A Critique of the Stochastic Discount Factor Methodology

RAYMOND KAN and GUOFU ZHOU*

ABSTRACT

In this paper, we point out that the widely used stochastic discount factor (SDF) methodology ignores a fully specified model for asset returns. As a result, it suffers from two potential problems when asset returns follow a linear factor model. The first problem is that the risk premium estimate from the SDF methodology is unreliable. The second problem is that the specification test under the SDF methodology has very low power in detecting misspecified models. Traditional methodologies typically incorporate a fully specified model for asset returns, and they can perform substantially better than the SDF methodology.

ASSET PRICING THEORIES, such as those of Sharpe (1964), Lintner (1965), Black (1972), Merton (1973), Ross (1976), and Breeden (1979), show that the expected return on a financial asset is a linear function of its covariances (or betas) with some systematic risk factors. This implication has been tested extensively in the finance literature by the so-called "traditional methodologies." In the traditional methodologies, a data-generating process is first proposed for the returns, and then the restrictions imposed by an asset pricing model are tested as parametric constraints on the return-generating process. The approach taken by the traditional methodologies has a potential problem, which is that when the proposed return-generating process is misspecified the test results could be misleading. Therefore, in applying the traditional methodologies, researchers typically have to justify that the proposed data-generating process provides a good description of the returns. For example, when the proposed return-generating process is a factor model, one would like the model to have high R^2 in explaining the returns on the test assets, especially when the test assets are well-diversified portfolios.

As many of the earlier theories are special cases of the stochastic discount factor (SDF) model, recent empirical asset pricing studies have been focused on testing the pricing restrictions in terms of the SDF model, rather

than on the traditional risk measures such as the beta and the Sharpe ratio. One of the most prominent papers in this line of research is Cochrane (1996), where the SDF methodology is fully explained. The formulation typically estimates the parameters and tests the pricing implications *without a fully specified model* of how the asset returns are generated in the economy. On the one hand, this appears very general and requires fewer assumptions and parameters than the traditional methodologies. On the other hand, it seems counterintuitive that one can be sure the pricing restrictions are true even if one knows little about the dynamics of the returns—that is, without a fully specified model (either parametric or non-parametric) of the returns.

This paper shows that if asset returns are generated by a linear factor model, then by ignoring the full dynamics of asset returns, as is currently done in empirical studies using the SDF methodology, two potential problems arise. The first problem is that the accuracy of the parameter estimation can be poor: the standard error of the estimated risk premium is often more than 40 times greater than that of the traditional methodologies, which should make one extra cautious when applying the SDF methodology. The second problem with the SDF methodology is that its specification test has very low power against misspecified models. With the usual sample size that we encounter in empirical studies, our simulation evidence suggests that the SDF methodology is not very reliable in detecting even gross misspecifications in an asset pricing model, especially when the proposed factors are not highly correlated with the returns.

The rest of the paper is organized as follows. The next section presents the traditional beta pricing model and the SDF model, and the empirical methodologies that are typically used to estimate and test such models. Though these are standard in the literature, the purpose here is to introduce notations and to facilitate later discussions. We also provide the intuition why the SDF methodology may not perform well when there is a lack of a fully specified model for the asset returns. In Sections II and III, we use asymptotic theory and Monte Carlo simulations to compare the performance of the traditional and SDF methodologies. The conclusions are in the final section.

I. Traditional and SDF Methodologies

A. *Tests of the Traditional Beta Pricing Model*

In order to make the results more easily understood, we present them in the simplest form. Let r_t be the excess return (in excess of the risk-free rate) on N risky assets at time t. Traditional methodologies begin by proposing a return-generating process for the excess returns, typically one that provides good explanatory power on the excess returns. For example, one may propose that excess returns are generated by a one-factor model

$$r_t = \alpha + \beta f_t + \varepsilon_t, \tag{1}$$

where f_t is the realized value of a systematic risk factor at time t, $\boldsymbol{\varepsilon}_t$ is the idiosyncratic risk of the assets with $E[\boldsymbol{\varepsilon}_t | f_t, \boldsymbol{\Phi}_{t-1}] = \mathbf{0}_N$ and $\text{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma}$, where $\mathbf{0}_N$ is an N-vector of zeros, $\boldsymbol{\Phi}_{t-1}$ is the information set at $t-1$, and $\boldsymbol{\beta} = \text{Cov}[\boldsymbol{r}_t, f_t | \boldsymbol{\Phi}_{t-1}] / \text{Var}[f_t | \boldsymbol{\Phi}_{t-1}]$ is the factor loadings of the returns with respect to the common factor. Since only unexpected shocks matter for unexpected returns, f_t can be modeled as a martingale difference sequence; that is, $E[f_t | \boldsymbol{\Phi}_{t-1}] = 0$. Under these assumptions, $\boldsymbol{\alpha} = E[\boldsymbol{r}_t | \boldsymbol{\Phi}_{t-1}]$ is the expected excess returns on the N assets. In the rest of the paper, the trivial case $\boldsymbol{\alpha} = \mathbf{0}_N$ is precluded. In general, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be functions of information variables at $t-1$, but for the purpose of simplifying technical details and focusing on the main point of this paper, we assume they are constants. Nevertheless, we do not assume $\text{Var}[\boldsymbol{\varepsilon}_t | \boldsymbol{\Phi}_{t-1}, f_t] = \boldsymbol{\Sigma}$, so conditional heteroskedasticity in $\boldsymbol{\varepsilon}_t$ is allowed in our setup.

A beta pricing model, in the exact form, suggests that the expected excess return of an asset is a linear function of its betas with respect to the systematic factors. In our one-factor case, the beta pricing model suggests

$$\boldsymbol{\alpha} = \boldsymbol{\beta}\lambda, \tag{2}$$

where λ is the risk premium. This clearly imposes a testable restriction on the parameters of the return-generating process in equation (1). Traditional tests of beta pricing model are basically done by carrying out various statistical tests of this restriction.

There are many alternatives to estimate the risk premium λ and test the beta pricing model. We describe two representative approaches here. If one is willing to make distributional assumptions on $\boldsymbol{\varepsilon}_t$, one can use the maximum likelihood approach. A popular choice is to assume conditional on f_t, $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma})$. Following Zhou (1991, 1995), we define $\boldsymbol{f} = [f_1, f_2, \ldots, f_T]'$, $\boldsymbol{X} = [\mathbf{1}_T, \boldsymbol{f}]$, $\boldsymbol{Y} = [\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_T]'$, where T is the number of time series observations, and $\mathbf{1}_T$ is a T-vector of ones. Let $\xi_1 \geq \xi_2 > 0$ be the two eigenvalues of

$$\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y})(\boldsymbol{Y}'\boldsymbol{Y})^{-1}(\boldsymbol{Y}'\boldsymbol{X}). \tag{3}$$

Under the normality assumption, the maximum likelihood estimator of λ is given by

$$\hat{\lambda}_{ML} = \frac{a_{12}}{\xi_1 - a_{11}}, \tag{4}$$

where a_{ij} are the (i,j)th elements of $\boldsymbol{A}$. The likelihood ratio test (with the Bartlett correction) of equation (2) is[1]

$$\text{LRT} = -\left(T - \frac{N+3}{2}\right)\log(1 - \xi_2) \overset{A}{\sim} \chi^2_{N-1}, \tag{5}$$

where $\overset{A}{\sim}$ means an asymptotic distribution.

---

[1] Based on simulation evidence, Zhou (1995) shows that the Bartlett correction can improve the small sample properties of the likelihood ratio test. The exact small sample distribution of the likelihood ratio test is also available from Zhou (1991, 1995).

If one does not wish to make any strong distributional assumptions on ε_t, then an alternative approach is to use the generalized method of moments (GMM) of Hansen (1982) to estimate the parameters and test the beta pricing model. Following, for example, MacKinlay and Richardson (1991) or Harvey and Zhou (1993), the GMM test of equation (2) uses the following moment conditions:

$$E[\boldsymbol{\varepsilon}_t] = E[\boldsymbol{r}_t - \boldsymbol{\alpha} - \boldsymbol{\beta}f_t] = E[\boldsymbol{r}_t - \boldsymbol{\beta}\lambda - \boldsymbol{\beta}f_t] = \boldsymbol{0}_N, \tag{6}$$

$$E[\boldsymbol{\varepsilon}_t f_t] = E[(\boldsymbol{r}_t - \boldsymbol{\alpha} - \boldsymbol{\beta}f_t)f_t] = E[(\boldsymbol{r}_t - \boldsymbol{\beta}\lambda - \boldsymbol{\beta}f_t)f_t] = \boldsymbol{0}_N. \tag{7}$$

To apply the GMM methodology, we define the sample moments as

$$\boldsymbol{g}_{1T}(\lambda, \boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{z}_t \otimes (\boldsymbol{r}_t - \boldsymbol{\beta}\lambda - \boldsymbol{\beta}f_t)], \tag{8}$$

where $\boldsymbol{z}_t = [1, f_t]'$. We assume f_t and $\boldsymbol{\varepsilon}_t$ are jointly stationary and ergodic with finite fourth moments, and under the true parameters,

$$\sqrt{T}\boldsymbol{g}_{1T} \overset{A}{\sim} N(\boldsymbol{0}_{2N}, \boldsymbol{S}_1), \tag{9}$$

where $\boldsymbol{S}_1$ is a $2N \times 2N$ positive definite constant matrix. This condition is much weaker than those assumed in other methods of testing asset pricing models. It allows for a variety of forms of autocorrelation and heteroskedasticity in $\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t$. In the GMM methodology, the estimators of the true parameters λ and $\boldsymbol{\beta}$ of the one-factor model, $\hat{\lambda}^*$ and $\hat{\boldsymbol{\beta}}^*$, are given by the solution of the following minimization problem,

$$\min_{\lambda, \boldsymbol{\beta}} \boldsymbol{g}_{1T}(\lambda, \boldsymbol{\beta})' \boldsymbol{W}_{1T} \boldsymbol{g}_{1T}(\lambda, \boldsymbol{\beta}), \tag{10}$$

where $\boldsymbol{W}_{1T}$ is a (possibly stochastic) $2N \times 2N$ positive definite weighting matrix with a limit $\boldsymbol{W}$ that is positive definite and nonstochastic. The standard approach is to choose an optimal weighting matrix equal to a consistent estimate of $\boldsymbol{S}_1^{-1}$.[2] Although there are $N + 1$ parameters in the beta pricing model and the optimization problem is a nonlinear one, it does not present as a serious problem to the estimation because, conditional on a given value of λ, the objective function is linear in $\boldsymbol{\beta}$ and the minimization problem can

---

[2] When the optimal weighting matrix depends on parameters, an iterative method has to be used. In the first round, a positive definite matrix, say, the identity matrix, is used as the weighting matrix to estimate the parameters. In the second round, the model is reestimated using the optimal weighting matrix based on the estimated parameters from the first round.

be solved analytically. As a result, the estimation problem can be written as a function of λ alone and a simple line search can be used to find the optimal $\hat{\lambda}^*$.[3]

A test of the traditional beta pricing model $\boldsymbol{\alpha} = \boldsymbol{\beta}\lambda$ can be carried out by using Hansen's (1982) overidentification test. Since we have $2N$ moment conditions and only $N + 1$ parameters, there are $N - 1$ overidentification conditions, and hence

$$J_1 \equiv T\boldsymbol{g}_{1T}(\hat{\lambda}^*,\hat{\boldsymbol{\beta}}^*)'\boldsymbol{W}_{1T}\boldsymbol{g}_{1T}(\hat{\lambda}^*,\hat{\boldsymbol{\beta}}^*) \overset{A}{\sim} \chi^2_{N-1}, \tag{11}$$

where $\boldsymbol{W}_{1T}$ is a consistent estimate of the optimal weighting matrix.[4] However, as Cochrane (1996) and Jagannathan and Wang (1996) suggest, it is sometimes desirable, for good economic reasons, to use a nonoptimal weighting matrix. In this case, J_1 will no longer have a simple chi-square distribution, but rather will be a weighted sum of chi-square distributions. Zhou (1994) provides a simple chi-square GMM test for an arbitrary weighting matrix, which can be used to bypass the difficulty of having to calculate a weighted sum of chi-square distributions. A numerically identical test is also proposed by Cochrane (1996). But an alternative optimal chi-square test can be obtained from the scoring algorithm, as presented by Newey (1985) and analyzed by Zhou (1994).

B. SDF Model

As discussed by Cochrane (1996), the beta pricing model is a special case of the SDF model. Under the SDF model, there exists a random variable m_t, the stochastic discount factor, such that

$$E[\boldsymbol{r}_t m_t] = \boldsymbol{0}_N. \tag{12}$$

When the exact one-factor asset pricing model in equation (2) holds, the stochastic discount factor is given by

$$m_t = \delta_0 - \delta_1 f_t \tag{13}$$

for some constants δ_0 and δ_1. As an econometric model, the parameters in equation (13) are not uniquely defined. If (δ_0, δ_1) satisfies the equation, so does any multiplier of it. Therefore, it is common to normalize the parameters by writing

$$E[\boldsymbol{r}_t(1 - f_t\lambda)] = \boldsymbol{0}_N, \tag{14}$$

---

[3] Details of the optimization are available upon request. For some special weighting matrices, Zhou (1994) even obtains an analytical solution to this optimization problem.

[4] Another way of testing $\boldsymbol{\alpha} = \boldsymbol{\beta}\lambda$ is to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in equations (6) and (7) as a fully specified model and test the nonlinear restriction on the parameters using a Wald test.

where $\lambda = \delta_1/\delta_0$. If $\text{Var}[f_t] = 1$, then $\boldsymbol{\beta} = E[\boldsymbol{r}_t f_t]$ and the λ in equation (14) is exactly the same as the λ in equation (2). For ease of comparison, we assume $\text{Var}[f_t] = 1$ in the following discussion.[5]

Intuitively, equation (14) only relates m_t to the asset returns in terms of covariances, not how they impact on each other. In theory, equation (14) is well established; there are no problems with the asset pricing restrictions at all. It is the empirical studies of equation (14) that give rise to the potential problems pointed out earlier. Current empirical studies in testing the SDF model typically focus on testing equation (14) alone without specifying the data-generating process in equation (1) that $\boldsymbol{r}_t$ follows. We argue in this paper that such a practice leads to serious problems.

Before we move on to discuss the estimation and test methodology of the SDF model, we point out that although equation (14) holds when we have the true systematic factor f_t, there are other factors that also allow equation (14) to hold exactly. We consider two classes of factors that have this property.

   1. *Noisy factor.* Suppose we define

$$g_t = \frac{f_t + n_t}{\sqrt{1 + \sigma_n^2}}, \tag{15}$$

where n_t is a pure measurement error with mean zero and finite variance σ_n^2 and it is uncorrelated with f_t and $\boldsymbol{\varepsilon}_t$.[6] By specifying g_t as the factor in the SDF model, then for

$$\lambda_g = \lambda\sqrt{1 + \sigma_n^2}, \tag{16}$$

we have

$$E[\boldsymbol{r}_t(1 - g_t\lambda_g)] = E\left[\boldsymbol{r}_t\left(1 - \frac{f_t}{\sqrt{1 + \sigma_n^2}}\lambda_g\right)\right] - \frac{E[\boldsymbol{r}_t n_t]\lambda_g}{\sqrt{1 + \sigma_n^2}}$$

$$= E[\boldsymbol{r}_t(1 - f_t\lambda)] = \boldsymbol{0}_N. \tag{17}$$

Therefore, the noisy factor g_t does the same job as the true factor f_t in pricing the assets. That pure measurement error does not affect the linear pricing relation is well known in the literature. It is discussed, for example, in Breeden, Gibbons, and Litzenberger (1989), and Co-

---

[5] In practice, standardizing macroeconomic factors is a nontrivial issue. The correct approach is to explicitly model their conditional distribution as in Cochrane (1996) and He et al. (1996), and include their estimation as part of the moment conditions. We ignore this issue here in order not to distract from the discussion of the main issue.

[6] The limiting case of $\sigma_n^2 \to \infty$ (i.e., $g_t = n_t/\sigma_n$) is the case that g_t is a useless factor, which is studied by Kan and Zhang (1999a, 1999b).

chrane (1996). Although the linear pricing relation is retained, the risk premium for the noisy factor is higher than that for the true factor. In fact, from equation (16), we can see that the noisier the factor, the higher is its risk premium. One may like to think that when σ_n^2 is large, the SDF model that uses the noisy factor is more likely to be rejected in finite samples than the one with the true factor. We will show with simulation that this view cannot be justified.

2. *Unsystematic factor.* We define

$$h_t = \frac{\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t}{\sqrt{\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}}, \tag{18}$$

and h_t is a linear combination of $\boldsymbol{\varepsilon}_t$. Therefore, h_t has mean zero and it is uncorrelated with f_t. By specifying h_t as the factor in the SDF model, then for

$$\lambda_h = \lambda \sqrt{\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}} \tag{19}$$

we have

$$E[\boldsymbol{r}_t(1 - h_t \lambda_h)] = E[\boldsymbol{r}_t] - \frac{E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}} \lambda_h$$

$$= \boldsymbol{\alpha} - \boldsymbol{\beta}\lambda = \mathbf{0}_N, \tag{20}$$

and h_t prices the N assets perfectly. Although h_t is an unsystematic factor by construction, we will still be tempted to conclude that it is "priced."

The fact that these two classes of "wrong" factors can satisfy equation (14) suggests the danger of attaching economic meaning to the test outcome of an SDF model. When one specifies a set of macroeconomic factors and finds that it satisfies equation (14), one really cannot tell whether it is the true factor f_t, the noisy factor g_t, or if it is just an unsystematic factor h_t. It should be pointed out that if g_t or h_t is proposed as the factor in the data-generating process, it is also difficult for the traditional methodologies to detect these "wrong" factors. However, because g_t and h_t typically do not possess good explanatory power on the returns of the test assets (especially when σ_n^2 is large and the test assets are well-diversified portfolios), they are less likely to be included as the systematic factors under the traditional methodologies. In contrast, the SDF methodology does not pay any attention to the return-generating process, hence g_t and h_t could easily be proposed and be mistaken as the "true" systematic factors.

Recognizing that there are countless SDFs that represent countless asset pricing models for a given set of asset returns, Hansen and Jagannathan (1991) solve explicitly the SDF that has the minimum variance among all the SDFs. Hansen and Jagannathan (1997) further show how to use SDFs to assess specification errors of asset pricing models. What we have shown here is that there are in fact many SDFs for a given linear factor model. Therefore, explicitly constructed "wrong" factors can potentially help to explain the failure of an asset pricing model. This highlights the danger of using factors in the SDF framework without a careful examination of the explanatory power of the factors.

C. GMM Estimation and Test of SDF Models

In estimating parameters and testing pricing restrictions of equation (14), the GMM is used almost exclusively. For illustrative purposes, we assume, as we did earlier for the traditional methodologies, that the model is estimated and tested without using the information/instrumental variables at $t-1$. The test of this simple form amounts to the so-called "unconditional test of the unconditional model" defined in Cochrane (1996). Let $\boldsymbol{u}_t = \boldsymbol{r}_t(1 - f_t \lambda)$ and $\boldsymbol{g}_{2T} = (1/T)\sum_{t=1}^{T} \boldsymbol{u}_t$. We assume under the true parameter that

$$\sqrt{T}\boldsymbol{g}_{2T} \stackrel{A}{\sim} N(\boldsymbol{0}_N, \boldsymbol{S}_2) \tag{21}$$

for some positive definite constant matrix $\boldsymbol{S}_2$. The true parameter λ is estimated by

$$\hat{\lambda} = \mathrm{argmin}_\lambda \boldsymbol{g}_{2T}(\lambda)' \boldsymbol{W}_{2T} \boldsymbol{g}_{2T}(\lambda), \tag{22}$$

where $\boldsymbol{W}_{2T}$ is typically a consistent estimate of $\boldsymbol{S}_2^{-1}$. The GMM estimation of the SDF model is very simple to implement because there is only one parameter, λ, to be estimated, and it can be analytically obtained as

$$\hat{\lambda} = (\boldsymbol{D}_{2T}' \boldsymbol{W}_{2T} \boldsymbol{D}_{2T})^{-1}(\boldsymbol{D}_{2T}' \boldsymbol{W}_{2T} \bar{\boldsymbol{r}}_T), \tag{23}$$

where $\boldsymbol{D}_{2T} = (1/T)\sum_{t=1}^{T} \boldsymbol{r}_t f_t$ and $\bar{\boldsymbol{r}}_T = (1/T)\sum_{t=1}^{T} \boldsymbol{r}_t$.

A test of the SDF model in equation (14) is usually carried out by using Hansen's (1982) overidentification test. Since we have N moment conditions and only one parameter, there are $N-1$ overidentification conditions; hence

$$J_2 \equiv T\boldsymbol{g}_{2T}(\hat{\lambda})' \boldsymbol{W}_{2T} \boldsymbol{g}_{2T}(\hat{\lambda}) \stackrel{A}{\sim} \chi^2_{N-1}, \tag{24}$$

when $\boldsymbol{W}_{2T}$ is a consistent estimate of the optimal weighting matrix $\boldsymbol{S}_2^{-1}$.

Therefore, if the beta pricing model is correct, both J_1 in equation (11) and J_2 have an asymptotic chi-square distribution and there are no strong reasons to prefer one test over the other. However, in finite samples, their per-

formance could differ. More importantly, when the model is misspecified, J_1 and J_2 could have very different power. We study these issues by simulation in Section III.

Although the estimation problem of the SDF methodology is very simple, ignoring the full dynamics of asset returns introduces serious problems. Intuitively, equation (14) is a restriction on part of the first and second moments between the asset returns and the factor. Testing equation (14) alone without using a fully specified model amounts to ignoring many other first and second moments entirely. As a result, it is not surprising that the estimation error of λ can be substantially large. It is also not surprising that a tested factor can be important in equation (14), but in fact may have little to do with the returns. This is the fundamental reason that causes the problems emphasized by this paper. In the following sections, we provide a comparison of the traditional methodologies with the SDF methodology in terms of the estimation accuracy of risk premium, and in terms of the size and the power of their tests of the asset pricing model.

II. Estimation Accuracy of Risk Premium

In this section, we demonstrate in two ways that there can be substantial loss of efficiency in estimating λ by using the SDF methodology. First, we provide theoretical results to show that the asymptotic variance of the estimated λ in the SDF methodology is greater than the variances of the traditional methodologies. Second, we provide Monte Carlo simulations to further illustrate that the standard error of the estimated λ in the SDF methodology is indeed very large (in small samples), and may not be reliable in applications. In contrast, the estimated λ for the traditional methodologies is very accurate even in small samples, making it better suited for estimating risk premia.

The consistency of $\hat{\lambda}_{ML}$, $\hat{\lambda}^*$, and $\hat{\lambda}$ is well known; that is, as sample size T increases, they all approach the true parameter λ. At a given finite sample size T, however, there will be an estimation error. In assessing the accuracy of $\hat{\lambda}$ in the SDF methodology with that of $\hat{\lambda}_{ML}$ and $\hat{\lambda}^*$ in the traditional methodologies, we can compare their asymptotic variances. The following proposition shows that $\hat{\lambda}^*$ is asymptotically more accurate than $\hat{\lambda}$ and it has the same efficiency as $\hat{\lambda}_{ML}$ under the normality assumption.[7]

PROPOSITION 1: *Suppose f_t is the true factor and it has a continuous distribution. We have*

$$\text{Avar}[\hat{\lambda}^*] < \text{Avar}[\hat{\lambda}]. \tag{25}$$

[7] Proposition 1 can be extended to the multifactor case to show that the vector of estimated risk premium is more accurate under the traditional methodologies than the SDF methodology. Results are available upon request.

For the case that $\boldsymbol{\varepsilon}_t$ has a multivariate normal distribution conditional on f_t, we have

$$\text{Avar}[\hat{\lambda}_{ML}] = \text{Avar}[\hat{\lambda}^*] = \frac{1 + \lambda^2}{\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}. \qquad (26)$$

Proposition 1 suggests that regardless of the distribution that $\boldsymbol{\varepsilon}_t$ follows, with or without conditional heteroskedasticity, traditional methodologies that incorporate the return-generating process will always provide an estimated risk premium that is asymptotically more efficient than that from the SDF methodology. One may like to think that the reason we achieve higher accuracy in $\hat{\lambda}^*$ is that we use more moment conditions than the SDF methodology. However, this is not the main reason for the improvement because, although we do have more moment conditions in the traditional methodology, we also have more parameters to estimate.

There are two main reasons why the full GMM estimator $\hat{\lambda}^*$ in the traditional methodology is more efficient than the SDF estimator $\hat{\lambda}$. The first reason is that the full GMM uses $\hat{\boldsymbol{\beta}}^*$ to explain average excess returns, whereas the SDF methodology uses $\boldsymbol{D}_{2T} = (1/T)\sum_{t=1}^{T} \boldsymbol{r}_t f_t$ to explain average excess returns. Both $\hat{\boldsymbol{\beta}}^*$ and $\boldsymbol{D}_{2T}$ are consistent estimates of $\boldsymbol{\beta}$ when $\text{Var}[f_t] = 1$; however, in general, $\hat{\boldsymbol{\beta}}^*$ is a more accurate estimator of $\boldsymbol{\beta}$ than $\boldsymbol{D}_{2T}$. For example, under the multivariate normality assumption on f_t and $\boldsymbol{\varepsilon}_t$, it can be shown that $\text{Avar}[\boldsymbol{D}_{2T}] = \boldsymbol{\Sigma} + \boldsymbol{\beta}\boldsymbol{\beta}'$, which is much larger than

$$\text{Avar}[\hat{\boldsymbol{\beta}}^*] = \frac{1}{1 + \lambda^2}\left[\boldsymbol{\Sigma} + \frac{\lambda^2 \boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}\right] < \boldsymbol{\Sigma}.^8$$

In the traditional methodology, the moment conditions in equation (7) and the restriction $\boldsymbol{\alpha} = \boldsymbol{\beta}\lambda$ allow us to obtain an estimate of $\boldsymbol{\beta}$ with high degree of accuracy. In contrast, the SDF methodology abandons the more accurate beta estimation and only relates the average returns to the average covariances.

The second reason the full GMM estimator $\hat{\lambda}^*$ in the traditional methodology is more efficient than the SDF estimator $\hat{\lambda}$ is that the realized return is a very noisy measure of expected return. The traditional methodology makes use of the factor structure of the return-generating process by taking away the systematic component $\boldsymbol{\beta}f_t$ from $\boldsymbol{r}_t$ in the moment conditions. When $\boldsymbol{\beta}f_t$ accounts for a significant portion of the variations of $\boldsymbol{r}_t$, then $\boldsymbol{r}_t - \boldsymbol{\beta}f_t$ is

---

[8] The expression for $\text{Avar}[\hat{\boldsymbol{\beta}}^*]$ can be obtained from the proof of Proposition 1. The inequality follows because

$$\frac{1}{1 + \lambda^2}\left[\boldsymbol{\Sigma} + \frac{\lambda^2 \boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}\right] = \boldsymbol{\Sigma} - \frac{\lambda^2}{1 + \lambda^2}\boldsymbol{\Sigma}^{1/2}[\boldsymbol{I}_N - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta}(\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1/2}]\boldsymbol{\Sigma}^{1/2}$$

and the second term is a positive semidefinite matrix.

a much less noisy measure of expected return than r_t. The SDF methodology, however, does not incorporate the return-generating process in its moment conditions and only relates realized excess returns r_t to realized covariance $r_t f_t$. When both of these two measures are very noisy, it is not surprising that the SDF methodology does not deliver a very accurate estimate of the risk premium.

The above analysis shows that the traditional methodology, which utilizes the fully specified asset return model, helps to substantially improve the estimation accuracy of λ. As a result, it may be tempting to estimate λ by using all of the moment conditions, those of the traditional ones in equations (6) and (7), and those of the SDF ones in equation (14). It turns out that there is some overlap between these moment conditions. Out of the $3N$ moment conditions, $N - 1$ of them are redundant. For example, if we know ε_t, $\varepsilon_t f_t$, and $u_{1t} = r_{1t}(1 - f_t \lambda)$ where $\beta_1 \neq 0$, then we can obtain the other elements of u_t by using the relation

$$\varepsilon_{it} - \lambda(\varepsilon_{it} f_t) + \frac{\beta_i}{\beta_1}[u_{1t} - \varepsilon_{1t} + \lambda(\varepsilon_{1t} f_t)]$$

$$= [r_{it} - \beta_i(\lambda + f_t)](1 - f_t \lambda) + \frac{\beta_i}{\beta_1}\beta_1(\lambda + f_t)(1 - f_t \lambda)$$

$$= r_{it}(1 - f_t \lambda)$$

$$= u_{it}, \quad \text{for } i = 2, \ldots, N. \tag{27}$$

Therefore, one can use at most $2N + 1$ moment conditions to estimate λ. Denote λ^{**} and $\boldsymbol{\beta}^{**}$ as the estimator of λ and $\boldsymbol{\beta}$ using any $2N + 1$ of the combined $3N$ moment conditions. The following proposition suggests that once the moment conditions in the traditional methodology are used, the additional one from the SDF model does not help to improve the accuracy of the estimation.

PROPOSITION 2: *Suppose f_t is the true factor. We have*

$$\text{Avar}\begin{bmatrix} \hat{\lambda}^* \\ \hat{\boldsymbol{\beta}}^* \end{bmatrix} = \text{Avar}\begin{bmatrix} \hat{\lambda}^{**} \\ \hat{\boldsymbol{\beta}}^{**} \end{bmatrix}. \tag{28}$$

Note that Proposition 2 does not suggest that any $2N$ out of the combined $3N$ moment conditions will do the same job as the traditional methodology. For example, if we combine the N moment conditions in equation (6) (or the N moment conditions in equation (7)) with the N moment conditions in equation (14) of the SDF methodology to estimate λ, it can be shown that the asymptotic variance of the estimated λ using these $2N$ moment conditions is still the same as that of $\hat{\lambda}$ from the SDF methodology. Therefore, it is important to choose the proper set of moment conditions to obtain a good estimate of λ. Proposition 2 suggests the moment conditions used by the

traditional methodology are the best and they are sufficient to learn almost everything about the parameters. Adding the SDF moment conditions into the traditional ones provides only redundant information.[9]

Although Proposition 1 suggests that the estimated risk premium in the traditional methodologies is asymptotically more accurate than that of the SDF methodology, it does not tell us the magnitude of improvement, nor does it tell us whether this result holds in finite samples. We address these issues by simulation. The setup of our simulation experiment is as follows. In our simulation, we generate excess returns on 10 assets using a one-factor model. The factor is generated independently from a standard normal distribution and it is designed to capture the behavior of the standardized excess return on the value-weighted market portfolio of the NYSE; that is,

$$f_t = \frac{r_{mt} - E[r_{mt}]}{\sigma_m} \sim N(0,1), \tag{29}$$

where r_{mt} is the excess return on the market portfolio and σ_m is its standard deviation. The betas of the 10 assets are set to equal the sample betas of the 10 size-ranked portfolios of the NYSE with respect to f_t, estimated using monthly returns from January 1926 to December 1997. The true risk premium is chosen to make the expected excess returns close to the average excess returns of the 10 size portfolios over the sample period; that is,

$$\lambda = \operatorname{argmin}_\lambda (\bar{r} - \boldsymbol{\beta}\lambda)'(\bar{r} - \boldsymbol{\beta}\lambda), \tag{30}$$

where $\bar{r}$ and $\boldsymbol{\beta}$ are the average returns and sample betas of the 10 size-ranked portfolios. Finally, the model disturbances are independently generated from a multivariate normal distribution

$$\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma}), \tag{31}$$

where $\boldsymbol{\Sigma}$ is chosen to be the sample covariance matrix of the market model residuals of the 10 size-ranked portfolios. In Table I, Panel A, we present the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and λ of the 10 assets we use in our simulation. Note that the value we choose for λ (0.1373) is very close to the sample Sharpe ratio for the value-weighted market portfolio of the NYSE, which is equal to 0.1248 for the period January 1926 to December 1997.[10]

We generate returns from this one-factor model for different lengths of time series and apply the traditional and the SDF methodologies to estimate the risk premium λ. In Panel B of Table I, we present a summary of the

---

[9] Similar to Proposition 1, Proposition 2 continues to hold for the multifactor case. When there are k-factors, only k of the SDF moment conditions can be added to the traditional moment conditions, and they do not improve the estimation accuracy of the risk premium and the betas. Results are available upon request.

[10] Under our definition of f_t, λ is equal to the Sharpe ratio of the market portfolio if the CAPM holds.

estimation results in 10,000 simulations. For the traditional methodologies, we report the average and standard deviation of the estimated risk premium using the maximum likelihood approach and the GMM approach.[11] Both the maximum likelihood approach and the GMM approach that uses the traditional moment conditions produce very reliable estimates of risk premium. Their estimated risk premia are almost unbiased and they are tightly distributed around the true λ. Although Proposition 1 suggests that under normality assumption on $\boldsymbol{\varepsilon}_t$, we have $\text{Avar}[\hat{\lambda}_{ML}] = \text{Avar}[\hat{\lambda}^*]$, the maximum likelihood estimator is better behaved than the full GMM estimator when the sample size T is small.

The last two columns of Panel B report the average and standard deviation of $\hat{\lambda}$, the estimated risk premium from the SDF methodology. The difference between the performance of the estimation risk premium in the SDF and the traditional methodologies is striking. The estimated risk premium using the SDF methodology is biased and volatile. For example, when $T = 120$, the average $\hat{\lambda}$ in our 10,000 simulations is 0.1497, quite far from the true value of $\lambda = 0.1373$. Furthermore, the standard deviation of $\hat{\lambda}$ is 0.1049, so the estimated risk premium from the SDF methodology could easily be negative. Although the bias and the standard deviation of $\hat{\lambda}$ reduce as T increases, $\hat{\lambda}$ is still volatile for T as large as 720. On average, the standard deviation of the estimated risk premium under the SDF methodology is more than 40 times larger than that of the traditional methodologies. Therefore, for the purpose of estimating risk premium, the traditional methodologies are much better suited for the job than the SDF methodology.

Before we move on to discuss the size and power of the tests in the traditional and the SDF methodologies, we should note that the excess returns in our simulation experiment for Panel B are generated in a way that is most favorable to the maximum likelihood approach. When $\boldsymbol{\varepsilon}_t$ is not normally distributed or its distribution is unknown, the maximum likelihood approach is difficult to apply. However, the results based on the GMM approach remain fairly robust to the distributional assumption on $\boldsymbol{\varepsilon}_t$. So the advantage of using the traditional moment conditions over the SDF moment conditions is still important, even when $\boldsymbol{\varepsilon}_t$ is not normally distributed. To illustrate this, we generate f_t and $\boldsymbol{\varepsilon}_t$ from a multivariate t-distribution with ν degrees of freedom and mean zero. The covariance matrix of f_t and $\boldsymbol{\varepsilon}_t$ stays the same as in the multivariate normal case (i.e., $\text{Var}[f_t] = 1$, $\text{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma}$), and they are uncorrelated with each other. The reason we choose the multivariate t-distribution is that it offers an opportunity for us to investigate the effect of conditional heteroskedasticity on our results. When f_t and $\boldsymbol{\varepsilon}_t$ have a multivariate t-distribution, the conditional variance of $\boldsymbol{\varepsilon}_t$ depends on f_t. More specifically, when $\nu > 2$, we have

$$\text{Var}[\boldsymbol{\varepsilon}_t | f_t] = \left(\frac{\nu - 2 + f_t^2}{\nu - 1} \right) \boldsymbol{\Sigma} \tag{32}$$

[11] The GMM estimation results are based on the second stage GMM with the identity matrix as the initial weighting matrix. Simulation results of the third and fourth stage GMM are mostly similar to the ones using the second stage GMM, therefore they are not separately reported.

Table I

Estimation Accuracy of Risk Premium under the Traditional Methodologies and the Stochastic Discount Factor Methodology

The table presents the performance of the estimated risk premium under traditional methodologies and the stochastic discount factor (SDF) methodology. Excess returns on 10 assets are simulated using a one-factor model

$$r_t = \alpha + \beta f_t + \varepsilon_t,$$

where the values of $\alpha = \beta\lambda$ (in percentage per month) and β are presented in Panel A. The parameters are chosen to mimic the returns on 10 size-ranked portfolios of the NYSE. The factor and the model disturbance are generated as $f_t \sim N(0,1)$ and $\varepsilon_t \sim N(\mathbf{0}_N, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is set to equal the sample covariance matrix of the market model residuals of the 10 size-ranked portfolios of the NYSE, estimated using monthly returns over the period January 1926 to December 1997. The estimation results of 10,000 simulations are reported in Panel B. For each length of time-series observations, T, we present the average and standard deviation of the estimated risk premium from the maximum likelihood method and the (second stage) GMM method using the traditional moment conditions and the SDF moment conditions. Panel C reports the same results as in Panel B but for the cases that (f_t, ε_t) are generated from a multivariate t-distribution with 5 and 10 degrees of freedom.

Panel A: Parameters of the One-Factor Pricing Model

Size-Ranked Portfolios

	1	2	3	4	5	6	7	8	9	10
α	1.129	1.070	0.993	0.954	0.923	0.911	0.875	0.833	0.810	0.709
β	0.082	0.078	0.072	0.069	0.067	0.066	0.064	0.061	0.059	0.052

$\lambda = 0.1373$

Panel B: Distribution of Estimated Risk Premium under Multivariate Normality Assumption

	Traditional Methodologies				SDF Methodology ($\hat{\lambda}$)	
	Maximum Likelihood ($\hat{\lambda}_{ML}$)		GMM ($\hat{\lambda}^*$)			
T	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
120	0.1374	0.0020	0.1375	0.0045	0.1497	0.1049
240	0.1373	0.0014	0.1374	0.0020	0.1438	0.0689
360	0.1373	0.0011	0.1373	0.0014	0.1417	0.0553
480	0.1373	0.0010	0.1373	0.0011	0.1404	0.0477
600	0.1373	0.0009	0.1373	0.0009	0.1399	0.0426
720	0.1373	0.0008	0.1373	0.0008	0.1396	0.0386

Panel C: Distribution of Estimated Risk Premium under Multivariate t-distribution Assumption

	Traditional Methodology ($\hat{\lambda}^*$)		SDF Methodology ($\hat{\lambda}$)	
T	Average	Standard Deviation	Average	Standard Deviation
5 degrees of freedom				
120	0.1374	0.0046	0.1611	0.1162
240	0.1374	0.0020	0.1513	0.0764
360	0.1374	0.0014	0.1480	0.0612
480	0.1374	0.0011	0.1456	0.0513
600	0.1374	0.0009	0.1440	0.0454
720	0.1373	0.0008	0.1432	0.0411

Table 1—*Continued*

	Traditional Methodology ($\hat{\lambda}^*$)			SDF Methodology ($\hat{\lambda}$)	
T	Average	Standard Deviation		Average	Standard Deviation
10 degrees of freedom					
120	0.1374	0.0045		0.1511	0.1088
240	0.1373	0.0020		0.1448	0.0713
360	0.1373	0.0014		0.1422	0.0567
480	0.1373	0.0011		0.1409	0.0486
600	0.1373	0.0009		0.1401	0.0431
720	0.1373	0.0008		0.1395	0.0394

and the conditional variance of $\boldsymbol{\varepsilon}_t$ is higher when the absolute value of f_t is large. In Panel C, we report the simulation results in 10,000 simulations for the case that f_t and $\boldsymbol{\varepsilon}_t$ are generated from a multivariate t-distribution with five degrees of freedom, and also for the case of 10 degrees of freedom. For the GMM estimated risk premium using the traditional moment conditions $\hat{\lambda}^*$, the results do not change much from those in Panel B; $\hat{\lambda}^*$ continues to be very accurate even in the presence of nonnormality and conditional heteroskedasticity. As for the GMM estimated risk premium using the SDF moment conditions, $\hat{\lambda}$ continues to be an unreliable estimator of λ. Therefore, the SDF methodology does not outperform the traditional methodology even when $\boldsymbol{\varepsilon}_t$ exhibits conditional heteroskedasticity. In fact, compared with the results in Panel B, we can see that both the bias and the standard deviation of $\hat{\lambda}$ are higher for the case of multivariate t-distribution, making the SDF methodology even less suitable for the purpose of estimating the risk premium in this case.

III. Size and Power of Overidentification Tests

Unlike the case of risk premium estimation where we can show that the traditional methodologies are superior, it is not entirely clear whether the traditional methodologies or the SDF methodology is better suited to test the asset pricing restriction $\boldsymbol{\alpha} = \boldsymbol{\beta}\lambda$. Both methodologies provide tests that have an asymptotic distribution of χ^2_{N-1} when the model is correct, and an asymptotic probability of 1 in rejecting the model when it is wrong. The real issue here is about their respective performance in finite samples. In this section, we rely on simulation evidence to assess whether these tests have the correct size in small samples and whether they have power in rejecting misspecified models.

To assess the size of the likelihood ratio test, LRT, and the two overidentification tests J_1 and J_2, we generate excess returns from a one-factor model as before.[12] We then compute LRT and J_1 of the traditional methodologies,

[12] Simulation results for the multivariate t-distribution are qualitatively similar to the case of multivariate normal distribution; therefore we do not separately report the results for the multivariate t-distribution case in this section.

and J_2 of the SDF methodology for three different models. In the first model, we use the true factor f_t to construct the sample moments and the test statistics. In the second model, we use a noisy factor $g_t = (f_t + n_t)/\sqrt{5}$ instead of f_t to compute the sample moment and the test statistics, where n_t is a measurement error that is generated from a normal distribution with mean 0 and variance 4. In the final model, we specify the unsystematic factor $h_t = \boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t/\sqrt{\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}$ as the true factor to perform the test. Although economically these three factors are very different, statistically they are all considered to be correctly specified models.[13] Therefore, asymptotically, all three tests should have an asymptotic distribution of χ^2_{N-1} for the three correctly specified models.

In Table II, we report the rejection rates of the LRT, J_1, and J_2 for the three models at the 10 percent, 5 percent, and 1 percent significance levels based on the χ^2_{N-1} distribution. For the case of the true factor, we observe in Table II that the probability of rejection in finite samples is very close to the size of the test for all three tests. This indicates that using the asymptotic distribution is a very good approximation when we have the true factor in the model. For the case of the noisy factor, the probability of rejection is typically less than the size of the test, especially when T is small. In this case, the performance of the three tests is roughly the same in small samples. For the case of unsystematic factors, the finite sample distribution of all three tests differs greatly from the asymptotic distribution of χ^2_{N-1} and all three tests underreject the null hypothesis. However, the problem of underrejection for J_2 is more serious than that of LRT and J_1. In summary, when the asymptotic distribution is used to make the acceptance and rejection decision, the traditional methodologies seem to do no worse than J_2 of the SDF methodology when we have the correctly specified model. However, when the proposed factor does not explain the returns well, we have to be more cautious in using the asymptotic distribution of the tests to make the acceptance and rejection decision.

Likelihood ratio tests and GMM overidentification tests are designed to detect misspecified models, so the major concern is on their power. Misspecification can take various forms; we focus here on the case in which there is a missing factor in the proposed model. In this case, the expected return of the assets is not a linear function of the beta of the proposed factor; that is, there does not exist a λ such that $\boldsymbol{\alpha} = \boldsymbol{\beta}\lambda$. To study the power of the tests, we simulate returns using a two-factor model. The two factors are independently generated from a bivariate normal distribution and are designed to capture the behavior of the standardized excess returns on the value-weighted market portfolio of the NYSE and the long-term Treasury bond; that is,

---

[13] This is because the moment conditions in equations (14), (6), and (7) can be satisfied with g_t or h_t, instead of f_t. Although the parameters $\boldsymbol{\beta}$ and λ are different for the three sets of factors, the exact linear pricing relation holds in all three cases.

<div align="center">

Table II

**Size of the Likelihood Ratio Test and the GMM
Overidentification Tests of Traditional and Stochastic
Discount Factor Methodologies**
</div>

The table presents the probability of rejecting three correctly specified models using the likelihood ratio test (LRT) and the (second stage) GMM overidentification tests using the traditional moment conditions and the SDF moment conditions. Excess returns on 10 assets are simulated using a one-factor model

$$r_t = \boldsymbol{\alpha} + \boldsymbol{\beta} f_t + \boldsymbol{\varepsilon}_t,$$

where the values of $\boldsymbol{\alpha} = \boldsymbol{\beta}\lambda$ (in percentage per month) and $\boldsymbol{\beta}$ are presented in Table I. The factor and the model disturbance are generated as $f_t \sim N(0,1)$ and $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is set to equal the sample covariance matrix of the market model residuals of the 10 size-ranked portfolios of the NYSE, estimated using monthly returns over the period January 1926 to December 1997. For each length of time-series observations, T, we present the probability of rejecting three different models at various significance levels in 10,000 simulations. The three models differ in terms of the factor they use. The first model uses the true factor f_t. The second model uses a noisy factor $g_t = (f_t + n_t)/\sqrt{5}$, where n_t is measurement error, distributed as $N(0,4)$. The third model uses an unsystematic factor $h_t = \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t/\sqrt{\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}$.

	True			Noisy			Unsystematic		
	Significance Level			Significance Level			Significance Level		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Maximum Likelihood Method (LRT)									
120	0.099	0.048	0.010	0.082	0.037	0.007	0.013	0.003	0.001
240	0.098	0.047	0.008	0.090	0.042	0.007	0.026	0.008	0.001
360	0.098	0.048	0.010	0.093	0.043	0.008	0.035	0.012	0.001
480	0.098	0.048	0.009	0.090	0.046	0.007	0.040	0.015	0.002
600	0.099	0.049	0.011	0.092	0.047	0.009	0.052	0.022	0.002
720	0.101	0.050	0.012	0.098	0.048	0.011	0.059	0.024	0.004
Panel B: GMM Using Traditional Moment Conditions (J_1)									
120	0.093	0.043	0.007	0.076	0.032	0.006	0.017	0.004	0.000
240	0.097	0.046	0.009	0.088	0.042	0.008	0.025	0.007	0.000
360	0.099	0.049	0.010	0.095	0.045	0.008	0.031	0.010	0.000
480	0.099	0.049	0.010	0.095	0.048	0.009	0.041	0.016	0.001
600	0.097	0.046	0.010	0.096	0.046	0.008	0.048	0.017	0.002
720	0.102	0.048	0.010	0.098	0.047	0.009	0.053	0.020	0.002
Panel C: Stochastic Discount Factor Methodology (J_2)									
120	0.097	0.046	0.007	0.079	0.037	0.006	0.017	0.008	0.001
240	0.098	0.047	0.009	0.086	0.041	0.007	0.013	0.006	0.001
360	0.102	0.048	0.008	0.092	0.043	0.008	0.010	0.006	0.001
480	0.101	0.048	0.009	0.094	0.043	0.007	0.012	0.007	0.002
600	0.101	0.051	0.011	0.093	0.046	0.010	0.012	0.007	0.003
720	0.104	0.051	0.009	0.098	0.047	0.010	0.011	0.008	0.004

$$f_{1t} = \frac{r_{mt} - E[r_{mt}]}{\sigma_m} \sim N(0,1), \tag{33}$$

$$f_{2t} = \frac{r_{bt} - E[r_{bt}]}{\sigma_b} \sim N(0,1), \tag{34}$$

$$\text{Cov}[\, f_{1t}, f_{2t}\,] = 0.2, \tag{35}$$

where r_{mt} and r_{bt} are the excess returns on the market portfolio and the long-term Treasury bond, and σ_m and σ_b are their standard deviations. The betas of the 10 assets are set to equal the sample betas of the 10 size-ranked portfolios of the NYSE with respect to f_{1t} and f_{2t}, estimated using monthly returns from January 1926 to December 1997. The true risk premia of the two factors are chosen so that the expected excess returns are close to the average excess returns on the 10 size portfolios over the sample period; that is,

$$(\lambda_1, \lambda_2) = \text{argmin}_{\lambda_1, \lambda_2} (\bar{r} - \boldsymbol{\beta}_1 \lambda_1 - \boldsymbol{\beta}_2 \lambda_2)'(\bar{r} - \boldsymbol{\beta}_1 \lambda_1 - \boldsymbol{\beta}_2 \lambda_2), \tag{36}$$

where $\bar{r}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the average returns and sample betas of the 10 size-ranked portfolios. Finally, the model disturbances are independently generated from a multivariate normal distribution

$$\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma}), \tag{37}$$

where $\boldsymbol{\Sigma}$ is chosen to be the sample covariance matrix of the residuals of the 10 size-ranked portfolios in the two-factor model. In Table III, Panel A, we present the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, λ_1, and λ_2 of the 10 assets that we use in our simulation. Under our simulation, the first factor is one that explains a lot of the time-series variations of the excess returns (with an average R^2 of 84.73 percent) and the second factor has a very low explanatory power on the excess returns (with an average R^2 of 2.51 percent). Nevertheless, neither $\boldsymbol{\beta}_1$ nor $\boldsymbol{\beta}_2$ alone can fully explain the expected excess return $\boldsymbol{\alpha}$.

In Table III, we report the rejection rates of LRT, J_1, and J_2 for two misspecified models at the 10 percent, 5 percent, and 1 percent significance levels based on the χ^2_{N-1} distribution. Panel B contains the results when only the first factor is included in the model, Panel C contains the results when only the second factor is included in the model. Since both models are misspecified models, we would like the test to reject them with high probability. For the case of the misspecified model that includes only the first factor, we can observe that all three tests have roughly the same power in rejecting the model. With T as large as 360, we can only reject the misspecified model at the 5 percent level roughly 12 percent of the time, but the power steadily increases as T goes up. This suggests that when the proposed factor has strong explanatory power on the returns and the model misspec-

ification is not serious, there is not much of a difference between the traditional methodologies and the SDF methodology.

Ironically, when the proposed factor in the model is a weak factor, the misspecified model becomes even more difficult to detect for the SDF methodology. This can be seen from our simulation results in Panel C of Table III. In this case, LRT and J_1 have reasonably good power in rejecting this misspecified model. When $T = 360$, these two tests reject the misspecified model at the 5 percent level for approximately 34 percent of the time. However, J_2 of the SDF methodology performs much worse than LRT and J_1. Even for $T = 360$, we still find that J_2 rejects the misspecified model less often than the size of the test, making it almost impossible to reject such a misspecified model. The poor performance of J_2 in finite samples is due to the fact that S_2 is unknown and has to be estimated. When the model is misspecified, the estimated S_2 will tend to be large because of the pricing error, and hence its inverse will be small. Since the inverse of estimated S_2 is used to compute J_2, the test statistic can be very small for grossly misspecified models, especially when the factor does not explain much of the return. Asymptotically, this is not a concern because eventually the pricing errors will dominate as T increases, but in finite samples, using an estimated S_2 makes the overidentification test J_2 very unreliable. Although the same problem also plagues LRT and J_1 of the traditional methodologies, we can see in Panel C that its impact on LRT and J_1 is much less severe. Therefore, if one has to pick a specification test to use, it appears that the ones from the traditional methodologies are superior to the one from the SDF methodology.

We should also note that J_2 of the SDF methodology seems to prefer models with a poor factor to the model with a good factor. This suggests, among other things, the danger of using the *p*-value of the likelihood ratio test or GMM overidentification test to choose models. In this regard, the traditional methodologies are superior because a poor factor is less likely to be proposed to be the only factor in the return-generating process. The SDF methodology does not specify a return-generating process and a poor factor could potentially be chosen as the only factor in the model. As our simulation experiment shows, such poor factors could make the model pass the GMM overidentification test of the SDF methodology easily, even though they do not explain much of the excess returns and their betas do not fully explain the expected excess returns.

As always, simulation evidence cannot be generalized to other scenarios, so our recommendation should be taken with caution. Nevertheless, from our simulation evidence, it does appear that there are compelling reasons to prefer the traditional methodologies to the SDF methodology. A more rigorous analysis of the size and power of these tests would go a long way in settling these issues.

Finally, we remark that even though nonstandard GMM overidentification tests, such as the one suggested by Jagannathan and Wang (1996), do not use the estimated covariance matrix of the sample moments to compute the test statistic, the estimated covariance matrix is still used in computing

Table III

Power of the Likelihood Ratio Test and the GMM Overidentification Tests of Traditional and Stochastic Discount Factor Methodologies

The table presents the probability of rejecting two misspecified models using the likelihood ratio test (LRT) and the (second stage) GMM overidentification tests using the traditional moment conditions and the SDF moment conditions. Excess returns on 10 assets are simulated using a two-factor model

$$r_t = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 f_{1t} + \boldsymbol{\beta}_2 f_{2t} + \boldsymbol{\varepsilon}_t,$$

where the values of $\boldsymbol{\alpha} = \lambda_1 \boldsymbol{\beta}_1 + \lambda_2 \boldsymbol{\beta}_2$ (in percentage per month), and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are presented in Panel A. The factors and the model disturbance are independently generated from a multivariate normal distribution with $f_{1t} \sim N(0,1)$, $f_{2t} \sim N(0,1)$, $\mathrm{Cov}[f_{1t}, f_{2t}] = 0.2$, and $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is set to equal the sample covariance matrix of the residuals of the 10 size-ranked portfolios of the NYSE in a two-factor model. The two factors are the excess return on the value-weighted market portfolio of the NYSE and the excess return on the long-term Treasury bond. The model is estimated using monthly returns over the period January 1926 to December 1997. For each length of time-series observations, T, Panels B and C present the probability of rejecting the two misspecified models at various significance levels in 10,000 simulations. Panel B presents the results for the first misspecified model which includes only the first factor f_{1t} in the model. Panel C presents the results for the second misspecified model which includes only the second factor f_{2t} in the model.

Panel A: Parameters of the Two-factor Pricing Model									
Size-Ranked Portfolios									
1	2	3	4	5	6	7	8	9	10
$\boldsymbol{\alpha}$ 1.280	1.163	1.001	0.940	0.926	0.891	0.836	0.748	0.703	0.628
$\boldsymbol{\beta}_1$ 0.084	0.079	0.073	0.070	0.068	0.067	0.064	0.061	0.059	0.052
$\boldsymbol{\beta}_2$ −0.0083	−0.0063	−0.0035	−0.0027	−0.0031	−0.0024	−0.0017	−0.0001	0.0006	0.0001

$$\lambda_1 = 0.1225, \ \lambda_2 = -0.3047$$

	Panel B: Misspecified Model with Only the First Factor								
	Likelihood Ratio Test			Traditional GMM (J_1)			SDF Methodology (J_2)		
	Significance Level			Significance Level			Significance Level		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.129	0.068	0.017	0.119	0.057	0.012	0.131	0.066	0.012
240	0.166	0.096	0.023	0.166	0.088	0.021	0.162	0.088	0.022
360	0.212	0.123	0.037	0.206	0.122	0.032	0.207	0.120	0.032
480	0.245	0.153	0.051	0.244	0.147	0.044	0.244	0.150	0.042
600	0.290	0.188	0.066	0.289	0.185	0.058	0.288	0.180	0.057
720	0.334	0.222	0.083	0.332	0.222	0.078	0.329	0.213	0.078

	Panel C: Misspecified Model with Only the Second Factor								
	Likelihood Ratio Test			Traditional GMM (J_1)			SDF Methodology (J_2)		
	Significance Level			Significance Level			Significance Level		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.111	0.052	0.008	0.110	0.050	0.006	0.038	0.017	0.002
240	0.295	0.185	0.053	0.299	0.182	0.047	0.059	0.023	0.003
360	0.472	0.338	0.135	0.474	0.341	0.139	0.081	0.037	0.007
480	0.613	0.481	0.246	0.617	0.487	0.248	0.134	0.062	0.011
600	0.730	0.608	0.362	0.731	0.613	0.367	0.219	0.110	0.020
720	0.816	0.717	0.484	0.815	0.710	0.480	0.319	0.176	0.039

the eigenvalues to construct the weights of the linear combination of χ_1^2 distribution that the test statistic is compared with. Therefore, the nonstandard GMM overidentification test does not escape from the problem that plagues the standard GMM. Although not reported, simulation evidence suggests that the nonstandard GMM overidentification test that uses the identity matrix as the weighting matrix generally has lower power than that of the standard GMM overidentification test in detecting our misspecified models.

IV. Conclusions

This paper exploits the fact that current empirical studies of asset pricing models using the SDF methodology typically ignore a fully specified model for asset returns. When asset returns are generated by a linear factor model, there are two potential problems associated with the use of the SDF methodology: (1) the accuracy of the estimated risk premium can be very poor and (2) its overidentification test has very little power in detecting misspecified models. These problems arise because the moment conditions the SDF methodology uses are very volatile, making accurate estimation and testing difficult under this methodology.

By specifying the return-generating process of the asset returns as in the traditional methodologies, these two potential problems can be mitigated. We demonstrate that, under the assumption that assets returns are generated by a linear factor model, the standard error of the risk premium under the traditional methodologies is much lower than that of the SDF methodology. The reason for such improvement is that the traditional methodologies use moment conditions that are much less volatile than that of the SDF methodology, and as a result they provide far more reliable inferences on the parameters. Moreover, the specification tests in the traditional methodologies generally have higher power in rejecting misspecified models than the SDF methodology. Our analysis focuses exclusively on linear factor models. This is not only due to their tractability, but also their premier importance in asset pricing. However, to the extent that any nonlinear model can be well approximated by a linear one, our results should also have implications on the use of the SDF methodology in nonlinear models where one must be cautious about the explanatory power of the factors, the parameter estimation error, the size, and the power of the tests.

Despite the fact that the SDF methodology has an interesting perspective to offer and a parsimonious model to estimate, there are costs associated with these benefits. In any event, it appears safe to say that the traditional methodologies are here to stay. In particular, traditional tests of asset pricing models will continue to play important roles in understanding the risks associated with investing, and perhaps even more so than the stochastic discount factor methodology for portfolio choice and performance evaluation problems.

Appendix

Proof of Proposition 1: We begin by deriving the asymptotic variance of $\hat{\lambda}^*$, which is given by the (1,1) element of $(\boldsymbol{D}_1' \boldsymbol{S}_1^{-1} \boldsymbol{D}_1)^{-1}$, where

$$\boldsymbol{D}_1 = E\left[\frac{\partial \boldsymbol{g}_{1T}}{\partial \lambda}, \frac{\partial \boldsymbol{g}_{1T}}{\partial \boldsymbol{\beta}'}\right] \tag{A1}$$

and

$$\boldsymbol{S}_1 = E[(\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t)(\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t)']. \tag{A2}$$

Define $\boldsymbol{a} = [1,0]'$ and $\boldsymbol{b} = [\lambda,1]'$; then we have

$$\boldsymbol{D}_1 = -[\boldsymbol{a} \otimes \boldsymbol{\beta}, \boldsymbol{b} \otimes \boldsymbol{I}_N]. \tag{A3}$$

Therefore, using the identity $\boldsymbol{a} \otimes \boldsymbol{\beta} = (\boldsymbol{a} \otimes \boldsymbol{I}_N)(1 \otimes \boldsymbol{\beta}) = (\boldsymbol{a} \otimes \boldsymbol{I}_N)\boldsymbol{\beta}$, we have

$$\begin{aligned}
\boldsymbol{D}_1' \boldsymbol{S}_1^{-1} \boldsymbol{D}_1 &= \begin{bmatrix} (\boldsymbol{a}' \otimes \boldsymbol{\beta}')\boldsymbol{S}_1^{-1}(\boldsymbol{a} \otimes \boldsymbol{\beta}) & (\boldsymbol{a}' \otimes \boldsymbol{\beta}')\boldsymbol{S}_1^{-1}(\boldsymbol{b} \otimes \boldsymbol{I}_N) \\ (\boldsymbol{b}' \otimes \boldsymbol{I}_N)\boldsymbol{S}_1^{-1}(\boldsymbol{a} \otimes \boldsymbol{\beta}) & (\boldsymbol{b}' \otimes \boldsymbol{\beta}')\boldsymbol{S}_1^{-1}(\boldsymbol{b} \otimes \boldsymbol{I}_N) \end{bmatrix} \\[2mm]
&= \begin{bmatrix} \boldsymbol{\beta}'(\boldsymbol{a}' \otimes \boldsymbol{I}_N)\boldsymbol{S}_1^{-1}(\boldsymbol{a} \otimes \boldsymbol{I}_N)\boldsymbol{\beta} & \boldsymbol{\beta}'(\boldsymbol{a}' \otimes \boldsymbol{I}_N)\boldsymbol{S}_1^{-1}(\boldsymbol{b} \otimes \boldsymbol{I}_N) \\ (\boldsymbol{b}' \otimes \boldsymbol{I}_N)\boldsymbol{S}_1^{-1}(\boldsymbol{a} \otimes \boldsymbol{I}_N)\boldsymbol{\beta} & (\boldsymbol{b}' \otimes \boldsymbol{I}_N)\boldsymbol{S}_1^{-1}(\boldsymbol{b} \otimes \boldsymbol{I}_N) \end{bmatrix}.
\end{aligned}$$

From the partitioned matrix inverse formula, the (1,1) element of $(\boldsymbol{D}_1' \boldsymbol{S}_1^{-1} \boldsymbol{D}_1)^{-1}$ is

$$\left(\boldsymbol{\beta}'[\boldsymbol{A}'\boldsymbol{S}_1^{-1}\boldsymbol{A} - \boldsymbol{A}'\boldsymbol{S}_1^{-1}\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{S}_1^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{S}_1^{-1}\boldsymbol{A}]\boldsymbol{\beta}\right)^{-1}, \tag{A4}$$

by writing $\boldsymbol{A} = \boldsymbol{a} \otimes \boldsymbol{I}_N$ and $\boldsymbol{B} = \boldsymbol{b} \otimes \boldsymbol{I}_N$. Defining $\boldsymbol{d} = [1, -\lambda]'$ and

$$\boldsymbol{U} = (\boldsymbol{d}' \otimes \boldsymbol{I}_N)\boldsymbol{S}_1(\boldsymbol{d} \otimes \boldsymbol{I}_N), \tag{A5}$$

we will show that the (1,1) element of $(\boldsymbol{D}_1' \boldsymbol{S}_1 \boldsymbol{D}_1)^{-1}$ can be simplified to $(\boldsymbol{\beta}'\boldsymbol{U}^{-1}\boldsymbol{\beta})^{-1}$. To prove this identity, we define a matrix $\boldsymbol{C} = [\boldsymbol{a}, \boldsymbol{b}]$ and consider the inverse of

$$(\boldsymbol{C}' \otimes \boldsymbol{I}_N)\boldsymbol{S}_1^{-1}(\boldsymbol{C} \otimes \boldsymbol{I}_N) = \begin{bmatrix} \boldsymbol{A}'\boldsymbol{S}_1^{-1}\boldsymbol{A} & \boldsymbol{A}'\boldsymbol{S}_1^{-1}\boldsymbol{B} \\ \boldsymbol{B}'\boldsymbol{S}_1^{-1}\boldsymbol{A} & \boldsymbol{B}'\boldsymbol{S}_1^{-1}\boldsymbol{B} \end{bmatrix}. \tag{A6}$$

Defining $D = [d, e]$, where $e = [0, 1]'$, it is easy to verify that $C^{-1} = D'$ and $(C')^{-1} = D$. Therefore, we have

$$
\begin{aligned}
&[(C' \otimes I_N) S_1^{-1}(C \otimes I_N)]^{-1} \\
&= (D' \otimes I_N) S_1 (D \otimes I_N) \\
&= \begin{bmatrix}
(d' \otimes I_N) S_1 (d \otimes I_N) & (d' \otimes I_N) S_1 (e \otimes I_N) \\
(e' \otimes I_N) S_1 (d \otimes I_N) & (e' \otimes I_N) S_1 (e \otimes I_N)
\end{bmatrix}.
\end{aligned}
\tag{A7}
$$

Note that the upper left block of $[(C' \otimes I_N) S_1^{-1}(C \otimes I_N)]^{-1}$ is just U. Another way to obtain this submatrix is to apply the partitioned matrix inverse formula to $(C' \otimes I_N) S_1^{-1}(C \otimes I_N)$, which gives the identity

$$
U = [(A' S_1^{-1} A) - (A' S_1^{-1} B)(B' S_1^{-1} B)^{-1}(B' S_1^{-1} A)]^{-1}.
\tag{A8}
$$

For the GMM estimation of the SDF model, we have $D_2 = E[\partial g_{2T}/\partial \lambda] = -\beta$, hence $\mathrm{Avar}[\hat{\lambda}] = (D_2' S_2^{-1} D_2)^{-1} = (\beta' S_2^{-1} \beta)^{-1}$. Using $r_t = \beta(\lambda + f_t) + \varepsilon_t$ and

$$
(1 - f_t \lambda) \varepsilon_t = (d' z_t) \varepsilon_t = (d' \otimes I_N)(z_t \otimes \varepsilon_t),
\tag{A9}
$$

we have

$$
\begin{aligned}
S_2 &= E[r_t r_t'(1 - f_t \lambda)^2] \\
&= E[\varepsilon_t \varepsilon_t'(1 - f_t \lambda)^2] + E[\beta \beta'(\lambda + f_t)^2(1 - f_t \lambda)^2] \\
&= (d' \otimes I_N) E[(z_t \otimes \varepsilon_t)(z_t \otimes \varepsilon_t)'](d \otimes I_N) + c \beta \beta' \\
&= U + c \beta \beta',
\end{aligned}
\tag{A10}
$$

where $c = E[(\lambda + f_t)^2(1 - f_t \lambda)^2]$. Note that $c > 0$ unless $P[f_t = -\lambda \text{ or } 1/\lambda] = 1$, which is impossible when f_t has a continuous distribution. Since

$$
S_2^{-1} = U^{-1} - \frac{U^{-1} \beta \beta' U^{-1}}{\beta' U^{-1} \beta + \dfrac{1}{c}},
\tag{A11}
$$

we have

$$
\beta' S_2^{-1} \beta = \beta' U^{-1} \beta - \frac{(\beta' U^{-1} \beta)^2}{\beta' U^{-1} \beta + \dfrac{1}{c}}
$$

$$
\Rightarrow (\beta' S_2^{-1} \beta)^{-1} = \frac{1}{\beta' U^{-1} \beta} + c > \frac{1}{\beta' U^{-1} \beta},
\tag{A12}
$$

which proves the inequality.

For the case that $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma})$ conditional on f, the log-likelihood function under the null is

$$\mathcal{L} = -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{t=1}^{T} (\boldsymbol{r}_t - \lambda\boldsymbol{\beta} - \boldsymbol{\beta}f_t)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{r}_t - \lambda\boldsymbol{\beta} - \boldsymbol{\beta}f_t).$$

(A13)

Hence, we have

$$\frac{\partial\mathcal{L}}{\partial\lambda} = \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1} \sum_{t=1}^{T} (\boldsymbol{r}_t - \lambda\boldsymbol{\beta} - \boldsymbol{\beta}f_t),$$

(A14)

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{\beta}} = \boldsymbol{\Sigma}^{-1} \sum_{t=1}^{T} (\boldsymbol{r}_t - \lambda\boldsymbol{\beta} - \boldsymbol{\beta}f_t)(f_t + \lambda).$$

(A15)

Then,

$$-\frac{1}{T} E \frac{\partial^2\mathcal{L}}{\partial\lambda^2} = \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta},$$

(A16)

$$-\frac{1}{T} E \frac{\partial^2\mathcal{L}}{\partial\lambda\partial\boldsymbol{\beta}'} = \frac{1}{T} E\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1} \sum_{t=1}^{T} (\lambda + f_t) = \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\lambda,$$

(A17)

$$-\frac{1}{T} E \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\beta}^2} = \frac{1}{T} E\boldsymbol{\Sigma}^{-1} \sum_{t=1}^{T} (f_t + \lambda)^2 = (1 + \lambda^2)\boldsymbol{\Sigma}^{-1}.$$

(A18)

Now, it is known that the asymptotic variance matrix of the maximum likelihood estimator of $(\lambda, \boldsymbol{\beta})$ should be the inverse of the Hessian matrix $\boldsymbol{H}$, where $\boldsymbol{H}$ is given by

$$\boldsymbol{H} = -\lim \frac{1}{T} E \begin{bmatrix} \dfrac{\partial^2\mathcal{L}}{\partial\lambda^2} & \dfrac{\partial^2\mathcal{L}}{\partial\lambda\partial\boldsymbol{\beta}'} \\[2mm] \dfrac{\partial^2\mathcal{L}}{\partial\boldsymbol{\beta}\partial\lambda} & \dfrac{\partial^2\mathcal{L}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta} & \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\lambda \\[2mm] \lambda\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta} & (1 + \lambda^2)\boldsymbol{\Sigma}^{-1} \end{bmatrix}.$$

(A19)

When $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma})$ conditional on f, we have $\boldsymbol{S}_1 = \boldsymbol{I}_2 \otimes \boldsymbol{\Sigma}$ and $\boldsymbol{H}$ has the same expression as $\boldsymbol{D}_1' \boldsymbol{S}_1^{-1} \boldsymbol{D}_1$. This completes the proof. Note that our proof only depends on $\boldsymbol{r}_t$ having a factor structure and the beta pricing model holding; it does not require the true factor f_t. Therefore, Proposition 1 continues to hold when g_t or h_t is used as the factor. ∎

Proof of Proposition 2: Without loss of generality, we assume the $2N + 1$ sample moment conditions used to estimate λ and $\boldsymbol{\beta}$ are

$$\boldsymbol{g}_{3T} = \begin{bmatrix} \boldsymbol{g}_{1T} \\ \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} r_{1t}(1 - f_t \lambda) \end{bmatrix}, \tag{A20}$$

where $\beta_1 \neq 0$. Define $\boldsymbol{d} = [1, -\lambda]'$ and $\boldsymbol{e}_1 = [1, \mathbf{0}_{N-1}']'$. Since

$$\begin{aligned} u_{1t} &= \boldsymbol{e}_1' \boldsymbol{u}_t \\ &= \boldsymbol{e}_1'(\boldsymbol{d}' \otimes \boldsymbol{I}_N)(\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t) + \boldsymbol{e}_1' \boldsymbol{\beta}(\lambda + f_t)(1 - f_t \lambda) \\ &= (\boldsymbol{d}' \otimes \boldsymbol{e}_1')(\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t) + \beta_1(\lambda + f_t)(1 - f_t \lambda), \end{aligned}$$

the asymptotic variance of $(\hat{\lambda}^{**}, \hat{\boldsymbol{\beta}}^{**})$ is given by $(\boldsymbol{D}_3' \boldsymbol{S}_3^{-1} \boldsymbol{D}_3)^{-1}$ where

$$\begin{aligned} \boldsymbol{D}_3 &= E\left[\frac{\partial \boldsymbol{g}_{3T}}{\partial \lambda}, \frac{\partial \boldsymbol{g}_{3T}}{\partial \boldsymbol{\beta}'} \right] \\ &= \begin{bmatrix} \boldsymbol{D}_1 \\ (\boldsymbol{d}' \otimes \boldsymbol{e}_1')\boldsymbol{D}_1 \end{bmatrix} \end{aligned} \tag{A21}$$

and

$$\begin{aligned} \boldsymbol{S}_3 &= E\begin{bmatrix} (\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t)(\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t)' & (\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t)u_{1t}' \\ u_{1t}(\boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t)' & u_{1t}^2 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{S}_1(\boldsymbol{d} \otimes \boldsymbol{e}_1) \\ (\boldsymbol{d}' \otimes \boldsymbol{e}_1')\boldsymbol{S}_1 & \boldsymbol{e}_1' \boldsymbol{S}_2 \boldsymbol{e}_1 \end{bmatrix}. \end{aligned} \tag{A22}$$

From the proof of Proposition 1, we know that $\boldsymbol{S}_2 = \boldsymbol{U} + c\boldsymbol{\beta}\boldsymbol{\beta}'$, where $c = E[(\lambda + f_t)^2(1 - f_t \lambda)^2]$. Using the partitioned matrix inverse formula, we have

$$\boldsymbol{S}_3^{-1} = \begin{bmatrix} \boldsymbol{S}_1^{-1} + (c\beta_1^2)^{-1}(\boldsymbol{d} \otimes \boldsymbol{e}_1)(\boldsymbol{d}' \otimes \boldsymbol{e}_1') & -(c\beta_1^2)^{-1}(\boldsymbol{d} \otimes \boldsymbol{e}_1) \\ -(c\beta_1^2)^{-1}(\boldsymbol{d}' \otimes \boldsymbol{e}_1') & (c\beta_1^2)^{-1} \end{bmatrix}. \tag{A23}$$

Therefore,

$$
\begin{aligned}
\boldsymbol{D}_3' \boldsymbol{S}_3^{-1} \boldsymbol{D}_3 &= \boldsymbol{D}_1' \boldsymbol{S}_1^{-1} \boldsymbol{D}_1 + (c\beta_1^2)^{-1} \boldsymbol{D}_1'(\boldsymbol{d} \otimes \boldsymbol{e}_1)(\boldsymbol{d}' \otimes \boldsymbol{e}_1')\boldsymbol{D}_1 \\
&\quad - (c\beta_1^2)^{-1} \boldsymbol{D}_1'(\boldsymbol{d} \otimes \boldsymbol{e}_1)(\boldsymbol{d}' \otimes \boldsymbol{e}_1')\boldsymbol{D}_1 \\
&\quad - (c\beta_1^2)^{-1} \boldsymbol{D}_1'(\boldsymbol{d} \otimes \boldsymbol{e}_1)(\boldsymbol{d}' \otimes \boldsymbol{e}_1')\boldsymbol{D}_1 \\
&\quad + (c\beta_1^2)^{-1} \boldsymbol{D}_1'(\boldsymbol{d} \otimes \boldsymbol{e}_1)(\boldsymbol{d}' \otimes \boldsymbol{e}_1')\boldsymbol{D}_1 \\
&= \boldsymbol{D}_1' \boldsymbol{S}_1^{-1} \boldsymbol{D}_1,
\end{aligned}
$$

and the asymptotic variance of $(\lambda^{**}, \boldsymbol{\beta}^{**})$ and $(\lambda^*, \boldsymbol{\beta}^*)$ are identical. This completes the proof. ∎

REFERENCES

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

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

Breeden, Douglas T., Michael R. Gibbons, and Robert Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231–262.

Cochrane, John H., 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.

Hansen, Lars Peter, 1982, Large sample properties of the generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hansen, Lars Peter, and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor model, *Journal of Finance* 52, 557–590.

Harvey, Campbell R., and Guofu Zhou, 1993, International asset pricing with alternative distributional specifications, *Journal of Empirical Finance* 48, 107–131.

He, Jia, Raymond Kan, Lilian Ng, and Chu Zhang, 1996, Tests of the relations among market-wide factors, firm-specific variables, and stock returns using a conditional asset pricing model, *Journal of Finance* 51, 1891–1908.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Kan, Raymond, and Chu Zhang, 1999a, GMM tests of stochastic discount factor models with useless factors, *Journal of Financial Economics*, forthcoming.

Kan, Raymond, and Chu Zhang, 1999b, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 203–235.

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

MacKinlay, A. Craig, and Matthew P. Richardson, 1991, Using generalized method of moments to test mean-variance efficiency, *Journal of Finance* 46, 511–527.

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

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

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

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

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

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

Zhou, Guofu, 1995, Small sample rank tests with applications to asset pricing, *Journal of Empirical Finance* 2, 71–93.