is the eigenvector corresponding to the smallest eigenvalue. Since it is invariance to scaling of $\mathbf{w}$, $Q(\gamma)$ also achieves the minimum with $\mathbf{w}=(1,w_2/w_1)'$. The eigenvalue f is clearly equal to $\lambda/(1-\lambda)$, and hence Eq. (15) follows.

To determine the distribution of λ_2, we rewrite $\mathbf{A}$ as

$$\mathbf{A}=\begin{pmatrix}\hat{\boldsymbol{\alpha}}'\\ -\hat{\boldsymbol{\xi}}'\end{pmatrix}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\alpha}},-\hat{\boldsymbol{\xi}}), \tag{A.1}$$

and hence the eigenvalues of Eq. (14) are the same as the eigenvalues of

$$\left|\sqrt{T}\mathbf{B}^{-1/2}\begin{pmatrix}\hat{\boldsymbol{\alpha}}'\\ -\hat{\boldsymbol{\xi}}'\end{pmatrix}(T\hat{\boldsymbol{\Sigma}})^{-1}(\hat{\boldsymbol{\alpha}},-\hat{\boldsymbol{\xi}})\sqrt{T}\mathbf{B}^{-1/2}-\frac{\lambda}{1-\lambda}\mathbf{I}\right|=0. \tag{A.2}$$

Consider now the distribution of $(\hat{\boldsymbol{\alpha}},-\hat{\boldsymbol{\xi}})$ and $\hat{\boldsymbol{\Sigma}}$. It is well known from multivariate analysis (e.g., Muirhead, 1982, p. 431) that $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\Sigma}}$ are statistically independent and have multivariate normal and Wishart distributions respectively (using the matrix normal notation of Muirhead (1982, p. 79)):

$$(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})'\sim \mathrm{N}\left[(\boldsymbol{\alpha},\boldsymbol{\beta})',(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}\right],\quad T\hat{\boldsymbol{\Sigma}}\sim W_N(T-K-1,\boldsymbol{\Sigma}), \tag{A.3}$$

where $\mathbf{X}=(1_T,\mathbf{R}_{\mathrm{p}})$ is a $T\times(K+1)$ matrix formed by ones and the returns on the K reference portfolios. Since $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}1_K)'=\mathbf{M}'(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})'$, with $\mathbf{M}$, a $(K+1)\times 2$ matrix, satisfying

$$\mathbf{M}'(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{M}=\begin{pmatrix}1 & 0\\ 0 & 1_K'\end{pmatrix}\begin{pmatrix}1+\overline{\mathbf{R}}_{\mathrm{p}}'\hat{\boldsymbol{\Delta}}^{-1}\overline{\mathbf{R}}_{\mathrm{p}} & -\overline{\mathbf{R}}_{\mathrm{p}}'\hat{\boldsymbol{\Delta}}^{-1}\\ -\hat{\boldsymbol{\Delta}}^{-1}\overline{\mathbf{R}}_{\mathrm{p}} & \hat{\boldsymbol{\Delta}}^{-1}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1_K\end{pmatrix}=\mathbf{B},$$

it follows that the covariance matrix of $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}1_K)'$ is $\mathbf{M}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{M}\otimes\boldsymbol{\Sigma}=(\mathbf{B}\otimes\boldsymbol{\Sigma})/T$, implying that $\sqrt{T}\mathbf{B}^{-1/2}(\hat{\boldsymbol{\alpha}},-\hat{\boldsymbol{\xi}}1_K)'$ has a mean $\sqrt{T}\mathbf{B}^{-1/2}(\boldsymbol{\alpha},-\boldsymbol{\xi})'$ and covariance matrix $\mathbf{I}_2\otimes\boldsymbol{\Sigma}$.

Now notice that $(T-K-1)\geq N\geq 2$, we have from Theorem 10.4.5 of Muirhead (1982, p. 454) that the joint density function of the eigenvalues of (A.2), λ_1 and λ_2, is a function of the hypergeometric function of matrix arguments with non-central parameter matrix $\boldsymbol{\Omega}=\mathbf{B}^{-1/2}(\boldsymbol{\alpha},-\boldsymbol{\xi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha},-\boldsymbol{\xi})\mathbf{B}^{-1/2}$ (see James, 1964 or Chapter 7 of Muirhead, 1982). Then, by using James (1964) or Theorem 10.4.2 of Muirhead (1982, p. 450), the joint density of λ_1 and λ_2 has exactly the same form as the joint density of the eigenvalues of $\mathbf{C}(\mathbf{C}+\mathbf{D})^{-1}$, where $\mathbf{C}\sim W_2(N,\mathbf{I}_2,\boldsymbol{\Omega})$ and $\mathbf{D}\sim W_2(T+1-N-K,\mathbf{I}_2)$. Then, Theorem 1 follows by using the fact that the non-zero eigenvalue of $\mathbf{\Omega}$ is ω_1 and by decomposing the Wishart matrices $\mathbf{C}$ and $\mathbf{D}$ into $\mathbf{C}=\mathbf{U}\mathbf{U}'$ and $\mathbf{D}=\mathbf{V}\mathbf{V}'$. Q.E.D.

Appendix B. An explicit expression for the exact density

Based on Theorem 1, we obtain, similar to Zhou (1991), an explicit expression for the exact density of λ_2.

Theorem 3. *Under the null hypothesis that the constraints (2) are true, if $N \geq 2$ and $T \geq N+K+1$, then the exact density of λ_2 is:*

$$f(\lambda_2) = k_1 \mathrm{e}^{-\frac{\omega_1}{2}} \sum_{j=0}^{\infty} \sum_{k=0}^{j} c_{j,k} \lambda_2^{n_1} (1-\lambda_2)^{2n_2+k+2} \times F(-n_1-j+k, n_2+1, n_2+k+3, 1-\lambda_2),$$

where $n_1=(N-3)/2$, $n_2=(T-K-N-2)/2$, $\Gamma(a)$ is the gamma function (evaluated at a), $k_1=\Gamma(T-K)/4\Gamma(N-1)\Gamma(T-K-N)$, and $c_{j,k}=(T/2)_j(-j)_k\ (1/2\,\omega_1)^j(1/2)_k\Gamma(n_2+1)\Gamma(k+2)/[(N/2)_j j!(k!)^2\Gamma(n_2+k+3)]$, with the factorial function $(a)_j \equiv a(a+1)\cdots(a+(j-1)), \forall a$, and $F(a,b,c;z)$ is the hypergeometric function (or Gaussian function), $F(a,b,c;z) \equiv 1+\sum_{j=1}^{\infty} (a)_j(b)_j z^j/[(c)_j j!]$.

References

Anderson, T.W., 1984, An Introduction of Multivariate Statistical Analysis, 2nd edn., Wiley, New York, NY.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444–454.

Breeden, D.T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7, 265–296.

Connor, G., 1984. A united beta pricing theory. Journal of Economic Theory 34, 13–31.

Connor, G., Korajczyk, R.A., 1995. The arbitrage pricing theory and multifactor models of asset returns. In: Jarrow, R.A. (Ed.), Handbooks in Operations Research and Management Science: Finance, Vol. 9, North-Holland, Amsterdam, 87–144.

Fama, E.F., 1991. Efficient capital markets: II. Journal of Finance 46, 1575–1617.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Geweke, J.F., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.

Gibbons, M.R., 1982. Multivariate tests of financial models: a new approach. Journal of Financial Economics 10, 3–27.

Gibbons, M.R., Ferson, W., 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. Journal of Financial Economics 14, 217–236.

Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Hansen, L.P., 1982. Large sample properties of the generalized method of moments estimators. Econometrica 50, 1029–1054.

Harvey, C.R., Zhou, G., 1993. International asset pricing with alternative distributional specifications. Journal of Empirical Finance 1, 107–131.

Harvey, C.R., Kirby, C., 1995. Analytic tests of factor pricing models, working paper, Duke University, Durham, NC.

He, J., Kan, R., Ng, L.K., Zhang, C., 1996. Tests of the relationships among economic forces, fundamental variables, and equity returns using multi-factor models. Journal of Finance 51, 1891–1908.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected return. Journal of Finance 51, 3–53.

Kandel, S., 1984. The likelihood ratio tests statistic of mean–variance efficiency without a riskless asset. Journal of Financial Economics 13, 575–592.

Kandel, S., 1986. The geometry of the likelihood estimator of the zero-beta return. Journal of Finance 31, 339–346.

Kandel, S., Stambaugh, R.F., 1989. A mean–variance framework for tests of asset pricing models. Review of Financial Studies 2, 125–156.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

MacKinlay, A.C., Richardson, M.P., 1991. Using generalized method of moments to test mean–variance efficiency. Journal of Finance 46, 511–527.

Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.

Muirhead, R.J., 1982, Aspects of Multivariate Statistical Theory, Wiley, New York.

Nanda, D.N., 1947. Distribution of a root of a determinantal equation. Annals of Mathematical Statistics 18, 47–57.

Newey, W.K., 1985. Generalized method of moments specification testing. Journal of Econometrics 29, 229–256.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Perlman, M.D., Oklin, I., 1980. Unbiasedness of invariant tests for MANOVA and other multivariate problems. Annals of Statistics 8, 1326–1341.

Press, W., Flannery, B., Teukolsky, S., Vettering, W., 1989. Numerical recipes, Cambridge Univ. Press.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Schott, J.R., 1984. Optimal bounds for the distributions of some test criteria for tests of dimensionality. Biometrika 71, 561–567.

Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. Journal of Financial Economics 14, 327–348.

Shanken, J., 1986. Testing portfolio efficiency when the zero-beta rate is unknown: a note. Journal of Finance 41, 269–276.

Shanken, J., 1992. On the estimation of beta-pricing models. Review of Financial Studies 5, 1–33.

Shanken, J., 1996. Statistical methods in tests of portfolio efficiency: a synthesis. In: Maddala, G.S., Rao, C.R. (Eds.), Handbooks of Statistics: Statistical Methods in Finance, Vol. 14, North-Holland, Amsterdam, 693–711.

Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

Zhou, G., 1991. Small sample tests of portfolio efficiency. Journal of Financial Economics 30, 165–191.

Zhou, G., 1994. Analytical GMM tests: asset pricing with time-varying risk premiums. Review of Financial Studies 7, 687–709.