Empirical Asset Pricing: the Beta Method versus the Stochastic Discount Factor Method

RAYMOND KAN and GUOFU ZHOU*

First draft: January 2000
This version: June 2002

*Kan is from the University of Toronto, Zhou is from Washington University in St. Louis. We thank Kerry Back, Long Chen, John Cochrane, Magnus Dahlquist, Heber Farnsworth, Ravi Jagannathan. Bob Korkie, Sergei Sarkissian, Gordon Sick, Tim Simin, Kevin Wang, Zhenyu Wang, Chu Zhang, and participants of the 2001 Northern Finance Meetings and 2001 Com^{2}MaC workshop on Mathematical and Quantitative Finance in South Korea, and especially Philip Dvbvig and Campbell Harvey, for helpful discussions and comments. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.

ABSTRACT

In a simple standardized factor model, Kan and Zhou (1999) show that the estimate of the parameter in the stochastic discount factor (SDF) method is much less efficient than the risk premium estimate in the beta method, when both are estimated using the generalized method of moments (GMM). Jagannathan and Wang (2001) and Cochrane (2000a,b) debate this conclusion in a nonstandardized factor model where the factor mean and variance have to be estimated, but their analysis relies on joint normality assumption for both the asset returns and the factors in an unconditional model. We make four contributions in this paper. First, we show that once the restrictive normality assumption is relaxed, the variance of the GMM estimate of the SDF parameter is highly sensitive to factor skewness and kurtosis whereas the variance of the GMM estimate of the risk premium is not. Second, we show that provide results for the general case and show that inference about the SDF parameter is highly sensitive to factor skewness and kurtosis, whereas inference about the factor premium is not. Therefore, even when the mean and the variance of the factor are unknown, inference based on the SDF parameter can still be less reliable than inference based on the risk premium parameter in realistic situations where the factors are leptokurtic. We also show that even for the case of independent and identically distributed returns and factors with conditional homoskedasticity, the GMM methodology can still be less efficient than the maximum likelihood method in estimating the parameters. In addition, we provide a theoretical power analysis of the GMM specification tests under both the SDF and the beta methods. Unlike the finding of Jagannathan and Wang (2001), we show that the two GMM specification tests have the same asymptotic local power. In addition, we show that the likelihood ratio test can be strictly more powerful than the GMM specification tests when the residuals are not normally distributed.

Traditional 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 security is a linear function of betas with respect to some common factors. This implication has been conveniently tested extensively in the finance literature by the so-called regression based "traditional method" or *beta method*, in which a regression model is proposed for the stock returns, and the theoretical implications are tested as hypotheses on the parameters of the regression model. However, it is now well known that linear asset pricing models and many nonlinear ones can be unified in a stochastic discount factor (SDF) framework. Different specifications of the SDF correspond to different asset pricing models. This framework has stimulated a different set of empirical tests based solely on the SDF formulation. This approach, the *SDF method*, has become extremely popular in the recent finance literature.

Although both the beta and the SDF methods are used by many researchers in many different contexts, usually only one of them is used in a given application. It is therefore important to know which of the two methods is better. In addition, as suggested by Jagannathan and Wang (2001, JW hereafter), the comparison can be so important that it might change the course of our empirical research on asset pricing models. If the traditional method performs better in linear models, it is natural to speculate that it can also perform better in situations that involve nonlinear models. This is because, in many cases, nonlinear SDF models are often linearized (see for example, Bansal, Hsieh, and Viswanathan (1993), Campbell (1993, 1996), and Cochrane (1996)), and we can study them also using the beta method too.

Comparison of the two methods is not an easy matter even for linear factor models, since the parameters of interest are different under the two setups. The beta method is formulated to analyze the factor risk premium, δ, and this is the primary parameter of interest. In contrast, the SDF is formulated to analyze the linear coefficient associated with the factor, λ, in the pricing kernel. In general, δ and λ are different parameters, so a direct comparison between them is difficult. The parameters do coincide for a standardized factor model, i.e., one in which the factor mean and variance are known in advance and the factor can be normalized to have mean zero and variance one. Using this fact, Kan and Zhou (1999, KZ hereafter) provide the first formal comparison between the two methods, assuming Generalized Method of Moments (GMM) estimation in both cases. The main findings are that the estimate of λ under the SDF method is far less efficient than the estimate of δ under the beta method and that the power of the GMM specification test in the SDF method is also much lower than it is in the beta method when the factor does not explain a lot of the returns.

The weakness of a standardized factor model is that it requires a known factor mean and variance. What happens if the mean and the variance of the factor are unknown constants?[1] As pointed out earlier, δ and λ will in general be different parameters, and a comparison between their estimation efficiency becomes more difficult. JW suggest augmenting the SDF and beta methods with two moments in their GMM estimation, one for the mean and one for the variance of the factor. Once these two moments are incorporated in the two methods, one can make an inference regarding both the SDF and the risk premium parameter using either method. As a result, one can compare the asymptotic variance of estimates of these two parameters in the two methods. For the case of independent and identically distributed (i.i.d.) jointly normal factors and residuals, JW show that the asymptotic variance of the estimates of λ and δ are identical in the GMM estimation of the two methods. In addition, they also show that there is no efficiency gain in using the maximum likelihood estimation of the beta model. Therefore, they conclude that, for parameter estimation, there is no advantage in using the beta method, whether it is estimated by GMM or maximum likelihood. Similar points are also made by Cochrane (2000b).

Known properties of stock returns make it clear that i.i.d. joint normality assumption is not a good description of the data. The first contribution of this paper is to provide a thorough analysis of the efficiency of GMM estimation without making strong distributional assumptions. We show that, for the GMM estimation, the results in JW are indeed general. If one augments the beta and SDF methods with the two additional moments conditions on estimating the mean and the variance of the factor, then the asymptotic variances of the estimates of δ and λ are always identical in both methods. This is because once these two additional moments are incorporated, the moment conditions used by the SDF method are just a subset of the moment conditions used

[1]When a macroeconomic variable is used as the factor, pre-processing is often performed to create a factor that is an innovation. One can obviously question whether such pre-processing actually creates a factor that has mean zero, or if it still leaves a nonzero mean in the prespecified factor. By the same token, one can also question the assumption that the prespecified factor has a constant conditional mean. One can even question the assumption that the prespecified factor is the correct factor. In this paper, we prefer not to debate the correctness of an assumption but instead focus our analysis on the more general case in which the conditional mean and the variance of the factor are unknown constants. Nevertheless, it is important to point out that inferences from the two-pass cross-sectional regression methodology used by, for example, Chen, Roll, and Ross (1986), are correct whether the factor has a zero or nonzero mean.

by the beta method. Therefore, estimation of δ or λ under the SDF method can never be more efficient than estimation under the beta method. On the other hand, while the beta method contains some additional moment conditions that are not in the SDF method, the beta method also requires estimation of the same number of additional parameters. As a result, no improvement on estimation efficiency of the parameters in the SDF method can be offered by the beta method. JW's conclusion that there is no efficiency gain in using the maximum likelihood estimation of the beta method, however, is specific to the joint normality assumption of the residuals and the factor. We show that, even for i.i.d. normal residuals, the maximum likelihood approach of the beta model can still offer important improvements over the GMM estimation of both the SDF and the beta model when the factor is not normally distributed.

While these results are interesting, they do not offer us a relevant comparison between the two methods. This is because, to our knowledge, no one in the finance literature uses the SDF method to make inferences on the factor premium parameter δ, and no one uses the beta method to make inferences on the SDF parameter λ. In fact, a crucial distinguishing feature between the two methods is that we make inferences on λ using the SDF method, and we make inferences on δ using the beta method. The relevant question to ask is whether it is better to make inferences on δ or on λ, when the mean and the variance of the factor need to be estimated. Since the most popular inference in the SDF method is to test $H_0 : \lambda = 0$, and the most popular inference in the beta method is to test $H_0 : \delta = 0$, we can compare the performance of these two estimates by comparing their coefficients of variation. Namely, we compare how many standard deviations that the mean of these estimates are away from zero. To facilitate this comparison, the second contribution of the paper is to provide explicit expressions of the asymptotic variance of GMM estimates of δ and λ, allowing for nonnormal factors and residuals, as well as for conditional heteroskedasticity. These explicit expressions allow us to show that the asymptotic variance of the GMM estimate of λ depends on both the third and the fourth moments of the factor, whereas the asymptotic variance of the GMM estimate of δ does not. In many empirical asset pricing studies, returns and factors exhibit significant positive kurtosis and this positive kurtosis renders the estimate of λ less reliable than the estimate of δ. It is in this sense that we believe the beta method is superior to the SDF method. The intuition of these results is simple. The risk premium δ is basically related to the first moment of the returns, whereas the SDF parameter can be viewed as approximately equal to the risk premium divided by the variance of the factor. This implies the variability of the estimate of λ depends also on how volatile the sample variance of the factor is. In general, the more leptokurtic is the factor, the more volatile is the sample variance of the factor.[2] As a result, inference using λ becomes less reliable in the presence of a leptokurtic factor.

Besides estimation accuracy, the size and power of specification tests are of importance in comparing the two methods. Unlike the case of estimation of risk premium, the specification tests in KZ are valid regardless of whether the mean and the variance of the factor are known or not. However, JW show that the size of the specification test under the beta method is highly distorted in finite samples. This finding is probably due to the fact that they use the specification test from the second stage GMM with the identity matrix as the initial weighting matrix. We show that, when the third stage GMM is used, or when a more appropriate initial weighting matrix is chosen, the size distortion problem of the specification test under the beta method disappears.

When the beta pricing model is misspecified, one would like the specification test to have power to detect such misspecifications. The third contribution of our paper is to provide a theoretical analysis of the power of the GMM specification tests under sequences of local alternatives. We find that the two methods offer exactly the same asymptotic local power. For the case of fixed alternatives, JW provide a theoretical analysis of the pricing errors under the two methods. We point out the problems in their theoretical analysis of pricing errors and provide simulation evidence showing that the GMM specification tests under the two methods are of similar power even under the fixed alternatives, provided that one uses an apporpriate weighting matrix. Simulation evidence also shows that the likelihood ratio test can dominate the GMM specification tests when the residuals are not normally distributed.

In a nutshell, the imprecision of the GMM estimate of the SDF parameter relative to the risk premium estimate, as pointed out by KZ, still exists even when the mean and the variance of the factor are unknown. However, if the factor explains a lot of the returns on the test assets, then there is no obvious advantage or disadvantage in using the GMM specification test under the beta method.

The rest of the paper is organized as follows. Section I presents the augmented SDF model

[2] It can be shown that under the i.i.d. setting, the asymptotic variance of the sample variance of the factor, $\hat{\sigma}^2$, is $(2+\gamma_2)\sigma^4$, where γ_2 is the kurtosis coefficient and σ^2 is the variance of the factor.

and the beta model, and discusses the parameter estimation and specification test under the GMM framework. Section II compares the asymptotic variance of the estimates of the risk premium and the SDF parameter under both methods. Section II also provides analytical expressions for the asymptotic variance of these two parameters under nonnormal factors and returns with possible conditional heteroskedasticity. These expressions allow us to compare the relative merits of inference using these two parameters. Maximum likelihood estimation is also discussed in this section. Section III presents the local power analysis of the specification tests under the two methods. This section also points out the problems in JW's analysis of pricing errors. Section IV presents simulation results. The final section concludes, and the Appendix contains proofs of all propositions.

I. The SDF and the Beta Models

In this section, we introduce the SDF and the beta models together with the GMM estimation of parameters and specification tests under these two models. Although the results hold for general multi-factor models, we limit our discussion to the case of a one-factor model in order to bring out the main points with the least technical burden on the readers.[3] Denote the excess returns (in excess of the risk-free rate) on N assets at time t as r_t. Under the one-factor beta pricing model, consistent with the CAPM or one-factor APT, the conditional expected excess return on each of N assets is a linear function of its conditional beta with respect to a common factor f_t:

$$E[r_t|I_{t-1}] = \beta_t \delta_t, \tag{1}$$

where $\beta_t = \mathrm{Cov}[r_t, f_t|I_{t-1}]/\mathrm{Var}[f_t|I_{t-1}]$ is the vector of conditional betas of the N assets at time t, δ_t is the conditional risk premium at time t, and I_{t-1} is the information variable available to the investors at time $t-1$.

Under the law of one price, there exists a random variable m_t, called the stochastic discount factor, the state-price density, or the pricing kernel, such that

$$E[r_t m_t|I_{t-1}] = 0_N, \tag{2}$$

where 0_N is an N-vector of zero. A beta pricing model can be cast in the SDF framework by

[3] A technical appendix of the results for general multi-factor models is available upon request.

specifying the stochastic discount factor as a linear function of f_t, and then we have

$$E[r_t(1 - f_t\lambda_t)|I_{t-1}] = 0_N. \tag{3}$$

Note that both δ_t and λ_t are in general functions of I_{t-1}. Without ancillary assumptions, (1) and (3) are too general to be tested.

In many empirical applications, β_t and δ_t are assumed to be constant in the beta model, and λ_t is assumed to be constant in the SDF model. It deserves emphasis that, although the beta model is equivalent to the SDF model on a theoretical basis, the two are not equivalent empirical models once ancillary assumptions are made. For example, assuming β is constant does not imply that the conditional covariance and conditional variance are constant over time.[4] Therefore, constant δ in the beta model does not imply constant λ in the SDF model, and vice versa. In order for both δ and λ to be constant, we need to make additional assumptions that the conditional mean and variance of the factor are also constant. Under these assumptions, it is easy to show that there is a one-to-one mapping between δ and λ

$$\delta = \frac{\sigma^2\lambda}{1 - \mu\lambda}, \tag{4}$$

$$\lambda = \frac{\delta}{\sigma^2 + \mu\delta}, \tag{5}$$

for given values of μ and σ^2, where μ and σ^2 are the mean and the variance of the common factor.[5] However, for δ to be finite, we need to assume $\lambda \neq 1/\mu$, and similarly we need to assume $\delta \neq -\sigma^2/\mu$ in order for λ to be finite.

When the factor is a general macroeconomic factor, there is no relation between δ and μ. For the case where the factor is the excess return on a portfolio, however, the risk premium δ is not a free parameter, but instead it is given by $\delta = \mu$. Similarly, the SDF parameter λ is also not a free parameter, and it is given by $\lambda = \mu/(\sigma^2 + \mu^2)$. In many empirical studies using the beta pricing models, the restriction on δ is not imposed even when excess portfolio return is used as the factor. This problem is even more pervasive for empirical work using the SDF models because

[4]For example, suppose (r_t, f_t, I_{t-1}) are jointly elliptically distributed, then β_t is constant over time, but conditional variance of (r_t, f_t) is a function of I_{t-1}.

[5]Another popular formulation of the stochastic discount factor is to define $m_t = 1 - (f_t - \mu)\lambda$ (e.g., Cochrane (1996)). This alternative formulation has the advantage that we can interpret λ as the reward-to-covariance parameter. Analysis of this SDF formulation is similar to the one that we present here, and therefore we do not provide a separate analysis for it.

the restriction on λ is almost never imposed. In the following discussion, we follow the common practice and do not place any restrictions on δ and λ. One could view our analysis as dealing with macroeconomic factors exclusively, or, alternatively, view the empirical asset pricing model as intended to price just the test assets, but not the factor portfolio in the case where the factor is the excess return on a portfolio.[6]

For statistical inference, we assume the excess returns follow a linear factor model,

$$r_t = \alpha + \beta(f_t - \mu) + \epsilon_t, \tag{6}$$

where ϵ_t is the residual, with $E[\epsilon_t] = 0_N$ and $E[\epsilon_t | f_t] = 0_N$. The exact beta pricing model implies $\alpha = E[r_t] = \beta\delta$, so we can write the return generating model as

$$r_t = \beta(\delta - \mu + f_t) + \epsilon_t. \tag{7}$$

In our analysis, we assume $\delta \neq 0$ to exclude the uninteresting case. To estimate the SDF model using GMM, researchers typically rely on the following moment conditions:

$$E[u_t(\lambda)] \equiv E[r_t(1 - f_t\lambda)] = 0_N, \tag{8}$$

and inference can only be made on its parameter λ, but not on δ. This has been a distinguishing feature of the SDF method.

JW (2001) suggest that one can also make inference on δ if one augments the SDF method with two additional moment conditions. They suggest using the following set of moment conditions:

$$E[g_t(\lambda, \mu, \sigma^2)] \equiv \begin{bmatrix} E[r_t(1 - f_t\lambda)] \\ E[f_t - \mu] \\ E[(f_t - \mu)^2 - \sigma^2] \end{bmatrix} = 0_{N+2}. \tag{9}$$

Under this set of moment conditions, one can estimate δ using the relation in (4). We shall denote the GMM estimate of δ and λ under this method as $\hat{\delta}_g$ and $\hat{\lambda}_g$ in order to highlight the fact that they are obtained from using the moments g_t.

Suppose we have T observations of r_t and f_t, then the GMM estimate of the parameters $\theta = (\lambda, \mu, \sigma^2)'$ is obtained by minimizing a quadratic form of the sample moment conditions as

$$\hat{\theta}_g = \mathrm{argmin}_\theta \bar{g}(\theta)' W_g \bar{g}(\theta), \tag{10}$$

[6] This implicitly assumes that the factor portfolio is not a linear combination of the test assets. Such an assumption is often required in empirical tests of asset pricing models to ensure that the covariance matrix of the residuals is nonsingular.

where $\bar{g}(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$ is the vector of sample moments, and W_g is a consistent estimate of the inverse of $S_g = \sum_{j=-\infty}^{\infty} E[g_t(\theta_0)g_{t+j}(\theta_0)']$, where θ_0 is the true parameter. Note that the augmented SDF moments are nonlinear functions of the parameters, so typically one uses a numerical optimization method to find the GMM estimates. In the Appendix, we show that an analytical solution to this particular minimization problem can be obtained. This analytical solution allows us to speed up our simulation but, more importantly, it allows us to obtain the global minimum solution.[7] As we show in the Appendix, there can be up to three sets of solutions to the first order condition of this minimization problem, so numerical methods cannot in general guarantee convergence to the global minimum. While the analytical solution of the GMM minimization problem is important for our simulations, it is not essential for our theoretical analysis. Therefore, we relegate its in-depth discussion to the Appendix.

For the GMM estimation of the beta pricing model, one typically relies on the moment conditions

$$\begin{aligned} E[r_t - \beta(\delta - \mu + f_t)] &= 0_N, &(11)\\ E[(r_t - \beta(\delta - \mu + f_t))f_t] &= 0_N, &(12)\\ E[f_t - \mu] &= 0. &(13) \end{aligned}$$

Note that this set of moment conditions can allow us only to make inference on δ, but not on λ. In order for us to also estimate λ using the beta method, JW suggest adding a moment condition to estimate the variance of the factor. Therefore, the moment conditions for the augmented beta method are

$$E[h_t(\lambda, \mu, \sigma^2, \beta)] \equiv \begin{bmatrix} E[r_t - \beta(\delta - \mu + f_t)] \\ E[(r_t - \beta(\delta - \mu + f_t))f_t] \\ E[f_t - \mu] \\ E[(f_t - \mu)^2 - \sigma^2] \end{bmatrix} = 0_{2N+2}, \qquad (14)$$

where δ in the moment conditions is given by (4). We shall denote the GMM estimate of δ and λ under this method as $\hat{\delta}_h$ and $\hat{\lambda}_h$, in order to highlight the fact that they are obtained from using the moments h_t. Note that, although in practice one estimates δ only using the beta method, we write h_t here as a function of λ just for easy comparison with the SDF method.

The GMM estimate of the parameters θ and β are obtained by minimizing a quadratic form of

[7] The analytical solution allows us to perform all our simulation with 10,000 repetitions. In contrast, a similar simulation study provided by JW is based on a simulation size of only 1,000.

the sample moment conditions,

$$\begin{bmatrix} \hat{\theta}_h \\ \hat{\beta} \end{bmatrix} = \text{argmin}_{\theta,\beta}\bar{h}(\theta,\beta)'W_h\bar{h}(\theta,\beta), \tag{15}$$

where $\bar{h}(\theta,\beta) = \frac{1}{T}\sum_{t=1}^{T} h_t(\theta,\beta)$ is the vector of sample moments, and W_h is a consistent estimate of the inverse of $S_h = \sum_{j=-\infty}^{\infty} E[h_t(\theta_0,\beta_0)h_{t+j}(\theta_0,\beta_0)']$, where θ_0 and β_0 are the true parameters. Note that the augmented SDF moments are nonlinear in the parameters. Although an analytical solution to this minimization problem is not available, we show in the Appendix that we can reduce the multi-dimensional minimization problem to a one-dimensional one, so that a simple line search will allow us to obtain the GMM estimates.

In both the SDF and the beta methods, one can test if the asset pricing model is correctly specified by performing a specification test. Under the SDF method, the GMM over-identification restriction test statistic is

$$J_g = T\bar{g}(\hat{\theta}_g)'W_g\bar{g}(\hat{\theta}_g). \tag{16}$$

The corresponding test statistic under the beta method is

$$J_h = T\bar{h}(\hat{\theta}_h,\hat{\beta})'W_h\bar{h}(\hat{\theta}_h,\hat{\beta}). \tag{17}$$

When the model is correctly specified, both J_g and J_h have an asymptotic distribution of χ^2_{N-1}. One can also test the beta pricing model by dropping the augmented moment conditions for the mean and the variance of the factor in both methods. Although μ appears in the first $2N$ moments under the beta method, we can still drop the augmented moment for the factor mean by simply defining a new parameter $\phi = \delta - \mu$ in (11) and (12). The over-identification test is then numerically identical to the one that is presented in KZ. This makes it clear that the over-identification test presented in KZ is valid, regardless of whether the mean and the variance of the factor are known or not. By dropping the two augmented moments, we have two fewer parameters, and the over-identification tests in the two methods still have an asymptotic distribution of χ^2_{N-1}.

II. A Comparison of SDF and Beta Methods

A. Asymptotic Variance of Parameter Estimates

Is the SDF method better, or is the beta method better? Under the assumption that (ϵ_t, f_t) are i.i.d. and jointly normally distributed, JW suggest that the two methods are the same when we compare the asymptotic variance of the estimates of λ (or δ) in both methods. We should at the outset emphasize that this is not a proper comparison of the two methods. This is because, to the best of our knowledge, no researchers actually use the SDF method to make inferences on δ, and no researchers use the beta method to make inferences on λ. Therefore, the proper comparison of the two methods should be focused on the question of whether or not inference based on δ is better than inference based on λ. We shall defer this question to the next subsection. In this subsection, we provide a proof of the JW's GMM estimation results, but in a general setting that does not require i.i.d. and joint normality assumption. We also provide an intuitive explanation.

To prepare for our key results, we start off our discussion with a linear transformation of the moment conditions used by the beta method h_t. Consider

$$m_t(\theta, \beta) = Ah_t(\theta, \beta) \tag{18}$$

where

$$A = \begin{bmatrix} I_N & -\lambda I_N & [1-(\mu+\delta)\lambda]\beta & -\lambda\beta \\ 0'_N & 0'_N & 1 & 0 \\ 0'_N & 0'_N & 0 & 1 \\ -\mu I_N & I_N & \delta\beta & 0 \end{bmatrix} \tag{19}$$

is a $(2N+2) \times (2N+2)$ nonsingular matrix. Asymptotically, there is no difference whether we conduct the GMM estimation and testing using h_t or using m_t, because the two sets of moments contain the same information. But what does m_t represent? Denote $\epsilon_t = r_t - \beta(\delta - \mu + f_t)$, the first N elements of m_t are

$$\begin{aligned} & \epsilon_t - \lambda\epsilon_t f_t + [1-(\mu+\delta)\lambda]\beta(f_t - \mu) - \lambda\beta[(f_t-\mu)^2 - \sigma^2] \\ = \;& [r_t - \beta(\delta - \mu + f_t)](1 - f_t\lambda) + [1-(\mu+\delta)\lambda]\beta(f_t-\mu) - \lambda\beta[(f_t-\mu)^2 - \sigma^2] \\ = \;& r_t(1 - f_t\lambda) - \beta\left[\delta(1-\mu\lambda) - \lambda\sigma^2\right] \\ = \;& r_t(1-f_t\lambda). \end{aligned} \tag{20}$$

Therefore, the first $N+2$ elements of m_t are simply g_t, the moment conditions used by the SDF method. This linear transformation makes it clear that the SDF moment conditions are just a subset of the moment conditions used by the beta method, so the SDF method can never attain better asymptotic efficiency in parameter estimation than the beta method. The linear transformation also suggests that one cannot add the SDF moment conditions to the beta method, since they are redundant.

Consider now the last N elements of m_t. They can be written as

$$-\mu\epsilon_t + \epsilon_t f_t + \delta\beta(f_t - \mu) = (f_t - \mu)[r_t - \beta(\delta - \mu + f_t)] + \beta\delta(f_t - \mu) = r_t(f_t - \mu) - \beta(f_t - \mu)^2. \quad (21)$$

Therefore, comparing with the SDF method, one can think of the beta method as having N additional moment conditions to estimate the β. In general, adding more moment conditions should improve the asymptotic efficiency of parameter estimation. In this particular case, however, the additional N moment conditions used by the beta method also bring in N additional parameters β, so one can no longer improve the estimation of θ using the beta method. The following lemma formalizes this result.

Lemma 1 *Denote $\phi = (\phi_1', \phi_2')'$ as a vector of $p = p_1 + p_2$ parameters where ϕ_1 is $p_1 \times 1$, and ϕ_2 is $p_2 \times 1$. Suppose we have $m = m_1 + m_2$ moment conditions*

$$E[g_t(\phi)] \equiv E\begin{bmatrix} g_{1t}(\phi_1) \\ g_{2t}(\phi_1, \phi_2) \end{bmatrix} = 0_m, \quad (22)$$

where g_{1t} is $m_1 \times 1$, and g_{2t} is $m_2 \times 1$, with $m_1 \geq p_1$ and $m_2 \geq p_2$. Denote the GMM estimator of ϕ_1 using just the moment conditions $E[g_{1t}(\phi_1)] = 0_{m_1}$ as $\hat{\phi}_1$ and the GMM estimator of ϕ_1, using all the moment conditions $E[g_t(\phi)] = 0_m$ as $\hat{\phi}_1^*$. Under the usual regularity conditions for GMM as in Hansen (1982), we have $\mathrm{Avar}[\hat{\phi}_1] \geq \mathrm{Avar}[\hat{\phi}_1^*]$.[8] *If $p_2 = m_2$, we have* $\mathrm{Avar}[\hat{\phi}_1] = \mathrm{Avar}[\hat{\phi}_1^*]$.

A direct application of Lemma 1 yields our first proposition.

Proposition 1: *Under the usual regularity conditions for GMM as in Hansen (1982), we have*

$$\begin{aligned} \mathrm{Avar}[\hat{\delta}_g] &= \mathrm{Avar}[\hat{\delta}_h], & (23) \\ \mathrm{Avar}[\hat{\lambda}_g] &= \mathrm{Avar}[\hat{\lambda}_h]. & (24) \end{aligned}$$

Therefore, the GMM estimation of δ and λ has the same asymptotic efficiency under both the SDF and the beta methods. JW prove this result for the case that (ϵ_t, f_t) is i.i.d. multivariate normal. Our Proposition 1 is significantly more general in that it allows for nonnormal residuals and factors. It also allows for non i.i.d. distribution as well as autocorrelated errors.[9]

[8] For two matrices A and B, we write $A \geq B$ when $A - B$ is a nonnegative definite matrix.

[9] It should be emphasized that unless overlapping returns are used for empirical tests, autocorrelation in g_t and h_t is hard to justify because it implies that the model is not correctly specified.

Lemma 1 also suggests that, for the estimation of λ under the SDF method, one can drop the last two moments in g_t without affecting the asymptotic efficiency. Similarly, for the estimation of δ under the beta method, one can drop the last moment in h_t without affecting the asymptotic efficiency. However, under the assumption that the mean of the factor is known as in KZ, the standard SDF moment conditions are no longer a subset of the moment conditions in the standard beta method, so Lemma 1 does not apply, and the parameter estimates under the beta method can be far superior to those under the SDF method.[10]

B. Comparison of Inference Using Risk Premium and SDF Parameter

The distinguishing feature of the SDF method is that it estimates the SDF parameter λ, and the distinguishing feature of the beta method is that it estimates the risk premium parameter δ. In fact, the current practice does not incorporate the augmented moment conditions in the SDF and the beta methods, so one cannot even make inferences on parameters that are specific to the other method. Therefore, to answer the question as to whether the beta method is better than the SDF method, one ought to focus on whether inference on δ is better than inference on λ. To address this question, we need explicit analytical expressions on $\mathrm{Avar}[\hat{\delta}]$ and $\mathrm{Avar}[\hat{\lambda}]$. For this purpose, we assume (ϵ_t, f_t) are i.i.d. Our distributional assumption is still more general than the i.i.d. multivariate normality assumption used by JW. Besides nonnormal residuals and factors, our distributional assumption also allows for conditional heteroskedasticity, so $\mathrm{Var}[\epsilon_t|f_t]$ does not have to be constant. The following proposition summarizes the asymptotic variance under the i.i.d. assumption. Due to Proposition 1, we simply denote the GMM estimators of δ and λ as $\hat{\delta}$ and $\hat{\lambda}$, without being explicit about whether we use g_t or h_t to estimate them.

Proposition 2: *When (ϵ_t, f_t) are independent and identically distributed over time, we have*

$$\mathrm{Avar}[\hat{\delta}] = \frac{(\sigma^2+\mu\delta)^2}{\sigma^4}(\beta' U^{-1}\beta)^{-1} + \sigma^2, \tag{25}$$

$$\mathrm{Avar}[\hat{\lambda}] = \frac{(\beta' U^{-1}\beta)^{-1}}{(\sigma^2+\mu\delta)^2} + \frac{\sigma^2(\sigma^4+\delta^4) + 2\delta(\delta^2-\sigma^2)\mu_3 + \delta^2(\mu_4 - 3\sigma^4)}{(\sigma^2+\mu\delta)^4}, \tag{26}$$

where $U = E[(1-f_t\lambda)^2\epsilon_t\epsilon_t']$, $\mu_3 = E[(f_t-\mu)^3]$, and $\mu_4 = E[(f_t-\mu)^4]$.

[10] When the mean and the variance of the factor are known, JW suggest one can improve parameter estimates under the SDF method by adding two somewhat unusual moment conditions that do not involve any parameters. Basically, by combining these two new moment conditions, one can transform the highly volatile SDF moment conditions into the more reliable beta moment conditions.

A couple of special cases deserve our attention. In the first case, we assume (ϵ_t, f_t) has a multivariate elliptical distribution. In this case, the two expressions can be simplified to

$$\mathrm{Avar}[\hat{\delta}] = \left[1 + \frac{(1+\frac{\gamma_2}{3})\delta^2}{\sigma^2}\right](\beta'\Sigma^{-1}\beta)^{-1} + \sigma^2, \quad (27)$$

$$\mathrm{Avar}[\hat{\lambda}] = \frac{\sigma^4}{(\sigma^2+\mu\delta)^4}\left[1 + \frac{(1+\frac{\gamma_2}{3})\delta^2}{\sigma^2}\right](\beta'\Sigma^{-1}\beta)^{-1} + \frac{\sigma^2(\sigma^4+\delta^4)+\gamma_2\delta^2\sigma^4}{(\sigma^2+\mu\delta)^4}, \quad (28)$$

where $\gamma_2 = \mu_4/\sigma^4 - 3$ is the kurtosis of f_t and $\Sigma = \mathrm{Var}[\epsilon_t]$ is the unconditional variance of ϵ_t.

In the second case, we assume $\mathrm{Var}[\epsilon_t|f_t] = \Sigma$ (i.e., conditional homoskedasticity). In this case, the two expressions are given by

$$\mathrm{Avar}[\hat{\delta}] = \left(1 + \frac{\delta^2}{\sigma^2}\right)(\beta'\Sigma^{-1}\beta)^{-1} + \sigma^2, \quad (29)$$

$$\mathrm{Avar}[\hat{\lambda}] = \frac{\sigma^2(\sigma^2+\delta^2)}{(\sigma^2+\mu\delta)^4}(\beta'\Sigma^{-1}\beta)^{-1} + \frac{\sigma^2(\sigma^4+\delta^4)+2\delta(\delta^2-\sigma^2)\mu_3+\delta^2(\mu_4-3\sigma^4)}{(\sigma^2+\mu\delta)^4}. \quad (30)$$

In the last case, we consider (ϵ_t, f_t) has a multivariate normal distribution and we have

$$\mathrm{Avar}[\hat{\delta}] = \left(1 + \frac{\delta^2}{\sigma^2}\right)(\beta'\Sigma^{-1}\beta)^{-1} + \sigma^2, \quad (31)$$

$$\mathrm{Avar}[\hat{\lambda}] = \frac{\sigma^2(\sigma^2+\delta^2)}{(\sigma^2+\mu\delta)^4}(\beta'\Sigma^{-1}\beta)^{-1} + \frac{\sigma^2(\sigma^4+\delta^4)}{(\sigma^2+\mu\delta)^4}, \quad (32)$$

which are the same as those given by JW. The expressions for the multivariate normality case can be obtained as a special case for the multivariate elliptical distribution by setting $\gamma_2 = 0$, or they can be obtained by setting $\mu_3 = 0$ and $\mu_4 = 3\sigma^4$ in the conditional homoskedasticity case.

In comparing the expressions for $\mathrm{Avar}[\hat{\delta}]$ with those for $\mathrm{Avar}[\hat{\lambda}]$, we note that the third and the fourth moments of the factor play an important role in determining $\mathrm{Avar}[\hat{\lambda}]$, but this is not the case for $\mathrm{Avar}[\hat{\delta}]$. The reason is that $\hat{\delta}$, as the estimated risk premium, is only related to the first moment of the return, so the third and the fourth moments of the factor do not contribute much to its asymptotic variance.[11] However, $\lambda = \delta/(\sigma^2 + \mu\delta)$, so when μ and σ^2 need to be estimated, the asymptotic variance of $\hat{\lambda}$ will be heavily influenced by the third and the fourth moments of the factor. In an extreme case, where the third or the fourth moment of the factor does

[11] Strictly speaking, this is only true for the conditional homoskedasticity case. When β needs to be estimated, asymptotic variance of $\hat{\beta}$ (and hence $\mathrm{Avar}[\hat{\delta}]$) can also be affected by the third and the fourth moments of the factor if there is conditional heteroskedasticity.

not exist, $\text{Avar}[\hat{\lambda}]$ is infinite, but $\text{Avar}[\hat{\delta}]$ can still be finite under the conditional homoskedasticity assumption.[12]

How do we compare $\text{Avar}[\hat{\delta}]$ with $\text{Avar}[\hat{\lambda}]$? The easiest way is to look at the case where the factor is standardized, i.e., $\mu = 0$ and $\sigma^2 = 1$. In this case, we have $\delta = \lambda$, and we can directly compare $\text{Avar}[\hat{\delta}]$ with $\text{Avar}[\hat{\lambda}]$ to determine whether or not inference using $\hat{\delta}$ is better than using $\hat{\lambda}$. It is important to note that unlike in KZ, we do not assume the econometrician has the knowledge that $\mu = 0$ and $\sigma^2 = 1$ here. These two moments still have to be estimated so the asymptotic variances of both $\hat{\delta}$ and $\hat{\lambda}$ fully take into account of the estimation errors of $\hat{\mu}$ and $\hat{\sigma}^2$. From Proposition 2, we have

$$\text{Avar}[\hat{\delta}] = (\beta' U^{-1}\beta)^{-1} + 1, \tag{33}$$

$$\text{Avar}[\hat{\lambda}] = (\beta' U^{-1}\beta)^{-1} + 1 + \delta^4 + 2\delta(\delta^2 - 1)\mu_3 + \delta^2(\mu_4 - 3), \tag{34}$$

when $\mu = 0$ and $\sigma^2 = 1$. Comparing (33) and (34), we can see that which estimator is more efficient depends in general on the third and the fourth moments of the factor. For the case where the standardized factor f_t is normal, we have $\mu_3 = 0$ and $\mu_4 = 3$ and hence

$$\text{Avar}[\hat{\lambda}] = (\beta' U^{-1}\beta)^{-1} + 1 + \delta^4 > \text{Avar}[\hat{\delta}]. \tag{35}$$

Therefore, under the case of a normal standardized factor, we can definitely conclude that using $\hat{\delta}$ to make inferences is asymptotically better than using $\hat{\lambda}$.[13] One must wonder why there is a difference in the asymptotic variance of $\hat{\delta}$ and $\hat{\lambda}$ when they both converge to the same limit. To understand this, we observe that $\hat{\lambda} = \hat{\delta}/(\hat{\sigma}^2 + \hat{\mu}\hat{\delta})$. Even though the denominator converges to one when $\mu = 0$ and $\sigma^2 = 1$, its sample fluctuations can make $\hat{\lambda}$ more volatile than $\hat{\delta}$. The efficiency gain from using $\hat{\delta}$ instead of $\hat{\lambda}$, however, depends on the magnitude of δ. For a standardized factor, the magnitude of δ depends on to what degrees the returns are correlated with the factor.[14] If the

[12] Note that the assumption of existence of fourth moment is only sufficient, but not necessary, for the asymptotic variance of $\hat{\delta}$ to exist.

[13] When the econometrician has the knowledge that $\mu = 0$, KZ show that $\text{Avar}[\hat{\delta}] = (\beta' U^{-1}\beta)^{-1} < \text{Avar}[\hat{\lambda}]$. Under the i.i.d. joint normality assumption on the returns and factor, Cochrane (2000a) points out KZ's expression understates the volatility of $\hat{\delta}$ if the econometrician does not know the mean of the factor. As we show here, the inequality of $\text{Avar}[\hat{\delta}] < \text{Avar}[\hat{\lambda}]$ still holds for the case of normal factor, even when the econometrician has to estimate the mean of the factor.

[14] To see this, we observe that the risk premium is given by $\delta = E[r_i]/\beta_i$, where $E[r_i]$ and β_i are the expected return and beta of the ith asset, respectively. For a standardized factor, we have $\beta_i = \rho_{if}\sigma_i$, where σ_i is the standard deviation of the ith asset and ρ_{if} is the correlation coefficient between the excess return on the ith asset and the common factor. Therefore, the less correlated is the factor with the returns, the lower is the beta and the higher is the risk premium.

factor explains a lot of the returns (for example, when the test assets and the factor are both returns from well diversified portfolios), then δ is small and the efficiency gain from using $\hat{\delta}$ is not much. If the factor does not explain a lot of the returns (for example, when the factor is a macroeconomic factor), then δ can be fairly large, and the efficiency gain from using $\hat{\delta}$ instead of $\hat{\lambda}$ is substantial.

For the general case, it depends on the sign of μ_3 and the magnitude of μ_4. For symmetric factors (i.e., $\mu_3 = 0$), Avar$[\hat{\delta}]$ is smaller than Avar$[\hat{\lambda}]$ if the factor is leptokurtic (i.e., $\mu_4 > 3$), but Avar$[\hat{\delta}]$ can be greater than Avar$[\hat{\lambda}]$ if $\mu_4 < 3$. Intuitively, this result follows because, under the i.i.d. assumption, the asymptotic variance of $\hat{\sigma}^2$ is $\mu_4 - \sigma^4$. Therefore, the more leptokurtic the factor is, the more volatile is $\hat{\sigma}^2$ and hence the more volatile is $\hat{\lambda}$.

In practice, commonly used factors and returns often exhibit very high kurtosis. This salient feature of the data suggests that the asymptotic variance of $\hat{\lambda}$ can be a lot higher than the asymptotic variance of $\hat{\delta}$ (after standardizing the factor). As an example, we report in Panel A of Table 1 the sample average, sample standard deviation, sample skewness and kurtosis of the excess return on the value-weighted market portfolio and the ten size-ranked portfolios of the combined NYSE-AMEX-NASDAQ; these were estimated using monthly data over the period January 1926 to December 1999. This is the same data set used by JW. Sample estimates of these ten size-ranked portfolios will also be used to determine the values of the parameters for our simulation experiment in Section IV.

Table I about here

As we can observe in Table I, there are significant positive skewness and kurtosis for the excess returns on the value-weighted market and the ten size-ranked portfolios.[15] The sample kurtosis of the data is particularly severe, and, in general, it is much higher than the sample skewness. To the extent that the common factor also exhibits this behavior, then we would expect Avar$[\hat{\lambda}]$ to be much higher than Avar$[\hat{\delta}]$, if the factor is standardized.

In general, factors are not standardized to have zero mean and unit variance, so we cannot directly compare Avar$[\hat{\delta}]$ with Avar$[\hat{\lambda}]$. Nevertheless, we can compare the performance of $\hat{\delta}$ and $\hat{\lambda}$

[15]Under the normality assumption, the standard deviation of the sample skewness in our data is roughly $\sqrt{6/888} = 0.0822$ and the standard deviation of the sample kurtosis is roughly $\sqrt{24/888} = 0.1644$.

by computing the asymptotic relative efficiency of $\hat{\delta}$ relative to $\hat{\lambda}$, which is defined as follows:[16]

$$\text{ARE} = \lim_{T\to\infty} \frac{\text{Var}[\hat{\lambda}]/E[\hat{\lambda}]^2}{\text{Var}[\hat{\delta}]/E[\hat{\delta}]^2} = \frac{\delta^2 \text{Avar}[\hat{\lambda}]}{\lambda^2 \text{Avar}[\hat{\delta}]} = (\sigma^2 + \mu\delta)^2 \frac{\text{Avar}[\hat{\lambda}]}{\text{Avar}[\hat{\delta}]}. \tag{36}$$

An ARE greater than one suggests that using $\hat{\delta}$ to make inference is asymptotically better than using $\hat{\lambda}$. The fact that ARE can be used as a measure of relative efficiency is because it is the limit of the ratio of squared coefficients of variation of $\hat{\lambda}$ to $\hat{\delta}$. A higher absolute value of coefficient of variation gives us a lower probability of rejecting the hypothesis that the parameter is equal to zero. To the extent that one is often interested in testing $H_0 : \delta = 0$ under the beta method and testing $H_0 : \lambda = 0$ under the SDF method, the ARE measure gives us a good indication of whether inference is better performed using $\hat{\delta}$ or $\hat{\lambda}$.

In general, one needs to know all the parameters, as well as μ_3 and μ_4 to determine whether ARE is greater than one or not. For values of μ_3 and μ_4 that are comparable with what we report in Table I and for a reasonable choice of other parameters, however, we generally find that ARE is greater than one. For the case of $\mu_3 = 0$ and $\mu_4 = 3\sigma^4$ (e.g., normal factor), expression (36) can be simplified to

$$\text{ARE} = \frac{\sigma^4}{(\sigma^2 + \mu\delta)^2} \left[\frac{(\beta' U^{-1}\beta)^{-1} + \frac{\sigma^2(\sigma^4+\delta^4)}{(\sigma^2+\mu\delta)^2}}{(\beta' U^{-1}\beta)^{-1} + \frac{\sigma^6}{(\sigma^2+\mu\delta)^2}} \right]. \tag{37}$$

The first term can be greater than or less than one, depending on whether $\mu\delta$ is less than or greater than zero. The second term is always greater than one. For the special case that $\mu = 0$, the first term is one and the ARE is always greater than one, so inference using $\hat{\delta}$ is asymptotically better than inference using $\hat{\lambda}$.[17]

In summary, inference using the risk premium parameter δ can be either superior or inferior to inference using the SDF parameter λ. The comparison depends heavily on the third and the fourth moment of the factor as well as the magnitude of δ. However, for a factor that exhibits significant kurtosis, as is commonly the case, we find inference on the risk premium parameter to be superior to inference on the SDF parameter. In the current finance literature, researchers make inference only on δ when using the beta method, and only on λ when using the SDF method. Therefore, to the extent that there is positive kurtosis in the common factor, one should consider inference based on the beta method to be more reliable than inference based on the SDF method.

[16]See Huber (1981) for a textbook treatment of comparing performance of alternative estimators using ARE.

[17]If we use the SDF formulation that assumes $m_t = 1 - (f_t - \mu)\lambda$ as in Cochrane (1996), then we have $\lambda = \delta/\sigma^2$ and it can be shown that ARE > 1 for any μ when $\mu_3 = 0$ and $\mu_4 \geq 3\sigma^4$.

C. Maximum Likelihood Estimation

Under the i.i.d. setting, it is well known that maximum likelihood estimation of parameters attains the Cramér-Rao lower bound, and it is asymptotically the most efficient. When (ϵ_t, f_t) has a multivariate normal distribution, JW show that the GMM estimators of δ and λ from either the beta method or the SDF method have the same asymptotic efficiency as the maximum likelihood estimators.[18] However, this conclusion is specific to the multivariate normality assumption. In general, even when ϵ_t is i.i.d. normal conditional on f_t, maximum likelihood estimators can still offer substantial efficiency gain over the GMM estimators when f_t is not normally distributed. In this subsection, we illustrate this with two examples. The first one is when f_t has a Student-t distribution, and the second one is when f_t has a normal mixture distribution. These two examples are interesting in their own right because both distributions of f_t are robust in the sense that they contain the normal distribution as a special case. More importantly, both distributions can give positive kurtosis that is more consistent with the data than the normal distribution. In the first example where f_t has a Student-t distribution with ν degrees of freedom, its kurtosis is $\gamma_2 = 6/(\nu - 4)$. Although we can get positive kurtosis, the magnitude of the kurtosis cannot be very large for a t-distribution with a finite fourth moment. Even for ν as low as five, we only have $\gamma_2 = 6$, which is still smaller than the sample kurtosis of 7.989 of the value-weighted market portfolio, as we report in Table I.[19] In order to better capture the high kurtosis in the data, our second example assumes the density function of f_t is a mixture of two normal distributions with the same mean but different variances,

$$f(f_t) = w\phi_1(f_t) + (1 - w)\phi_2(f_t) \tag{38}$$

where $\phi_i(\cdot)$ is the density function of a normal distribution with mean μ and variance σ_i^2, $i = 1, 2$. We assume $0 < w < 0.5$ and $\sigma_1^2 \neq \sigma_2^2$ so that the parameters μ, w, σ_1^2 and σ_2^2 can be uniquely identified. This normal mixture distribution has a kurtosis of

$$\gamma_2 = 3\left[\frac{w\sigma_1^4 + (1 - w)\sigma_2^4}{(w\sigma_1^2 + (1 - w)\sigma_2^2)^2} - 1\right], \tag{39}$$

[18] Under the i.i.d. setting, the maximum likelihood estimation can be considered as a special case of the GMM when the score function is used as the moment, so strictly speaking, one can call maximum likelihood GMM too. However, we follow the common practice, which is to label GMM in the case for which we do not explicitly model the distribution of (ϵ_t, f_t), and to label maximum likelihood in the case for which we explicitly model the distribution of (ϵ_t, f_t).

[19] If we do not restrict ν to be an integer, then we can also have high kurtosis in the t-distribution for ν that is close to four.

and it can allow for much higher kurtosis than the Student-t distribution. It also has the advantage that all of its moments exist, which is not the case for the Student-t distribution.

In the Appendix, we discuss the maximum likelihood estimation problem when ϵ_t is multivariate normal, but f_t is nonnormal. We also present the analytical solution to the likelihood ratio test of the beta pricing model and show that the likelihood ratio test in KZ is indeed valid, even when the factor is not normal and its mean and variance are unknown. The following proposition provides the asymptotic variance of the maximum likelihood estimator of the risk premium, $\hat{\delta}_{ML}$, under the two examples of a nonnormal factor.

Proposition 3: *Suppose, conditional on f_t, ϵ_t is independent and identically distributed as $N(0_N, \Sigma)$. (1) When f_t has a Student-t distribution with ν degrees of freedom, we have*

$$\mathrm{Avar}[\hat{\delta}_{ML}] = \left(1+\frac{\delta^2}{\sigma^2}\right)(\beta'\Sigma^{-1}\beta)^{-1} + \left(1-\frac{6}{\nu^2+\nu}\right)\sigma^2 < \left(1+\frac{\delta^2}{\sigma^2}\right)(\beta'\Sigma^{-1}\beta)^{-1} + \sigma^2 = \mathrm{Avar}[\hat{\delta}]. \tag{40}$$

(2) When f_t has a normal mixture distribution with parameters w, μ, σ_1^2 and σ_2^2, we have

$$\mathrm{Avar}[\hat{\delta}_{ML}] = \left(1+\frac{\delta^2}{\sigma^2}\right)(\beta'\Sigma^{-1}\beta)^{-1} + \frac{1}{c} < \left(1+\frac{\delta^2}{\sigma^2}\right)(\beta'\Sigma^{-1}\beta)^{-1} + \sigma^2 = \mathrm{Avar}[\hat{\delta}], \tag{41}$$

where

$$c = \frac{w}{\sigma_1^2} + \frac{1-w}{\sigma_2^2} - 2w(1-w)\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right)^2 \int_0^\infty \frac{\phi_1(z)\phi_2(z)z^2}{w\phi_1(z)+(1-w)\phi_2(z)}\mathrm{d}z > \frac{1}{\sigma^2}. \tag{42}$$

One can also estimate λ under the maximum likelihood method. Although we do not present the asymptotic variance of $\hat{\lambda}_{ML}$ here because its expression is more complex, the strict inequality $\mathrm{Avar}[\hat{\lambda}_{ML}] < \mathrm{Avar}[\hat{\lambda}]$ also holds for these two cases. Details of these results are available upon request.

Proposition 3 shows that even when only f_t is not normally distributed, there can be significant efficiency gain in using the maximum likelihood estimator. When both ϵ_t and f_t are not normally distributed, the efficiency gain in using maximum likelihood estimation can be even more significant.[20] Therefore, incorporating information on the distribution of (ϵ_t, f_t) can result in better estimation of the risk premium and the SDF parameter than of the GMM. The disadvantage of

[20]See Kan and Zhou (2001a) for an analysis of maximum likelihood estimation under the multivariate elliptical distribution assumption on (ϵ_t, f_t).

using the maximum likelihood estimation is that one can make a wrong distributional assumption that may lead to an erroneous inference. This trade-off is particularly important when there is high kurtosis in the residuals and the factors because, in those situations, the GMM estimators can be very volatile relative to the maximum likelihood estimators. One may wish to make some robust distributional assumptions in order to extract more information from the data.

III. Specification Tests

A. Choice of Weighting Matrix

Under both the SDF and the beta methods, the GMM over-identification test can serve as a specification test of the asset pricing model. Although both J_g and J_h in (16) and (17) have an asymptotic χ^2_{N-1} distribution when the model is correctly specified, the distributions of these two test statistics do not conform exactly to their asymptotic distribution in finite samples. If one relies solely on the asymptotic distribution to make inferences, then it is desirable that the finite sample distribution of the test statistic used be well approximated by its asymptotic distribution. The analysis of the finite sample performance of the GMM specification test requires us to specify how the weighting matrix is estimated, however. In general, one needs to choose an initial weighting matrix to come up with a consistent estimate of the parameters in a first stage GMM, and then uses the parameter estimate to obtain a consistent estimate of the optimal weighting matrix to be used for the second stage GMM. Such a procedure can also be iterated.[21] In addition, there are issues related to the estimation of the spectral density matrix about whether we should use sample covariance or sample second moments, and also about whether we should account for autocorrelations. Therefore, a discussion of the finite sample distribution of the GMM specification test cannot be complete without specifying how the weighting matrix is estimated.

Using simulation evidence, JW report that the GMM specification test in the beta method, J_h, grossly under-rejects in finite samples, whereas the GMM specification test in the SDF method, J_g, has roughly the correct size in finite samples. Therefore, they conclude that the GMM specification test under the SDF method is more reliable. However, JW do not describe how the weighting matrix is estimated in their GMM specification tests. According to our analysis, it seems very

[21] Alternatively, one can avoid specifying the initial weighting matrix by using the continuous-updating GMM method suggested by Hansen, Heaton, and Yaron (1996).

likely that they use the identity matrix as the initial weighting matrix and report the test statistics from the second stage GMM.

The choice of the initial weighting matrix is crucial in determining the finite sample performance of the GMM specification test. Ideally, one would like the initial weighting matrix to be close to proportional to the optimal weighting matrix. In that case, the first stage GMM estimate will be more accurate, and the resulting estimate of the weighting matrix for the second stage GMM will also be more reliable. If the initial weighting matrix is not close to the optimal weighting matrix, however, then the first stage GMM estimate is volatile, and the result is that the estimate of the spectral density matrix in the first stage is also volatile. Since the weighting matrix in the second stage is the inverse of the spectral density matrix obtained in the first stage, a volatile estimate of the spectral density matrix often leads to a small weighting matrix for the second stage. As a result, the specification test statistic in the second stage GMM will tend to be small and this gives rise to the under-rejection problem.

Then why is the identity matrix a poor choice for the beta method, but not for the SDF method? To understand this, we consider the conditional homoskedasticity case. In this case, the asymptotic variance of the first N elements of the sample moments in the beta method, $\bar{h}$, is Σ, but the asymptotic variance of the second N elements of $\bar{h}$ is $(\mu^2 + \sigma^2)\Sigma$. When the factor has a mean and a variance that are similar to those of the excess return on a portfolio (as in the case of JW), $\mu^2 + \sigma^2$ is a very small number relative to one.[22] This implies that the second N elements of $\bar{h}$ are substantially less volatile than its first N elements. Therefore, a good initial weighting matrix should place significantly more weight on the second N elements than on the first N elements of $\bar{h}$. The identity matrix does not have this property and hence leads to a poor weighting matrix for the second stage GMM specification test.[23] On the other hand, while the identity matrix is not the same as the inverse of the asymptotic variance of the sample moments of the SDF method, $\bar{g}$, we do not have as serious differences in the scales of its elements as in the beta method, especially when the factor explains a lot of the variations of the returns. Therefore, using the identity matrix as the initial weighting matrix for the SDF method does not cause as serious a problem as it does

[22]Note that if the factor is standardized as in KZ, then we do not have this scaling problem in the moments because $\mu^2 + \sigma^2 = 1$. This explains why KZ do not find the under-rejection problem in the second stage GMM specification test of the beta method, even though they use the identity matrix as the initial weighting matrix.

[23]Another problem with using the identity matrix as the initial weighting matrix is that the test statistics are not invariant to rescaling of the factors.

for the beta method.[24]

We have two suggestions to take care of this under-rejection problem in the GMM specification test for the beta method. The first one is to use the third stage GMM specification test if the identity matrix is used as the initial weighting matrix. This is because the parameter estimates from the second stage GMM are more reliable than those from the first stage. As a result, the weighting matrix used in the third stage GMM will be less volatile than the weighting matrix used in the second stage GMM.[25] The second suggestion is to use a sample estimate of the optimal weighting matrix under the conditional homoskedasticity assumption as the initial weighting matrix. While such an initial weighting matrix is only optimal under the conditional homoskedasticity assumption, its use as the initial weighting matrix can still help to remove the under-rejection problem in the second stage GMM, even when there is conditional heteroskedasticity. Our simulation evidence in Section IV shows that both of our suggestions are quite effective in restoring the correct size for the GMM specification test under the beta method.

B. Local Power Analysis of Specification Tests

GMM specification tests are designed to detect model misspecifications. If the exact beta pricing model does not hold, one would prefer the GMM specification test that rejects the model with higher probability. In order to compare the relative performance of the GMM specification tests under the beta and the SDF methods, we need to examine the asymptotic power of the two tests when the model does not hold. To this end, we introduce the following sequences of local alternatives to the beta pricing model:

$$H_1 : E[r_t] = \beta\delta + T^{-\frac{1}{2}}\eta, \tag{43}$$

where η is an $N \times 1$ nonzero vector. Under this sequence of local alternatives, the mispricing disappears as T increases. However, the limiting distribution of the GMM specification test is no longer a central χ^2_{N-1} distribution, but instead a noncentral $\chi^2_{N-1}(\omega)$ distribution where ω is the

[24] Cochrane (2000b) shows by simulation that when the identify matrix is used as the initial weighting matrix, the first and the second stage GMM estimates of the parameter in the SDF method have about the same variance. This conclusion, however, does not hold true if the factor does not explain a lot of the variations in the returns on the test assets. In those situations, KZ demonstrates by simulation that the performance of the second stage GMM specification test in the SDF method can also be very poor.

[25] Ferson and Foerster (1994) find that, for the specification test, the one from the iterative GMM is better behaved than the one from the second stage GMM. Kan and Zhang (1999) show that the third stage GMM has better power than the second stage GMM in rejecting a misspecified model that contains a useless factor.

noncentrality parameter. The noncentrality parameter here is a natural measure of the asymptotic local power of the GMM specification test. The use of local alternatives for asymptotic power analysis is common in the statistics and econometrics literature. It has the advantage that the parameters remain well defined under the alternative. In addition, the limits of the weighting matrix and the derivative matrix remain the same across the null and the alternative.[26]

The following Proposition demonstrates that there is no difference in the asymptotic local power of the two GMM specification tests.

Proposition 4: *Under the sequences of local alternatives $H_1 : E[r_t] = \beta\delta + T^{-\frac{1}{2}}\eta$, where η is an $N \times 1$ nonzero vector, the GMM specification tests J_g and J_h both have a limiting noncentral $\chi^2_{N-1}(\omega)$ distribution with noncentrality parameter*

$$\omega = \frac{\sigma^4(\eta' S_u^{-1}\eta)}{(\sigma^2 + \mu\delta)^2}\left[1 - \frac{(\eta' S_u^{-1}\beta)^2}{(\eta' S_u^{-1}\eta)(\beta' S_u^{-1}\beta)}\right], \tag{44}$$

where S_u is the spectral density matrix of $u_t(\lambda) = r_t(1 - f_t\lambda)$ under the true parameter.

Based on the proof of Proposition 4, we can also show that, by dropping the augmented moments in the SDF and the beta methods as in KZ, the limiting distribution of the GMM specification tests has the same noncentrality parameter under the local alternatives. Therefore, the GMM specification tests presented in KZ also have the same asymptotic local power as J_g and J_h, and this does not depend on whether the mean and the variance of the factor are known.

To gain more insights into what determines the asymptotic local power of the GMM specification test, we consider the case where (ϵ_t, f_t) is i.i.d. and conditional homoskedastic. In this case, the noncentrality parameter can be simplified to

$$\omega = \frac{\sigma^2(\eta'\Sigma^{-1}\eta)}{\sigma^2 + \delta^2}\left[1 - \frac{(\eta'\Sigma^{-1}\beta)^2}{(\beta'\Sigma^{-1}\beta)(\eta'\Sigma^{-1}\eta)}\right] = \frac{\sigma^2(\eta'\Sigma^{-1}\eta)}{\sigma^2 + \delta^2}\left[1 - R^2_{GLS}(\eta, \beta)\right], \tag{45}$$

where $R^2_{GLS}(\eta, \beta)$ is the noncentral generalized least squares (GLS) coefficient of determination between η and β. From this expression, we can see that there are two determinants of the asymptotic local power: (1) the magnitude of $\Sigma^{-\frac{1}{2}}\eta$ as measured by its norm, (2) how close is $\Sigma^{-\frac{1}{2}}\eta$ to $\Sigma^{-\frac{1}{2}}\beta$. Keeping $\eta'\Sigma^{-1}\eta$ constant, the GMM specification test has little power in rejecting the beta pricing model if η is almost proportional to β. On the other hand, if $\Sigma^{-\frac{1}{2}}\eta$ is orthogonal to $\Sigma^{-\frac{1}{2}}\beta$, then

[26] See Newey (1985) and Hall (1999) for a discussion of local power analysis in the GMM setting.

the GMM specification test will have very good power to reject the asset pricing model. However, it is extremely important to emphasize that the distance between η and β here must be measured by the R^2_{GLS} and not by the usual correlation.[27]

Note that although the third and the fourth moments of the factor are important in determining the asymptotic variance of $\hat{\lambda}$, they do not play an important role here in determining the asymptotic local power of the GMM specification tests. This is because the alternatives that we study here have misspecifications on only the beta pricing model but not on the constant mean and variance assumption of the factor.

C. Specification Tests Under Fixed Alternatives

Although the analysis of local alternatives is important, one may consider this kind of alternatives quite limited in their applications, and choose to perform a power analysis under a fixed alternative. The fixed alternative that we are most interested in is $H_2 : \alpha \equiv E[r_t] \neq \beta\delta$ for any δ, where α is a fixed vector. This is also the fixed alternative studied by JW. Asymptotically, both J_g and J_h diverge to infinity under H_2 instead of having a noncentral χ^2_{N-1} distribution as in the case of local alternatives. This is because, for any choice of parameters θ and β, we have $E[g_t(\theta)] \neq 0_{N+2}$ and $E[h_t(\theta,\beta)] \neq 0_{2N+2}$, and they are equal to some nonzero constant vectors. Therefore, for any weighting matrix W_g that is $O_p(1)$, we have $\bar{g}(\theta)'W_g\bar{g}(\theta) = O_p(1)$ for any θ and hence $J_g = T\bar{g}(\theta)'W_g\bar{g}(\theta) = O_p(T)$. Similarly, we have $J_h = O_p(T)$ under H_2. Therefore, both GMM specification tests will reject the asset pricing model with probability one asymptotically.

In order to have a meaningful comparison of the power of the two GMM specification tests under fixed alternatives, one needs to compare their performance in finite samples. In contrast to our analytical analysis of the local alternatives, simulation appears to be the only tractable approach for fixed alternatives. While simulation evidence on the relative power of J_g and J_h will be provided in Section IV, we point out some general issues regarding the power comparison here. In a manner similar to the analysis of the size of the two tests, the choice of initial weighting matrix is crucial in determining the power of the tests in finite samples. When the second stage GMM specification test is used with the identity matrix as the initial weighting matrix, we often find that $J_g > J_h$ in

[27] Chen, Kan, and Zhang (1999) provide a more complete analysis on this difference and its impact on model selection in multivariate regressions.

finite samples. However, we have to emphasize that this inequality by no means implies the GMM specification test in the SDF method is more powerful than the one in the beta method. This is because the inequality holds even under the null hypothesis. Hence, the inequality only implies that the two tests have different sizes when we use the asymptotic χ^2_{N-1} distribution. In evaluating the power of J_g and J_h, it is important to ensure that both of them have the correct size under the null hypothesis. Therefore, instead of reporting the unadjusted power function based on the asymptotic distribution, we report the size-adjusted power function of J_g and J_h in our simulations so we can have a fair comparison of their power in finite samples.

Under fixed alternatives, JW suggest another way to detect model misspecifications is to look at the pricing errors of the test assets. Their theoretical analysis suggests that the pricing errors from the beta method are less volatile than the ones from the SDF method. Their simulation results do not provide support for their analytical results, however. It turns out that pricing error is not a well defined concept under the setup of their empirical asset pricing model. This is because, unlike the calculation of Jensen's alpha, that the risk premium is defined as the excess return on the market portfolio, the risk premium and the SDF parameter in their model are not specified on an *ex ante* basis. Under fixed alternatives, there is no choice of parameter λ that will make the SDF moment conditions correct, so it is not entirely clear how we define λ under the alternative. Similarly, there is no δ that will satisfy the moment conditions in the beta method. Therefore, defining population pricing errors under both methods is problematic without first defining these two parameters under the fixed alternatives.

Under a fixed alternative, one could define δ and λ as the limit of their sample estimates. Such a limit will depend on the choice of the weighting matrix, however. To make our discussion more concrete, we consider the case of estimating λ using the standard SDF moment conditions $E[r_t(1-f_t\lambda)] = 0_N$. For a weighting matrix W_T with nonstochastic limit W, the sample estimate of $\hat{\lambda}$ is simply

$$\hat{\lambda} = (\overline{rf}' W_T \overline{rf})^{-1}(\overline{rf}' W_T \bar{r}), \tag{46}$$

where $\overline{rf} = \frac{1}{T}\sum_{t=1}^T r_t f_t$ and $\bar{r} = \frac{1}{T}\sum_{t=1}^T r_t$. Under the alternative $H_2 : E[r_t] = \alpha$, it is easy to verify that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T r_t f_t = E[r_t f_t] = E[(\alpha + \beta(f_t - \mu) + \epsilon_t) f_t] = \mu\alpha + \sigma^2\beta, \tag{47}$$

and hence

$$\lim_{T\to\infty} \hat{\lambda} = \frac{(\mu\alpha + \sigma^2\beta)'W\alpha}{(\mu\alpha + \sigma^2\beta)'W(\mu\alpha + \sigma^2\beta)} \equiv \lambda^W. \tag{48}$$

When $\alpha = \beta\delta$, then we have $\lambda^W = \lambda$, and it is independent of the choice of W. However, when $\alpha \neq \beta\delta$ for any δ, then the limit of $\hat{\lambda}$ depends on the choice of the weighting matrix. Therefore, the limit of $\hat{\lambda}$ in the first stage GMM is in general not the same as its limit in the second and the subsequent stages of GMM. Similarly, the optimal weighting matrix is also not well defined under the fixed alternatives. Using $\hat{\lambda}$ from different stages of GMM will get us different weighting matrices that converge to different limits. In general, under a fixed alternative, (4) and (5) will not hold for the limit of the estimated parameters. Therefore, JW's theoretical analysis of pricing errors, which crucially depends on the validity of (4) and (5), is problematic.[28] Under the multivariate normality assumption on (ϵ_t, f_t), we have worked out the limiting values of the pricing errors for different stages of GMM under both methods. However, in view of the difficulty of justifying pricing errors analysis under our framework, we do not present the results here. They are available upon request.

IV. Simulation Results

A. Choice of Parameters

In this section, we perform a simulation experiment to study the finite sample performance of the SDF and the beta methods. In our simulation, we generate excess returns on ten test assets using a one-factor model

$$r_t = \alpha + \beta(f_t - \mu) + \epsilon_t. \tag{49}$$

The parameters of the ten test assets and the common factor are chosen to match the corresponding sample moments of the ten size-ranked portfolios and the value-weighted market portfolio of the combined NYSE-AMEX-NASDAQ estimated over the period January 1926 to December 1999. Panel B of Table I summarizes our choice of parameters. Under the alternative hypothesis, we generate excess returns on the ten assets by setting $\alpha = \bar{r}$, where $\bar{r}$ is the average excess return on the ten size-ranked portfolios, reported in Panel A of Table I. Under the null hypothesis, we set

[28]It should be noted that JW's comparison is still problematic under the correct model. This is because the pricing errors under the beta and the SDF methods are not the same object, so simply comparing the asymptotic variance of pricing errors under the two methods does not tell us which method is better, just as one cannot simply compare Avar$[\hat{\delta}]$ with Avar$[\hat{\lambda}]$ to determine whether inference on δ is better than inference on λ.

$\alpha = \beta\delta$, where $\delta = 0.7712$ is chosen as

$$\delta = \text{argmin}_\delta (\bar{r} - \beta\delta)'(\bar{r} - \beta\delta), \tag{50}$$

and the SDF parameter λ is determined by (5). Our simulation experiment is almost identical to the one performed by JW, except for our choices of δ and λ. JW choose δ_0 and δ to minimize

$$(\bar{r} - \delta_0 - \beta\delta)'(\bar{r} - \delta_0 - \beta\delta). \tag{51}$$

Their choice of δ is 1.3740, which is almost twice as big as ours. However, with their choice of δ, the expected excess returns on their test assets (which are set to $\beta\delta$, but not $\delta_0 + \beta\delta$) are much higher than the average excess returns on the ten size-ranked portfolios.

B. Parameter Estimation

In this subsection, we present the parameter estimation results under the correctly specified model. We present the maximum likelihood and GMM estimation results under the beta method, as well as the GMM estimation results under the SDF method. In all our GMM estimations, we compute the estimated spectral density matrix as the sample covariance matrix of the moments, and we do not make any adjustment for autocorrelations. For example, we compute the weighting matrix W_g under the SDF method as

$$W_g = \left[\frac{1}{T} \sum_{t=1}^{T} (g_t(\hat{\theta}_g) - \bar{g}(\hat{\theta}_g))(g_t(\hat{\theta}_g) - \bar{g}(\hat{\theta}_g))' \right]^{-1} \tag{52}$$

where $\bar{g}(\hat{\theta}_g) = \frac{1}{T} \sum_{t=1}^{T} g_t(\hat{\theta}_g)$. Although we consider this estimation method quite reasonable in our context, it is entirely possible that other variations of estimating the weighting matrix could give totally different results from what we report here. Therefore, when interpreting our simulation results, one should bear in mind that they could be specific to our particular choice of estimating the weighting matrix.

In Table II, we provide the parameter estimation results for the case that (ϵ_t, f_t) are simulated from a multivariate normal distribution. We report the variance of estimates of δ and λ under the beta and the SDF methods for different lengths of time series observations in 10,000 simulations. For the GMM estimations, we report the variance of three estimators. The first two estimators are from the second and the third stage GMM, and use the identity matrix as the initial weighting matrix. The third estimator is from the second stage GMM, and uses the sample estimates of the optimal weighting matrix as the initial weighting matrix, where the optimal weighting matrix is derived under the assumption that (ϵ_t, f_t) is i.i.d. and conditional homoskedastic. The details on how to compute the sample estimates of the optimal weighting matrix are given in the Appendix.

Table II about here

Although under the joint normality assumption on the residuals and the factor, the asymptotic variance of various estimators of δ and λ are the same, their finite sample performance could differ significantly from each other. Table II shows that, in finite samples, the maximum likelihood estimator dominates all the GMM estimators, especially when T is small. As for the GMM estimators, we find that, when T is small, the second stage GMM estimator under the beta method is particularly volatile if the identity matrix is used as the initial weighting matrix. Therefore, one should be cautious in using this GMM estimator. Once we use the third stage GMM or change the initial weighting matrix to a sample estimate of the optimal weighting matrix, however, then there is very little difference between the performance of the GMM estimates from the beta and from the SDF methods.

Table II also reports the asymptotic relative efficiency of the GMM estimates of δ to λ. It shows that, under our choice of parameters and the assumption of joint normality on residuals and factor, inference using $\hat{\delta}$ is 3.45% less efficient than inference using $\hat{\lambda}$ when T is large. This is because, for a normal factor, it is possible that $\hat{\lambda}$ is more efficient than $\hat{\delta}$, if δ and μ have the same sign. If μ is chosen to be a smaller number or negative, then we will find inference using $\hat{\delta}$ to be more efficient than inference using $\hat{\lambda}$.

Table III reports a set of results similar to Table II, except that we replace the normal factor in Table II with a factor that has a Student-t distribution with five degrees of freedom. Under this assumption, the maximum likelihood estimators of δ and λ are asymptotically more efficient than their GMM estimators. This is also true in finite samples; we typically find that the maximum likelihood estimators are about 20% less volatile than the corresponding GMM estimators. As in Table II, we find that when T is small, the second stage GMM estimators of δ and λ under the beta method are more volatile if the identity matrix is used as the initial weighting matrix. Except in this case, there is no important difference between the performance of various GMM estimators.

Table III about here

By comparing the variance of the GMM estimates of λ in Table III with those in Table II, we observe that the kurtosis of the t-distributed factor makes the estimate of λ less reliable. By comparing the variance of GMM estimates of δ in Tables II and III, however, we find that the kurtosis of the factor does not have much impact on the GMM estimates of δ. Asymptotically, when the factor has a t-distribution with five degrees of freedom, we find that inference using $\hat{\delta}$ is about 7.92% more efficient than inference using $\hat{\lambda}$.

It should be noted that a t-distribution with five degrees of freedom has a kurtosis of six, but such kurtosis is still smaller than the sample kurtosis of 7.989 for the excess return on the value-weighted market portfolio as reported in Table I. To increase the kurtosis, one could reduce the degrees of freedom of the t-distribution. However, when the degrees of freedom are less than or equal to four, the kurtosis of the Student-t distribution is infinity. In this case, the asymptotic variance of $\hat{\delta}$ remains the same as in the normal factor case, but the asymptotic variance of $\hat{\lambda}$ becomes infinity. Simulation results (not reported) show that $\hat{\delta}$ has roughly the same distribution as in Table II, but $\hat{\lambda}$ is totally unreliable in finite samples. This case illustrates that, although the beta model and the linear SDF model are theoretically equivalent, their empirical setup can make a big difference when it comes to inference on the parameters.

To allow for a higher kurtosis in the common factor, we choose to model the factor as a mixture of two normal distributions with a common mean and different variances. The parameters of w, σ_1 and σ_2 are chosen to match exactly the sample absolute moment, the sample variance, and the sample kurtosis of the excess return on the value-weighted market portfolio of the NYSE-AMEX-NASDAQ. The parameter values we choose are $w = 0.0782$, $\sigma_1 = 0.1413$, and $\sigma_2 = 0.0399$. One could think that the monthly excess returns on the market portfolio are drawn from two different populations, with 7.82% of the time that the market has a high volatility month, and 92.18% of the time that the market has a low volatility month. Table IV reports the estimation results when the residuals have a multivariate normal distribution and the factor has the chosen normal mixture distribution.

Table IV about here

The higher kurtosis of the estimation results of the normal mixture factor has two effects on the estimation results. The first effect is that it allows for more efficient maximum likelihood estimation. In both asymptotic and finite samples, we find the maximum likelihood estimators of δ and λ to be at least 30% less volatile than their GMM estimators. The second effect is that the GMM estimators of λ become more volatile than the normal factor case. When T is large, inference using $\hat{\delta}$ is 11.64% more efficient than inference using $\hat{\lambda}$. Similar to the results in Tables II and III, we find that when T is small, the second stage GMM estimates under the beta methods are more volatile if the identity matrix is used as the initial weighting matrix. All the other GMM estimators have a similar performance.

To gain a better understanding of how kurtosis determines the asymptotic relative efficiency of $\hat{\delta}$ to $\hat{\lambda}$, we plot the asymptotic relative efficiency of the two GMM estimators as a function of the kurtosis of the factor, assuming conditional homoskedasticity and a symmetric factor (i.e., $\mu_3 = 0$), with the parameters μ, σ^2, β and Σ given in Table I. We plot the function for two different values of δ. One value of δ is 0.7712, which is what we use in our simulations. The other value of δ is 1.3740, which is what JW use in their simulations.

Figure 1 about here

As we can see from Figure 1, the asymptotic relative efficiency of $\hat{\delta}$ to $\hat{\lambda}$ is a linear and increasing function of the kurtosis of the factor. For a factor with similar kurtosis as the value-weighted market portfolio, we have an ARE of 1.116 for $\delta = 0.7712$ but an ARE of 1.294 for $\delta = 1.4140$. However, if the factor we use has the same kurtosis as the excess return of the smallest size decile, then the ARE is 1.373 for $\delta = 0.7712$, and 2.203 for $\delta = 1.3740$.

In summary, our simulation results suggest that, for parameter estimation, the maximum likelihood estimators of δ and λ are superior to the GMM estimators. For the estimation of the same parameter, the GMM estimation of the beta and the SDF methods have similar properties, except for the second stage GMM under the beta method, when the identity matrix is used as the initial weighting matrix. As for the relative efficiency of making inference on δ to inference on λ, we find that inference based on λ is slightly superior under our choice of parameters, if the factor is normally distributed. However, when the factor has a kurtosis similar to what we find in real data, making inference on δ is more reliable than making inference on λ.

C. Size and Power of Specification Tests

In this subsection, we present simulation results related to the size and the power of the likelihood ratio and the GMM specification tests. Since the simulation results for the nonnormal factor cases are very similar to the results for the normal factor, we do not report those results here but they are available upon request. Table V presents the actual sizes of the tests under the correctly specified model when the asymptotic sizes of the test are 5% and 10% respectively. The results are based on 10,000 simulations, with (ϵ_t, f_t) drawn from a multivariate normal distribution. As we can see in Table V, the likelihood ratio test has a size that is closest to the level of significances of the test in small samples. All the GMM specification tests have similar size properties except for the second stage GMM under the beta method, when the identity matrix is used as the initial weighting matrix. In that case, the GMM specification test grossly under-rejects, and this problem does not disappear even when T is as high as 720. As discussed earlier in the text, the identity matrix is a poor choice of the initial weighting matrix for the beta method, and the second stage GMM specification test, which uses the weighting matrix obtained after the first stage, is not very reliable. Therefore, one should avoid using the second stage GMM under the beta method if the identity matrix is used as the initial weighting matrix. As suggested earlier and shown here by simulations, using the third stage GMM or using the sample estimate of the optimal weighting matrix as the initial weighting matrix will take care of this under-rejection problem.

Table V about here

Table VI reports the power of the likelihood ratio and the GMM specification tests under the i.i.d. multivariate normality assumption of (ϵ_t, f_t) in 10,000 simulations. The returns are now generated from a one-factor model, but the expected returns are not set to $\beta\delta$. Instead, the expected returns are set to the average returns on the ten size-ranked portfolios. In presenting the power of the tests, we use the empirical distribution of the test statistics obtained from the simulation in Table V to make the acceptance/rejection decision. Using the empirical distribution instead of the asymptotic χ^2_{N-1} distribution has the advantage that the tests have the correct size under the null hypothesis. From Table VI, we can see that, for the fixed alternative that we choose, the power of different tests is remarkably similar, with the exception of the second stage GMM under the beta method, when the identity matrix is used as the initial weighting matrix.

Table VI about here

In Table VI, we find that under the multivariate normality assumption of (ϵ_t, f_t), the likelihood ratio test is not any more powerful than the GMM specification tests in rejecting a misspecified model. A similar pattern is also observed for cases with nonnormal factors, as long as ϵ_t is still multivariate normally distributed. However, for situations where ϵ_t is not normally distributed, the likelihood ratio test starts to show its dominance over the GMM specification tests. To illustrate this, we report in Table VII the power of the likelihood ratio and the GMM specification tests when (ϵ_t, f_t) are sampled from a multivariate Student-t distribution with five degrees of freedom, using the same fixed alternative as in Table VI.[29] As we can observe from Table VII, the likelihood ratio test clearly dominates the GMM specification tests when (ϵ_t, f_t) are multivariate Student-t distributed. Therefore, besides parameter estimation, there can also be advantages in using the maximum likelihood method to detect model misspecifications.

Table VII about here

In summary, our simulation results suggest one should stay away from the second stage GMM specification test under the beta method if the identity matrix is used as the initial weighting matrix and the factor is not standardized. This test has the wrong size and the lowest power. However, if we use the third stage GMM or a more appropriate initial weighting matrix, then we find that there is no obvious disadvantage in using the GMM specification test in the beta method as compared with the GMM specification test in the SDF method. In addition, under the conditional homoskedasticity assumption, nonnormality of factors have very little influence on the size and power of GMM specification tests in the two methods. When the residuals are multivariate normally distributed, likelihood ratio test has similar power to the GMM specification tests, but as long as the residuals are not multivariate normally distributed, the likelihood ratio test can be more powerful than the GMM specification tests.

[29] Details of the computation of the likelihood ratio test for multivariate Student-t distribution are discussed in Kan and Zhou (2001a).

V. Conclusions

As a theoretical model, there is no doubt that the SDF framework is elegant and general. As empirical models, however, the SDF and the beta models both require ancillary assumptions to be tested, and one is not any more general than the other. In many situations, they make different assumptions and it is not possible to compare inferences under the two methods directly. Under the mainstream constant beta pricing model, with a constant mean and variance for the common factor, the two methods can be compared, and we show that inference based on the risk premium parameter in the beta method can be superior to inference based on the SDF parameter in the SDF method, especially when the factor is leptokurtic. While it is true that, when the mean of the factor is unknown, we no longer have the same overwhelming advantage in the beta method as demonstrated in KZ, the advantage of using the beta method for inference can still be substantial. In addition, we show that the maximum likelihood estimation of the beta model can still strictly dominate the GMM estimation of the beta and the SDF models even when the residuals are i.i.d. multivariate normally distributed. When it comes to detecting misspecifications in the linear beta pricing model, the GMM specification test in the beta method has a similar power to the one in the SDF method but the likelihood ratio test can still dominate the GMM specification tests when the residuals are not normally distributed.

In general, we do not see the SDF and the beta methods as competing but complementary. We see the SDF method as just an alternative empirical model but not as the Holy Grail. While the SDF method can incorporate information and deal with nonlinear models, we note that the beta method can incorporate information just as easily, and can also take care of nonlinear models after linearization (as is frequently done under the SDF framework).[30] It is important for future research to compare the performance of these two methods in dealing with conditional and nonlinear asset pricing models. In any event, it appears safe to say that the traditional methodologies are here to stay.

[30]Examples of incorporating information in the beta method include Shanken (1990), Ferson and Harvey (1991, 1993), and Ferson and Korajczyk (1995).

Appendix

Analytical solution to the GMM estimator under the SDF method

The GMM estimator of $\theta = (\lambda, \mu, \sigma^2)'$ is the solution to the minimization problem

$$\min_{\theta} Q(\theta) = \min_{\theta} \bar{g}(\theta)' W_g \bar{g}(\theta), \tag{A1}$$

where W_g is a positive definite matrix and $\bar{g}(\theta)$ is given by

$$\bar{g}(\theta) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}(r_t - r_t f_t \lambda) \\ \frac{1}{T}\sum_{t=1}^{T}(f_t - \mu) \\ \frac{1}{T}\sum_{t=1}^{T}(f_t^2 - 2f_t\mu + \mu^2 - \sigma^2) \end{bmatrix}. \tag{A2}$$

Conditional on a given value of μ, $\bar{g}$ is a linear function of $\theta_1 = (\lambda, \sigma^2)'$, so we can obtain an analytical solution for them. Denote

$$X = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} r_t f_t & 0_N \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad Y = \begin{bmatrix} \bar{r} \\ \bar{f} - \mu \\ \frac{1}{T}\sum_{t=1}^{T} f_t^2 - 2\bar{f}\mu + \mu^2 \end{bmatrix}, \tag{A3}$$

where $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$ and $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$, we can write $\bar{g}(\theta_1|\mu) = Y - X\theta_1$. Conditional on a given value of μ, the optimal solution of θ_1 is given by

$$\hat{\theta}_1(\mu) = (X'W_gX)^{-1}(X'W_gY), \tag{A4}$$

and the objective function can then be written as a function of μ alone as

$$Q(\mu) = Y'HY, \tag{A5}$$

where $H = W_g - W_gX(X'W_gX)^{-1}X'W_g$. Writing $Y = Y_0 + Y_1\mu + Y_2\mu^2$, where

$$Y_0 = \begin{bmatrix} \bar{r} \\ \bar{f} \\ \frac{1}{T}\sum_{t=1}^{T} f_t^2 \end{bmatrix}, \qquad Y_1 = \begin{bmatrix} 0_N \\ -1 \\ -2\bar{f} \end{bmatrix}, \qquad Y_2 = \begin{bmatrix} 0_{N+1} \\ 1 \end{bmatrix}, \tag{A6}$$

the objective function can now be written as a quartic polynomial of μ

$$Q(\mu) = Y_0'HY_0 + 2Y_0'HY_1\mu + (Y_1'HY_1 + 2Y_0'HY_2)\mu^2 + 2Y_1'HY_2\mu^3 + Y_2'HY_2\mu^4. \tag{A7}$$

Taking derivative, we have $\hat{\mu}$ as one of the solutions to the equation

$$Q'(\mu) = 2Y_0'HY_1 + 2(Y_1'HY_1 + 2Y_0'HY_2)\mu + 6Y_1'HY_2\mu^2 + 4Y_2'HY_2\mu^3 = 0 \tag{A8}$$

Note that this is a cubic equation in μ and there can be as many as three real roots, so one has to find out all the real roots and check which one is the global minimum.

Near analytical solution to the GMM estimator under the beta method

Although numerically identical, we reparameterize the parameters to $\vartheta = (\phi, \beta', \mu, \sigma^2)'$ where $\phi = \delta - \mu$ for convenience. The GMM estimator of ϑ is the solution to the minimization problem

$$\min_{\vartheta} Q(\vartheta) = \min_{\vartheta} \bar{h}(\vartheta)' W_h \bar{h}(\vartheta), \tag{A9}$$

where W_h is a positive definite matrix and $\bar{h}(\vartheta)$ is given by

$$\bar{h}(\vartheta) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}(r_t - (\phi + f_t)\beta) \\ \frac{1}{T}\sum_{t=1}^{T}(r_t - (\phi + f_t)\beta) f_t \\ \frac{1}{T}\sum_{t=1}^{T}(f_t - \mu) \\ \frac{1}{T}\sum_{t=1}^{T}(f_t^2 - 2f_t\mu + \mu^2 - \sigma^2) \end{bmatrix}. \tag{A10}$$

Conditional on a given value of ϕ and μ, $\bar{h}$ is a linear function of $\vartheta_1 = (\beta', \sigma^2)'$, so we can obtain an analytical solution for them. Denote

$$X = \begin{bmatrix} (\phi + \bar{f}) I_N & 0_N \\ \left(\phi \bar{f} + \frac{1}{T}\sum_{t=1}^{T} f_t^2\right) I_N & 0_N \\ 0_N' & 0 \\ 0_N' & 1 \end{bmatrix}, \qquad Y = \begin{bmatrix} \bar{r} \\ \frac{1}{T}\sum_{t=1}^{T} r_t f_t \\ \bar{f} - \mu \\ \frac{1}{T}\sum_{t=1}^{T} f_t^2 - 2\bar{f}\mu + \mu^2 \end{bmatrix}, \tag{A11}$$

where $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$, $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$ and I_N is an N-dimensional identity matrix, we can write $\bar{h}(\vartheta_1|\phi, \mu) = Y - X\vartheta_1$. Conditional on a given value of $(\phi,\ \mu)$, the optimal solution of ϑ_1 is given by

$$\hat{\vartheta}_1(\phi, \mu) = (X' W_h X)^{-1}(X' W_h Y), \tag{A12}$$

and the objective function can then be written as a function of ϕ and μ alone as

$$Q(\phi, \mu) = Y' H Y, \tag{A13}$$

where $H = W_h - W_h X(X' W_h X)^{-1} X' W_h$. Note that H is only a function of ϕ and Y is only a function of μ. Writing $Y = Y_0 + Y_1\mu + Y_2\mu^2$, where

$$Y_0 = \begin{bmatrix} \bar{r} \\ \frac{1}{T}\sum_{t=1}^{T} r_t f_t \\ \bar{f} \\ \frac{1}{T}\sum_{t=1}^{T} f_t^2 \end{bmatrix}, \qquad Y_1 = \begin{bmatrix} 0_{2N} \\ -1 \\ -2\bar{f} \end{bmatrix}, \qquad Y_2 = \begin{bmatrix} 0_{2N+1} \\ 1 \end{bmatrix}, \tag{A14}$$

the objective function can now be written as a quartic polynomial of μ when conditional on a given value of ϕ. Similar to the case of the SDF method, there can be as many as three real roots of μ to the first order condition, so one has to find out all the real roots and check which one is the global minimum. Nevertheless, conditional on a given value of ϕ, we can analytically minimize $Q(\vartheta_1, \mu|\phi)$. A line search on ϕ will then allow us to find the minimum of $Q(\vartheta)$.

Proof of Lemma 1: Denote

$$S = \sum_{j=-\infty}^{\infty} E[g_t(\phi)g_{t+j}(\phi)'] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \tag{A15}$$

and its inverse as

$$S^{-1} = \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix}, \tag{A16}$$

where the partition corresponds to the two blocks of moments in g_t. The derivative of the moment conditions is

$$D = E\left[\frac{\partial g_t(\phi)}{\partial \phi'}\right] = \begin{bmatrix} D_{11} & \mathrm{O}_{m_1 \times p_2} \\ D_{21} & D_{22} \end{bmatrix}, \tag{A17}$$

and it is assumed to be of full column rank under the usual regularity condition. If we use just the first m_1 moments to estimate ϕ_1, the asymptotic variance of the estimator $\hat{\phi}_1$ is

$$\text{Avar}[\hat{\phi}_1] = (D_{11}' S_{11}^{-1} D_{11})^{-1}. \tag{A18}$$

To obtain the asymptotic variance of $\hat{\phi}_1^*$, we write

$$D'S^{-1}D = \begin{bmatrix} A & D_{11}'S^{12}D_{22} + D_{21}'S^{22}D_{22} \\ D_{22}'S^{21}D_{11} + D_{22}'S^{22}D_{21} & D_{22}'S^{22}D_{22} \end{bmatrix}, \tag{A19}$$

where

$$A = D_{11}'S^{11}D_{11} + D_{21}'S^{21}D_{11} + D_{11}'S^{12}D_{21} + D_{21}'S^{22}D_{21}. \tag{A20}$$

Defining $H = D_{11}'S^{12}(S^{22})^{-\frac{1}{2}} + D_{21}'(S^{22})^{\frac{1}{2}}$ and using the identity $S_{11}^{-1} = S^{11} - S^{12}(S^{22})^{-1}S^{21}$ from the inverse of partitioned matrix formula, we can write

$$D'S^{-1}D = \begin{bmatrix} D_{11}'S_{11}^{-1}D_{11} + HH' & H(S^{22})^{\frac{1}{2}}D_{22} \\ D_{22}'(S^{22})^{\frac{1}{2}}H' & D_{22}'S^{22}D_{22} \end{bmatrix}. \tag{A21}$$

Therefore,

$$\begin{aligned} \text{Avar}[\hat{\phi}_1^*] &= \left[D_{11}'S_{11}^{-1}D_{11} + HH' - H(S^{22})^{\frac{1}{2}}D_{22}(D_{22}'S^{22}D_{22})^{-1}D_{22}'(S^{22})^{\frac{1}{2}}H'\right]^{-1} \\ &= \left(D_{11}'S_{11}^{-1}D_{11} + HMH'\right)^{-1}, \end{aligned} \tag{A22}$$

where

$$M = I_{m_2} - (S^{22})^{\frac{1}{2}}D_{22}(D_{22}'S^{22}D_{22})^{-1}D_{22}'(S^{22})^{\frac{1}{2}} \tag{A23}$$

is idempotent with rank $m_2 - p_2$. Since $\text{Avar}[\hat{\phi}_1^*]^{-1} \geq \text{Avar}[\hat{\phi}_1]^{-1}$, we have $\text{Avar}[\hat{\phi}_1^*] \leq \text{Avar}[\hat{\phi}_1]$. When $p_2 = m_2$, D_{22} is a square matrix and its inverse exists because D is assumed to have full column rank. In this case, M is a zero matrix and hence $\text{Avar}[\hat{\phi}_1^*] = \text{Avar}[\hat{\phi}_1]$. This completes the proof.

Proof of Proposition 1: We first show that the GMM estimator of $\phi = (\theta', \beta')'$ using m_t and h_t have the same asymptotic variance. Since

$$D_m \equiv E\left[\frac{\partial m_t(\phi)}{\partial \phi'}\right] = AE\left[\frac{\partial h_t(\phi)}{\partial \phi'}\right] = AD_h, \tag{A24}$$

and

$$S_m \equiv \sum_{j=-\infty}^{\infty} E[m_t(\phi) m_{t+j}(\phi)'] = A\left[\sum_{j=-\infty}^{\infty} E[g_t(\phi) g_{t+j}(\phi)']\right] A' = AS_h A', \tag{A25}$$

we have

$$\text{Avar}[\hat{\phi}_m] = (D_m' S_m^{-1} D_m)^{-1} = (D_h' A' (AS_h A')^{-1} AD_h)^{-1} = (D_h' S_h^{-1} D_h)^{-1} = \text{Avar}[\hat{\phi}_h], \tag{A26}$$

where $\hat{\phi}_m$ and $\hat{\phi}_h$ are the GMM estimators of ϕ using m_t and h_t, respectively. The equality follows because A is nonsingular and its inverse exists. Then, by writing $\phi_1 = \theta$ and $\phi_2 = \beta$, we have $\text{Avar}[\hat{\theta}_m] = \text{Avar}[\hat{\theta}_g]$ from Lemma 1, where $\hat{\theta}_m$ is the GMM estimator of θ using the moment conditions m_t. Therefore, we have $\text{Avar}[\hat{\theta}_h] = \text{Avar}[\hat{\theta}_m] = \text{Avar}[\hat{\theta}_g]$. This implies $\text{Avar}[\hat{\lambda}_h] = \text{Avar}[\hat{\lambda}_g]$ and applying the delta method, we also have $\text{Avar}[\hat{\delta}_h] = \text{Avar}[\hat{\delta}_g]$. This completes the proof.

Proof of Proposition 2: Due to Proposition 1, we need to provide only the derivation of $\text{Avar}[\hat{\theta}_g]$ under the SDF method here. Direct derivation of $\text{Avar}[\hat{\theta}_h]$ under the beta method is available upon request. Under the return generating model, we have

$$E[(1 - f_t\lambda)^2 r_t r_t'] = E[\epsilon_t \epsilon_t' (1 - f_t\lambda)^2] + E[(\delta - \mu + f_t)^2 (1 - f_t\lambda)^2]\beta\beta' = U + c\beta\beta', \tag{A27}$$

where

$$c = E[(\delta - \mu + f_t)^2 (1 - f_t\lambda)^2] = \frac{\sigma^2(\sigma^4 + \delta^4) + 2\delta(\delta^2 - \sigma^2)\mu_3 + \delta^2(\mu_4 - 3\sigma^4)}{(\sigma^2 + \mu\delta)^2} \tag{A28}$$

is obtained using the identities $E[(f_t - \mu)^2 f_t] = \mu_3 + \mu\sigma^2$, $E[(f_t - \mu) f_t^2] = \mu_3 + 2\mu\sigma^2$, and $E[(f_t - \mu)^2 f_t^2] = \mu_4 + \mu(2\mu_3 + \mu\sigma^2)$. Using these results and the i.i.d. assumption, the spectral density matrix of $g_t(\theta)$ is given by

$$S_g = E[g_t(\theta) g_t(\theta)'] = \begin{bmatrix} U + c\beta\beta' & \beta e' \\ e\beta' & H \end{bmatrix}, \tag{A29}$$

where

$$e = \begin{bmatrix} \frac{\sigma^2(\sigma^2-\delta^2)-\delta\mu_3}{\sigma^2+\mu\delta} \\ \frac{(\sigma^2-\delta^2)\mu_3-\delta(\mu_4-\sigma^4)}{\sigma^2+\mu\delta} \end{bmatrix}, \qquad H = \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4-\sigma^4 \end{bmatrix}. \tag{A30}$$

The expectation of $\partial g_t(\theta)/\partial\theta'$ is given by

$$D_g = \begin{bmatrix} -b\beta & \mathrm{O}_{N\times 2} \\ 0_2 & -I_2 \end{bmatrix}, \tag{A31}$$

where $b = \sigma^2 + \mu\delta$. With some matrix algebra, we can verify that

$$\mathrm{Avar}[\hat{\theta}_g] = (D_g' S_g^{-1} D_g)^{-1} = \begin{bmatrix} \frac{(\beta' U^{-1}\beta)^{-1}+c}{b^2} & \frac{e'}{b} \\ \frac{e}{b} & H \end{bmatrix}. \tag{A32}$$

$\mathrm{Avar}[\hat{\lambda}_g]$ is the $(1,1)$ element of $\mathrm{Avar}[\hat{\theta}_g]$ and $\mathrm{Avar}[\hat{\delta}_g]$ can be obtained using the delta method as

$$\mathrm{Avar}[\hat{\delta}_g] = \begin{bmatrix} \frac{b^2}{\sigma^2} & \frac{\delta^2}{\sigma^2} & \frac{\delta}{\sigma^2} \end{bmatrix} \begin{bmatrix} \frac{(\beta' U^{-1}\beta)^{-1}+c}{b^2} & \frac{e'}{b} \\ \frac{e}{b} & H \end{bmatrix} \begin{bmatrix} \frac{b^2}{\sigma^2} \\ \frac{\delta^2}{\sigma^2} \\ \frac{\delta}{\sigma^2} \end{bmatrix} = \frac{(\sigma^2+\mu\delta)^2}{\sigma^4}(\beta' U^{-1}\beta)^{-1} + \sigma^2. \tag{A33}$$

This completes the proof.

Proof of (27) and (28): Under the multivariate elliptical distribution, we have $\mu_3 = 0$, and by definition, we can write $\mu_4 - 3\sigma^4 = \gamma_2\sigma^4$. For the matrix U, we have

$$\begin{aligned} U &= E[(1-2\lambda f_t + \lambda^2 f_t^2)\epsilon_t\epsilon_t'] \\ &= \Sigma - 2\lambda E[f_t\epsilon_t\epsilon_t'] + \lambda^2 E[f_t^2\epsilon_t\epsilon_t'] \\ &= \Sigma - 2\lambda\mu\Sigma + \lambda^2\left[\mu^2\Sigma + \left(1+\frac{\gamma_2}{3}\right)\sigma^2\Sigma\right] \\ &= \left[(1-\lambda\mu)^2 + \left(1+\frac{\gamma_2}{3}\right)\lambda^2\sigma^2\right]\Sigma \\ &= \frac{\sigma^4}{(\sigma^2+\mu\delta)^2}\left[1 + \frac{\delta^2\left(1+\frac{\gamma_2}{3}\right)}{\sigma^2}\right]\Sigma. \end{aligned} \tag{A34}$$

The third equality uses the identity $E[(f_t-\mu)\epsilon_t\epsilon_t'] = \mathrm{O}_{N\times N}$, which follows from the symmetry property of multivariate elliptical distribution. It also uses the identity

$$E[f_t^2\epsilon_t\epsilon_t'] = \mu^2\Sigma + \left(1+\frac{\gamma_2}{3}\right)\sigma^2\Sigma, \tag{A35}$$

that is derived in Kan and Zhou (2001b). This completes the proof.

Proof of (29) and (30): Under conditional homoskedasticity, we have

$$U = E[(1-f_t\lambda)^2\epsilon_t\epsilon_t'] = E[(1-f_t\lambda)^2]\Sigma = [1-2\mu\lambda + (\mu^2+\sigma^2)\lambda^2]\Sigma = \frac{\sigma^2(\sigma^2+\delta^2)}{(\sigma^2+\mu\delta)^2}\Sigma. \tag{A36}$$

This completes the proof.

Maximum likelihood estimation under normal residuals and nonnormal factor

Denote the density function of f_t as $f(f_t;\theta_f)$, where θ_f is a vector of parameters. Note that μ and σ^2 are not necessarily the parameters of the distribution, but they can always be written as functions of the parameters θ_f. For convenience, we do a reparameterization of $\phi = \delta - \mu$. Let $\vartheta = (\phi, \beta', \mathrm{vech}(\Sigma)', \theta_f')'$ be the parameters; then we can write the log-likelihood function of $(r_1, f_1, \ldots, r_T, f_T)$ as

$$\begin{aligned}\mathcal{L}(\vartheta | r_1, f_1, \ldots, r_T, f_T) &= -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(r_t - \beta(\phi + f_t))'\Sigma^{-1}(r_t - \beta(\phi + f_t)) \\ &\quad + \sum_{t=1}^{T}\ln(f(f_t;\theta_f)). \end{aligned} \tag{A37}$$

Note that the likelihood function has two parts, the first part is only a function of ϕ, β and Σ, and the second part is only a function of θ_f. Therefore, we can maximize each of them individually. The first maximization problem is exactly the same as in KZ. Define $F = [f_1, f_2, \ldots, f_T]'$, $X = [1_T,\ F]$, $Y = [r_1,\ r_2,\ \ldots, r_T]'$, where 1_T is a T-vector of ones. Let $\xi_1 \geq \xi_2 > 0$ be the two eigenvalues of the 2×2 matrix $A = (X'X)^{-1}(X'Y)(Y'Y)^{-1}(Y'X)$. Under the normality assumption on ϵ_t, the maximum likelihood estimator of ϕ is given by

$$\hat{\phi}_{ML} = \frac{a_{12}}{\xi_1 - a_{11}}, \tag{A38}$$

where a_{ij} is the (i,j)th element of A. The maximum likelihood estimator of δ is then given by $\hat{\delta}_{ML} = \hat{\phi}_{ML} + \hat{\mu}_{ML}$, where $\hat{\mu}_{ML}$ is determined by the maximum likelihood estimator of θ_f obtained from maximizing the second part of the likelihood function. In general, maximization of the second part requires numerical optimization, but fast algorithms are available for many distributions of f_t.[31] For the likelihood ratio test of the model $E[r_t] \equiv \alpha = \beta\delta$, we note that the estimates of θ_f are the same under both the null and the alternative. The only difference between the null and the alternative is in the first part of the likelihood function. Therefore, the same likelihood ratio test in KZ applies, and it is given by

$$\mathrm{LRT} = -T\log(1 - \xi_2), \tag{A39}$$

[31] For Student-t distribution with unknown degrees of freedom, we use the ECME algorithm in Liu and Rubin (1995). For the normal mixture distribution, we use the EM algorithm by Dempster, Laird, and Rudin (1977). Details are available upon request.

which has an asymptotic χ^2_{N-1} distribution under the null hypothesis $H_0 : \alpha = \beta\delta$. This also makes it clear that the likelihood ratio test in KZ only depends on the normality assumption of ϵ_t but not on f_t, and it does not depend on whether the mean and the variance of the factor are known.

Proof of Proposition 3: As discussed above, the likelihood function has two parts, the first part depends on only ϕ, β and Σ, and the second part depends on only θ_f. Therefore, $\hat{\phi}_{ML}$ and $\hat{\mu}_{ML}$ are asymptotically independent, and we have $\text{Avar}[\hat{\delta}_{ML}] = \text{Avar}[\hat{\phi}_{ML}] + \text{Avar}[\hat{\mu}_{ML}]$. Under the normality assumption on ϵ_t, we have

$$\text{Avar}[\hat{\phi}_{ML}] = \left(1 + \frac{\delta^2}{\sigma^2}\right)(\beta'\Sigma^{-1}\beta)^{-1} \tag{A40}$$

using the same proof as in KZ. For Student-t distribution, simple manipulation of the results in Lange, Little, and Taylor (1989) gives

$$\text{Avar}[\hat{\mu}_{ML}] = \left(1 - \frac{6}{\nu^2 + \nu}\right)\sigma^2. \tag{A41}$$

For normal mixture distribution, it is easy to verify that $\partial\mathcal{L}/\partial\mu$ is an odd function of $f_t - \mu$, but all the other partial derivatives are even functions of $f_t - \mu$. Therefore, $\hat{\mu}_{ML}$ is asymptotically independent of $\hat{w}_{ML}$, $\hat{\sigma}^2_{1,ML}$ and $\hat{\sigma}^2_{2,ML}$, so the asymptotic variance of $\hat{\mu}_{ML}$ is simply given by the inverse of $-E[\partial^2\mathcal{L}/\partial\mu^2] = c$. Some tedious algebra can then show that

$$\int_0^\infty \frac{\phi_1(z)\phi_2(z)z^2}{w\phi_1(z) + (1-w)\phi_2(z)}\mathrm{d}z \le \frac{\sigma_1^2\sigma_2^2}{2\sigma^2}, \tag{A42}$$

where $\sigma^2 = w\sigma_1^2 + (1-w)\sigma_2^2$ is the variance of the factor, and equality holds if and only if $w = 0$ or 1, or $\sigma_1^2 = \sigma_2^2$.[32] With this result, it is easy to verify that $c > 1/\sigma^2$ if $0 < w < 0.5$ and $\sigma_1^2 \neq \sigma_2^2$. This completes the proof.

Proof of Proposition 4: We start off by proving the limiting distribution of the GMM specification test under the standard SDF method, which uses the moment condition $E[u_t(\lambda)] = E[r_t(1 - f_t\lambda)] = 0_N$. Under the sequences of local alternatives, we have

$$E[r_t(1 - f_t\lambda)] = E[(\beta(\delta - \mu + f_t) + T^{-\frac{1}{2}}\eta + \epsilon_t)(1 - f_t\lambda)] = (1 - \lambda\mu)T^{-\frac{1}{2}}\eta = \frac{\sigma^2}{\sigma^2 + \mu\delta}T^{-\frac{1}{2}}\eta. \tag{A43}$$

Invoking the results in Newey (1985), the GMM specification test of the standard SDF method is $\chi^2_{N-1}(\omega)$, where

$$\omega = \frac{\sigma^4}{(\sigma^2 + \mu\delta)^2}\eta'\left[S_u^{-1} - S_u^{-1}D_u(D_u'S_u^{-1}D_u)^{-1}D_u'S_u^{-1}\right]\eta, \tag{A44}$$

[32] Proof of this inequality is available upon request.

and D_u is given by

$$D_u = E\left[\frac{\partial u_t(\lambda)}{\partial \lambda}\right] = -E[r_t f_t] = -(\sigma^2 + \mu\delta)\beta. \tag{A45}$$

Substituting D_u into the expression for ω, we obtain (44).

To show that J_g and J_h also have a limiting distribution of $\chi^2_{N-1}(\omega)$, we need the following extension to Lemma 1, which is due to Eichenbaum, Hansen and Singleton (1988, equation (C.2)).

Lemma 2 *Under the same conditions as in Lemma 1, and defining $Q = [I_{m_1},\ \mathrm{O}_{m_1\times m_2}]$, we have*

$$S^{-1} - S^{-1}D(D'S^{-1}D)^{-1}D'S^{-1} = Q'[S_{11}^{-1} - S_{11}^{-1}D_{11}(D_{11}'S_{11}^{-1}D_{11})^{-1}D_{11}'S_{11}^{-1}]Q \tag{A46}$$

when $p_2 = m_2$.

Asymptotically, there is no difference using $m_t = Ah_t$ or using h_t, so we need only to show that the GMM specification test using m_t has a limiting distribution of $\chi^2_{N-1}(\omega)$. As we show in the main text, the first N elements of m_t are just u_t. Therefore, under the sequence of local alternatives, we have

$$E[m_t] = \begin{bmatrix} \frac{\sigma^2}{\sigma^2+\mu\delta}T^{-\frac{1}{2}}\eta \\ 0_{N+2} \end{bmatrix}. \tag{A47}$$

Using results in Newey (1985), the GMM specification test has a noncentrality parameter

$$\begin{aligned} \omega &= \frac{\sigma^4}{(\mu\delta+\sigma^2)^2}[\eta',\ 0_{N+2}']\left[S_m^{-1} - S_m^{-1}D_m(D_m'S_m^{-1}D_m)^{-1}D_m'S_m^{-1}\right][\eta',\ 0_{N+2}']' \\ &= \frac{\sigma^4}{(\mu\delta+\sigma^2)^2}[\eta',\ 0_{N+2}']Q'\left[S_u^{-1} - S_u^{-1}D_u(D_u'S_u^{-1}D_u)^{-1}D_u'S_u^{-1}\right]Q[\eta',\ 0_{N+2}']' \\ &= \frac{\sigma^4}{(\mu\delta+\sigma^2)^2}\eta'\left[S_u^{-1} - S_u^{-1}D_u(D_u'S_u^{-1}D_u)^{-1}D_u'S_u^{-1}\right]\eta, \end{aligned} \tag{A48}$$

where $Q = [I_N,\ \mathrm{O}_{N\times(N+2)}]$. The second equality follows from Lemma 2 because D_m is block triangular, and m_t has $m_2 = N+2$ more moment conditions than u_t, but also $p_2 = N+2$ more parameters. Proof for the case of J_g is identical. This completes the proof.

Proof of (45): Under the i.i.d. conditional heteroskedasticity assumption, we have $S_u = a\Sigma + c\beta\beta'$ from (A27) and (A36), where $a = \sigma^2(\sigma^2+\delta^2)/(\sigma^2+\mu\delta)^2$ and c is a constant scalar. Using

$$S_u^{-1} = \frac{1}{a}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\beta\beta'\Sigma^{-1}}{\beta'\Sigma^{-1}\beta + \frac{a}{c}}\right), \tag{A49}$$

it is easy to show that $(\beta' S_u^{-1}\beta)^{-1} = a(\beta'\Sigma^{-1}\beta)^{-1} + c$, and $(\beta' S_u^{-1}\beta)^{-1}\beta' S_u^{-1} = (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}$. Therefore, using (A45), we have

$$
\begin{aligned}
S_u^{-1} - S_u^{-1}D_u(D_u'S_u^{-1}D_u)^{-1}D_u'S_u^{-1} &= S_u^{-1}\left[I_N - \beta(\beta'S_u^{-1}\beta)^{-1}\beta'S_u^{-1}\right] \\
&= \frac{1}{a}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\beta\beta'\Sigma^{-1}}{\beta'\Sigma^{-1}\beta + \frac{a}{c}}\right)\left[I_N - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\right] \\
&= \frac{1}{a}\left[\Sigma^{-1} - \Sigma^{-1}\beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\right]. \qquad \text{(A50)}
\end{aligned}
$$

This completes the proof.

Computation of Sample Estimates of Optimal Weighting Matrix: Under the i.i.d. conditional homoskedasticity assumption, the spectral density weighting matrix under the beta method is given by

$$
S_h = \begin{bmatrix} \Sigma & \mu\Sigma & \mathrm{O}_{N\times 2} \\ \mu\Sigma & (\mu^2+\sigma^2)\Sigma & \mathrm{O}_{N\times 2} \\ \mathrm{O}_{2\times N} & \mathrm{O}_{2\times N} & H \end{bmatrix}, \qquad \text{(A51)}
$$

where H is defined in (A30). By replacing Σ, μ, σ^2 and H by their sample estimates, we obtain a consistent estimate of S_h, and hence its inverse can be used as a consistent estimate of the optimal weighting matrix.

Under the i.i.d. conditional homoskedasticity assumption, the spectral density weighting matrix under the SDF method is given by

$$
S_g = \begin{bmatrix} a\Sigma + c\beta\beta' & \beta e' \\ e\beta' & H \end{bmatrix}, \qquad \text{(A52)}
$$

where $a = \sigma^2(\sigma^2+\delta^2)/(\sigma^2+\mu\delta)^2$, c is defined in (A28), and e and H are defined in (A30). This presents a problem because a, c, and e all involve the parameter δ that cannot be directly estimated. However, it turns out that, under the SDF method, the following weighting matrix is just as optimal as S_g^{-1}

$$
W_g = \begin{bmatrix} \Sigma^{-1} & \mathrm{O}_{2N\times 2} \\ \mathrm{O}_{2\times 2N} & H^{-1} \end{bmatrix}. \qquad \text{(A53)}
$$

It is because with some algebra, we can verify

$$
(D_g'W_gD_g)^{-1}(D_g'W_gS_gW_gD_g)(D_g'W_gD_g)^{-1} = (D_g'S_g^{-1}D_g)^{-1}, \qquad \text{(A54)}
$$

so the asymptotic variance of the estimator is the same whether we use S_g^{-1} or W_g. Therefore, for the SDF method, our sample estimates of the optimal weighting matrix can be obtained by simply replacing Σ and H in (A53) with their sample estimates.

References

Bansal, Ravi, David A. Hsieh, and S. Viswanathan, 1993, A new approach to international arbitrage pricing, *Journal of Finance* 48, 1719–1747.

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.

Campbell, John Y., 1993, Intertemporal asset pricing without consumption data, *American Economic Review* 83, 487–512.

Campbell, John Y., 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.

Chen, Nai-fu, Richarad R. Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

Chen, Nai-fu, Raymond Kan, and Chu Zhang, 1999, A critique of the use of t-ratios in model selection, working paper, University of Toronto.

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

Cochrane, John H., 2000a, A rehabilitation of stochastic discount methodology, working paper, Graduate School of Business, University of Chicago.

Cochrane, John H., 2000b, A resurrection of the stochastic discount factor/GMM methodology, working paper, Graduate School of Business, University of Chicago.

Dempster, Arthur P., Nan M. Laird, Donald B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society* Series B 39, 1–38.

Eichenbaum, Martin S., Lars Peter Hansen, and Kenneth J. Singleton, 1988, A time series analysis of representative agent models of consumption and leisure choice under uncertainty, *Quarterly Journal of Economics* 103, 51–78.

Ferson, Wayne E., and Stephen Foerster, 1994, Finite sample properties of the generalized method of moments in tests of conditional asset pricing models, *Journal of Financial Economics* 36, 29–55.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

Ferson, Wayne E., and Campbell R. Harvey, 1993, The risk and predictability of international equity returns. *Review of Financial Studies* 6, 527–566.

Ferson, Wayne E., and Robert A. Korajczyk, 1995, Do arbitrage pricing models explain the predictability of stock returns, *Journal of Business* 68, 309–350.

Hall, Alastair R., 1999, Hypothesis testing in models estimated by GMM, in Laszlo Matyas, ed.: *Generalized Method of Moments Estimation* (Cambridge University Press, Cambridge).

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

Hansen, Lars Peter, John Heaton, and Amir Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262–280.

Huber, Peter, 1981, *Robust Statistics*, (Wiley, New York).

Jagannathan, Ravi, and Zhenyu Wang, 2001, Empirical evaluation of asset pricing models: a comparison of the SDF and beta methods, working paper, Northwestern University and Columbia University.

Kan, Raymond, and Chu Zhang, 1999, GMM tests of stochastic discount factor models with useless factors, *Journal of Financial Economics* 54, 103–127.

Kan, Raymond, and Guofu Zhou, 1999, A critique of the stochastic discount factor methodology, *Journal of Finance* 54, 1021–1048.

Kan, Raymond, and Guofu Zhou, 2001a, Tests of asset pricing models under elliptical distributions, working paper, University of Toronto and Washington University in St. Louis.

Kan, Raymond, and Guofu Zhou, 2001b, Tests of mean-variance spanning, working paper, University of Toronto and Washington University in St. Louis.

Lange, Kenneth L., Roderick J. A. Little, and Jeremy M. G. Taylor, 1989, Robust statistical modeling using the t distribution, *Journal of the American Statistical Association* 84, 881–896.

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

Liu Chuanhai, and Donald B. Rubin, 1995, ML estimation of the t distribution using EM and its extensions, ECM and ECME, *Statistica Sinica* 5, 19–39.

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.

Shanken, Jay, 1990, Intertemporal asset pricing: an empirical investigation, *Journal of Econometrics* 45, 99–120.

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

Zhou, Guofu, 1993, Asset pricing test under alternative distributions, *Journal of Finance* 48, 1925–1942.

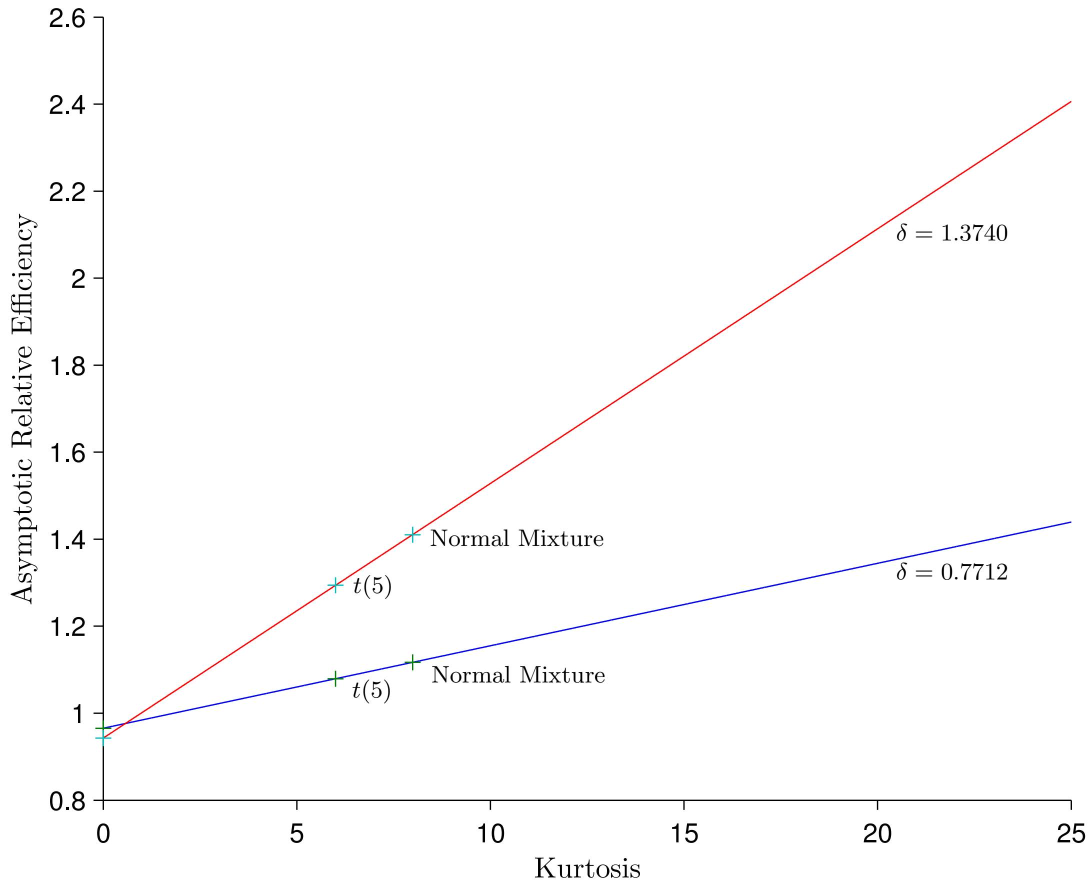

Figure 1
Asymptotic Relative Efficiency as a Function of Factor Kurtosis
The figure plots the asymptotic relative efficiency of $\hat{\delta}$ to $\hat{\lambda}$ as a function of kurtosis of the factor, where $\hat{\delta}$ is the GMM estimator of the risk premium parameter, and $\hat{\lambda}$ is the GMM estimator of the SDF parameter. The asymptotic relative efficiency is computed under two different values of risk premium parameter, δ, and it is based on the i.i.d. conditional homoskedasticity assumption with symmetrically distributed factors. The parameters for μ, σ^2, β and Σ are given in Table I.

Table I
Sample Statistics of the Value-Weighted Market and the Ten Size-Ranked Portfolios of the Combined NYSE-AMEX-NASDAQ

Panel A of the table presents the sample average, sample standard deviation, sample skewness ($\hat{\gamma}_1 = \hat{\mu}_3/\hat{\sigma}^3$) and kurtosis ($\hat{\gamma}_2 = \hat{\mu}_4/\hat{\sigma}^4 - 3$) of the excess returns (in excess of one-month T-bill rate) on the value-weighted market and ten size-ranked portfolios (from smallest to largest) of the combined NYSE-AMEX-NASDAQ. The sample statistics are obtained using monthly data over the period January 1926 to December 1999 from the Center for Research in Security Prices. Average excess return and sample standard deviation are reported in percentage per month. Panel B of the table presents the values of the parameters μ, σ, δ, λ, β and Σ, which are used in our simulation experiment. μ, σ and δ are reported as percentages. For the values of Σ, we present the standard deviation (σ_ϵ, in percentage) and the correlation matrix of the market model residuals from the 10 portfolios.

Panel A: Sample Statistics

		Size-Ranked Portfolios									
	VW	1	2	3	4	5	6	7	8	9	10
Average	0.703	1.537	1.120	0.958	0.893	0.879	0.867	0.847	0.754	0.797	0.668
Std. Dev.	5.501	10.914	9.413	8.517	7.880	7.470	7.270	6.872	6.437	6.151	5.235
Skewness	0.230	2.852	2.079	1.930	1.196	0.922	0.777	0.655	0.393	0.428	0.137
Kurtosis	7.989	21.543	18.232	19.673	11.890	10.636	9.273	10.024	8.560	9.228	7.264

Panel B: Values of Parameters in Simulations

$\mu = 0.703$ $\quad\sigma = 5.501$ $\quad\delta = 0.7712$ $\quad\lambda = 2.5037$

	Portfolios									
	1	2	3	4	5	6	7	8	9	10
β	1.449	1.392	1.311	1.264	1.239	1.226	1.182	1.122	1.091	0.947
σ_ϵ	7.457	5.482	4.535	3.714	3.069	2.726	2.234	1.830	1.366	0.557

Correlation Matrix of Market Model Residuals

1	1.000									
2	0.910	1.000								
3	0.848	0.910	1.000							
4	0.807	0.887	0.902	1.000						
5	0.764	0.839	0.853	0.899	1.000					
6	0.715	0.805	0.835	0.887	0.904	1.000				
7	0.608	0.681	0.714	0.807	0.832	0.851	1.000			
8	0.468	0.556	0.598	0.715	0.755	0.789	0.824	1.000		
9	0.382	0.436	0.457	0.536	0.596	0.651	0.664	0.748	1.000	
10	−0.672	−0.735	−0.752	−0.816	−0.835	−0.860	−0.837	−0.832	−0.763	1.000

Table II
Variance of Estimates of Risk Premium and SDF Parameters Under the Beta and the SDF Methods when Residuals and Factors are Normally Distributed

The table presents the variance of estimates of risk premium (δ, in percentage) and SDF parameter (λ) under the beta and the SDF methods. The returns and factors are generated under the null hypothesis by a one-factor model with parameters given in Table I, and with the factors and the residuals sampled from a multivariate normal distribution. For the GMM estimations, we present variance of three different GMM estimators for δ and λ. The first two estimators are from the second and third stage GMM when the identity matrix is used as the initial weighting matrix. The third estimator is from the second stage GMM when the sample estimate of the optimal weighting matrix is used as the initial weighting matrix. Results for the variance of the estimators are presented for different lengths of time series observations (T), and they are based on 10,000 simulations.

Variance of $\hat{\delta}$ ($\delta = 0.7712$)

	Beta Method				SDF Method		
		GMM Estimation			GMM Estimation		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.2518	0.6652	0.3734	0.3471	0.3448	0.3503	0.3490
240	0.1261	0.2140	0.1492	0.1472	0.1469	0.1477	0.1474
360	0.0836	0.1082	0.0927	0.0925	0.0924	0.0927	0.0926
480	0.0630	0.0721	0.0679	0.0680	0.0679	0.0680	0.0680
600	0.0503	0.0542	0.0533	0.0534	0.0534	0.0535	0.0534
720	0.0420	0.0440	0.0441	0.0442	0.0442	0.0442	0.0442
Asymptotic	30.288	30.288	30.288	30.288	30.288	30.288	30.288

Variance of $\hat{\lambda}$ ($\lambda = 2.5037$)

	Beta Method				SDF Method		
		GMM Estimation			GMM Estimation		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	2.6026	6.4832	3.9251	3.4845	3.4602	3.5600	3.4993
240	1.2931	2.1061	1.5406	1.4976	1.4940	1.5066	1.4989
360	0.8530	1.0637	0.9491	0.9399	0.9388	0.9425	0.9403
480	0.6430	0.7135	0.6942	0.6918	0.6914	0.6929	0.6919
600	0.5135	0.5391	0.5448	0.5440	0.5438	0.5446	0.5441
720	0.4285	0.4392	0.4501	0.4497	0.4496	0.4500	0.4498
Asymptotic	308.20	308.20	308.20	308.20	308.20	308.20	308.20

Asymptotic Relative Efficiency of $\hat{\delta}$ to $\hat{\lambda} = 0.9655$

Table III
Variance of Estimates of Risk Premium and SDF Parameters Under the Beta and the SDF Methods when Residuals are Normally Distributed and Factors are Student-t Distributed

The table presents the variance of estimates of risk premium (δ, in percentage) and SDF parameter (λ) under the beta and the SDF methods. The returns and factors are generated under the null hypothesis by a one-factor model with parameters given in Table I, with the factors sampled from a Student-t distribution with five degrees of freedom, and with the residuals sampled from a multivariate normal distribution. For the GMM estimations, we present variance of three different GMM estimators for δ and λ. The first two estimators are from the second and third stage GMM when the identity matrix is used as the initial weighting matrix. The third estimator is from the second stage GMM when the sample estimate of the optimal weighting matrix is used as the initial weighting matrix. Results for the variance of the estimators are presented for different lengths of time series observations (T), and they are based on 10,000 simulations.

Variance of $\hat{\delta}$ ($\delta = 0.7712$)

	Beta Method				SDF Method		
		GMM Estimation			GMM Estimation		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.2070	0.6461	0.3594	0.3408	0.3389	0.3440	0.3417
240	0.1018	0.2040	0.1471	0.1458	0.1455	0.1465	0.1459
360	0.0675	0.1040	0.0913	0.0921	0.0920	0.0923	0.0921
480	0.0501	0.0679	0.0658	0.0668	0.0667	0.0669	0.0668
600	0.0400	0.0514	0.0516	0.0524	0.0524	0.0525	0.0524
720	0.0332	0.0416	0.0427	0.0433	0.0433	0.0434	0.0433
Asymptotic	24.236	30.288	30.288	30.288	30.288	30.288	30.288

Variance of $\hat{\lambda}$ ($\lambda = 2.5037$)

	Beta Method				SDF Method		
		GMM Estimation			GMM Estimation		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	2.2730	9.4160	4.5588	3.8034	3.7593	3.8772	3.7954
240	1.1365	2.3654	1.6954	1.6362	1.6291	1.6461	1.6338
360	0.7577	1.3262	1.2764	1.0286	1.0266	1.0317	1.0281
480	0.5682	0.7989	0.8514	0.7462	0.7455	0.7475	0.7459
600	0.4546	0.6322	0.5737	0.5849	0.5847	0.5855	0.5849
720	0.3788	0.5460	0.4967	0.4845	0.4843	0.4848	0.4844
Asymptotic	272.75	344.50	344.50	344.50	344.50	344.50	344.50

Asymptotic Relative Efficiency of $\hat{\delta}$ to $\hat{\lambda} = 1.0792$

Table IV

Variance of Estimates of Risk Premium and SDF Parameters Under the Beta and the SDF Methods when Residuals are Normally Distributed and Factors are Normal Mixture Distributed

The table presents the variance of estimates of risk premium (δ, in percentage) and SDF parameter (λ) under the beta and the SDF methods. The returns and factors are generated under the null hypothesis by a one-factor model with parameters given in Table I, with the factors sampled from a normal mixture distribution with parameters $w = 0.0782$, $\sigma_1 = 0.1413$, $\sigma_2 = 0.0399$, and with the residuals sampled from a multivariate normal distribution. For the GMM estimations, we present variance of three different GMM estimators for δ and λ. The first two estimators are from the second and third stage GMM when the identity matrix is used as the initial weighting matrix. The third estimator is from the second stage GMM when the sample estimate of the optimal weighting matrix is used as the initial weighting matrix. Results for the variance of the estimators are presented for different lengths of time series observations (T), and they are based on 10,000 simulations.

Variance of $\hat{\delta}$ ($\delta = 0.7712$)

	Beta Method				SDF Method		
		GMM Estimation			GMM Estimation		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.1647	0.5933	0.3432	0.3241	0.3270	0.3321	0.3263
240	0.0817	0.1943	0.1445	0.1452	0.1454	0.1465	0.1452
360	0.0532	0.0982	0.0887	0.0915	0.0915	0.0919	0.0915
480	0.0393	0.0639	0.0644	0.0670	0.0670	0.0672	0.0670
600	0.0315	0.0495	0.0517	0.0536	0.0536	0.0537	0.0536
720	0.0263	0.0403	0.0423	0.0438	0.0438	0.0439	0.0438
Asymptotic	19.196	30.288	30.288	30.288	30.288	30.288	30.288

Variance of $\hat{\lambda}$ ($\lambda = 2.5037$)

	Beta Method				SDF Method		
		GMM Estimation			GMM Estimation		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	2.4292	7.8960	5.4024	4.0716	4.0626	4.2296	4.0521
240	1.1302	2.6804	1.9595	1.7727	1.7705	1.7945	1.7675
360	0.7179	1.3737	1.1229	1.0952	1.0931	1.1017	1.0931
480	0.5258	0.8978	0.7916	0.7979	0.7966	0.8007	0.7969
600	0.4194	0.6919	0.6276	0.6367	0.6360	0.6382	0.6361
720	0.3489	0.5607	0.5139	0.5247	0.5241	0.5256	0.5243
Asymptotic	247.95	356.39	356.39	356.39	356.39	356.39	356.39

Asymptotic Relative Efficiency of $\hat{\delta}$ to $\hat{\lambda} = 1.1164$

Table V

Actual Size of Maximum Likelihood Ratio and GMM Specification Tests Under the Beta and the SDF Methods when Residuals and Factors are Normally Distributed

The table presents the actual probabilities of rejection for maximum likelihood ratio and GMM specification tests under the beta and the SDF methods, when the asymptotic level of significance of the tests is 5% and 10%, respectively. The returns and factors are generated under the null hypothesis by a one-factor model with parameters given in Table I, and with the factors and the residuals sampled from a multivariate normal distribution. For the GMM estimations, we present three specification tests. The first two tests are from the second and third stage GMM when the identity matrix is used as the initial weighting matrix. The third test is from the second stage GMM when the sample estimate of the optimal weighting matrix is used as the initial weighting matrix. Results are presented for different lengths of time series observations (T), and they are based on 10,000 simulations.

Level of Significance = 5%

	Beta Method				SDF Method		
		GMM Specification Test			GMM Specification Test		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.0667	0.0003	0.0891	0.0900	0.0903	0.0905	0.0913
240	0.0578	0.0003	0.0672	0.0686	0.0684	0.0688	0.0687
360	0.0539	0.0002	0.0596	0.0604	0.0604	0.0605	0.0606
480	0.0545	0.0002	0.0583	0.0591	0.0590	0.0591	0.0592
600	0.0526	0.0001	0.0567	0.0571	0.0570	0.0570	0.0571
720	0.0504	0.0001	0.0531	0.0534	0.0532	0.0533	0.0534

Level of Significance = 10%

	Beta Method				SDF Method		
		GMM Specification Test			GMM Specification Test		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.1257	0.0008	0.1523	0.1548	0.1550	0.1557	0.1562
240	0.1126	0.0007	0.1242	0.1262	0.1258	0.1265	0.1266
360	0.1070	0.0005	0.1150	0.1166	0.1167	0.1165	0.1167
480	0.1046	0.0005	0.1107	0.1116	0.1113	0.1115	0.1116
600	0.1034	0.0005	0.1081	0.1087	0.1087	0.1087	0.1087
720	0.1032	0.0008	0.1069	0.1073	0.1071	0.1073	0.1073

Table VI
Power of Maximum Likelihood Ratio and GMM Specification Tests Under the Beta and the SDF Methods when Residuals and Factors are Normally Distributed

The table presents the probabilities of rejection for maximum likelihood ratio and GMM specification tests under the beta and the SDF methods when the level of significance is 5% and 10%, and the acceptance/rejection decision is based on the empirical distribution under the null hypothesis. The returns and factors are generated under the alternative hypothesis by a one-factor model with parameters given in Table I, and with the factors and the residuals sampled from a multivariate normal distribution. For the GMM estimations, we present three specification tests. The first two tests are from the second and third stage GMM when the identity matrix is used as the initial weighting matrix. The third test is from the second stage GMM when the sample estimate of the optimal weighting matrix is used as the initial weighting matrix. Results are presented for different lengths of time series observations (T), and they are based on 10,000 simulations.

Level of Significance = 5%

	Beta Method				SDF Method		
		GMM Specification Test			GMM Specification Test		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.1290	0.0756	0.1225	0.1292	0.1295	0.1289	0.1289
240	0.2416	0.1306	0.2401	0.2411	0.2408	0.2413	0.2411
360	0.3753	0.2102	0.3740	0.3742	0.3744	0.3738	0.3741
480	0.5020	0.3098	0.4999	0.4996	0.4996	0.4994	0.4995
600	0.6226	0.4217	0.6215	0.6215	0.6217	0.6214	0.6214
720	0.7238	0.5317	0.7229	0.7231	0.7236	0.7233	0.7233

Level of Significance = 10%

	Beta Method				SDF Method		
		GMM Specification Test			GMM Specification Test		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.2186	0.1429	0.2104	0.2161	0.2159	0.2147	0.2144
240	0.3614	0.2184	0.3586	0.3587	0.3598	0.3589	0.3586
360	0.5048	0.3190	0.5035	0.5036	0.5032	0.5032	0.5033
480	0.6321	0.4373	0.6320	0.6316	0.6319	0.6318	0.6316
600	0.7405	0.5547	0.7392	0.7395	0.7397	0.7394	0.7392
720	0.8179	0.6628	0.8174	0.8174	0.8175	0.8174	0.8174

Table VII
Power of Maximum Likelihood Ratio and GMM Specification Tests Under the Beta and the SDF Methods when Residuals and Factors are Jointly t-Distributed

The table presents the probabilities of rejection for maximum likelihood ratio and GMM specification tests under the beta and the SDF methods when the level of significance is 5% and 10%, and the acceptance/rejection decision is based on the empirical distribution under the null hypothesis. The returns and factors are generated under the alternative hypothesis by a one-factor model with parameters given in Table I, and with the factors and the residuals sampled from a multivariate Student-t distribution with five degrees of freedom. For the GMM estimations, we present three specification tests. The first two tests are from the second and third stage GMM when the identity matrix is used as the initial weighting matrix. The third test is from the second stage GMM when the sample estimate of the optimal weighting matrix is used as the initial weighting matrix. Results are presented for different lengths of time series observations (T), and they are based on 10,000 simulations.

Level of Significance = 5%

	Beta Method				SDF Method		
		GMM Specification Test			GMM Specification Test		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.1789	0.0880	0.1476	0.1530	0.1468	0.1466	0.1508
240	0.3719	0.1415	0.2758	0.2764	0.2746	0.2779	0.2768
360	0.5471	0.2123	0.3938	0.3959	0.3924	0.3963	0.3957
480	0.7023	0.3148	0.5202	0.5154	0.5141	0.5176	0.5155
600	0.8222	0.4154	0.6355	0.6303	0.6265	0.6316	0.6305
720	0.9009	0.5348	0.7301	0.7269	0.7239	0.7291	0.7264

Level of Significance = 10%

	Beta Method				SDF Method		
		GMM Specification Test			GMM Specification Test		
T	Maximum Likelihood	Identity Stage 2	Identity Stage 3	Optimal Stage 2	Identity Stage 2	Identity Stage 3	Optimal Stage 2
120	0.2874	0.1477	0.2477	0.2457	0.2444	0.2467	0.2442
240	0.4968	0.2317	0.3827	0.3856	0.3848	0.3874	0.3853
360	0.6718	0.3205	0.5258	0.5263	0.5237	0.5274	0.5260
480	0.8015	0.4391	0.6453	0.6409	0.6406	0.6414	0.6412
600	0.8898	0.5539	0.7493	0.7426	0.7424	0.7442	0.7424
720	0.9470	0.6782	0.8228	0.8203	0.8185	0.8200	0.8196