

NBER WORKING PAPER SERIES

EMPIRICAL EVALUATION OF ASSET PRICING MODELS: A COMPARISON OF THE
SDF AND BETA METHODS

Ravi Jagannathan
Zhenyu Wang

Working Paper 8098
<http://www.nber.org/papers/w8098>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2001

For helpful comments, we thank John Cochrane, Wayne Ferson, Bob Hodrick, Narayana Kocherlakota, René Stulz (editor), Guofu Zhou, the anonymous referee and the seminar participants at the Federal Reserve Bank of New York, Columbia University, the University of Southern California, Washington University at St. Louis, the NBER Conference on Asset Pricing and Portfolio Choice held in May 2000, and the Western Finance Association Meeting held in June 2000. We especially benefited from discussions with Kent Daniel, Lars Hansen and John Heaton. Adam Kolasinski provided excellent research assistance and Lolotte Olkowski provided editorial help. First draft: September 12, 1999. Earlier draft has been circulated under the title “Efficiency of the stochastic discount factor method for estimating risk premiums.” The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2001 by Ravi Jagannathan and Zhenyu Wang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Empirical Evaluation of Asset Pricing Models: A Comparison of the SDF and Beta Methods
Ravi Jagannathan and Zhenyu Wang
NBER Working Paper No. 8098
January 2001
JEL No. G0, C5

ABSTRACT

The stochastic discount factor (SDF) method provides a unified general framework for econometric analysis of asset pricing models. It has recently been pointed out that the generality of the SDF method may come at the cost of estimation efficiency. We show that there is no need for this concern. The SDF method is as efficient as the classical beta method for estimating risk premia. In addition, the SDF method has an advantage – the classical beta method, unlike the SDF method, substantially understates the effect of sampling errors when the estimated unanticipated changes in macroeconomic variables are used as pervasive factors.

Ravi Jagannathan
Kellogg Graduate School of Management
Northwestern University
2001 Sheridan Road
Leverone/Anderson Complex
and NBER
rjaganna@northwestern.edu

Zhenyu Wang
Graduate School of Business
Columbia University

1 Introduction

The use of the *stochastic discount factor (SDF) method* for econometric evaluation of asset pricing models has become common in the recent empirical finance literature. A SDF has the following property: the value of a financial asset equals the expected value of the product of the payoff on the asset and the SDF. An asset pricing model identifies a particular SDF that is a function of observable variables and model parameters. Hence, each asset pricing model can be identified using its SDF, which is the SDF representation of the model. For example, a linear factor pricing model identifies a specific linear function of the factors as a SDF. The SDF method involves estimating the asset pricing model using its SDF representation and the generalized method of moments (GMM). As Cochrane (2000) points out, the SDF method is sufficiently general that it can be used for analysis of linear as well as non linear asset-pricing models, including pricing models for derivative securities.

The generality of the framework provides a deeper understanding of the economic issues involved in evaluating asset-pricing models using financial market data. Hence, it is not surprising that the SDF method is becoming the preferred method even for studying linear factor pricing models, replacing the classical *beta method*. The latter method estimates the parameters of a linear factor pricing model using its beta representation and cross-sectional regressions. Under the beta representation of a linear factor pricing model, the expected return on an asset is a linear function of its factor betas.

In spite of its wide use, little is known about the estimation efficiency of the SDF method relative to the classical beta method. A question that arises is whether the generality of the SDF framework comes at the cost of estimation efficiency, especially for risk premiums, in linear factor pricing models. Estimated values of factor risk premiums in linear factor pricing models are frequently used for calculating the cost of capital while making capital budgeting decisions and monitoring regulated monopolies. Applications of such models in capital budgeting decisions can be found in standard textbooks on corporate finance. In order to calculate the cost of capital for the regulation of New York public utility companies, Elton, Gruber and Mei (1994) estimate the risk premiums for macro economic factors, and Schink and Bower (1994) estimate the risk premiums for book-to-market and size factors. Fama and French (1997) focus on the precision with which industry cost of capital can be estimated using various factor models. Hence precise estimation of factor risk premia has

received attention in the cost of capital literature. In view of this in this paper we compare the precision with which risk premiums in linear asset pricing models can be estimated using the two methods.

When returns and factors are jointly normally distributed and independent over time, the classical beta method provides the most efficient unbiased estimator of factor risk premia for linear models. If the SDF method turns out to be inefficient relative to the classical beta method for linear models under these assumptions, some variation of the beta method may well dominate the SDF method for nonlinear models as well in terms of estimation efficiency. This is because a nonlinear model can be locally approximated by a linear model. On the other hand, if the SDF method is as efficient as the beta method, it would become the preferred method because of its generality.

The beta method and the corresponding SDF method are not nested within a more general econometric model. Hence comparing the two methods is not straight forward. We therefore establish the link between the risk premium estimators under the two methods. We then show that asymptotically the SDF method provides as precise an estimate of the risk premium as the beta method. Using Monte Carlo simulations, we show that the two methods provide equally precise estimates in finite samples as well. The sampling errors in the two methods are also similar in the presence of conditional heteroscedasticity. Therefore, linearizing nonlinear asset pricing models and estimating risk premiums using the beta method will not lead to increase in estimation efficiency.

In a special case, Kan and Zhou (1999) compared the estimation efficiency of the risk premium for the SDF method and the beta method. They showed that the SDF method is far less efficient than the beta method. The sampling error in the SDF method was 40 times as large as that in the beta method. For the purpose of establishing the link between the two estimators and comparing their efficiency, they made the simplifying assumption that the economy wide pervasive factor has a zero mean and unit variance. This may not seem unreasonable given the common practice of defining factors as the estimated unanticipated changes in macroeconomic variables and ignoring the associated estimation errors (e.g., Chen, Roll and Ross, 1986, and Breeden, Gibbons and Litzenberger, 1989).

We show that Kan and Zhou's (1999) conclusion about the relative inefficiency of the SDF method critically depends on this simplifying assumption. In particular, their result

holds as long as the factor mean and factor variance are predetermined without estimation. It is not necessary for the factor to have a zero mean and unit variance. This is because predetermining the mean and the variance of the factor leads to more precise estimate of the risk premium when the beta method is used but not when the standard SDF method is used. We show that the additional information about the mean and the variance of the factor is automatically taken into account in the beta method but not in the SDF method. It is therefore important to explicitly incorporate the restrictions on the first two moments of the factor while using the SDF method when the factor moments are predetermined. When this is done, the SDF method becomes asymptotically as efficient as the beta method.

In most situations the mean and the variance of the factors will have to be estimated and not predetermined. In such cases it is important to take into account the corresponding estimation errors involved. Suppose unanticipated changes in macroeconomic factors, measured using auxiliary time-series models, are used as factors in linear factor pricing models, as is commonly done in empirical studies in the literature. Then the classical beta method will substantially overstate the precision of estimated risk premiums for such macroeconomic factors, since the estimation errors associated with the factor mean and factor variance are not automatically taken into account. In contrast, the standard SDF method will give the correct standard errors for the estimated risk premiums. This is because the SDF method automatically takes the sampling errors associated with estimated factor innovations. This has not been realized in the empirical finance literature.

We also examine the specification tests associated with the two methods. An intuitive test for model mis-specification would be to examine whether the model assigns the correct expected return to every asset – i.e., whether the vector of pricing errors for the model is zero. For the SDF method, this test is algebraically equivalent to Hansen’s (1982) J test. This is not the case for the beta method. The covariance matrix of the vector of average pricing errors for the beta method is smaller than that for the SDF method. The Wald test statistic for examining whether the expected value of the vector of average pricing errors equals the zero vector has the same asymptotic central χ^2 distribution for both methods when the model holds. However, when the model is misspecified and the vector of pricing errors does not equal the zero vector the Wald test statistic has an asymptotic noncentral χ^2 distribution. Since the vector of pricing errors is estimated more precisely under the beta method, the noncentrality parameter is larger for the beta method when compared to the

SDF method. Hence asymptotically the beta method has a higher power. However, this advantage of the beta method does not always hold well in finite samples.

We organize the rest of the paper as follows. In section 2, we describe the SDF and the beta methods for estimating linear factor pricing models. In section 3, we develop the analytical results for the comparison of the two methods. In section 4 we compare the finite sample properties of the estimators using Monte Carlo simulations. We conclude in section 5.

2 Description of the two methods

In this section, we review the beta and SDF methods separately. We also take this opportunity to set up the necessary notations. In the first subsection, we start by describing the beta representation and review the estimation methods, which are the cross-sectional regression, the maximum likelihood method and the generalized method of moments (GMM). The cross-sectional regression and the maximum likelihood method can be viewed as special cases of the GMM, which allows conditional heteroscedasticity and serial correlation in asset returns. In the second subsection, we describe the advantages of the SDF method as well as the concerns raised in the recent finance literature. The main advantage is that the SDF method provides a unified treatment of both linear and nonlinear models and sets up the estimation problem for the GMM naturally. This allows us to explore and compare a class of rich and complicated linear and nonlinear models. The concern is that the generality of this methodology might come at a cost in the efficiency of parameter estimation and the power of specification tests.

2.1 The Beta Method

Let r_t be a vector of n asset returns in excess of the risk-free rate. To reduce notational complexity, we assume that there is only one economy-wide pervasive risk factor f_t . Let μ and σ^2 be the mean and variance of the factor f_t . Then the asset pricing model under the beta representation is given by:

$$E[r_t] = \delta\beta , \tag{1}$$

where β , defined as $\text{Cov}[r_t, f_t]/\sigma^2$, is the sensitivity of asset returns to the factor, and δ is the factor risk premium.

When the economy-wide pervasive factor f_t is the return on a portfolio of traded assets, we call it a traded factor. An example of a traded factor would be the return on the value weighted portfolio of all exchange traded stocks typically used in empirical studies of Sharpe's (1963) capital asset pricing model. However, a pervasive risk factor need not be directly traded. Examples of nontraded factors can be found in Chen, Roll and Ross (1986) who use the growth rate of industrial production and the rate of inflation as factors and Gibbons and Litzenberger (1989) who use the growth rate in per capita consumption as the factor.

When the factor is the excess return on a traded asset, it should also satisfy equation (1). Hence in this case $\mu = \delta$, i.e., the risk premium is the mean of the factor. This restriction allows us to use the sample mean of the factor as an estimator of the risk premium. If the factor is not traded, this restriction does not hold, and we have to estimate the risk premium using returns on traded assets. We focus on this latter case where the factor is not traded. However, all our results continue to hold even when the factor is traded.

Note that the vector β can be consistently estimated using the time-series regression: $r_t = \phi + \beta f_t + \epsilon_t$. The residual ϵ_t has zero mean and is uncorrelated with the factor f_t . The asset pricing model (1) imposes a restriction on the intercept, ϕ , i.e., $\phi = (\delta - \mu)\beta$. By substituting this expression for ϕ in the regression equation we obtain the following decomposition of the excess return on stocks:

$$r_t = (\delta - \mu + f_t)\beta + \epsilon_t . \tag{2}$$

For convenience, we refer to equation (2) as the *factor model*.

The beta method uses the beta representation (1) that gives rise to the factor model (2) to estimate the risk premium. The classic two-step approach proposed by Black, Jensen and Scholes (1972) is to estimate the beta of each asset i from the time-series regression

$$r_{it} = \phi_i + \beta_i f_t + \epsilon_{it} , \quad i = 1, \dots, T ,$$

and then estimate the risk premium δ from the cross-sectional regression

$$\bar{r}_i = \gamma + \delta \hat{\beta}_i + e_i ,$$

where $\bar{r}_i = (1/T) \sum_{t=1}^T r_{it}$ and $\hat{\beta}_i$ is the beta of asset i estimated with the time-series regression. The residual e_i in the cross-sectional regression is in general a function of the residual ϵ_{it} in the time-series regression. If the beta representation (1) holds, the intercept γ should be zero. This implication allows economists to test the model by examining statistically the null hypothesis $\gamma = 0$. If the factor is the excess return on a traded asset, the risk premium should equal the mean of the factor, i.e., $\delta = \mu$, which is another null hypothesis for a statistical test of the model. Using this approach, Black, Jensen and Scholes (1972) reject the CAPM because the intercept γ is found to be too high relative to zero and the risk premium δ is found to be too low relative to the average return on the S&P500 portfolio. This led Black (1972) to develop the zero-beta CAPM.

The precision with which the parameters γ and δ are estimated through cross-sectional regression is usually calculated using the procedure suggested by Fama and MacBeth (1973). For each period t , this involves estimating the parameters in the following cross-sectional regression:

$$r_{it} = \gamma_t + \delta_t \hat{\beta}_i + e_{it} .$$

Let the estimates of γ_t and δ_t be denoted by $\hat{\gamma}_t$ and $\hat{\delta}_t$. Fama and MacBeth (1973) suggest treating $\hat{\gamma}_t$ and $\hat{\delta}_t$ as independent samples to obtain estimates of γ and δ :

$$\hat{\gamma} = \frac{1}{T} \sum_{t=1}^T \hat{\gamma}_t , \quad \hat{\delta} = \frac{1}{T} \sum_{t=1}^T \hat{\delta}_t .$$

Then the variance of the above estimators is

$$\text{Var}[\hat{\gamma}] = \frac{1}{T^2} \sum_{t=1}^T (\hat{\gamma}_t - \hat{\gamma})^2 , \quad \text{Var}[\hat{\delta}] = \frac{1}{T^2} \sum_{t=1}^T (\hat{\delta}_t - \hat{\delta})^2 .$$

This two-step approach is often referred to as Fama-MacBeth method and widely used in the finance literature. For example, Fama and French (1992) use this approach to show that there is no relationship between expected return on stocks and their betas and Chen, Roll and Ross (1986) use this approach to study a linear multifactor asset pricing model.

A shortcoming of the cross-sectional regression approach is that it ignores the sampling errors associated with the estimated betas.¹ Shanken (1992) shows that the Fama-MacBeth

¹Black, Jensen and Scholes (1972) suggest minimizing the estimation error associated with betas by applying the cross-sectional regression method to beta-sorted portfolios. Since portfolio betas are estimated substantially more precisely than betas of individual stocks, the errors in variables problem becomes less of an issue.

method over states the precision of the estimated parameters when returns and factors are conditionally homoscedastic and temporally independent. Jagannathan and Wang (1998) argue that this is not always the case when returns and factors exhibit conditional heteroscedasticity. Shanken (1992) and Jagannathan Wang (1998) provide formulas for calculating the precision of the estimated parameters.

When returns and factors exhibit conditional homoscedasticity, we can apply the maximum likelihood procedure to the beta representation and thereby avoid the shortcomings associated with the two-step Fama-MacBeth method. When the observations have identical and independent normal distributions, the beta model implies

$$r_t = (\delta - \mu + f_t)\beta + \epsilon_t, \quad \epsilon_t|f_t \sim N(0_{n \times 1}, \Omega), \quad f_t \sim N(\mu, \sigma^2), \quad (3)$$

where $0_{n \times 1}$ is an $n \times 1$ vector of zeros and Ω is an $n \times n$ positive definite matrix. It is important to notice that the conditional mean and variance of ϵ_t is independent of f_t . The likelihood function is given by:

$$L(\delta, \beta, \Omega, \mu, \sigma) = |\Omega|^{-\frac{T}{2}} \left(\prod_{t=1}^T \exp \left\{ -\frac{1}{2} [r_t - (\delta - \mu + f_t)\beta]' \Omega^{-1} [r_t - (\delta - \mu + f_t)\beta] \right\} \right) \cdot \left(\sigma^{-T} \prod_{t=1}^T \exp \left\{ -\frac{1}{2\sigma^2} (f_t - \mu)^2 \right\} \right).$$

The estimated values of the parameters maximize the likelihood function. When the factor is the excess return on a traded asset, Gibbons, Ross and Shanken (1989) show that this approach is equivalent to estimating the parameters using standard linear multivariate time-series regression. Shanken (1992) shows that the cross-sectional method is equivalent to the maximum likelihood method if the estimation errors in betas are properly taken into account.

While applying the maximum likelihood procedure to the beta representation provides the most efficient estimates for risk premiums, the assumption of conditional homoscedasticity is a major limitation. This is because returns on financial assets often exhibit conditional heteroscedasticity.² Hence it is more appealing to use Hansen's (1982) generalized method of moments (GMM) that allows conditional heteroscedasticity and serial correlation to estimate the parameters in the beta representation. For using the GMM, we make use of the following

²See French, Schwert and Stamabaugh (1987) and Glosten, Jagannathan and Runkle (1992), and Bollerslev, Chou and Kroner (1992) for evidence.

four moment restrictions implied by the factor model (2)³ — the zero mean of the residuals, the zero covariance between the residuals and the factor, and the definition of the mean and the variance of the factor:

$$\mathbb{E}[r_t - (\delta - \mu + f_t)\beta] = \mathbf{0}_{n \times 1} \quad (4)$$

$$\mathbb{E}[(r_t - (\delta - \mu + f_t)\beta)f_t] = \mathbf{0}_{n \times 1} \quad (5)$$

$$\mathbb{E}[f_t - \mu] = 0 \quad (6)$$

$$\mathbb{E}[(f_t - \mu)^2 - \sigma^2] = 0 . \quad (7)$$

If the factor is the return on a tradable asset, we have $\delta = \mu$ and the first two moment restrictions become

$$\mathbb{E}[r_t - f_t\beta] = \mathbf{0}_{n \times 1} \quad (8)$$

$$\mathbb{E}[(r_t - f_t\beta)f_t] = \mathbf{0}_{n \times 1} , \quad (9)$$

which do not depend on the mean and variance of the factor.⁴

The GMM estimation can be proceeded as follows. Define $\theta = (\delta, \beta', \mu, \sigma^2)'$ and $x_t = (r_t', f_t)'$. The sample analog of the left-hand side of the moment restrictions can be constructed as

$$e_T(\theta) = \frac{1}{T} \sum_{t=1}^T g(x_t, \theta) , \quad \text{where} \quad g(x_t, \theta) = \begin{pmatrix} r_t - (\delta - \mu + f_t)\beta \\ (r_t - (\delta - \mu + f_t)\beta)f_t \\ f_t - \mu \\ (f_t - \mu)^2 - \sigma^2 \end{pmatrix} .$$

Then the expected value of $g_t(x_t, \theta)$ is zero. We assume that the vector $x_t = (r_t', f_t)'$ follows a jointly stationary and ergodic process. Since $e_T(\theta)$ is the sample mean of a process whose population mean is zero, this implies that e_T will satisfy the Central Limit Theorem (see Hamilton, 1994). Thus, we have:

$$\lim_{T \rightarrow \infty} \sqrt{T} e_T(\theta) \sim N(\mathbf{0}_{2n+2 \times 1}, S) ,$$

where S is the spectral density matrix of $g(x_t, \theta)$, i.e.,

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}[g(x_t, \theta)g(x_{t+j}, \theta)'] .$$

³When it is not necessary to estimate the variance, σ^2 , of the factor the last moment restrictions can be ignored.

⁴In this case, we can estimate β from the moment restrictions (8) – (9) and separately estimate μ and σ^2 from the moment restrictions (6) and (7). The estimate of μ is also the estimate of the risk premium.

Let W_T be a consistent estimator of S^{-1} . The GMM estimator of θ is given by:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} e_T(\theta)' W_T e_T(\theta) .$$

When the regularity conditions mentioned in Hansen (1982) are satisfied, which is typically the case, the Law of Large Numbers implies that $\operatorname{plim}_{T \rightarrow \infty} E[\partial e_T(\theta)/\partial \theta'] = D$ exists. We assume that the rank of D equals the dimension of θ . It follows from Hansen (1982) that the asymptotic distribution of θ^* is given by:

$$\lim_{T \rightarrow \infty} \sqrt{T}(\theta^* - \theta) \sim N(0_{(n+3) \times 1}, (D' S^{-1} D)^{-1}) .$$

Let $\theta^* = (\delta^*, \beta^{*'}, \mu^*, \sigma^{*2})'$ denote the parameters estimated with the GMM, and $\operatorname{Avar}[\delta^*]$ denote the asymptotic variance of the estimated risk premium δ^* . Then $\operatorname{Avar}[\delta^*]$ is the first element of the matrix $(D' S^{-1} D)^{-1}$.

When returns and factors exhibit conditional homoscedasticity and independence over time, MacKinlay and Richardson (1991) show that the GMM estimator is equivalent to the multivariate-regression-based estimator suggested by Gibbons, Ross and Shanken (1989). For large samples the two are also equivalent to the maximum likelihood estimator. Ferson and Harvey (1997) extend the equivalence argument to models where betas are linear functions of observable variables. The advantage of the GMM estimator is that it is robust to the presence of conditional heteroscedasticity. Hence MacKinlay and Richardson (1991) recommend estimating the parameters using the GMM and the beta representation. We refer to the combination of the GMM and the beta representation as the *beta method*.

Hansen's J -statistic is often used as a specification test to examine whether the data are consistent with the model. When the linear factor pricing model holds, the J -statistic, given by,

$$J^* = T e_T(\theta^*)' W_T e_T(\theta^*) ,$$

converges to a central χ^2 distribution as T becomes large. Another way to examine the validity of the pricing model is to test if Jensen's alpha given by, $\alpha = E[r] - \delta\beta$, is zero. Jensen's alpha measures the deviation of the vector of excess returns from what the corresponding object should be according the pricing model. Although Jensen suggested the use of alpha for evaluating the performance of the standard capital asset pricing model, it can be used to evaluate any linear asset pricing model. The sample estimate of Jensen's alpha is given by:

$$\alpha^* = \bar{r} - \delta^* \beta^* , \quad \text{where} \quad \bar{r} = \frac{1}{T} \sum_{t=1}^T r_t . \quad (10)$$

When the pricing model holds the estimated Jensen's alpha should not be different from the zero vector after allowing for sampling errors.

2.2 The SDF Method

By substituting the expression for β into equation (1) and rearranging the terms, we obtain the SDF representation of the linear asset pricing model given below:

$$E[r_t m_t] = 0_{n \times 1}, \quad \text{where } m_t = 1 - \lambda f_t, \quad (11)$$

In the above expression the parameter λ is the following linear transformation of the risk premium δ :

$$\lambda = \frac{\delta}{\sigma^2 + \mu\delta}, \quad (12)$$

or, equivalently,

$$\delta = \frac{\lambda\sigma^2}{1 - \mu\lambda}. \quad (13)$$

According to equation (11), the expected value of the excess return on an asset discounted by m_t equals zero. Hence any random variable, m_t , that satisfies equation (11) is referred to as a *stochastic discount factor* (SDF).

When the law of one price holds, there always exists a random variable m_t that satisfies equation (11). In general, a number of random variables satisfying equation (11) exist and hence there will be more than one stochastic discount factors. An asset pricing model takes the stand that a particular random variable is a stochastic discount factor. It follows that the linear factor pricing model (11) identifies the random variable, $m_t = 1 - \lambda f_t$, as a stochastic discount factor.

In fact, equation (11) can be viewed as the *Euler equation* (the first order condition) for the portfolio choice problem faced by an investor whose inter-temporal marginal rate of substitution is a linear function of the factor. Hansen and Richard (1987) observed that an asset-pricing model could be written in the form

$$E_{t-1}[m_t y_t] = p_{t-1},$$

where y_t is the $n \times 1$ vector of payoffs from n securities at the end of period t , p_{t-1} is the $n \times 1$ vector of prices for the securities at the end of period $t - 1$ and $E_{t-1}[\cdot]$ is the expectation

conditional on the information at the end of period $t - 1$. When the utility function in the economic model depends on a vector of k parameters, denoted by θ , the stochastic discount factor m_t will in general also be a function of θ and a vector of economic variables, denoted by f_t , i.e., the stochastic discount factor can be written as $m_t = m(\theta, f_t)$. For example, in the representative-agent and endowment economy of Lucas (1978), $m(\theta, f_t) = \rho c_t^{-a}$, where $f_t = c_t$, the growth rate of aggregate consumption, and $\theta = (\rho, a)'$. The single factor linear pricing model implies $m(\theta, f_t) = 1 - \lambda f_t$, where $\theta = \lambda$. Hence the SDF representation of an asset pricing model can be written as follows,

$$E_{t-1}[m(\theta, f_t)y_t] = p_{t-1} . \quad (14)$$

When y_t is a vector of gross returns, p_{t-1} is a vector of 1s. When y_t is a vector of excess returns, p_{t-1} is a vector of zeros. Most derivative pricing models can also be written in the form given in (14). Ingersoll (1987) derives the stochastic discount factor representation for a number of theoretical asset pricing models. However, he does not use the term ‘stochastic discount factor’ — Hansen and Richard (1987) coined the term. Hansen and Jagannathan (1991, 1997) develop diagnostic tests for asset pricing models based on the SDF representation. Ferson (1995), Campbell, Lo and MacKinlay (1997) and Cochrane (1999) provide comprehensive introductions to the stochastic discount factor framework.

As Hansen and Singleton (1982) demonstrate, the SDF representation makes it easier to take conditioning information into account while econometrically evaluating an asset pricing model. Let z_{t-1} be an $l \times 1$ vector of random variables that are realized at the end of period $t - 1$. Multiplying z_{t-1} to both sides of equation (14), we obtain

$$E_{t-1}[m(\theta, f_t)(y_t \otimes z_{t-1})] = p_{t-1} \otimes z_{t-1} ,$$

where \otimes represents the Kronecker product in matrix algebra. Let $x_t = (y_t', p_{t-1}', z_{t-1}')'$. The above equation implies

$$E[g(x_t, \theta)] = 0_{nl \times 1} , \quad \text{where } g(x_t, \theta) = (m(\theta, f_t)y_t - p_{t-1}) \otimes z_{t-1} .$$

This gives the moment conditions for estimating the model parameters using the GMM. Notice that $y_t \otimes z_{t-1}$ can be viewed as the vector of payoffs on a set of managed portfolios and $p_{t-1} \otimes z_{t-1}$ the corresponding prices.

In order to apply the GMM, we assume that all the assumptions in Hansen (1982) hold for g and x_t . Let

$$e_T(\theta) = T^{-1} \sum_{t=1}^T g(x_t, \theta) \quad \text{and} \quad S = \sum_{t=-\infty}^{\infty} \text{E}[g(x_t, \theta)g(x_{t+j}, \theta)'] .$$

If W_T is a consistent estimator of S^{-1} , the GMM estimator of θ is

$$\hat{\theta} = \underset{\theta}{\text{argmin}} e_T(\theta)' W_T e_T(\theta) .$$

Then, the asymptotic variance of the estimator is

$$\text{Avar}[\hat{\theta}] = (D' S^{-1} D)^{-1} ,$$

where $D = \text{plim}_{T \rightarrow \infty} \text{E}[\partial e_T(\theta) / \partial \theta']$. For specification tests, economists usually examine the J -statistics, which is defined as

$$\hat{J} = T e_T(\hat{\theta})' W_T e_T(\hat{\theta}) .$$

Under the null hypothesis that the model holds, the J -statistics converges to a χ^2 distribution as T becomes infinitely large.

When the model does not hold, we are interested in studying the pricing error given by,

$$\pi = \text{E}[m(\theta, f_t) y_t - p_{t-1}] .$$

Hansen and Jagannathan (1997) develop diagnostic tests based on the pricing error. The sample analog of the pricing error is given by:

$$\hat{\pi} = \frac{1}{T} \left(\sum_{t=1}^T m(\hat{\theta}, f_t) y_t - p_{t-1} \right) .$$

Classical tests of asset pricing models are equivalent to examining whether the vector of pricing errors are no different from zero after allowing for sampling errors.

For the linear one-factor pricing model of excess returns, the moment restriction corresponding to $z_{t-1} = 1$ is given by:

$$\text{E}[r_t(1 - \lambda f_t)] = 0_{n \times 1} . \tag{15}$$

Using moment restriction (15) and the GMM we obtain an estimate of the parameter $\hat{\lambda}$. The sample analog of the vector of pricing errors, $\pi = \text{E}[r_t] - \lambda \text{E}[r_t f_t]$, is given by:

$$\hat{\pi} = \bar{r} - \hat{\lambda}(\bar{r}f) , \quad \text{where} \quad \bar{r} = \frac{1}{T} \left(\sum_{t=1}^T r_t \right) \quad \text{and} \quad \bar{r}f = \frac{1}{T} \left(\sum_{t=1}^T r_t f_t \right) . \tag{16}$$

Notice that the J -statistic we discussed earlier is a quadratic function of the estimated pricing error. Also, Jensen's alpha is a scaled version of the pricing error. Substituting equation (12) into $\pi = E[r_t] - \lambda E[r_t f_t]$, we obtain

$$\pi = \left(\frac{\sigma^2}{\sigma^2 + \mu\delta} \right) \alpha, \quad (17)$$

i.e.,

$$\alpha = \left(\frac{\sigma^2 + \mu\delta}{\sigma^2} \right) \pi. \quad (18)$$

Hence Jensen's alpha will be zero if and only if the pricing error is zero. In the special situation where the mean of the factor is zero, i.e., $\mu = 0$, Jensen's alpha equals the pricing error.

This discussion reveals that the SDF method, which is the combination of the SDF representation and the GMM, provides a convenient and general framework for analyzing linear and nonlinear asset pricing model. As pointed out in the introduction, a natural question that arises is whether the generality of the method comes at the cost of estimation efficiency for specific classes of models when compared to the beta method. This is a difficult question to answer since two methods are not nested within a more general econometric specification.

Recently, Kan and Zhou (1999) provides a comparison of the two methods under certain restrictive assumptions. They find that the estimation error for the risk premium is even asymptotically larger for the SDF method when compared to the beta method. It is about 40 times as large in finite samples. They also find the specification test under the SDF method has a lower power. Based on this they conclude that the classical beta method is preferable to the SDF method, at least for linear factor pricing models.

Kan and Zhou's findings, if correct, have important implications for empirical evaluation of asset pricing models. Notice that a nonlinear model can often be well approximated by a linear one. For example, Cochrane (1996) advocates that non-linear stochastic factors be approximated by linear functions of macroeconomic variables. Campbell (1993, 1996) uses linear approximation for the consumption-based CAPM. Linearizing a non-linear model and applying the beta method may provide more precise risk premium estimates and a more powerful test than the SDF method, if the approximation error is sufficiently small.

In the next section we show that Kan and Zhou’s (1999) comparison of the beta and SDF methods is valid only when estimated innovations to macro economic variables are used as pervasive risk factors, ignoring the estimation errors involved. We provide the link between the two risk premium estimators even though the two methods are not nested within a more general econometric specification. In addition, we also examine the ability of the two methods to detect model mis-specification. For that purpose we establish the relation between the vector of pricing errors ($\hat{\pi}$) estimated using the SDF method and the vector of Jensen’s alpha (α^*) estimated using the beta method. Estimated pricing errors and Jensen’s alpha have been extensively used to examine mis-specification in asset pricing models. For example, Fama and French (1993) study the performance of their linear three-factor model using Jensen’s alpha. Hodrick and Zhang (2000) compare the performance of several linear models using their pricing errors.

3 Analytical Results

In this section, we compare the risk premium estimates obtained using the SDF and the beta methods. First, we show that the two methods have the same estimation efficiency for estimating risk premia. We also show that the estimated pricing error obtained using the SDF method is potentially larger than that obtained using the beta method. Next, we consider the special case where a researcher constructs macroeconomic factors using time series models and uses them to evaluate the factor pricing model — but ignores the sampling errors associated with the factors so constructed, as is common in this literature. In that case, the beta method overstates the precision with which risk premia are estimated, whereas the SDF method provides the correct standard errors. This result points to a serious shortcoming of the beta method.

When the mean and the variance of the factor is known the beta method provides more precise estimates of risk premia. This is because this additional information about the mean and the variance of the factor is automatically incorporated in the estimate when the beta method is used. This is not the case when the SDF method is used — this information must be incorporated through additional moment conditions. When that is done, the SDF method is as efficient as the beta method.

3.1 Comparison of the Two Methods

First consider the estimates of the risk premium obtained using the two methods. The beta method gives the GMM estimate δ^* from the moment restrictions of the beta model, while the SDF method gives the GMM estimate $\hat{\lambda}$ from the moment restriction of the SDF model. We cannot compare the precision of the two estimates directly because δ and λ will not in general be equal. We therefore first transform the estimate of δ obtained with the beta method into an estimate of λ as given below:

$$\lambda^* = \frac{\delta^*}{\sigma^{*2} + \mu^* \delta^*} , \quad (19)$$

We then compare the asymptotic variances of λ^* and $\hat{\lambda}$.

Notice that the transformation requires the estimates of μ and σ^2 . The estimation errors in μ^* and σ^{*2} affect the asymptotic variance of the estimated λ^* . In order to calculate the asymptotic variance of λ^* , we need to make use of the joint distribution of δ^* , μ^* and σ^{*2} . Equivalently, we can substitute equation (13) into moment restrictions (4) and (5) to express them in terms of λ and then estimate λ , β , μ and σ^2 jointly. The standard errors for the estimate of λ will then automatically take into account the estimation errors in μ^* and σ^{*2} .

Next consider the estimated pricing error $\hat{\pi}$ obtained using the SDF method with the estimated Jensen's alpha α^* obtained using the beta method. In general π and α will not be equal. The pricing error π and Jensen's α are related to each other as by equation (17). The estimated pricing error $\hat{\pi}$ is calculated using equation (16), while the estimated Jensen's alpha α^* is calculated using equation (10). In view of this, using (17), we transform the estimated Jensen's α^* to get an estimate of the pricing error π^* as follows:

$$\pi^* = \left(\frac{\sigma^{*2}}{\sigma^{*2} + \mu^* \delta^*} \right) \alpha^* . \quad (20)$$

The estimator π^* can be viewed as the estimated pricing error obtained with the beta method. We can then compare the asymptotic variances of π^* and $\hat{\pi}$.

The analytical results of our comparison of the two methods are summarized in the next theorem. The proof of the theorem is presented in the Appendix.

Theorem 1 *Assume that the beta representation (1) and the equivalent SDF representation (11) hold. Then Hansen's J-statistic for the beta method as well as the SDF method have an*

asymptotic χ^2 distribution with $n - 1$ degrees of freedom, i.e., $\hat{J} \xrightarrow{d} \chi_{n-1}^2$ and $J^* \xrightarrow{d} \chi_{n-1}^2$ as $T \rightarrow \infty$. If the vector, $(r'_t, f_t)'$, is drawn from an i.i.d. joint Normal distribution, then the risk premium estimated using the SDF method has the same asymptotic variance as the risk premium estimated using the beta method, i.e., $\text{Avar}[\hat{\lambda}] = \text{Avar}[\lambda^*]$. However, the asymptotic variance of the pricing error estimated using the SDF method is at least as large as that estimated using the beta method, i.e., $\text{Avar}[\hat{\pi}] - \text{Avar}[\pi^*]$ is positive semi-definite.

In order to compare the asymptotic variances of the parameters, in Theorem 1, we assume that the asset returns and the factors are generated by an i.i.d. joint Normal process. This implies that the asset returns and the factor are conditionally homoscedastic and serially independent. We need this assumption to analytically derive the results in Theorem 1. The results in Theorem 1 can be readily extended to the case where the factor and the asset returns are serially correlated — but notational complexity increases substantially. However, we are unable to provide analytical results in the presence of conditional homoscedasticity. We will have to rely on Monte Carlo simulations to examine the case with conditional heteroscedasticity. Finally, under the i.i.d. joint Normal assumption, the beta-method estimator for the risk premium is equivalent to the maximum-likelihood estimator. It is well known that the latter is asymptotically the most efficient consistent estimator. Therefore, when $\text{Avar}[\hat{\lambda}] = \text{Avar}[\lambda^*]$, the SDF method estimator of the risk premium is also asymptotically the most efficient consistent estimator. This has not shown elsewhere in the literature so far.

In Theorem 1, the factor can be either a return on a tradable asset or a non-tradable macroeconomic variable. As we have discussed in the earlier section, if the factor is a tradable asset return, the asset pricing model implies the restriction $\delta = \mu$. It can be shown that Theorem 1 continues to hold even when this restriction is imposed during estimation. In fact, under this restriction the asymptotic variances of $\text{Avar}(\hat{\delta})$ and $\text{Avar}(\delta^*)$ are both equal to the asymptotic variance of the sample average of f_t . Therefore, if imposing the restriction $\delta = \mu$, we only need to estimate the risk premium from the observations of the factor. Incorporation of asset returns does not improve estimation efficiency.

Theorem 1 indicates that the J -statistics in the two methods have the same asymptotic distribution. This holds even when asset returns exhibit conditional heteroscedasticity and serial correlation. The degrees of freedom in the SDF method is $n - 1$ because there are n restrictions and 1 parameter in equation (15). The degrees of freedom in the beta method

is also $n - 1$ because there are $2n + 2$ restrictions in equations (4), (5), (6) and (7) and $n + 3$ parameters. Therefore, the asymptotic variances of \hat{J} and J^* must be the same. The distributions of \hat{J} and J^* can only be different in finite samples. Later, we will use Monte Carlo simulation to address the issue of finite-sample distributions.

When the model is misspecified, the J -statistic will have a noncentral χ^2 distribution. The noncentrality parameter will depend on the variance of the estimated pricing errors. The last part of Theorem 1 demonstrates that the beta method estimates the pricing error with possibly higher precision than the SDF method. Therefore, the noncentrality parameter of the J -statistic for the beta method will be at least as large as that for the SDF method when assets are mispriced by a constant amount through time. The conventional wisdom is that evaluating model mispricing using Jensen's alpha obtained using the beta method is equivalent to evaluating the model using the SDF method. Theorem 1 points out that this conventional wisdom is wrong.

Now consider the parameter δ and Jensen's alpha α . In the beta method, the estimates δ^* and α^* are obtained as described in Section 2.1. For proper comparison we transform the estimates $\hat{\lambda}$ and $\hat{\pi}$ obtained using the SDF method as described in Section 2.2 to get estimates $\hat{\delta}$ and $\hat{\alpha}$ using formula (13) and (18). It can be readily shown, by mimicking the derivations given in the proof of Theorem 1, that $\text{Avar}[\hat{\delta}] = \text{Avar}[\delta^*]$ and that $\text{Avar}[\hat{\alpha}] - \text{Avar}[\alpha^*]$ is positive semi-definite. The detailed derivation can be found in Jagannathan and Wang (2000).

3.2 Effect of Ignoring the Sampling Errors Associated with Estimated Factors

According to Theorem 1 the beta method and the SDF method provide risk premium estimates that are equally precise. This contradicts Kan and Zhou's (1999) conclusion. In this section, we examine why this is so. When comparing the two methods, Kan and Zhou (1999) assume that the pervasive risk factor has zero mean and unit variance, i.e., $\mu = 0$ and $\sigma^2 = 1$. Under this assumption, it follows from equation (12) that $\delta = \lambda$. This allows Kan and Zhou to compare directly the errors in the estimates of δ and λ . Since this assumption is common for both the beta method as well as the SDF method one may expect that it would not give an advantage to one method over the other. In this subsection we show that this is not the

case. Predetermining the mean and the variance of the factor increases the efficiency of the estimator under the beta method but not under the SDF method. However, by adding additional moment conditions to incorporate the information brought in through predetermining the moments of the factor, the SDF method estimator can be made as efficient as the beta method estimator.

Since the mean and the variance of the factor are predetermined Kan and Zhou (1999) drop equations (6) and (7) used in the GMM. For the beta method, Kan and Zhou obtain the estimator of the parameter λ from the moment restriction

$$\begin{aligned} \mathbb{E}[r_t - (\lambda + f_t)\beta] &= \mathbf{0}_{n \times 1} \\ \mathbb{E}[(r_t - (\lambda + f_t)\beta)f_t] &= \mathbf{0}_{n \times 1}. \end{aligned}$$

For the SDF method, they obtain the estimator of the parameter λ from the moment restriction

$$\mathbb{E}[r_t(1 - \lambda f_t)] = \mathbf{0}_{n \times 1}.$$

The most important aspect of the assumption made by Kan and Zhou (1999) is that the mean and variance of the factor are predetermined without estimation. This is equivalent to ignoring the sampling errors associated with the estimates of μ and σ^2 . This is a common practice in the empirical finance literature — examples include Breeden, Gibbons and Litzenberger (1989) and Chen, Roll and Ross (1986). In order to fully understand how this assumption affects the precision of the risk premium estimators, we set μ and σ^2 to be some arbitrary constants, μ_o and σ_o^2 . Since our focus is on the effect this has on the precision with which the risk premium parameter is estimated but not in the bias, without loss of generality we can choose these two constants to equal the values of μ and σ^2 in the population. Notice that it is not necessary for μ and σ^2 to equal 0 and 1 respectively as in Kan and Zhou (1999).

In general, predetermining a subset of the parameters reduces the sampling error of the remaining estimates. The following lemma summarizes this effect.

Lemma 1 *Let x_t be the observed data in period t and $g(x_t, \theta_1, \theta_2)$ be a vector function of $(x_t, \theta_1, \theta_2)$, where θ_1 is the vector of parameters of our interests and θ_2 the parameters that are not of our interests. Suppose that $\mathbb{E}[g(x_t, \theta_1, \theta_2)] = 0$. Let $(\hat{\theta}'_1, \hat{\theta}'_2)'$ be the GMM estimator of $(\theta'_1, \theta'_2)'$. When θ_2 is pre-determined, let $\tilde{\theta}_1$ be the GMM estimator of θ_1 . Then,*

$\text{Avar}[\hat{\theta}_1] - \text{Avar}[\tilde{\theta}_1]$ is positive semi-definite. If $\lim_{T \rightarrow \infty} \text{E}[T^{-1} \sum_{t=1}^T \partial g(x_t, \theta_1, \theta_2) / \partial \theta_2'] = 0$, then $\text{Avar}[\hat{\theta}_1] = \text{Avar}[\tilde{\theta}_1]$.

When the factor moments are predetermined, the asymptotic variance of the estimated risk premium becomes smaller. In this case, the beta method uses the following moment conditions to estimate the risk premium, δ , using the GMM.

$$\begin{aligned} \text{E}[r_t - (\delta - \mu_o + f_t)\beta] &= 0_{n \times 1} \\ \text{E}[(r_t - (\delta - \mu_o + f_t)\beta)f_t] &= 0_{n \times 1} \\ \text{E}[f_t - \mu_o] &= 0 \\ \text{E}[(f_t - \mu_o)^2 - \sigma_o^2] &= 0 \end{aligned}$$

According to Lemma 1, the variance of the risk premium estimator, denoted by δ_o^* , when factor moments are predetermined should not be larger than the variance of the estimator when factor moments are estimated jointly with the risk premium, i.e.,

$$\text{Avar}[\delta^*] \geq \text{Avar}[\delta_o^*]. \quad (21)$$

In their analysis, following common practice, Kan and Zhou (1999) do not use the two moment conditions that restrict the sample first and second moments of the factor to equal their predetermined counter parts in expectation. It would therefore appear that they do not make use of the information about the first two moments of the factor in their estimation.

In that case, for the beta method, the risk premium δ , J -statistics and Jensen's α are calculated from the first two moment conditions reproduced below for convenience:⁵:

$$\text{E}[r_t - (\delta - \mu_o + f_t)\beta] = 0_{n \times 1} \quad (22)$$

$$\text{E}[(r_t - (\delta - \mu_o + f_t)\beta)f_t] = 0_{n \times 1}. \quad (23)$$

We denote the estimates by δ^\dagger , J^\dagger and α^\dagger respectively. It follows from equation (12) and (17) that the estimators of the parameter λ and pricing error π should be, respectively,

$$\lambda^\dagger = \frac{\delta^\dagger}{\sigma_o^2 + \mu_o \delta^\dagger} \quad (24)$$

$$\pi^\dagger = \left(\frac{\sigma_o^2}{\sigma_o^2 + \mu_o \delta^\dagger} \right) \alpha^\dagger. \quad (25)$$

⁵In fact, Kan and Zhou (1999) did not examine Jensen's α and the pricing error π .

In general, dropping the moment restrictions that do not depend on the parameters in the GMM will increase the sampling error of the estimated parameters. However, when certain conditions are met, moment conditions can be dropped without affecting the sampling error of the estimated parameters. It turns out that these conditions are satisfied for the beta method. The following Lemma makes these statements precise.

Lemma 2 *Let x_t denote the vector of variables that are observed at the end of date t , $g_1(x_t, \theta)$ be a vector function of (x_t, θ) , where θ is the vector of parameters, and $g_2(x_t)$ be a vector function of x_t . Let $g(x_t, \theta)' = (g_1(x_t, \theta)', g_2(x_t)')'$. Suppose $E[g(x_t, \theta)] = 0$. Let $\hat{\theta}$ be the GMM estimator from the moment restriction $E[g(x_t, \theta)] = 0$. Let $\tilde{\theta}$ be the GMM estimator of θ from the moment restriction $E[g_1(x_t, \theta)] = 0$. Then, $\text{Avar}[\hat{\theta}] - \text{Avar}[\tilde{\theta}]$ is negative semi-definite. If however*

$$\sum_{j=-\infty}^{\infty} E[g_1(x_t, \theta)g_2(x_{t+j})'] = 0, \quad (26)$$

then $\text{Avar}[\hat{\theta}] = \text{Avar}[\tilde{\theta}]$.

Suppose the data, x_t , are serially independent. Then equation (26) holds if $g_1(x_t, \theta)$ and $g_2(x_t)$ are uncorrelated. In this case, dropping the moment restriction $E[g_2(x_t)] = 0$ does not affect the asymptotic variance of the estimated parameters.

In order to apply Lemma 2 to the set up in Kan and Zhou (1999) let

$$g_1(x_t, \theta) = \begin{pmatrix} r_t - (\delta - \mu_o + f_t)\beta \\ (r_t - (\delta - \mu_o + f_t)\beta)f_t \end{pmatrix} \quad \text{and} \quad g_2(x_t) = \begin{pmatrix} f_t - \mu_o \\ (f_t - \mu_o)^2 - \sigma^2 \end{pmatrix}.$$

Kan and Zhou (1999) assume that $(r_t', f_t)'$ follows an i.i.d. joint Normal process. Therefore it follows from equation (34) in the Appendix g_1 that g_2 are uncorrelated. Hence, the condition given in the last part of Lemma 2 is satisfied, and dropping the factor moment restrictions does not affect the variance of the risk premium estimator in the beta method. This gives,

$$\text{Avar}[\delta^\dagger] = \text{Avar}[\delta_o^{*}].$$

It then follows from inequality (21) that

$$\text{Avar}[\delta^\dagger] \leq \text{Avar}[\delta^*]. \quad (27)$$

In the SDF method, however, predetermining μ and σ^2 does not affect any of the estimators of λ , π or the J -statistic. When μ and σ^2 are predetermined, the parameter λ ,

the J -statistic and the pricing error π in the SDF method are calculated using the moment restriction:

$$\mathbf{E}[r_t(1 - \lambda f_t)] = \mathbf{0}_{n \times 1}, \quad (15')$$

which is the same as equation (15) because μ and σ^2 are not involved in the moment restriction. We therefore still denote the estimates by $\hat{\lambda}$, \hat{J} and $\hat{\pi}$ respectively when the factor moments are predetermined. It follows from Theorem 1 and inequality (27) that

$$\text{Avar}[\lambda^\dagger] \leq \text{Avar}[\hat{\lambda}]. \quad (28)$$

To summarize, the estimated risk premiums in the SDF and beta methods have the same sampling error when the factor moments are not predetermined. Predetermining the factor moments reduces the sampling error of the estimate in the beta method, even though the moment conditions corresponding to the information brought in through predetermining are ignored. In the SDF method, however, the variance of the estimator is not affected when factor moments are predetermined. This explains why Kan and Zhou (1999) find that the SDF method is less efficient.

In fact, we can show that the strict inequality in (28) holds. We can also compare the sampling distribution of the J -statistics and the pricing error. The following theorem summarizes the results. The proof of the theorem can be found in the Appendix.

Theorem 2 *Assume that the beta representation (1) and the equivalent SDF representation (11) hold. Also, assume that the factor moments μ and σ^2 are predetermined. The asymptotic distributions of the J -statistics in the two methods are the χ^2 distribution with $n - 1$ degrees of freedom, i.e., $\hat{J} \xrightarrow{d} \chi_{n-1}^2$ and $J^\dagger \xrightarrow{d} \chi_{n-1}^2$ as $T \rightarrow \infty$. If $(r'_t, f_t)'$ has a joint normal distribution, identical and independent across time t , then the estimated risk premium obtained with the SDF method has larger asymptotic variances than that obtained with the beta method, i.e., $\text{Avar}[\hat{\lambda}] > \text{Avar}[\lambda^\dagger]$, and the asymptotic variance of the estimated pricing error in the SDF method is at least as large as that in the beta method, i.e., $\text{Avar}[\hat{\pi}] - \text{Avar}[\pi^\dagger]$ is positive semi-definite.*

We showed in Theorem (1) that SDF method is as efficient as in the beta method when the factor moments are not known and have to be estimated. Hence, from Theorem (2) we conclude that Kan and Zhou reached a different conclusion because they ignored the

estimation error associated with the first two moments of the factor and treat them as though they are known with certainty.

If the factor is the stock market index, fairly precise information about the variance of the factor can be obtained using prices of stock index options. Variances can also be precisely estimated using high frequency data. Hence it may not be entirely unreasonable to predetermine the variance of the factor and assume that it is known with certainty. However, it would be hard to justify predetermining the mean of the factor. Suppose we predetermine only the mean or the variance of the factor but not both at the same time. In either case, Theorem 2 continue to hold. However, our calculation (not reported) shows that most of the gain in efficiency for the beta method is associated with the predetermination of the mean. Predetermining the variance of the factor results in only a small gain.

Theorem 2 has the following important implication. Unanticipated changes in macroeconomic variables are often used as factors in empirical applications involving factor-pricing models. Since unanticipated changes are unobserved, time series models are frequently used to construct such unanticipated changes. This way of constructing factors is equivalent to predetermining their means, i.e., assuming that factor means are known with certainty. Consequently the beta method will substantially overstate the precision with which factor risk premia are estimated. This is not the case with the standard SDF method.

In what follows, we demonstrate that the SDF method can be made as efficient as the beta method even when the factor moments are predetermined by properly incorporating the factor moment information. From Lemma 2 we know that the restriction $E[g_2(x_t)] = 0$ might affect the estimation efficiency if $g_1(x_t, \theta)$ and $g_2(x_t)$ are correlated. For this reason, we add the factor moment restrictions to the moment restriction (15) in the SDF method. Thus, we obtain the GMM estimate of the parameter λ from the following moment restrictions:

$$E[r_t(1 - \lambda f_t)] = 0 \tag{29}$$

$$E[f_t - \mu_o] = 0 \tag{30}$$

$$E[(f_t - \mu_o)^2 - \sigma_o^2] = 0. \tag{31}$$

We denote the estimate of λ , the J -statistics and the pricing error obtained from these moment restrictions as $\bar{\lambda}$, \bar{J} and $\bar{\pi}$ respectively. Let us refer this approach as *the modified SDF method* in order to distinguish it from the SDF method without factor-moment restrictions.

The sampling properties of the estimates in the modified SDF method, in comparison with the estimates in the beta method, are summarized in the next theorem. The proof of the theorem is delegated to the Appendix.

Theorem 3 *Assume that the beta representation (1) and the equivalent SDF representation (11) hold. Also, assume that the factor moments μ and σ^2 are predetermined. The asymptotic distribution of the J-statistic in the modified SDF method is the χ^2 distribution with $n + 1$ degrees of freedom, i.e., $\bar{J} \xrightarrow{d} \chi_{n+1}^2$ as $T \rightarrow \infty$. If $(r'_t, f_t)'$ has a joint Normal distribution, identical and independent across time t , then the estimated risk premia obtained with the modified SDF and the beta methods have the same asymptotic variance, i.e., $\text{Avar}[\bar{\lambda}] = \text{Avar}[\lambda^\dagger]$, but the asymptotic variance of the estimated pricing error obtained using the modified SDF method is at least as large as that obtained using the beta method, i.e., $\text{Avar}[\bar{\pi}] - \text{Avar}[\pi^\dagger]$ is positive semi-definite.*

Therefore, a simple remedy to the inefficiency of SDF method when factor moments are predetermined is to incorporate the factor-moment restrictions. Nevertheless, the pricing error obtained using the modified SDF method is potentially larger than that obtained with the beta method.

When the factor moments are not predetermined, we do not need such remedy to the SDF method. As shown in Theorem 1, the SDF method is already as efficient as the beta method for estimating risk premiums. A natural question that arises is whether we can further increase the efficiency by incorporating the factor moment restrictions, i.e., whether the asymptotic variances of the GMM estimates of λ , μ and σ^2 obtained from the moment restrictions,

$$\begin{aligned} \text{E}[r_t(1 - \lambda f_t)] &= 0_{n \times 1} & (15') \\ \text{E}[f_t - \mu] &= 0 \\ \text{E}[(f_t - \mu)^2 - \sigma^2] &= 0. \end{aligned}$$

are smaller than the asymptotic variance of the corresponding estimators that are obtained from the first moment restriction (15') alone. The answer is no because $\text{Avar}[\hat{\lambda}] = \text{Avar}[\lambda^*]$ and the beta method is equivalent to the maximum likelihood method, which is the most efficient. Following the proof of Theorem 1, the readers can also directly show that in-

incorporating the factor moment restrictions does not reduce the asymptotic variance of the estimated λ . The proof can also be found in Jagannathan and Wang (2000).

4 Monte Carlo Simulation

The formulae for the asymptotic variance of the parameters estimated using the beta and SDF methods given in the theorems are valid only when the length of the time series of observations is sufficiently large. Theory does not tell us what the length of the time series of observations should be for it to be considered sufficiently large. Hence, in this section, we examine this issue using Monte Carlo simulation methods. In the first subsection, we examine the finite sample properties of the factor risk premiums estimated using the beta and SDF methods. In the second subsection, we examine the relative size and power of specification tests using the two methods.

4.1 Estimation Efficiency

In the first set of simulations, we assume that the returns and the factor are drawn from a multivariate Normal distribution, as assumed in our theorems. To be consistent with the beta model and its equivalent SDF model, we generate excess returns from the following process.

$$r_t = (\delta - \mu + f_t)\beta + \epsilon_t, \quad \epsilon_t|f_t \sim N(0_{n \times 1}, \Omega), \quad f_t \sim N(\mu, \sigma^2).$$

We choose the values for the parameters to match the sample moments of the data.

Monthly returns from January 1926 to December 1998 for the ten size-decile portfolios, the value-weighted market index portfolio of stocks traded on the NYSE, AMEX and NASDAQ and the one-month Treasury Bills are from the Center for Research on Security Prices (CRSP). Excess returns are obtained by subtracting the Treasury Bill returns. We set μ and σ^2 to equal the sample mean and the sample variance of the excess return on the market index portfolio. We set β for each size-decile portfolio to equal the slope coefficient in the OLS regressions of the size-decile excess returns on the excess market return. The covariance matrix, $\Omega = E[\epsilon_t \epsilon_t' | f_t]$, is set equal to the sample covariance matrix of the residuals obtained in the ten OLS regressions. We set the risk premium δ to be the value of the slope coefficient

obtained from the cross-sectional regression of the historical average excess returns on betas. The parameter λ satisfies $\lambda = \delta/(\sigma^2 + \mu\delta)$. Table 1 provides all the parameters used in the simulation.

In the population, we set risk premium $\delta = 1.3640$ for monthly returns. This implies $\lambda = 4.3790$. This large risk premium is in fact the negative of the firm-size premium since β is strongly negatively correlated with firm size in the sample we use for calibration purposes. If we were to construct portfolios by ranking firms on both market capitalization and estimated beta, the betas will not be as highly negatively correlated with firm size. In that case, as in Fama and French (1992) and Jagannathan and Wang (1996), the risk premium on beta would be close to zero, and the risk premium on firm size would be large in magnitude and negative. Also, notice that we do not set the values of δ and μ to be equal. Therefore, we do not impose the restriction that the factor is the return on a portfolio of tradable assets.⁶ The asymptotic variances of the estimated λ implied by the parameters are reported in Table 2.

We consider the following six different time horizons in our simulations: 60 months; 120 months; 360 months; 600 months. The first two horizons are often used for estimating betas and risk premia using rolling windows. The third horizon is similar to that used in many related empirical studies⁷. The fourth is a half-century, which is often the length of period considered in studies of the equity premium puzzle. For each time series length, $T = 60, 120, 360, \text{ or } 600$, we draw 1000 independent samples of the time series and estimated the parameters 1000 times. To see the effects of time horizons on the estimators, Table 2 gives the asymptotic standard deviation of the estimated risk premium for different sample sizes.

We draw repeated samples of the excess returns and the factor from the multivariate Normal distribution with the mean vector and covariance matrix chosen as above. The simulation results for λ is reported in the top part of Table 3. For each estimator of λ , the table gives the standard deviation of the 1000 estimated risk premiums. When the mean and variance of the factor are estimated along with the risk premium parameter, the estimator $\hat{\lambda}$ in the SDF method and the estimator λ^* in the beta method have the same precision for any of the sample size T considered. This is true even when the sample size is as small as 60 months. Therefore, there is no efficiency gain from the use of the beta method over the

⁶Although our focus is on the general macroeconomic factors, our simulation results also hold well when we set $\delta = \mu$.

⁷For example, Fama and French (1992) and Jagannathan and Wang (1996) use 330 monthly observations. Fama and French (1993) use 342 monthly observations.

SDF method. Our simulation also show similar results for the estimated δ — we do not report them in order to save space. Those who are interested in the numerical results of δ are referred to Jagannathan and Wang (2000).

If we predetermine the mean and variance of the factor and ignore the related moment restrictions, the SDF method is much less efficient than the beta method. This is true in both small as well as large samples. When we have 30 years of monthly observations, the standard deviation of $\hat{\lambda}$ obtained with the SDF method is several times as high as the standard deviation of λ^\dagger with the beta method. These results are similar to those reported by Kan and Zhou (1999). This also confirms that when macroeconomic factor innovations estimated using time-series models are used as factors, the classical beta method will tend to substantially overstate the precision of the estimated risk premia.

When the factor moment restrictions are added to the SDF method, the efficiency of the SDF method improves substantially. Notice that the standard error of $\bar{\lambda}$ is nearly the same as that of λ^\dagger . The increase in efficiency of the modified SDF method, relative to the beta method, is expected because the two methods have the same asymptotic variance for the estimated risk premium. Thus, the sharp disadvantage of the SDF method to the beta method reported by Kan and Zhou (1999) is mainly due to ignoring the estimation error in the factor moments.

In our theoretical derivation of the asymptotic variance, we assume that the variance of the returns do not depend on the realized value of the factor. This may be a rather restrictive assumption, as pointed out by MacKinlay and Richardson (1991). We therefore examine the applicability of our results when returns exhibit conditional heteroscedasticity. For this purpose, following MacKinlay and Richardson (1991), we make independent draws of the returns and the factor from a multivariate t distribution rather than a joint Normal distribution. When the multivariate t distribution has ν degrees of freedom, the conditional covariance matrix of the residuals in the regression equation, conditional on the realized value of the factor, is given by (see MacKinlay and Richardson, 1991, equation 14):

$$\text{Var}[\epsilon_t|f_t] = \frac{\nu - 2 + (f_t - \mu)^2/\sigma^2}{\nu - 1} \Omega .$$

Notice that dependence of the conditional covariance on the realized value of the factor increases as the degree of freedom ν decreases. However, there is a lower bound for the number of degrees of freedom for us to choose. The asymptotic distribution theory for

the GMM requires that returns and factors have finite fourth moments. Hence the degree of freedom has to be higher than 4. In view of this we use 5 degrees of freedom for the multivariate t distribution, following MacKinlay and Richardson (1991).

The middle part of Table 3 gives the simulation results corresponding to the t distribution. As can be seen, the standard errors computed using the asymptotic theory are about the same as that we get through simulation. It is also true that predetermining the factor mean and variance causes the estimator λ^\dagger in the beta method to be more efficient than the estimator $\hat{\lambda}$ in the SDF method. Further, incorporation of the factor moment restrictions makes the estimator $\bar{\lambda}$ in the modified SDF method as efficient as the estimator λ^\dagger in the beta method. When the mean and variance of the factor are not predetermined, which is a more realistic and correct approach, the two estimators λ^* and $\hat{\lambda}$ in the two methods have almost the same precision in every sample size. This indicates that our results are robust to the existence of conditional heteroscedasticity.

As an alternative to the multi-variate t distribution, we also considered the joint empirical distribution of the excess returns and the factor. The monthly observations of the return on the value-weighted index of NYSE, AMEX and NASDAQ are used as the data of f . The residuals in the regression of decile returns on the index return are used as the data of ϵ . Independent samples $\{(f_t, \epsilon_t)'\}_{t=1, \dots, T}$ are drawn from an estimated empirical distribution of the data.⁸ Excess returns on 10 portfolios are constructed to satisfy $r_t = (\delta - \mu + f_t)\beta + \epsilon_t$ for $t = 1, \dots, T$. The parameters, λ , δ , β , μ , σ and Ω , are set to those in Table 1. Each estimator is then calculated based on the T samples to obtain a sample of the estimator. We repeat this independently 1000 times to obtain 1000 independent samples of each estimator.

The simulation results are given in the last part of Table 3. Again, when the mean and variance of the factor are estimated together with the risk premium, the sampling errors for the risk premium estimated using the SDF method and the beta method are almost identical. Ignoring the estimation error in the innovations of factors causes the beta method to be far more precise than the SDF method. When the factor mean and variance are predetermined, incorporating the factor moment restrictions greatly improves the precision of the SDF method when the factor mean and variance are predetermined.

⁸We can estimate the empirical distribution either by the Bootstrap method or the method described by Taylor and Thompson (1986). The results are the same, and we only report the former.

4.2 Specification Tests

It is common to evaluate model mis-specification by examining the pricing error of the model. A better method would provide a more precise estimate of the model pricing error. In the theorems presented earlier, we showed that the estimated pricing error has a smaller variance in the beta method when compared to that in the SDF method. To assess the quantitative importance of the difference between the two methods, we set λ , δ , β , μ and σ to the values shown in Table 1 and calculate the asymptotic standard deviations of the estimated pricing errors for the ten size-decile portfolios. The calculations are based on the formulae given in the Appendix under the null hypothesis of $E[r] = \delta\beta$. The observations of the returns and the factor are assumed, as in the theorems, to have a joint Normal distribution, identical and independent across the observations. The results are reported in the rows corresponding to $T = \infty$ in Table 4. The other rows are the asymptotic standard deviations divided by the square root of the time-series length T . We only consider the general case where the factor mean and variance are not predetermined. As claimed by the theorems in this paper, the standard deviations of the pricing errors estimated in the SDF method are larger than those in the beta method. However, the differences are quite small. Such small differences would be difficult to detect using Monte Carlo methods.

If we use the same parameter values as in the above theoretical calculation, the standard deviations of the estimated pricing errors obtained from Monte Carlo simulations cannot be very different for the two methods. Table 5 reports the results of our Monte Carlo simulation for the pricing errors. The design of the simulation is exactly the same as the design for Table 3. To save space, we only report the results for Normal distribution. As can be seen in Table 5, the standard deviations of the pricing errors obtained in the two methods are very similar for all the assets. This is similar to what Cochrane (2000) obtained. Based on Monte Carlo simulations, it would be tempting to conclude that the beta and SDF methods give the same standard deviation of the estimated pricing errors. Our theorems show that this conclusion is wrong. The difference can be large for some parameter values. However, for parameter values that are likely to be encountered in most empirical studies, the differences between the two methods are negligible.

A convenient way of examining model specification is the test based on the Hansen's J -statistic. In Theorem 1 we have pointed out that the J -statistics in the SDF and the

beta methods have the same asymptotic distribution. Therefore, the sampling distribution of the J -statistics can be different only in finite samples or in misspecified models. In what follows we therefore examine the size and power of the tests under the two methods in finite samples.

To examine the test size in the two methods, we use Monte Carlo simulations to compute the rejection rates under the null hypothesis that $E[r_t] = \delta\beta$. For this purpose, independent samples $\{(f_t, \epsilon_t')\}_{t=1, \dots, T}$ are drawn from a Normal distribution as we have described earlier.⁹ Excess returns on the 10 portfolios are then generated to satisfy $r_t = (\delta + f_t - \mu)\beta + \epsilon_t$ for $t = 1, \dots, T$ so that the asset pricing model holds. The parameters, λ , δ , β , μ , σ and Ω , are set to the values in Table 1. We only consider the general case where the factor moments are not predetermined. We obtain the parameter estimates and the J -statistic for each method based on a sample size of T . We repeat this 1000 times to obtain 1000 independent samples of the J statistic. We calculate the rejection rates for the 40%, 30%, 20% and 10% significance levels.

The simulation results are reported in Panel A of Table 6. For the size of the samples we considered, the rejection rates in the SDF method are all close to the theoretical p -values based on the χ^2 distribution. However, the rejection rates in the beta method are all too small relative to the theoretical values based on the asymptotic distribution. Thus, the test size of the beta method is not correct in small samples. When the sample size T increase, the test size moves closer to the theoretical p -value. Our simulations (not reported) indicates that we need more than a century of monthly observations to bring the test size to the level of the theoretical p -values. In this case, it is clear that the SDF method works better than the beta method.

To examine the power of the two methods, we conduct similar Monte Carlo simulations, except that we add a nonzero Jensen's alpha to the model for generating excess returns. The nonzero Jensen's alpha makes the asset-pricing model mis-specified. To be specific, excess returns are constructed to satisfy $r_t = \alpha + (\delta + f_t - \mu)\beta + \epsilon_t$. We use Jensen's α s that are close to what we obtain in the sample used for calibration purposes. They are listed in Table 1. As is to be expected based on empirical findings reported in the finance literature, portfolios of small stocks are associated with larger pricing errors. The power is the fraction of the

⁹We also have also drawn samples from t distribution and the empirical distribution. The results are not reported because they are similar to the results obtained from the Normal distribution.

simulated J -statistics that are larger than the critical value based on the χ^2 distribution. The results are reported in Panel B of Table 6. Given the smaller size of the beta method, it is not surprising that it has much smaller power as well when compared to the SDF method in finite samples.

Our simulation results on the size and power of the two methods are strikingly different than those reported by Kan and Zhou (1999). Their simulation shows that the two methods have the same test size but the beta method has higher test power than the SDF method. To understand why we obtain different results, we repeated the simulations using a set of parameters similar to those used by Kan and Zhou (1999) — in particular, we set $\mu = 0$ and $\sigma^2 = 1$. In that case, we find that the beta and SDF methods have the same test size in finite samples. The simulation results are reported in Panel C of Table 6. We then examined the power by adding Jensen’s alpha to the data generating process. The rejection rates in these simulations are reported in Panel D of Table 6. It is clear that the two methods have similar power. Kan and Zhou (1999) find higher power for the beta method because they introduce mis-specification using some artificially constructed factors that are “noisy” and “unsystematic” (see Kan and Zhou (1999) for the definition of these factors). In contrast, we find that the two methods have similar power when we introduce model mis-specification by adding Jensen’s alpha’s calibrated to match what we observe in the data.

5 Conclusion

The stochastic discount factor (SDF) method has received wide attention in the theoretical and empirical asset pricing literature. The main attraction of the SDF method is its generality. It provides an elegant framework for econometric evaluation of both linear and nonlinear asset pricing models, including pricing models for derivative securities. We examine whether the generality of the SDF framework comes at the cost of estimation efficiency for risk premiums. For that purpose we compare the classical beta method with the SDF method for linear factor-pricing models. For such models the classical beta method is equivalent to the maximum likelihood method under suitable assumptions regarding the statistical properties of returns and factors. Hence the classical beta method has a natural advantage for such models. If the SDF method provides as precise an estimate of factor risk premiums even for linear factor-pricing models, then there would be less need for concern that the generality of

the SDF method comes at a cost.

We show that in spite of its generality, the SDF method has the same asymptotic precision as the beta method for estimating risk premiums in linear factor-pricing models. Monte Carlo simulations suggest that they provide estimates with similar precision even in finite samples. If our findings were otherwise there would have been some advantage to applying the beta method to nonlinear asset pricing models through linear approximations. Our results suggest that there may be no such gains.

In empirical studies of asset pricing models that use macroeconomic factors, the common practice is to measure factors as unanticipated changes in the corresponding variables by using time-series methods. Although the factors so constructed will have estimation errors, it is often difficult to take them into account when evaluating asset-pricing models using the classical beta method. Therefore, empirical studies often ignore the estimation errors associated with the factors, treating them as being of second order of importance. It follows from our results that this practice is likely to lead to substantial overstatement of the precision with which risk premiums are estimated using the classical beta method. In contrast, the reported standard errors will be correct when the SDF method is used, which is an important advantage.

We find that the relative size and power of the beta and SDF methods in finite samples depend to a large extent on the data generating process. For example, for one set of values for the parameters, we find that the beta method has the wrong test size for the correctly specified model and lower power for the incorrectly specified model. However, for an alternative set of values of the parameters, that are similar to those used in Kan and Zhou (1999), the two methods have similar test size and power. We also find that the beta method provides a more precise estimate of the vector of model pricing errors and hence has a higher asymptotic power. However, for the values of parameters we considered in our simulations, this advantage is not of practical importance.

6 Appendix

6.1 Proof of Theorem 1

It follows from Lemma 4.2 in Hansen (1982) that each of the J -Statistics has an asymptotic χ^2 distribution. The number of degrees of freedom of the χ^2 distribution equals the difference between the number of moment restrictions and the number of parameters that are estimated. There are n moment restrictions and 1 parameters in the SDF method, and there are $2n + 2$ moment restrictions and $n + 3$ parameters in the regression method. Thus, the asymptotic distribution of the J -statistics in the SDF method and the beta method have the same χ^2 distribution with $n - 1$ degrees of freedom.

In order to calculate the asymptotic variance of the estimator $\hat{\lambda}$ obtained with the SDF method, let us consider the following vector of random variables

$$g(r_t, f_t, \lambda) = r_t(1 - \lambda f_t).$$

Substituting r_t by equation (3), which is implied by the i.i.d.-normal assumption, and λ by equation (12), we obtain the covariance matrix of g as

$$\begin{aligned} S &= E[g(r_t, f_t, \lambda)g(r_t, f_t, \lambda)'] = \\ &= E \left[\left(1 - \frac{\delta}{\sigma^2 + \mu\delta} f_t \right)^2 (\delta + f_t - \mu)^2 \right] \beta\beta' + E \left[\left(1 - \frac{\delta}{\sigma^2 + \mu\delta} f_t \right)^2 \right] \Omega \\ &= \frac{\sigma^2(\sigma^4 + \delta^4)}{(\sigma^2 + \mu\delta)^2} \beta\beta' + \frac{\sigma^2(\sigma^2 + \delta^2)}{(\sigma^2 + \mu\delta)^2} \Omega, \end{aligned}$$

and its inverse is

$$S^{-1} = \frac{(\sigma^2 + \mu\delta)^2}{\sigma^2(\sigma^2 + \delta^2)} \Omega^{-1} - \frac{(\sigma^2 + \mu\delta)^2}{\sigma^2(\sigma^2 + \delta^2)} \left(\beta'\Omega^{-1}\beta + \frac{\sigma^4 + \delta^4}{\sigma^2 + \delta^2} \right)^{-1} \Omega^{-1} \beta\beta'\Omega^{-1}.$$

The expected value of the derivative of g with respect to λ is

$$D = E \left[\frac{\partial g}{\partial \lambda} \right] = -(\sigma^2 + \mu\delta)\beta.$$

The asymptotic variance of the estimator of λ is $(D'S^{-1}D)^{-1}$, which gives

$$\text{Avar}(\hat{\lambda}) = \frac{\sigma^2(\sigma^2 + \delta^2)}{(\sigma^2 + \mu\delta)^4} (\beta'\Omega^{-1}\beta)^{-1} + \frac{\sigma^2(\sigma^4 + \delta^4)}{(\sigma^2 + \mu\delta)^4}. \quad (32)$$

The estimated pricing error in the SDF method is

$$\hat{\pi} = e(\hat{\lambda}) = \frac{1}{T} \left(\sum_{t=1}^T g(r_t, f_t, \hat{\lambda}) \right).$$

It follows from Hansen (1982) that

$$\text{Avar}[e(\hat{\lambda})] = S - D[D'S^{-1}D]^{-1}D',$$

which gives

$$\text{Avar}[\hat{\pi}] = \frac{\sigma^2(\sigma^2 + \delta^2)}{(\sigma^2 + \mu\delta)^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta']. \quad (33)$$

Next, let us calculate the asymptotic variance of λ^* . In the beta method, the vector of unknown parameters is $\theta = (\delta, \beta', \mu, \sigma^2)'$. Let us define a random variable $g(r_t, f_t, \theta)$ as

$$g(r_t, f_t, \theta) = \begin{pmatrix} r_t - (\delta + f_t - \mu)\beta \\ (r_t - (\delta + f_t - \mu)\beta)f_t \\ f_t - \mu \\ (f_t - \mu)^2 - \sigma^2 \end{pmatrix} = \begin{pmatrix} \epsilon_t \\ \epsilon_t f_t \\ f_t - \mu \\ (f_t - \mu)^2 - \sigma^2 \end{pmatrix}.$$

The covariance matrix of g is

$$S = \begin{pmatrix} \Omega & \mu\Omega & 0 & 0 \\ \mu\Omega & (\mu^2 + \sigma^2)\Omega & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 2\sigma^4 \end{pmatrix}, \quad (34)$$

and its inverse is

$$S^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} (\sigma^2 + \mu^2)\Omega^{-1} & -\mu\Omega^{-1} & 0 & 0 \\ -\mu\Omega^{-1} & \Omega^{-1} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma^2} \end{pmatrix}.$$

The expected value of the derivative of g with respect to θ is

$$D = E \left[\frac{\partial g}{\partial \theta'} \right] = \begin{pmatrix} -\beta & -\delta I_n & \beta & 0 \\ -\mu\beta & -(\sigma^2 + \mu\delta)I_n & \mu\beta & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

It follows that

$$D'S^{-1}D = \begin{pmatrix} \beta'\Omega^{-1}\beta & \delta\beta'\Omega^{-1} & -\beta'\Omega^{-1}\beta & 0 \\ \delta\Omega^{-1}\beta & (\sigma^2 + \delta^2)\Omega^{-1} & -\delta\Omega^{-1}\beta & 0 \\ -\beta'\Omega^{-1}\beta & -\delta\beta'\Omega^{-1} & \frac{1}{\sigma^2} + \beta'\Omega^{-1}\beta & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

The asymptotic covariance matrix of the estimator θ^* is $(D'S^{-1}D)^{-1}$. We find that the inverse matrix of $D'S^{-1}D$ is

$$\begin{pmatrix} \frac{\sigma^2 + \delta^2}{\sigma^2}(\beta'\Omega^{-1}\beta)^{-1} + \sigma^2 & -\frac{\delta}{\sigma^2}(\beta'\Omega^{-1}\beta)^{-1}\beta' & \sigma^2 & 0 \\ -\frac{\delta}{\sigma^2}(\beta'\Omega^{-1}\beta)^{-1}\beta & \frac{1}{\sigma^2 + \delta^2}\Omega + \frac{\delta^2}{\sigma^2(\sigma^2 + \delta^2)}(\beta'\Omega^{-1}\beta)^{-1}\beta\beta' & 0 & 0 \\ \sigma^2 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 2\sigma^4 \end{pmatrix}.$$

The asymptotic variance of the estimator for δ is the first element of the above matrix, which is then

$$\text{Avar}(\delta^*) = \frac{\sigma^2 + \delta^2}{\sigma^2}(\beta'\Omega^{-1}\beta)^{-1} + \sigma^2.$$

Using equation (19) and the Delta method, we obtain that

$$\text{Avar}(\lambda^*) = \frac{\sigma^2(\sigma^2 + \delta^2)}{(\sigma^2 + \mu\delta)^4}(\beta'\Omega^{-1}\beta)^{-1} + \frac{\sigma^2(\sigma^4 + \delta^4)}{(\sigma^2 + \mu\delta)^4}. \quad (35)$$

To calculate the asymptotic variance of the pricing error obtained with the beta method, we define

$$e(\theta) = \frac{1}{T} \left(\sum_{t=1}^T g(r_t, f_t, \theta) \right).$$

It follows from Hansen (1982) that

$$\begin{aligned} \text{Avar}[e(\theta^*)] &= S - D(D'S^{-1}D)^{-1}D' \\ &= \begin{pmatrix} \frac{\sigma^2}{\sigma^2 + \delta^2}[\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'] & A_{12} & 0_{n \times 1} & 0_{n \times 1} \\ A_{21} & A_{22} & 0_{n \times 1} & 0_{n \times 1} \\ 0_{1 \times n} & 0_{1 \times n} & 0 & 0 \\ 0_{1 \times n} & 0_{1 \times n} & 0 & 0 \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} A_{12} &= A'_{21} = \frac{(\mu(\sigma^2 + \delta^2) + \delta(\mu^2 + \sigma^2))}{\sigma^2 + \delta^2}[\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'] \\ A_{22} &= \frac{(\delta^2 - \mu^2)(\mu^2 + \sigma^2)}{\sigma^2 + \delta^2}\Omega - \frac{(\mu + \delta)^2(\sigma^2 + \mu\delta)^2}{\sigma^2(\sigma^2 + \delta^2)}(\beta'\Omega^{-1}\beta)^{-1}\beta\beta' \end{aligned}$$

The estimated Jensen's alpha in the beta method is

$$\alpha^* = Q^*e(\theta^*), \quad \text{where } Q^* = [I_n, 0_{n \times n}, -\beta^*, 0_{n \times 1}].$$

Thus,

$$\text{Avar}[\alpha^*] = Q[S - D(D'S^{-1}D)^{-1}D]Q' \quad \text{where } Q = [I_n, 0_{n \times n}, -\beta, 0_{n \times 1}],$$

which gives

$$\begin{aligned}\text{Avar}[\alpha^*] &= Q[S - D(D'S^{-1}D)^{-1}D']Q' \\ &= \frac{\sigma^2}{\sigma^2 + \delta^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta']\end{aligned}$$

It then follows from equation (20) that the asymptotic variance of the estimated pricing error is

$$\begin{aligned}\text{Avar}[\pi^*] &= \left(\frac{\sigma^2}{\sigma^2 + \mu\delta}\right)^2 \text{Avar}[\alpha^*] \\ &= \frac{\sigma^6}{(\sigma^2 + \delta^2)(\sigma^2 + \mu\delta)^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'].\end{aligned}\tag{36}$$

Finally, the equality $\text{Avar}[\hat{\lambda}] = \text{Avar}[\lambda^*]$ follows from equations (32) and (35). The matrix $\text{Avar}[\hat{\pi}] - \text{Avar}[\pi^*]$ is positive semi-definite because equations (33) and (36) imply

$$\text{Avar}[\hat{\pi}] - \text{Avar}[\pi^*] = \frac{\sigma^2\delta^2(2\sigma^2 + \delta^2)}{(\sigma^2 + \delta^2)(\sigma^2 + \mu\delta)^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta']$$

and because $\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'$ is positive semi-definite. The proof is then complete.

6.2 Proof of Theorem 2

In the SDF method, the sampling distributions of the estimated λ , J -statistics and the pricing error are exactly the same as in Theorem 1 because the moment restriction in the SDF method is independent of the factor mean and variance. In the beta method, the asymptotic distribution of the J -statistics is the χ^2 distribution with $n - 1$ degrees of freedom, same as the J -statistics in the SDF method, because there are $2n$ moment restrictions and $n + 1$ parameters to estimate. Similarly to the derivation of $\text{Avar}[\lambda^*]$ and $\text{Avar}[\pi^*]$ in the proof of Theorem 1, one obtain the asymptotic variances of λ^\dagger and π^\dagger as

$$\text{Avar}(\lambda^\dagger) = \frac{\sigma^2(\sigma^2 + \delta^2)}{(\sigma^2 + \mu\delta)^4} (\beta'\Omega^{-1}\beta)^{-1},\tag{37}$$

$$\text{Avar}[\pi^\dagger] = \frac{\sigma^6}{(\sigma^2 + \delta^2)(\sigma^2 + \mu\delta)^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'].\tag{38}$$

The inequality $\text{Avar}[\hat{\lambda}] > \text{Avar}[\lambda^\dagger]$ is obtained by comparing the equations (32) and (37). The matrix $\text{Avar}[\hat{\pi}] - \text{Avar}[\pi^\dagger]$ is positive semi-definite because equations (33) and (38) imply

$$\text{Avar}[\hat{\pi}] - \text{Avar}[\pi^\dagger] = \frac{\sigma^2\delta^2(2\sigma^2 + \delta^2)}{(\sigma^2 + \delta^2)(\sigma^2 + \mu\delta)^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta']$$

and because $\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'$ is positive semi-definite. The proof is then complete.

6.3 Proof of Theorem 3

In the SDF method, the asymptotic distribution of the J -statistics is the χ^2 distribution with $n + 1$ degrees of freedom because there are $n + 2$ moment restrictions and 1 parameter to estimate.

In order to calculate the asymptotic variance of the estimator $\hat{\lambda}$ obtained with the SDF method, let us consider the following vector of random variables

$$g(r_t, f_t, \lambda) = \begin{pmatrix} r_t(1 - \lambda f_t) \\ f_t - \mu \\ (f_t - \mu)^2 - \sigma^2 \end{pmatrix}.$$

Substituting r_t by equation (3), which is implied by the i.i.d.-normal assumption, and λ by equation (12), we obtain the covariance matrix of g as

$$S = \begin{pmatrix} \frac{\sigma^2}{(\sigma^2 + \mu\delta)^2} [(\sigma^4 + \delta^4)\beta\beta' + (\sigma^2 + \delta^2)\Omega] & \frac{\sigma^2(\sigma^2 - \delta^2)}{\sigma^2 + \mu\delta}\beta & -\frac{2\delta\sigma^4}{\sigma^4 + \mu\delta}\beta \\ \frac{\sigma^2(\sigma^2 - \delta^2)}{\sigma^2 + \mu\delta}\beta' & \sigma^2 & 0 \\ -\frac{2\delta\sigma^4}{\sigma^2 + \mu\delta}\beta' & 0 & 2\sigma^4 \end{pmatrix}.$$

The expected value of the derivative of g with respect to λ is

$$D = \text{E} \left[\frac{\partial g}{\partial \lambda} \right] = \begin{pmatrix} -(\sigma^2 + \mu\delta)\beta \\ 0 \\ 0 \end{pmatrix}.$$

After some algebraic manipulation, we obtain

$$D'S^{-1}D = \frac{(\sigma^2 + \mu\delta)^4}{\sigma^2(\sigma^2 + \delta^2)}(\beta'\Omega^{-1}\beta).$$

The asymptotic variance of the estimator of λ is $(D'S^{-1}D)^{-1}$, which gives

$$\text{Avar}(\bar{\lambda}) = \frac{\sigma^2(\sigma^2 + \delta^2)}{(\sigma^2 + \mu\delta)^4}(\beta'\Omega^{-1}\beta)^{-1}. \quad (39)$$

To calculate the asymptotic variance of the estimated pricing error in the SDF method, we define

$$e(\lambda) = \frac{1}{T} \left(\sum_{t=1}^T g(r_t, f_t, \lambda) \right).$$

It follows from Hansen (1982) that

$$\text{Avar}[e(\bar{\lambda})] = S - D[D'S^{-1}D]^{-1}D'.$$

The estimated pricing error in the SDF method is

$$\bar{\pi} = Qe(\bar{\lambda}), \quad \text{where } Q = [I_n, \mathbf{0}_{n \times 2}].$$

We therefore have

$$\text{Avar}[\bar{\pi}] = Q[S - D[D'S^{-1}D]^{-1}D']Q',$$

which gives

$$\text{Avar}[\bar{\pi}] = \frac{\sigma^2(\sigma^2 + \delta^2)}{(\sigma^2 + \mu\delta)^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'] + \frac{\sigma^2(\sigma^4 + \delta^4)}{(\sigma^2 + \mu\delta)^2}. \quad (40)$$

Finally, we obtain the equality $\text{Avar}[\bar{\lambda}] = \text{Avar}[\lambda^\dagger]$ by comparing equations (39) and (37). The matrix $\text{Avar}[\bar{\pi}] - \text{Avar}[\pi^\dagger]$ is positive semi-definite because equations (40) and (38) imply

$$\text{Avar}[\hat{\pi}] - \text{Avar}[\pi^\dagger] = \frac{\sigma^2\delta^2(2\sigma^2 + \delta^2)}{(\sigma^2 + \delta^2)(\sigma^2 + \mu\delta)^2} [\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'] + \frac{\sigma^2(\sigma^4 + \delta^4)}{(\sigma^2 + \mu\delta)^2}.$$

and because $\Omega - (\beta'\Omega^{-1}\beta)^{-1}\beta\beta'$ is positive semi-definite. This completes the proof.

6.4 Proof of Lemma 1

Let

$$D_1 = \lim_{T \rightarrow \infty} \text{E} \left[\frac{1}{T} \sum_{t=1}^T \frac{\partial g(x_t, \theta_1, \theta_2)}{\partial \theta'_1} \right]$$

$$D_2 = \lim_{T \rightarrow \infty} \text{E} \left[\frac{1}{T} \sum_{t=1}^T \frac{\partial g(x_t, \theta_1, \theta_2)}{\partial \theta'_2} \right].$$

It follows that

$$D = \lim_{T \rightarrow \infty} \text{E} \left[\frac{1}{T} \sum_{t=1}^T \frac{\partial g(x_t, \theta_1, \theta_2)}{\partial \theta'} \right] = (D_1, D_2).$$

Let S be the spectral density matrix of $g(x_t, \theta_1, \theta_2)$. We thus have $\text{Avar}[\tilde{\theta}_1] = (D'_1 S^{-1} D_1)^{-1}$ and $\text{Avar}[(\hat{\theta}'_1, \hat{\theta}'_2)'] = (D' S^{-1} D)^{-1}$. The inverse of the asymptotic variance of $(\hat{\theta}'_1, \hat{\theta}'_2)'$ is

$$D' S^{-1} D = \begin{pmatrix} D'_1 S^{-1} D_1 & D'_1 S^{-1} D_2 \\ D'_2 S^{-1} D_1 & D'_2 S^{-1} D_2 \end{pmatrix}.$$

It follows from the formula of the inverse of partitioned matrix that

$$\begin{aligned} \text{Avar}(\hat{\theta}_1) &= (D'_1 S^{-1} D_1)^{-1} + (D'_1 S^{-1} D_1)^{-1} (D'_1 S^{-1} D_2) [(D'_2 S^{-1} D_2) \\ &\quad - (D'_2 S^{-1} D_1) (D'_1 S^{-1} D_1)^{-1} (D'_1 S^{-1} D_2)]^{-1} (D'_2 S^{-1} D_1) (D'_1 S^{-1} D_1)^{-1}. \end{aligned} \quad (41)$$

We must have $\text{Avar}[\hat{\theta}_1] - \text{Avar}[\tilde{\theta}_1]$ being positive semi-definite because the first term on the right-hand side of equation (41) equals $\text{Avar}[\tilde{\theta}_1]$ and the other term is positive semi-definite. In view of equation (41), it is obvious that

$$\lim_{T \rightarrow \infty} \text{E} \left[\frac{1}{T} \sum_{t=1}^T \partial g(x_t, \theta_1, \theta_2) / \partial \theta'_2 \right] = 0$$

implies that $D_2 = 0$ and thus $\text{Avar}[\hat{\theta}_1] = (D_1 S^{-1} D_1)^{-1}$, which is the same as $\text{Avar}[\tilde{\theta}_1]$.

6.5 Proof of Lemma 2

Suppose that the dimensions of $g_1(x_t, \theta)$ and $g_2(x_t)$ are m and n respectively and that the dimensions of θ is k . Let

$$S_{ij} = \sum_{j=-\infty}^{\infty} \text{E}[g_i(x_t, \theta) g_j(x_{t+j}, \theta)'], \quad \forall i, j = 1, 2.$$

The spectral density matrix of $g(x_t, \theta)$ is

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

Let us denote the inverse of S by

$$S^{-1} = \begin{pmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{pmatrix}.$$

Define

$$D_1 = \lim_{T \rightarrow \infty} \text{E} \left[\frac{1}{T} \sum_{t=1}^T \frac{\partial g_1(x_t, \theta)}{\partial \theta'} \right],$$

and $D = \begin{pmatrix} D_1 \\ 0_{n \times k} \end{pmatrix}.$

The asymptotic variances of $\hat{\theta}$ and $\tilde{\theta}$ are, respectively,

$$\text{Avar}[\hat{\theta}] = (D' S^{-1} D)^{-1} \quad \text{and} \quad \text{Avar}[\tilde{\theta}] = (D'_1 S_{11}^{-1} D_1)^{-1}.$$

By the formula of the inverse of partitioned matrix, we have

$$S^{11} = S_{11}^{-1} + S_{11}^{-1} S_{12} (S_{22} - S_{21} S_{11}^{-1} S_{12})^{-1} S_{21} S_{11}^{-1},$$

which implies

$$D' S^{-1} D = D'_1 S^{11} D_1 = D'_1 S_{11}^{-1} D_1 + D'_1 S_{11}^{-1} S_{12} (S_{22} - S_{21} S_{11}^{-1} S_{12})^{-1} S_{21} S_{11}^{-1} D_1.$$

The inverse is then

$$\begin{aligned} (D'S^{-1}D)^{-1} = & (D_1'S_{11}^{-1}D_1)^{-1} - (D_1'S_{11}^{-1}D_1)^{-1}D_1'S_{11}^{-1}S_{12}[S_{22} - S_{21}S_{11}^{-1}S_{12} \\ & + S_{21}S_{11}^{-1}D_1(D_1'S_{11}^{-1}D_1)^{-1}D_1'S_{11}^{-1}S_{12}]S_{21}S_{11}^{-1}D_1(D_1'S_{11}^{-1}D_1)^{-1} \end{aligned} \quad (42)$$

Since both $S_{22} - S_{21}S_{11}^{-1}S_{12}$ and $S_{21}S_{11}^{-1}D_1(D_1'S_{11}^{-1}D_1)^{-1}D_1'S_{11}^{-1}S_{12}$ are positive semi-definite, it follows from equation (42) that $(D'S^{-1}D)^{-1} - (D_1'S_{11}^{-1}D_1)^{-1}$ is negative semi-definite, which implies that $\text{Avar}[\hat{\theta}] - \text{Avar}[\tilde{\theta}]$ is negative semi-definite. If

$$\sum_{j=-\infty}^{\infty} \text{E}[g_1(x_t, \theta)g_2(x_{t+j})'] = 0,$$

then $S_{12} = 0$ and $S_{21} = 0$, which imply $\text{Avar}[\hat{\theta}] = \text{Avar}[\tilde{\theta}]$ by equation (42).

7 References

- Black, F., 1972, “Capital market equilibrium with restricted borrowing,” *Journal of Business*, 45, 444–454.
- Black, F., M. Jensen, and M. Scholes, 1972, “The capital asset pricing model: some empirical tests,” in Jensen, M. (ed.), *Studies in the Theory of Capital Markets*, Praeger, New York.
- Bollerslev, Tim, Ray Chou and Kenneth Kroner, 1992, “ARCH Modeling in Finance,” *Journal of Econometrics*, 52, 5-59.
- Breeden, D., M. Gibbons, and R. Litzenberger, 1989, “Empirical tests of the consumption-oriented CAPM,” *Journal of Finance*, 44, 231–262.
- Campbell, J., 1993, “Intertemporal asset pricing without consumption data,” *American Economic Review*, 83, 487–512.
- Campbell, J., 1996, “Understanding risk and return,” *Journal of Political Economy*, 104, 298–345.
- Campbell, J., A. Lo, and A. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.
- Chen, N., R. Roll, and S. Ross, 1986, Economic forces and the stock market, *Journal of Business*, 59, 383–404.
- Cochrane, J., 1996, “A cross-sectional test of an investment-based asset pricing model,” *Journal of Political Economy*, 104, 572–621.
- Cochrane, J., 1999, “Asset pricing,” manuscript, Graduate School of Business, University of Chicago.
- Cochrane, J., 2000, “A resurrection of the stochastic discount factor/GMM methodology,” manuscript, Graduate School of Business, University of Chicago.
- Dybvig, P., and J. Ingersoll, 1982, “Mean-variance theory in complete markets,” *Journal of Business*, 55, 233–252.
- Elton, E., and M. Gruber, and J. Mei, 1994, “Cost of capital using arbitrage pricing theory: A case study of nine New York utilities,” *Financial markets, Institutions & Instruments*, 3, 46–73.
- Fama, E., and J. MacBeth, 1973, “Risk, return, and equilibrium: empirical tests,” *Journal*

of *Political Economy*, 71, 607–636.

Fama, E., and K. French, 1992, “The cross-section of expected stock returns,” *Journal of Finance*, 47, 427–465.

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

E. Fama and K. French, 1997, “Industry costs of equity,” *Journal of Financial Economics*, 43, 153–193.

Ferson, W., 1995, “Theory and empirical testing of asset pricing models,” *Handbooks in Operations Research and Management Science* edited by Jarrow, Maksimovic and Ziemba, 145–200.

Ferson, W., and C. Harvey, 1997, “Fundamental determinants of national equity market returns: A perspective on conditional asset pricing,” *Journal of Banking and Finance* 21, 1625–1665.

French, Ken, William G. Schwert, and Robert F. Stambaugh, 1987, “Expected stock returns and volatility,” *Journal of Financial Economics*, 17, 5–26.

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

Glosten, Lawrence, Ravi Jagannathan, and David Runkle, 1992, “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks,” *Journal of Finance*, 48, 1779–1801.

Hamilton, J., 1994, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

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

Hansen, L., and R. Richard, 1987, “The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models,” *Econometrica*, 55, 587–613.

Hansen, L., and K. Singleton, 1982, “Generalized instrumental variables estimation of non-linear rational expectations models,” *Econometrica*, 50, 1269–1286.

Hansen, L., and R. Jagannathan, 1991, “Implications of security market data for models of dynamic economies,” *Journal of Political Economy*, 99, 225–262.

Hansen, L., and R. Jagannathan, 1997, “Assessing specification errors in stochastic discount

- factor models,” *Journal of Finance*, 52, 557–590.
- Hodrick, R., and X. Zhang, 2000, “Evaluating the specification errors of asset pricing models,” Working paper #7661, *NBER*.
- Ingersoll, J., 1987, *Theory of Financial Decision Making*, Rowman & Littlefield, Totowa, NJ.
- Jagannathan, R., and Z. Wang, 1996, “The conditional CAPM and the cross-section of expected returns,” *Journal of Finance*, 51, 3–53.
- Jagannathan, R., and Z. Wang, 1998, “An asymptotic theory for estimating beta-pricing models using cross-sectional regression,” *Journal of Finance*, 53, 1285–1309.
- Jagannathan, R., and Z. Wang, 2000, “Efficiency of the stochastic discount factor method for estimating risk premium,” Working paper, Columbia University.
- Kan, R., and G. Zhou, 1999, “A critique of the stochastic discount factor methodology,” *Journal of Finance*, 54, 1221–1248.
- Lucas, R., 1978, Asset prices in an exchange economy, *Econometrica*, 46, 1429–1446.
- MacKinlay, A., and M. Richardson, 1991, “Using generalized method of moments to test mean-variance efficiency,” *Journal of Finance*, 46, 511–527.
- Ross, S., 1976, “The arbitrage theory of capital asset pricing,” *Journal of Economics Theory*, 13, 341–360.
- Schink, G., and R. Bower, 1994, “Application of the Fama-French model to utility stocks,” *Financial markets, Institutions & Instruments*, 3, 74–96.
- Shanken, J., 1992, “On the estimation of beta-pricing models,” *Review of Financial Studies*, 5, 1–34.
- Sharpe, W., 1964, “Capital asset prices: a theory of market equilibrium under conditions of risk,” *Journal of Finance*, 19, 425–422.
- Taylor, M., and J. Thompson, 1986, “Data based random number generation for a multivariate distribution via stochastic simulation,” *Computational Statistics & Data Analysis*, 4, 93–101.

Table 1: Parameter Values used in Monte Carlo Simulations

This table presents the parameters used in our Monte Carlo simulations. The choice of the parameters are based on monthly historical observations (from January 1926 to December 1998) of returns (in excess of returns on one-month Treasury Bills) on decile portfolios and the value-weighted market index of NYSE, AMEX, and NASDAQ. The data are obtained from the Center for Research on Security Prices (CRSP). The mean (μ) and standard deviation (σ) of the factor is set to be the sample mean and standard deviation of returns on the market index. The betas (β) are set to be the slopes in the time-series regression of the decile returns on the market index return. Jensen's α is set to be the intercept in the same regression. The sample covariance of the residuals in this regression is chosen to be the covariance matrix Ω . The risk premium δ is set to be the slope in the cross-sectional regression of the decile's historical average return on beta. The parameter λ satisfies $\lambda = \delta/(\sigma^2 + \mu\delta)$. The numbers reported for μ , σ and α are multiplied by 100 while the numbers for Ω are multiplied by 10,000.

$\mu = 0.6914$		$\sigma = 5.5154$		$\delta = 1.3740$		$\lambda = 4.3790$			
Decile portfolios									
Small	2	3	4	5	6	7	8	9	Large
β									
1.46	1.40	1.31	1.26	1.24	1.23	1.18	1.12	1.09	0.95
α									
0.494	0.126	0.000	-0.004	0.000	-0.002	0.020	-0.042	0.023	0.003
Ω									
55.68	37.11	28.82	22.30	17.39	14.45	10.16	6.38	3.89	-2.81
37.11	29.83	22.62	17.88	13.96	11.87	8.30	5.47	3.17	-2.24
28.82	22.62	20.48	15.18	11.86	10.30	7.27	4.93	2.78	-1.92
22.30	17.88	15.18	13.67	10.12	8.86	6.64	4.77	2.63	-1.69
17.39	13.96	11.86	10.12	9.29	7.43	5.62	4.13	2.40	-1.42
14.45	11.87	10.30	8.86	7.43	7.31	5.10	3.84	2.33	-1.30
10.16	8.30	7.27	6.64	5.62	5.10	4.90	3.29	1.94	-1.04
6.38	5.47	4.93	4.77	4.13	3.84	3.29	3.25	1.78	-0.84
3.89	3.17	2.78	2.63	2.40	2.33	1.94	1.78	1.78	-0.57
-2.81	-2.24	-1.92	-1.69	-1.42	-1.30	-1.04	-0.84	-0.57	0.31

Table 2: Asymptotic Standard Deviations of the Estimated λ

The first row of numbers are the theoretical asymptotic standard deviations of various estimators of λ under the null hypothesis of $E[r_t] = \delta\beta$. The observations of the returns and the factor are assumed to have identical and independent joint Normal distributions. The true parameters, λ , δ , β , μ , σ and Ω , are assumed to be those given in Table 1. The numbers in the row corresponding to ∞ are the asymptotic standard deviations calculated according to the formulae derived in the Appendix. The numbers in the other rows are the asymptotic standard deviations divided by the square-root of T .

T	λ^*	$\hat{\lambda}$	λ^\dagger	$\bar{\lambda}$
∞	17.0897	17.0897	0.5055	0.5055
5×12	2.2063	2.2063	0.0653	0.0653
10×12	1.5601	1.5601	0.0461	0.0461
30×12	0.9007	0.9007	0.0266	0.0266
50×12	0.6977	0.6977	0.0206	0.0206

Table 3: Simulation Results for λ

This table provides the results of Monte Carlo simulations on various estimators λ under the null hypothesis of $E[r_t] = \delta\beta$. Independent samples $\{(f_t, \epsilon_t)'\}_{t=1, \dots, T}$ are drawn from a Normal, t or empirical distribution. Excess returns on 10 portfolios are constructed to satisfy $r_t = (\delta + f_t - \mu)\beta + \epsilon_t$ for $t = 1, \dots, T$. The parameters, λ , δ , β , μ , σ and Ω , are set to those in Table 1. Each estimator is then calculated based on the T samples to obtain a sample of the estimator. We repeat this independently 1000 times to obtain 1000 independent samples of the estimators. The table presents the standard deviation of the 1000 samples of each estimator.

T	λ^*	$\hat{\lambda}$	λ^\dagger	$\bar{\lambda}$
Normal distribution				
5×12	3.3097	2.8389	0.8618	0.6292
10×12	1.7533	1.8302	0.4432	0.2567
30×12	0.9684	0.9855	0.1327	0.0863
50×12	0.7798	0.7779	0.0833	0.0547
Student- t distribution				
5×12	3.3817	3.1348	0.7092	0.7581
10×12	2.0429	1.9996	0.4465	0.3899
30×12	0.9881	1.0452	0.1667	0.1494
50×12	0.7854	0.8005	0.1044	0.1002
Empirical distribution				
5×12	2.9727	2.8697	0.7416	0.6528
10×12	1.7542	1.7769	0.4334	0.2702
30×12	0.9250	0.9679	0.1395	0.0773
50×12	0.7148	0.7437	0.0830	0.0455

Table 4: Asymptotic Standard Deviations of the Estimated Pricing Errors

The first row of numbers are the theoretical asymptotic standard deviations (multiplied by 100) of various estimators of the pricing errors under the null hypothesis of $E[r_t] = \delta\beta$. The observations of the returns and the factor are assumed to have identical and independent joint Normal distributions. The true parameters, λ , δ , β , μ , σ and Ω , are assumed to be those given in Table 1. The numbers in the row corresponding to ∞ are the asymptotic standard deviations calculated according to the formulae derived in the Appendix. The numbers in the other rows are the asymptotic standard deviations divided by the square-root of T .

T		Small	2	3	4	5
∞	Beta	7.0180	5.1350	4.2538	3.4739	2.8620
	SDF	7.4535	5.4537	4.5178	3.6895	3.0396
5×12	Beta	0.9060	0.6629	0.5492	0.4485	0.3695
	SDF	0.9622	0.7041	0.5832	0.4763	0.3924
10×12	Beta	0.6407	0.4688	0.3883	0.3171	0.2613
	SDF	0.6804	0.4979	0.4124	0.3368	0.2775
30×12	Beta	0.3699	0.2706	0.2242	0.1831	0.1508
	SDF	0.3928	0.2874	0.2381	0.1945	0.1602
50×12	Beta	0.2865	0.2096	0.1737	0.1418	0.1168
	SDF	0.3043	0.2226	0.1844	0.1506	0.1241
T		6	7	8	9	Large
∞	Beta	2.5378	2.0756	1.6891	1.2430	0.5060
	SDF	2.6953	2.2044	1.7939	1.3202	0.5374
5×12	Beta	0.3276	0.2680	0.2181	0.1605	0.0653
	SDF	0.3480	0.2846	0.2316	0.1704	0.0694
10×12	Beta	0.2317	0.1895	0.1542	0.1135	0.0462
	SDF	0.2460	0.2012	0.1638	0.1205	0.0491
30×12	Beta	0.1338	0.1094	0.0890	0.0655	0.0267
	SDF	0.1421	0.1162	0.0945	0.0696	0.0283
50×12	Beta	0.1036	0.0847	0.0690	0.0507	0.0207
	SDF	0.1100	0.0900	0.0732	0.0539	0.0219

Table 5: Simulation Results for Pricing Errors

This table provides the results of Monte Carlo simulations on various estimators of pricing errors under the null hypothesis of $E[r_t] = \delta\beta$. Independent samples $\{(f_t, \epsilon_t')\}_{t=1, \dots, T}$ are drawn from a Normal distribution. Excess returns on 10 portfolios are constructed to satisfy $r_t = (\delta + f_t - \mu)\beta + \epsilon_t$ for $t = 1, \dots, T$. The parameters, $\lambda, \delta, \beta, \mu, \sigma$ and Ω , are set to those in Table 1. Each estimator is then calculated based on the T samples to obtain a sample of the estimator. We repeat this independently 1000 times to obtain 1000 independent samples of the estimators. The table presents the standard deviation (multiplied by 100) of the 1000 samples of each estimator.

T		Small	2	3	4	5
5×12	Beta	0.6987	0.5994	0.5346	0.5229	0.5208
	SDF	0.8913	0.6540	0.5484	0.4664	0.4116
10×12	Beta	0.4814	0.3612	0.3123	0.2697	0.2439
	SDF	0.6622	0.4868	0.4065	0.3377	0.2901
30×12	Beta	0.3277	0.2549	0.2159	0.1920	0.1677
	SDF	0.3660	0.2756	0.2227	0.1852	0.1548
50×12	Beta	0.2847	0.2191	0.1858	0.1593	0.1370
	SDF	0.2999	0.2206	0.1808	0.1473	0.1175
T		6	7	8	9	Large
5×12	Beta	0.5203	0.5255	0.5124	0.5142	0.4869
	SDF	0.3787	0.3460	0.3268	0.2995	0.2650
10×12	Beta	0.2336	0.2197	0.2153	0.2097	0.2145
	SDF	0.2616	0.2212	0.1983	0.1686	0.1327
30×12	Beta	0.1621	0.1504	0.1400	0.1309	0.1135
	SDF	0.1402	0.1190	0.0997	0.0781	0.0500
50×12	Beta	0.1315	0.1179	0.1090	0.1005	0.0836
	SDF	0.1080	0.0893	0.0742	0.0576	0.0335

Table 6: Simulation Results for the J -Statistics

This table provides the results of Monte Carlo simulations on the rejection rate of J -statistics. Independent samples $\{(f_t, \epsilon_t')\}_{t=1, \dots, T}$ are drawn from a Normal distribution. In panel A and C, excess returns on 10 portfolios are generated to satisfy $r_t = (\delta + f_t - \mu)\beta + \epsilon_t$ for $t = 1, \dots, T$. In panel B and D, excess returns are generated to satisfy $r_t = \alpha + (\delta + f_t - \mu)\beta + \epsilon_t$. In panel A and B, the parameters, $\lambda, \delta, \beta, \mu, \sigma$ and Ω as well as α , are set to those in Table 1. In panel C and D, we use a set of alternative parameters that are similar to those used in Kan and Zhou (1999). Each estimator is calculated based on the T samples to obtain a sample of the J -statistics. We repeat this independently 1000 times to obtain 1000 independent samples of the estimators. The table presents the fraction of the simulated J -statistics that are larger than the critical point at the significance levels of 40%, 30%, 20% and 10%.

T	$p = 0.40$		$p = .30$		$p = 0.20$		$p = .10$	
	Beta	SDF	Beta	SDF	Beta	SDF	Beta	SDF
A. Without pricing errors								
5×12	0.010	0.422	0.007	0.311	0.003	0.194	0.001	0.076
10×12	0.006	0.412	0.000	0.316	0.000	0.206	0.000	0.087
30×12	0.076	0.385	0.037	0.282	0.011	0.201	0.000	0.111
50×12	0.175	0.425	0.092	0.305	0.027	0.202	0.002	0.092
B. With pricing errors								
5×12	0.010	0.518	0.006	0.394	0.002	0.284	0.000	0.138
10×12	0.016	0.621	0.007	0.507	0.005	0.373	0.003	0.212
30×12	0.386	0.833	0.258	0.773	0.124	0.665	0.027	0.509
50×12	0.819	0.933	0.703	0.899	0.531	0.849	0.290	0.731
C. Without pricing errors								
5×12	0.435	0.443	0.329	0.334	0.204	0.210	0.077	0.088
10×12	0.413	0.415	0.304	0.306	0.211	0.212	0.100	0.101
30×12	0.405	0.405	0.312	0.313	0.205	0.205	0.109	0.109
50×12	0.405	0.405	0.309	0.309	0.196	0.196	0.096	0.096
D. With pricing errors								
5×12	0.528	0.536	0.416	0.426	0.270	0.284	0.131	0.143
10×12	0.628	0.629	0.511	0.512	0.389	0.393	0.217	0.221
30×12	0.841	0.843	0.780	0.780	0.686	0.685	0.541	0.542
50×12	0.938	0.938	0.920	0.920	0.873	0.873	0.763	0.763