

Model Comparison Using the Hansen-Jagannathan Distance

RAYMOND KAN and CESARE ROBOTTI*

*Kan is from the University of Toronto; Robotti is from the Federal Reserve Bank of Atlanta. We thank Foussemi Chabi-Yo, Long Chen, Yufeng Han, Joel Hasbrouck (the editor), Yaxuan Qi, Sergei Sarkissian, Jay Shanken, Halbert White, Hong Zhang, Guofu Zhou, an anonymous referee, seminar participants at the Federal Reserve Bank of Chicago, Hong Kong University of Science and Technology, National University of Ireland, Singapore Management University, Syracuse University, University of Southampton, and participants at the 2006 All Georgia Finance Conference, 2007 Asian Finance Association Conference, 2007 China International Conference in Finance, 2007 Northern Finance Meetings, and 2008 Society for Financial Econometrics Conference for helpful discussions and comments. Kan gratefully acknowledges financial support from the National Bank Financial of Canada. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Corresponding author: Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6; Tel: (416) 978-4291; Fax: (416) 978-5433; e-mail:kan@chass.utoronto.ca.

Model Comparison Using the Hansen-Jagannathan Distance

ABSTRACT

Although it is of interest to test whether or not a particular asset pricing model is literally true, a more useful task for empirical researchers is to determine how wrong a model is and to compare the performance of competing asset pricing models. In this paper, we propose a new methodology to test whether or not two competing linear asset pricing models have the same Hansen-Jagannathan distance. We show that the asymptotic distribution of the test statistic depends on whether the competing models are correctly specified or misspecified, and on whether the competing models are nested or non-nested. In addition, given the increasing interest in misspecified models, we propose a simple methodology for computing the standard errors of the estimated stochastic discount factor parameters that are robust to model misspecification. Using monthly data on 25 size and book-to-market ranked portfolios and the one-month T-bill, we show that the commonly used returns and factors are, for the most part, too noisy for us to conclude that one model is superior to the other models in terms of Hansen-Jagannathan distance. Specifically, there is little evidence that conditional and intertemporal capital asset pricing model (CAPM)-type specifications outperform the simple unconditional CAPM. In addition, we show that many of the macroeconomic factors commonly used in the literature are no longer priced once potential model misspecification is taken into account.

Asset pricing models are, at best, approximations of reality. Although it is of interest to test whether or not a particular asset pricing model is literally true, a more useful task for empirical researchers is to determine how wrong a model is and to compare the performance of competing asset pricing models. The latter task requires a scalar measure of model misspecification. While there are many reasonable measures that can be used, the one introduced by Hansen and Jagannathan (1997) has gained tremendous popularity in the empirical asset pricing literature. Their proposed measure, called the Hansen-Jagannathan distance (HJ-distance), has been used both as a model diagnostic and as a tool for model selection by many researchers. Examples include Jagannathan and Wang (1996); Jagannathan, Kubota, and Takehara (1998); Campbell and Cochrane (2000); Lettau and Ludvigson (2001); Hodrick and Zhang (2001); Farnsworth, Ferson, Jackson, and Todd (2002); Dittmar (2002); and Chen and Ludvigson (2004), among others.

While the HJ-distance is an attractive tool for comparing competing asset pricing models, no formal model comparison test using the HJ-distance has yet been proposed.¹ The existing tests proposed by Hansen, Heaton, and Luttmer (1995), Jagannathan and Wang (1996), and Hansen and Jagannathan (1997) allow us to test only whether a given model has a particular HJ-distance value but not whether or not two competing models have the same HJ-distance.² Because the p -values from this kind of test are not a good way to compare models, researchers typically focus on the values of the sample HJ-distances of competing models and conclude that the model with the lowest sample HJ-distance is the best model. However, this practice is not entirely satisfactory because the difference in sample HJ-distances of two models is subject to statistical variations, so that a model with lower sample HJ-distance may not significantly outperform its competitor. The first methodological contribution of this paper is the proposal of a methodology to formally test whether or not two competing linear asset pricing models have the same HJ-distance. We provide the asymptotic distribution of our test statistic under general distributional assumptions and show that the asymptotic distribution of the test statistic depends on whether the competing models are correctly specified or misspecified, and on whether the competing models are nested or non-nested.

In addition to model comparisons, researchers are also interested in whether or not a particular

¹Wang and Zhang (2005) use simulation-based methods to compare the HJ-distances of two competing models, but they do not provide a formal model comparison tool.

²The asymptotic distribution of the squared sample HJ-distance presented in Hansen, Heaton, and Luttmer (1995) and Hansen and Jagannathan (1997) is valid when the HJ-distance of the model is nonzero, whereas the one presented in Jagannathan and Wang (1996) is valid when the model is correctly specified.

factor in an asset pricing model is “priced.” This is typically determined by testing if the stochastic discount factor (SDF) parameter associated with that factor is significantly different from zero. All existing studies perform this test using a standard error that assumes that the model is correctly specified. It is difficult to justify this assumption when estimating the SDF parameters for many different models because some (if not all) of the models are bound to be misspecified. The second methodological contribution of this paper is the proposal of robust standard errors for the estimates of the SDF parameters that are applicable to both correctly specified and misspecified models. We find that the asymptotic variances of the SDF parameter estimates tend to be larger under a misspecified model than under a correctly specified model. The difference depends on the extent of model misspecification as well as on the correlation between factors and returns. We show that the misspecification adjustment term can be very large when the underlying factor is poorly mimicked by asset returns, a situation that typically arises when the factors are macroeconomic variables.

After describing the econometric methodology, we provide an in-depth empirical analysis to demonstrate the relevance of our new tests. We focus on the empirical performance of several unconditional and conditional asset pricing models using monthly data and two different sets of test assets. First, we investigate whether model misspecification substantially affects the properties of the SDF parameter estimates. Statistically significant SDF parameter estimates are often interpreted as evidence that the underlying factors are important sources of systematic risk. Consistent with our theoretical results, we find that the t -ratios and the p -values under correctly specified and potentially misspecified models are about the same for factors that are returns on well diversified portfolios, while they differ greatly for factors that are not traded, such as macroeconomic factors. For non-traded factors, the evidence that the t -ratios under potentially misspecified models are substantially smaller than the t -ratios under correctly specified models is overwhelming. Therefore, by ignoring model misspecification and using the traditional way of computing standard errors (i.e., assuming that the model is correctly specified), one might mistakenly conclude that a factor is priced. Second, we empirically investigate whether different asset pricing models exhibit significantly different HJ-distance measures. Overall, our econometric analysis suggests that the commonly used returns and factors are too noisy for us to conclude that one model clearly outperforms the others. For example, we find little evidence that conditional and intertemporal capital asset pricing model (CAPM)-type specifications such as the Campbell (1996) and Jagannathan and

Wang (1996) models outperform the simple unconditional CAPM in terms of HJ-distance.

The rest of the paper is organized as follows. Section 1 presents an asymptotic analysis of the sample HJ-distance under correctly specified and misspecified models. In addition, we provide an asymptotic analysis of the estimates of the SDF parameters under potentially misspecified models. Section 2 introduces tests of equality of HJ-distances for two competing models and provides the asymptotic distributions of the test statistics for different scenarios. Section 3 presents the empirical analysis. The final section summarizes our findings, and the Appendix contains proofs of all propositions.

1. Asymptotic Analysis under Potentially Misspecified Models

1.1 Pricing Errors and HJ-Distance

Let y be a proposed SDF and R be a vector of gross returns on N test portfolios. If y correctly prices the N portfolios, the pricing errors, e , of the N portfolios are

$$e \equiv E[Ry] - 1_N = 0_N, \quad (1)$$

where 1_N is an N -vector of ones and 0_N is an N -vector of zeros.³ However, if y is a misspecified model, then the pricing errors of the model are nonzero. In most cases, the proposed discount factor y involves some unknown parameters λ , and it is customary to suggest that $y(\lambda)$ is a misspecified model if for all values of λ

$$e(\lambda) = E[Ry(\lambda)] - 1_N \neq 0_N. \quad (2)$$

When an asset pricing model is misspecified, researchers are often interested in obtaining a scalar measure of the magnitude of the misspecification. The popular HJ-distance is defined as the square root of a quadratic form of the pricing errors

$$\delta = [e(\lambda)'U^{-1}e(\lambda)]^{\frac{1}{2}}, \quad (3)$$

where $U = E[RR']$ is the second moment matrix of R . Equation (3) shows that δ depends on the parameters λ . When the model is misspecified, it is customary to choose λ to minimize the

³We assume that the elements of R are all gross returns so that their costs are given by the vector 1_N . If some of the elements of R are returns on zero net investment portfolios, we replace 1_N with q , where $q \neq 0_N$ is a vector of initial costs of the N test assets. A separate appendix (available upon request) shows the necessary modifications of our analysis when all the elements of R are excess returns (i.e., $q = 0_N$).

HJ-distance. Under this choice of λ , the HJ-distance is defined as

$$\delta = \left[\min_{\lambda} e(\lambda)' U^{-1} e(\lambda) \right]^{\frac{1}{2}}. \quad (4)$$

It is important to note that δ is also well defined for other choices of λ , but we focus on Equation (4) in this paper because this is the most popular way of calculating the HJ-distance in the empirical literature.

In this paper, we focus on linear asset pricing models because they are the most popular models in the empirical asset pricing literature. A linear factor asset pricing model suggests that y is a linear function of K systematic factors f

$$y(\lambda_0, \lambda_1) = \lambda_0 + \lambda_1' f = \lambda' x, \quad (5)$$

where $x = [1, f']'$ and $\lambda = [\lambda_0, \lambda_1']'$.

To prepare for our analysis, we define $Y = [f', R']'$ and its mean and covariance matrix as

$$\mu = E[Y] \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad (6)$$

$$V = \text{Var}[Y] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \quad (7)$$

Under the linear SDF, the pricing errors of the N assets are given by

$$e(\lambda) = E[Ry] - 1_N = E[Rx'\lambda] - 1_N = D\lambda - 1_N, \quad (8)$$

where $D = E[Rx'] = [\mu_2, V_{21} + \mu_2\mu_1']$. Although the standard definition of the HJ-distance uses U^{-1} as the weighting matrix, Kan and Zhou (2004) show that for linear factor models, using V_{22}^{-1} as the weighting matrix would produce mathematically identical results for both the SDF parameters and the HJ-distance. Using V_{22}^{-1} as the weighting matrix, the squared HJ-distance is given by

$$\delta^2 = \min_{\lambda} (D\lambda - 1_N)' V_{22}^{-1} (D\lambda - 1_N) = 1_N' V_{22}^{-1} 1_N - 1_N' V_{22}^{-1} D (D' V_{22}^{-1} D)^{-1} D' V_{22}^{-1} 1_N. \quad (9)$$

We assume that V_{21} is of full column rank (which implies that D is also of full column rank). Hence, there exists a unique λ that minimizes $e(\lambda)' V_{22}^{-1} e(\lambda)$, which we denote by

$$\lambda_{HJ} = (D' V_{22}^{-1} D)^{-1} (D' V_{22}^{-1} 1_N). \quad (10)$$

In the subsequent analysis, we drop the subscript from λ_{HJ} for brevity. In addition, when it is clear from the context, we write the pricing errors $e(\lambda_{HJ})$ simply as e and the SDF $y(\lambda_{HJ}) = \lambda_{HJ}' x$ simply as y .

1.2 Asymptotic Distribution of the Sample HJ-Distance under Correctly Specified and Misspecified Models

In practice, the population HJ-distance of a model is unobservable and has to be estimated using the sample HJ-distance. In this subsection, we summarize the asymptotic distribution of the sample HJ-distance for the case of linear factor models. Let $Y_t = [f_t', R_t']'$, where f_t is a vector of proposed factors at time t and R_t is a vector of gross returns on N test assets at time t . Suppose that we have T observations on Y_t and denote the sample moments of Y_t by

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \frac{1}{T} \sum_{t=1}^T Y_t, \quad (11)$$

$$\hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} = \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \quad (12)$$

The sample squared HJ-distance and the SDF parameter estimates are simply the sample counterparts of Equations (9) and (10),

$$\hat{\delta}^2 = 1_N' \hat{V}_{22}^{-1} 1_N - 1_N' \hat{V}_{22}^{-1} \hat{D} (\hat{D}' \hat{V}_{22}^{-1} \hat{D})^{-1} \hat{D}' \hat{V}_{22}^{-1} 1_N, \quad (13)$$

$$\hat{\lambda} = (\hat{D}' \hat{V}_{22}^{-1} \hat{D})^{-1} (\hat{D}' \hat{V}_{22}^{-1} 1_N), \quad (14)$$

where $\hat{D} = [\hat{\mu}_2, \hat{V}_{21} + \hat{\mu}_2 \hat{\mu}_1']$. Under a correctly specified model ($\delta = 0$), the asymptotic distribution of $\hat{\delta}^2$ is well known. For linear factor models, Jagannathan and Wang (1996) show that when $\delta = 0$,

$$T \hat{\delta}^2 \overset{A}{\approx} \sum_{i=1}^{N-K-1} \xi_i x_i, \quad (15)$$

where the x_i 's are independent χ_1^2 random variables and the weights ξ_i 's are equal to the nonzero eigenvalues of

$$S^{\frac{1}{2}} V_{22}^{-1} S^{\frac{1}{2}} - S^{\frac{1}{2}} V_{22}^{-1} D (D' V_{22}^{-1} D)^{-1} D' V_{22}^{-1} S^{\frac{1}{2}}, \quad (16)$$

where S is the asymptotic covariance matrix of

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T (R_t x_t' \lambda - 1_N). \quad (17)$$

The asymptotic distribution of $\hat{\delta}$ under a misspecified model is also well known. Hansen, Heaton, and Luttmer (1995) and Hansen and Jagannathan (1997) show that when $\delta \neq 0$,

$$\sqrt{T}(\hat{\delta}^2 - \delta^2) \overset{A}{\approx} N(0, v), \quad (18)$$

$$\sqrt{T}(\hat{\delta} - \delta) \overset{A}{\approx} N\left(0, \frac{v}{4\delta^2}\right), \quad (19)$$

where v is the asymptotic variance of $\frac{1}{\sqrt{T}} \sum_{t=1}^T q_t$ and

$$q_t = y_t^2 - (y_t - \eta' R_t)^2 - 2\eta' 1_N - \delta^2 = 2\eta' R_t y_t - (\eta' R_t)^2 - 2\eta' 1_N - \delta^2, \quad (20)$$

with $\eta = U^{-1}e$. When the SDF is linear in the factors and λ is chosen to minimize the HJ-distance, the first-order condition suggests that $D'V_{22}^{-1}e = 0_{K+1}$. It follows that $\eta = V_{22}^{-1}e$ and $\eta' 1_N = e'V_{22}^{-1}(D\lambda - e) = -\delta^2$. Then, we can simplify q_t to

$$q_t = 2u_t y_t - u_t^2 + \delta^2, \quad (21)$$

where $u_t = e'V_{22}^{-1}R_t$.

In conducting statistical tests, we need a consistent estimator of v . This can be accomplished by using one of the frequency zero spectral density estimators described by Newey and West (1987) or Andrews (1991) and replacing q_t with

$$\hat{q}_t = 2\hat{u}_t \hat{y}_t - \hat{u}_t^2 + \hat{\delta}^2, \quad (22)$$

where $\hat{u}_t = \hat{e}'\hat{V}_{22}^{-1}R_t$, $\hat{y}_t = \hat{\lambda}'x_t$, with $\hat{\lambda} = (\hat{D}'\hat{V}_{22}^{-1}\hat{D})^{-1}\hat{D}'\hat{V}_{22}^{-1}1_N$, and $\hat{e} = \hat{D}\hat{\lambda} - 1_N$.

There are situations in which we may like to construct a confidence interval for δ^2 (or equivalently for δ). However, this exercise is nontrivial since the asymptotic variance of $\sqrt{T}(\hat{\delta}^2 - \delta^2)$ depends on δ^2 . Consequently, one cannot simply use $\hat{\delta}^2 \pm 2 \times \text{s.e.}(\hat{\delta}^2)$ to obtain a 95% confidence interval for δ^2 . One way of constructing a confidence interval for δ^2 is to use the statistical method (see, for example, Casella and Berger 1990, Section 9.2.3).⁴ Using this methodology, we first plot the 2.5 and 97.5 percentiles of the distribution of $\hat{\delta}^2$ for different values of δ^2 . We then draw a horizontal line at the observed value of $\hat{\delta}^2$. This horizontal line will intersect the 97.5 percentile line first and then the 2.5 percentile line of $\hat{\delta}^2$. The interval between these two intersection points gives us a 95% confidence interval for δ^2 .

In implementing this method, there is one more issue to overcome: the asymptotic variance of $\hat{\delta}^2$ depends not only on δ^2 but also on other nuisance parameters. For example, when the factors and the returns are i.i.d. multivariate normally distributed, it can be easily shown that

$$v = 4(\mu_y^2 + \sigma_y^2)\delta^2 + 2\delta^4, \quad (23)$$

⁴Lewellen, Nagel, and Shanken (2006) use the statistical method to construct a confidence interval for the adjusted generalized least squares (GLS) R^2 in a cross-sectional regression, and Kan and Robotti (2007) use the statistical method to construct a confidence interval for the Hansen-Jagannathan bound.

where $\mu_y = E[y_t] = \lambda_0 + \lambda_1' \mu_1$ and $\sigma_y^2 = \text{Var}[y_t] = \lambda_1' V_{11} \lambda_1$. As a result, one needs to know μ_y and σ_y^2 to construct the confidence interval for δ^2 . In practice, we replace μ_y and σ_y^2 by their consistent estimators. For the more general case, the asymptotic distribution of $\hat{\delta}^2$ depends on many more nuisance parameters, and we need to estimate these nuisance parameters to come up with a consistent estimator of v . In our empirical work, we replace \hat{q}_t with

$$\tilde{q}_t = 2\tilde{u}_t \hat{y}_t - \tilde{u}_t^2 + \delta^2 \quad (24)$$

to construct a consistent estimator of v for a given value of δ^2 , where $\tilde{u}_t = \delta \hat{u}_t / \hat{\delta}$. By scaling \hat{u}_t to \tilde{u}_t , we can ensure that $\sum_{t=1}^T \tilde{u}_t^2 / T = \delta^2$. This amounts to scaling the pricing errors to make sure that the model has the desired δ^2 .

1.3 Asymptotic Distribution of the SDF Parameter Estimates under Potentially Misspecified Models

In many empirical studies, interest lies in the point estimates of the SDF parameters λ . A statistically significant $\hat{\lambda}$ associated with a given factor is often interpreted as evidence that the factor is priced. However, when computing the standard error of $\hat{\lambda}$, researchers typically rely on the asymptotic distribution under the assumption that the model is correctly specified. This practice is difficult to justify, especially when the model is rejected by the data. In this subsection, we study the asymptotic distribution of $\hat{\lambda}$ under potentially misspecified models. Our analysis closely follows those of Hall and Inoue (2003) and Kan and Robotti (2008).⁵ A similar analysis is also performed by Hou and Kimmel (2006) in the context of two-pass GLS cross-sectional regressions.

Proposition 1. *Under a potentially misspecified model*

$$\sqrt{T}(\hat{\lambda} - \lambda) \overset{A}{\rightsquigarrow} N(0_{K+1}, V(\hat{\lambda})), \quad (25)$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \quad (26)$$

⁵Gallant and White (1988) first considered GMM with misspecified models, but, as Hall and Inoue (2003) note, they did not treat the important case of a stochastic weighting matrix. However, Theorem 6.10 of White (1994) can be used to obtain asymptotic results under misspecified GMM with a stochastic weighting matrix. It should be noted that Hansen, Heaton, and Luttmer (1995, Appendix C) also present the asymptotic distribution of the SDF parameters for a misspecified model. However, their results do not contain an explicit expression of the asymptotic covariance matrix.

with

$$h_t = -HD'V_{22}^{-1}R_t y_t + H[D'V_{22}^{-1}(R_t - \mu_2) - x_t]u_t + \lambda, \quad (27)$$

where $H = (D'V_{22}^{-1}D)^{-1}$ and $u_t = e'V_{22}^{-1}R_t$. When the model is correctly specified, $e = 0_N$, $u_t = 0$, and h_t can be simplified to

$$h_t = -HD'V_{22}^{-1}R_t y_t + \lambda. \quad (28)$$

It is easily verified that under the linear SDF, Proposition 1 coincides with Theorem 2 in Hall and Inoue (2003). When estimating the standard errors of $\hat{\lambda}$, it is advisable to use the sample counterpart of Equation (27) instead of the sample counterpart of Equation (28). This is because the latter is valid only when the model is correctly specified, whereas the former is valid for both correctly specified and misspecified models.

Note that when the model is misspecified, h_t has an extra term of

$$H[D'V_{22}^{-1}(R_t - \mu_2) - x_t]u_t. \quad (29)$$

There are situations in which this term is relatively unimportant. The first case arises when model misspecification (δ^2) is small. In this case, $u_t = e'V_{22}^{-1}R_t$ has a very small variance because $\frac{1}{\sqrt{T}} \sum_{t=1}^T u_t \overset{A}{\sim} N(0, \delta^2)$ and the term in Equation (29) tends to be very small. The second case occurs when the constant term and the factors can be very well mimicked by the returns. In this case, $D'V_{22}^{-1}(R_t - \mu_2) - x_t$ is very small, and misspecification does not have much of an impact on the asymptotic variance of $\hat{\lambda}$.

While we cannot prove that misspecification always increases the asymptotic variance of $\hat{\lambda}$ for the general case, we can show that this is true for the special case when factors and returns are multivariate elliptically distributed. Lemma 1 presents this result.

Lemma 1. *Suppose $Y_t = [f'_t, R'_t]'$ is i.i.d. multivariate elliptically distributed with finite fourth moments and its multivariate kurtosis parameter is κ . Let $\mu_y = \lambda_0 + \lambda'_1 \mu_1$ and $\sigma_y^2 = \lambda'_1 V_{11} \lambda_1$. The asymptotic variance of $\sqrt{T}(\hat{\lambda} - \lambda)$ is given by*

$$\begin{aligned} V(\hat{\lambda}) = & [\mu_y^2 + (1 + \kappa)\sigma_y^2]H + \begin{bmatrix} \sigma_y^2 - \mu_y^2 + \lambda_0^2 + 2\kappa(\mu'_1 \lambda_1)^2 & (\lambda_0 - 2\kappa\mu'_1 \lambda_1)\lambda'_1 \\ (\lambda_0 - 2\kappa\mu'_1 \lambda_1)\lambda_1 & (1 + 2\kappa)\lambda_1 \lambda'_1 \end{bmatrix} \\ & + \delta^2 H \left([1 + (1 + \kappa)\mu'_2 V_{22}^{-1} \mu_2] \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix} \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix}' + \begin{bmatrix} 0 & 0'_K \\ 0_K & (1 + \kappa)(V_{11} - V_{12}V_{22}^{-1}V_{21}) \end{bmatrix} \right) H, \end{aligned} \quad (30)$$

where $H = (D'V_{22}^{-1}D)^{-1}$.

When the model is correctly specified (i.e., $\delta = 0$),

$$V(\hat{\lambda}) = [\mu_y^2 + (1 + \kappa)\sigma_y^2]H + \begin{bmatrix} \sigma_y^2 - \mu_y^2 + \lambda_0^2 + 2\kappa(\mu'_1\lambda_1)^2 & (\lambda_0 - 2\kappa\mu'_1\lambda_1)\lambda'_1 \\ (\lambda_0 - 2\kappa\mu'_1\lambda_1)\lambda_1 & (1 + 2\kappa)\lambda_1\lambda'_1 \end{bmatrix}. \quad (31)$$

We call the last term in Equation (30) the misspecification adjustment term since it exists only when the model is misspecified (i.e., $\delta > 0$). Since $1 + \kappa > 0$, this adjustment term is positive semidefinite.⁶ Consequently, misspecification always increases the asymptotic variance of $\hat{\lambda}$.

Besides depending on the magnitude of misspecification (δ^2), the misspecification adjustment term also depends crucially on how well the factors can be mimicked by the returns. Specifically, $V_{11} - V_{12}V_{22}^{-1}V_{21}$ is the covariance matrix of the residuals from regressing factors on returns. When the factors are portfolio returns, this term tends to be very small and the misspecification adjustment is relatively minor. However, if the factors are macroeconomic factors, then the matrix $V_{11} - V_{12}V_{22}^{-1}V_{21}$ tends to be very large and misspecification can have a serious impact on the asymptotic variance of $\hat{\lambda}$. Ignoring model misspecification and using the traditional way of computing standard errors (i.e., assuming that the model is correctly specified), one can mistakenly conclude that a factor is priced. A similar point was also made by Hou and Kimmel (2006) and Kan and Robotti (2008) for the case of excess returns. The only difference is that in the case of gross returns, we still have a misspecification adjustment term even when the factors are fully mimicked by the returns (i.e., when $V_{11} - V_{12}V_{22}^{-1}V_{21}$ is a zero matrix). This is because the constant term in the SDF cannot be written as a linear combination of the returns unless V_{22} is singular. As a result, we cannot fully mimic the SDF with the returns, and there is still some room for model misspecification to affect the asymptotic variance of $\hat{\lambda}$.

2. Tests of Equality of the HJ-Distances of Two Models

Our analysis in this section is similar in spirit to the model selection methodology of Vuong (1989); Rivers and Vuong (2002); and Golden (2003). Vuong's (1989) model selection methodology is based on the likelihood function and is limited to the i.i.d. situation. Hence, it is not directly applicable here. The analyses of Rivers and Vuong (2002) and Golden (2003) allow for more general

⁶Bentler and Berkane (1986) show that $\kappa > -2/(N + K + 2)$ for multivariate elliptical distributions, which implies that $1 + \kappa > 0$.

model selection criteria as well as less restrictive distributional assumptions, and their results are applicable to our problem. In particular, Golden's methodology can be used directly to study our problem. Besides specializing Golden's analysis to linear models, the additional contribution of this section is to provide a refinement of Golden's analysis and to present a simpler sequential test of $\delta_1^2 = \delta_2^2$ for the case of non-nested models. In addition, because our models are linear, we are able to provide explicit expressions for all the test statistics, making our results readily accessible to finance researchers.

We consider two competing models. Let $x_1 = [1, f'_1, f'_2]'$ and $x_2 = [1, f'_1, f'_3]'$, where f_1 to f_3 are three sets of distinct factors, and f_i is of dimension $K_i \times 1$, $i = 1, 2, 3$. We assume that the SDF of model 1 is linear in x_1 and is given by $y_1 = \eta'x_1$, whereas the SDF of model 2 is linear in x_2 and is given by $y_2 = \lambda'x_2$. Let $D_1 = E[Rx'_1]$ and $D_2 = E[Rx'_2]$ and assume that both D_1 and D_2 have full column rank, so that the SDF parameters that minimize the HJ-distances of the two models are uniquely identified as

$$\eta = (D'_1 V_{22}^{-1} D_1)^{-1} D'_1 V_{22}^{-1} 1_N, \quad (32)$$

$$\lambda = (D'_2 V_{22}^{-1} D_2)^{-1} D'_2 V_{22}^{-1} 1_N. \quad (33)$$

It follows that the pricing errors and the squared HJ-distances of the two models are given by

$$e_i = D_i (D'_i V_{22}^{-1} D_i)^{-1} D'_i V_{22}^{-1} 1_N - 1_N, \quad i = 1, 2, \quad (34)$$

$$\delta_i^2 = 1'_N V_{22}^{-1} 1_N - 1'_N V_{22}^{-1} D_i (D'_i V_{22}^{-1} D_i)^{-1} D'_i V_{22}^{-1} 1_N, \quad i = 1, 2. \quad (35)$$

When $K_1 = 0$, the two models do not share a common factor. When $K_2 = 0$, the second model nests the first model as a special case. Similarly, when $K_3 = 0$, the first model nests the second model as a special case. When both $K_2 > 0$ and $K_3 > 0$, the two models are not nested.⁷ We study the nested models case in the next subsection and deal with the non-nested models case in Section 2.2.

2.1 Nested Models

Without loss of generality, we assume $K_2 = 0$, so that model 2 nests model 1 as a special case. For the nested models case, the following lemma shows that $\delta_1^2 = \delta_2^2$ implies some restrictions on the

⁷Vuong (1989) defines this case as the overlapping models case. He also deals with a separate case of strictly non-nested models in which x_1 and x_2 do not share a common element. Since linear SDFs always contain a constant term, we do not have to deal with the case of strictly non-nested models here.

SDF parameters of model 2.

Lemma 2. $\delta_1^2 = \delta_2^2$ if and only if $\lambda_2 = 0_{K_3}$, where λ_2 is a vector of the last K_3 elements of λ .

Note that Lemma 2 is applicable even when the models are misspecified. In order to test the equality of HJ-distances of the two models, Lemma 2 suggests that one can simply perform a test of $H_0 : \lambda_2 = 0_{K_3}$ in model 2. Suppose that $\hat{V}(\hat{\lambda}_2)$ is a consistent estimator of the asymptotic variance of $\sqrt{T}(\hat{\lambda}_2 - \lambda_2)$. Then, under the null hypothesis $H_0 : \lambda_2 = 0_{K_3}$,

$$T\hat{\lambda}_2'\hat{V}(\hat{\lambda}_2)^{-1}\hat{\lambda}_2 \overset{A}{\sim} \chi_{K_3}^2, \quad (36)$$

which can be used for testing $H_0 : \delta_1^2 = \delta_2^2$. However, it is important to note that, in general, we cannot conduct this test using the usual standard error of $\hat{\lambda}_2$, which assumes that model 2 is correctly specified. Instead, we need to rely on the misspecification robust standard errors of $\hat{\lambda}_2$ based on Equation (27) to perform the test of $H_0 : \lambda_2 = 0_{K_3}$.

Alternatively, we can derive the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ and use it for the purpose of testing $H_0 : \delta_1^2 = \delta_2^2$. Proposition 2 presents the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$.

Proposition 2. Partition $H_2 = (D_2'V_{22}^{-1}D_2)^{-1}$ as

$$H_2 = \begin{bmatrix} H_{2,11} & H_{2,12} \\ H_{2,21} & H_{2,22} \end{bmatrix}, \quad (37)$$

where $H_{2,22}$ is $K_3 \times K_3$. Under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2$

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \overset{A}{\sim} \sum_{i=1}^{K_3} \xi_i x_i, \quad (38)$$

where the x_i 's are independent χ_1^2 random variables and the ξ_i 's are the eigenvalues of $H_{2,22}^{-1}V(\hat{\lambda}_2)$, with $V(\hat{\lambda}_2)$ being the asymptotic variance of $\sqrt{T}(\hat{\lambda}_2 - \lambda_2)$.

Again, it should be emphasized that the misspecification robust version of $V(\hat{\lambda}_2)$ should be used to test $H_0 : \delta_1^2 = \delta_2^2$. This is because model misspecification tends to create additional sampling variation in $\hat{\delta}_1^2 - \hat{\delta}_2^2$. Without taking into account potential model misspecification, one might mistakenly reject $H_0 : \delta_1^2 = \delta_2^2$. In actual testing, we replace ξ_i with its sample counterpart $\hat{\xi}_i$, where the $\hat{\xi}_i$'s are the eigenvalues of $\hat{H}_{2,22}^{-1}\hat{V}(\hat{\lambda}_2)$, and $\hat{H}_{2,22}$ and $\hat{V}(\hat{\lambda}_2)$ are consistent estimators of $H_{2,22}$ and $V(\hat{\lambda}_2)$, respectively.

2.2 Non-Nested Models

For the nested models case, Lemma 2 suggests that $\delta_1^2 = \delta_2^2$ holds if and only if $\lambda_2 = 0_{K_3}$ (i.e., $y_1 = y_2$). In contrast, $\delta_1^2 = \delta_2^2$ can occur under two different scenarios for non-nested models. The first scenario is $y_1 = y_2$, which clearly implies $e_1 = e_2$ and $\delta_1^2 = \delta_2^2$. The second scenario is $y_1 \neq y_2$ (i.e., $e_1 \neq e_2$), but the aggregate pricing errors in the two models are the same— $e_1' V_{22}^{-1} e_1 = e_2' V_{22}^{-1} e_2$ —so that δ_1^2 is still equal to δ_2^2 . As it turns out, the asymptotic distributions of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ under these two scenarios are very different and we have to deal with them separately.

2.2.1 Tests of Equality of Two Stochastic Discount Factors

The condition $y_1 = y_2$ imposes parametric restrictions on η and λ . Suppose we partition η and λ as $\eta = [\eta_1', \eta_2']'$ and $\lambda = [\lambda_1', \lambda_2']'$, where η_1 and λ_1 are the first $K_1 + 1$ elements of η and λ , respectively. At first sight, it may appear that $y_1 = y_2$ holds if and only if $\eta_1 = \lambda_1$, $\eta_2 = 0_{K_2}$ and $\lambda_2 = 0_{K_3}$. The following lemma shows that the restriction $\eta_1 = \lambda_1$ is redundant because it is implied by the other two restrictions.

Lemma 3. *For non-nested models, $y_1 = y_2$ if and only if $\eta_2 = 0_{K_2}$ and $\lambda_2 = 0_{K_3}$.*

Note that Lemma 3 is applicable even when the models are misspecified. It suggests that we can test $H_0 : y_1 = y_2$ by simply testing the parametric hypothesis $H_0 : \eta_2 = 0_{K_2}, \lambda_2 = 0_{K_3}$. Let $\psi = [\eta_2', \lambda_2']'$ and $\hat{\psi} = [\hat{\eta}_2', \hat{\lambda}_2']'$. Using the same proof of Proposition 1, we can establish that the asymptotic distribution of $\hat{\psi}$ under potentially misspecified models is given by

$$\sqrt{T}(\hat{\psi} - \psi) \overset{A}{\rightsquigarrow} N(0_{K_2+K_3}, V(\hat{\psi})), \quad (39)$$

where

$$V(\hat{\psi}) = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t \tilde{h}_{t+j}'], \quad (40)$$

with

$$\tilde{h}_t \equiv \begin{bmatrix} \tilde{h}_{1t} \\ \tilde{h}_{2t} \end{bmatrix} = \begin{bmatrix} -H_{1b} D_1' V_{22}^{-1} R_t y_{1t} + H_{1b} [D_1' V_{22}^{-1} (R_t - \mu_2) - x_{1t}] u_{1t} + \eta_2 \\ -H_{2b} D_2' V_{22}^{-1} R_t y_{2t} + H_{2b} [D_2' V_{22}^{-1} (R_t - \mu_2) - x_{2t}] u_{2t} + \lambda_2 \end{bmatrix}, \quad (41)$$

where $u_{1t} = e_1' V_{22}^{-1} R_t$, $u_{2t} = e_2' V_{22}^{-1} R_t$, H_{1b} is the last K_2 rows of $(D_1' V_{22}^{-1} D_1)^{-1}$, and H_{2b} is the last K_3 rows of $(D_2' V_{22}^{-1} D_2)^{-1}$.

Suppose that $\hat{V}(\hat{\psi})$ is a consistent estimator of $V(\hat{\psi})$. Then under the null hypothesis $H_0 : \psi = 0_{K_2+K_3}$,

$$T\hat{\psi}'\hat{V}(\hat{\psi})^{-1}\hat{\psi} \stackrel{A}{\sim} \chi_{K_2+K_3}^2, \quad (42)$$

and this can be used as a statistic for testing $H_0 : y_1 = y_2$.⁸ Just like in the nested models case, it is important that we conduct this test using the robust standard error of $\hat{\psi}$ based on Equations (40) and (41).

When $y_1 = y_2$, the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ is given by the following proposition.

Proposition 3. *Let $H_1 = (D_1'V_{22}^{-1}D_1)^{-1}$ and $H_2 = (D_2'V_{22}^{-1}D_2)^{-1}$, and partition them as*

$$H_1 = \begin{bmatrix} H_{1,11} & H_{1,12} \\ H_{1,21} & H_{1,22} \end{bmatrix}, \quad H_2 = \begin{bmatrix} H_{2,11} & H_{2,12} \\ H_{2,21} & H_{2,22} \end{bmatrix}, \quad (43)$$

where $H_{1,11}$ and $H_{2,11}$ are of dimension $(K_1+1) \times (K_1+1)$. Under the null hypothesis $H_0 : y_1 = y_2$, we have

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \stackrel{A}{\sim} \sum_{i=1}^{K_2+K_3} \xi_i x_i, \quad (44)$$

where the x_i 's are independent χ_1^2 random variables and the ξ_i 's are the eigenvalues of

$$\begin{bmatrix} -H_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & H_{2,22}^{-1} \end{bmatrix} V(\hat{\psi}). \quad (45)$$

Note that Equation (44) allows us to construct a test of $H_0 : y_1 = y_2$ using $\hat{\delta}_1^2 - \hat{\delta}_2^2$. However, it should be pointed out that unlike the Wald test in Equation (42), there are cases (as we shall see later) in which $y_1 \neq y_2$ but yet Equation (44) fails to reject $H_0 : y_1 = y_2$ with probability one as T goes to infinity.

Before moving on to the case of $y_1 \neq y_2$, a couple of remarks are in order. The first remark is that we can think of the results of the nested models case as a special case of testing $H_0 : y_1 = y_2$ with $K_2 = 0$. The only difference is that the ξ_i 's in Proposition 2 are all positive, whereas some of the ξ_i 's in Proposition 3 are negative. As a result, we need to perform a two-sided test for the non-nested models case when we use Equation (44) to test $H_0 : y_1 = y_2$. The second remark is more subtle. Unlike Equations (36) and (38), which are tests of $H_0 : \delta_1^2 = \delta_2^2$ for the nested models

⁸Note that we should not perform a Wald test of $H_0 : \eta_1 = \lambda_1$, $\psi = 0_{K_2+K_3}$. This is because the asymptotic variance of $\sqrt{T}[\hat{\eta}_1 - \hat{\lambda}_1, \hat{\psi}']'$ is singular under H_0 , and the Wald test statistic does not have the standard asymptotic $\chi_{K_1+K_2+K_3+1}^2$ distribution. The proof of this result is available upon request.

case, Equations (42) and (44) for the non-nested models case are only tests of $H_0 : y_1 = y_2$. They should not be interpreted as pure tests of $H_0 : \delta_1^2 = \delta_2^2$. This is because $y_1 = y_2$ is a sufficient but not a necessary condition for $\delta_1^2 = \delta_2^2$. We can have $\delta_1^2 = \delta_2^2$ even when $y_1 \neq y_2$, and these cases are taken up in the next subsection.

2.2.2 Tests of Equality of the HJ-Distances of Two Distinct Stochastic Discount Factors

For non-nested distinct SDFs (i.e., $y_1 \neq y_2$), the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2$ depends on whether (1) both models are correctly specified, or (2) both models are misspecified.

The first case is a little peculiar and it requires some explanation. In the likelihood ratio setting of Vuong (1989), we cannot have two distinct non-nested models that are both correctly specified. One may wonder how two distinct SDFs can be both correctly specified. Two asset pricing models are considered to be correctly specified when they both produce zero pricing errors. This occurs when the vector 1_N is in the span of D_1 as well as in the span of D_2 . A simple example of this is when the first model is the correctly specified model and the second model has $f_3 = f_2 + \epsilon$, where ϵ is a vector of pure measurement errors with mean zero and independent of the returns. In this case, $D_2 = E[Rx_2'] = E[Rx_1'] = D_1$ and the second model also produces zero pricing errors even though $y_1 \neq y_2$.

The following proposition presents a simple chi-squared test for testing if both models 1 and 2 are correctly specified.

Proposition 4. *Let $n_1 = N - K_1 - K_2 - 1$ and $n_2 = N - K_1 - K_3 - 1$. Also, let P_1 be an $N \times n_1$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}}D_1$ and P_2 be an $N \times n_2$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}}D_2$. Define*

$$g_t(\theta) = \begin{bmatrix} g_{1t}(\eta) \\ g_{2t}(\lambda) \end{bmatrix} = \begin{bmatrix} R_t x'_{1t} \eta - 1_N \\ R_t x'_{2t} \lambda - 1_N \end{bmatrix}, \quad (46)$$

where $\theta = [\eta', \lambda']'$, and

$$S = \sum_{j=-\infty}^{\infty} E[g_t(\theta)g_{t+j}(\theta)'] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}. \quad (47)$$

When $y_1 \neq y_2$ and under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2 = 0$,

$$T \begin{bmatrix} \hat{P}'_1 \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_1 \\ \hat{P}'_2 \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_2 \end{bmatrix}' \begin{bmatrix} \hat{P}'_1 \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{11} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_1 & \hat{P}'_1 \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{12} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_2 \\ \hat{P}'_2 \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{21} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_1 & \hat{P}'_2 \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{22} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{P}'_1 \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_1 \\ \hat{P}'_2 \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_2 \end{bmatrix} \stackrel{A}{\sim} \chi_{n_1+n_2}^2, \quad (48)$$

where \hat{e}_1 and \hat{e}_2 are the sample pricing errors of models 1 and 2, and \hat{P}_1 , \hat{P}_2 , \hat{S} are consistent estimators of P_1 , P_2 , and S , respectively.

When $y_1 \neq y_2$, the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ when both models are correctly specified is given in the following proposition.

Proposition 5. *Using the notation in Proposition 4, when $y_1 \neq y_2$ and under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2 = 0$,*

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \stackrel{A}{\sim} \sum_{i=1}^{n_1+n_2} \xi_i x_i, \quad (49)$$

where the x_i 's are independent χ_1^2 random variables and the ξ_i 's are the eigenvalues of

$$\begin{bmatrix} P'_1 V_{22}^{-\frac{1}{2}} S_{11} V_{22}^{-\frac{1}{2}} P_1 & P'_1 V_{22}^{-\frac{1}{2}} S_{12} V_{22}^{-\frac{1}{2}} P_2 \\ -P'_2 V_{22}^{-\frac{1}{2}} S_{21} V_{22}^{-\frac{1}{2}} P_1 & -P'_2 V_{22}^{-\frac{1}{2}} S_{22} V_{22}^{-\frac{1}{2}} P_2 \end{bmatrix}. \quad (50)$$

Note that the ξ_i 's are not all positive, because $\hat{\delta}_1^2 - \hat{\delta}_2^2$ can be negative. Therefore, we need to perform a two-sided test of $H_0 : \delta_1^2 = \delta_2^2$ instead of a one-sided test, as in the nested models case. Comparing Propositions 3 and 5, we see that $\hat{\delta}_1^2 - \hat{\delta}_2^2 = O_P(T^{-1})$ when $y_1 = y_2$ as well as when $y_1 \neq y_2$ and $\delta_1^2 = \delta_2^2 = 0$. Therefore, the test in Proposition 3 is not consistent against the alternative of $y_1 \neq y_2$ when both models are correctly specified.

Finally, similar to the asymptotic distribution of $\hat{\delta}^2$, the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ changes when the models are misspecified. Consequently, we cannot use Proposition 5 to test $H_0 : \delta_1^2 = \delta_2^2$ when the models are misspecified. Proposition 6 presents the appropriate asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ when both non-nested models are misspecified and $y_1 \neq y_2$.

Proposition 6. *Suppose $y_1 \neq y_2$. Let $d_t = q_{1t} - q_{2t}$, where*

$$\begin{aligned} q_{1t} &= 2u_{1t}y_{1t} - u_{1t}^2 + \delta_1^2, \\ q_{2t} &= 2u_{2t}y_{2t} - u_{2t}^2 + \delta_2^2, \end{aligned}$$

with $u_{1t} = e_1' V_{22}^{-1} R_t$ and $u_{2t} = e_2' V_{22}^{-1} R_t$. When $\delta_1 \neq 0$ and $\delta_2 \neq 0$,

$$\sqrt{T}(\hat{\delta}_1^2 - \hat{\delta}_2^2 - (\delta_1^2 - \delta_2^2)) \overset{A}{\rightsquigarrow} N(0, v_d), \quad (51)$$

where

$$v_d = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]. \quad (52)$$

Under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2 \neq 0$,

$$\sqrt{T}(\hat{\delta}_1^2 - \hat{\delta}_2^2) \overset{A}{\rightsquigarrow} N(0, v_d) \quad (53)$$

and d_t can be simplified to

$$d_t = 2u_{1t}y_{1t} - u_{1t}^2 - 2u_{2t}y_{2t} + u_{2t}^2. \quad (54)$$

The expression of d_t in Equation (54) reveals that there are situations in which one cannot use the normal test in Proposition 6 to test $H_0 : \delta_1^2 = \delta_2^2$. This can happen when (1) $y_{1t} = y_{2t}$, which implies $u_{1t} = u_{2t}$ and hence $d_t = 0$; or (2) $y_{1t} \neq y_{2t}$ but both models are correctly specified—i.e., $u_{1t} = u_{2t} = 0$, which also leads to $d_t = 0$. Golden (2003) presents a test of $v_d = 0$ that can be used to determine whether the normal test should be used or not. His test is a weighted chi-squared test based on the sample estimate of v_d , and it effectively combines the two tests in our Propositions 3 and 5 without the need to distinguish the two reasons for $d_t = 0$. Nevertheless, we prefer to keep these two cases separate because we believe that researchers can benefit from learning the underlying reason for $d_t = 0$. In addition, by separating the two cases of $d_t = 0$, we can obtain a chi-squared test, which is much easier to implement than the weighted chi-squared test.

2.3 Summary and Discussion

Under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2$, $\hat{\delta}_1^2 - \hat{\delta}_2^2$ can be either $O_P(T^{-1})$ or $O_P(T^{-\frac{1}{2}})$. For the nested models case, the situation is quite clear: $T(\hat{\delta}_1^2 - \hat{\delta}_2^2)$ is asymptotically distributed as a linear combination of χ_1^2 random variables with positive weights. For the non-nested models case, the asymptotic distribution depends on whether $y_1 = y_2$ or not. If $y_1 = y_2$, $T(\hat{\delta}_1^2 - \hat{\delta}_2^2)$ is asymptotically distributed as a linear combination of χ_1^2 random variables with both positive and negative weights. If $y_1 \neq y_2$ and both models are correctly specified, $T(\hat{\delta}_1^2 - \hat{\delta}_2^2)$ is still asymptotically distributed as a linear combination of χ_1^2 random variables, but the weights and the number of χ_1^2 random variables

are different from the case of $y_1 = y_2$. Finally, if $y_1 \neq y_2$ and both models are misspecified, $\sqrt{T}(\hat{\delta}_1^2 - \hat{\delta}_2^2)$ is asymptotically normally distributed.

The three different asymptotic distributions of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ in the non-nested models case present a significant challenge in testing $H_0 : \delta_1^2 = \delta_2^2$. Namely, which asymptotic distribution should we use to perform the test of $H_0 : \delta_1^2 = \delta_2^2$? One approach is to perform a sequential test, as suggested by Vuong (1989). In our context, this procedure involves first testing $H_0 : y_1 = y_2$ using Equation (42). If we reject $H_0 : y_1 = y_2$, then we use Equation (48) to test $H_0 : \delta_1^2 = \delta_2^2 = 0$. If this hypothesis is rejected, then we use Equation (53) to test $H_0 : \delta_1^2 = \delta_2^2 \neq 0$.⁹ Suppose α_1 , α_2 , and α_3 are the asymptotic significance levels used in these three tests. Then the sequential test has a significance level that is asymptotically bounded above by $\max[\alpha_1, \alpha_2, \alpha_3]$. Thus, if $\alpha_1 = \alpha_2 = \alpha_3 = 0.05$, the significance level of this procedure, as a test of $H_0 : \delta_1^2 = \delta_2^2$, is asymptotically no larger than 5%.

Another approach is to just perform the normal test in Proposition 6. This amounts to assuming that $y_1 \neq y_2$ and that both models are misspecified. The first assumption seems reasonable since most of our models have only the constant term in common. Consequently, $y_1 = y_2$ implies that the risk premia in both models are jointly equal to zero, a very unlikely scenario. The second assumption is sensible because asset pricing models are approximations of reality and we do not expect them to be correctly specified.

We implement both tests in our empirical analysis. Out of the 126 pairwise non-nested model comparisons that we perform, the sequential test produces only seven rejections of equality of HJ-distances at the 5% level, even less than the 14 rejections produced by the normal test. Therefore, the results from using the sequential test strengthen our claim that the data are too noisy for us to conclude that one model clearly outperforms the others. Nevertheless, for simplicity and ease of comparison, we report only the results based on the normal test for non-nested models in our tables.¹⁰

⁹We can also use a sequential test based on Golden's (2003) test of $v_d = 0$. However, unlike Golden's test, which involves a weighted chi-squared distribution, our test is much easier to implement since it involves only a chi-squared distribution.

¹⁰The results from using the sequential test are available upon request.

3. Empirical Analysis

We illustrate the relevance of our methodology with an empirical application. First, we describe the data used in the empirical analysis and outline the different specifications of the linear SDFs considered. Second, we present our results.

3.1 Data

3.1.1 Asset Returns and Conditioning Variables

We use monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate and the gross one-month T-bill rate (26 portfolios) to compare various asset pricing models. The returns on the size and book-to-market ranked portfolios are from Kenneth French's Web site. The one-month T-bill rate is from Ibbotson Associates (SBBI module) and pertains to a bill with at least one month to maturity.¹¹ For most of our time series, the data are from January 1952 to December 2006 (660 monthly observations). Following Hodrick and Zhang (2001), we consider unconditional as well as conditional models. For conditional models, the conditioning variables are either the cyclical part of the natural logarithm of the industrial production index lagged one period (*Lag IP*) or a January dummy (*JAN*). The industrial production index is from the Board of Governors of the Federal Reserve System. To initialize the cyclical series we use the Hodrick-Prescott (1997) filter on the five years of data that precede the starting date of our sample.¹²

3.1.2 Economic Variables and Asset Pricing Models

In our empirical analysis, we analyze six asset pricing models. These are the same models that were considered by Hodrick and Zhang (2001). The first model is the CAPM, which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{vw} r_t^{vw}, \quad (55)$$

where r_t^{vw} is the excess return on the value-weighted combined NYSE-AMEX-NASDAQ index from the Center for Research in Security Prices (CRSP).

¹¹In an earlier version of the paper, we also perform an additional analysis using quarterly data. The central messages of the paper are not affected by the choice of the return horizon.

¹²We set the smoothing parameter equal to 4,800, a fairly common value when using monthly data.

The second model is a linearized consumption CAPM (C-CAPM) that assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{cg} r_t^{cg}, \quad (56)$$

where r_t^{cg} is the growth rate in real nondurables consumption from the Bureau of Economic Analysis, U.S. Department of Commerce. For the C-CAPM, we have only monthly data starting in February 1959 (575 monthly observations).

The third model (JW) is the conditional CAPM of Jagannathan and Wang (1996), which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{jvw} r_t^{jvw} + \lambda_{prem} r_{t-1}^{prem} + \lambda_{lab} r_t^{lab}, \quad (57)$$

where r_t^{jvw} is the return on the valued-weighted combined NYSE-AMEX-NASDAQ index from CRSP, r_{t-1}^{prem} is the lagged yield spread between BAA and AAA rated corporate bonds (from the Board of Governors of the Federal Reserve System), and r_t^{lab} is the growth rate in per capita labor income. Per capita labor income, L , is defined as the difference between total personal income and dividend payments divided by the total population (from the Bureau of Economic Analysis, U.S. Department of Commerce). Following Jagannathan and Wang (1996), we use a two-month moving average to construct the growth rate in per capita labor income, $r_t^{lab} = (L_{t-1} + L_{t-2}) / (L_{t-2} + L_{t-3}) - 1$, for the purpose of minimizing the influence of measurement error.

The fourth model (CAMP) is a linearized version of Campbell's (1996) intertemporal capital asset pricing model that assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{rvw} r_t^{rvw} + \lambda_{clab} r_t^{clab} + \lambda_{div} r_{t-1}^{div} + \lambda_{rtb} r_{t-1}^{rtb} + \lambda_{trm} r_{t-1}^{trm}, \quad (58)$$

where r_t^{rvw} is the real return on the CRSP value-weighted index, r_t^{clab} is the monthly growth rate in real labor income (constructed differently from the JW labor series), r_{t-1}^{div} is the dividend yield on the CRSP value-weighted market portfolio, r_{t-1}^{rtb} is the difference between the one-month T-bill rate and its one-year backward moving average, and r_{t-1}^{trm} is the yield spread between long-term and short-term government bonds. The last three variables are lagged variables for forecasting returns, and they are known to the market at the end of month $t-1$. Following Campbell (1996), the factors used in the model are in fact innovations (in percentage points per month) in these five variables from a first-order vector autoregression. For the CAMP model, the data are obtained directly from

Campbell, and we have only monthly data covering the period from February 1952 to December 1990 (467 monthly observations).

The fifth model (FF3) is the Fama-French (1993) three-factor model, which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{vw}r_t^{vw} + \lambda_{smb}r_t^{smb} + \lambda_{hml}r_t^{hml}, \quad (59)$$

where r_t^{smb} is the return difference between portfolios of small and large stocks and r_t^{hml} is the return difference between portfolios of high and low book-to-market ratios. The Fama-French factors are from Kenneth French's Web site.

The sixth model (FF5) is the Fama-French (1993) five-factor model, which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{vw}r_t^{vw} + \lambda_{smb}r_t^{smb} + \lambda_{hml}r_t^{hml} + \lambda_{term}r_t^{term} + \lambda_{def}r_t^{def}, \quad (60)$$

where r_t^{term} is the return spread between a 30-year Treasury bond and the one-month T-bill (from Ibbotson Associates), and r_t^{def} is the return spread between long-term corporate and long-term government bonds (from Ibbotson Associates).

To form conditional models, we assume that the λ 's are linear functions of a conditioning variable (either *Lag IP* or *JAN*). This is equivalent to scaling the factors of the unconditional monthly models described above by a constant and the conditioning variable. Consequently, in the conditional case, the smallest model will have four factors and the biggest model will have twelve factors.¹³ Scaling factors by instruments is one popular way of allowing factor risk premia to vary over time. Examples of this type of practice are found in Ferson and Harvey (1991, 1999) and Campbell (1996), among others.

3.2 Results

First, we provide a summary of the different asset pricing models considered. Second, we analyze the impact of potential model misspecification on the statistical properties of the estimated SDF parameters. Third, we present the results of our tests of equality of the HJ-distances of two models.

¹³Although the JW and the CAMP models are already unconditional versions of conditional models, we follow Hodrick and Zhang (2001) and scale their factors by a constant and either *Lag IP* or *JAN*.

3.2.1 Summary of the Models

Table 1 provides a summary of the estimation results of different asset pricing models. The estimates of the HJ-distance are denoted by $\hat{\delta}$. The p -value of the test of $H_0 : \delta = 0$ from Equation (15) is $p(\hat{\delta} = 0)$. The standard error of the sample HJ-distance from Equation (19) computed under the alternative hypothesis that $\delta \neq 0$ is $se(\hat{\delta})$.¹⁴ The 95% confidence interval for δ based on the statistical method is $CI(\hat{\delta})$. “No. of par.” is the number of parameters in each asset pricing model.

Table 1 about here

In Panel A, we present the estimation results for the unconditional asset pricing models, and we find that all models are rejected by the data at the 5% level. This provides compelling evidence to incorporate model misspecification into our statistical analysis. Despite having the lowest HJ-distance, the CAMP model still does not pass the test of $H_0 : \delta = 0$ at the 5% level. Moreover, an examination of the 95% confidence interval for its δ also indicates that the HJ-distance of this model is far from zero. At this point, it is necessary to enter a caveat: although in Panel A the specification tests and the confidence intervals analyses produce outcomes that are consistent with each other, this does not always have to be the case. There can be cases where the specification test cannot reject $H_0 : \delta = 0$, but the asymptotic confidence interval for δ does not cover zero. The reason behind possibly different outcomes provided by the specification tests and confidence intervals analyses is that they are based on different asymptotic distributions. The p -value from testing $H_0 : \delta = 0$ is computed under the hypothesis that the model is correctly specified, whereas the confidence interval for δ is constructed using the asymptotic distribution of $\hat{\delta}$ under misspecified models. It is important to emphasize that this type of behavior arises because of the discontinuity of the asymptotic distribution of $\hat{\delta}$ at $\delta = 0$.¹⁵

In addition to the rejection of the models, the confidence intervals for δ of different models significantly overlap with each other, possibly suggesting that, after accounting for sampling variability,

¹⁴The $se(\hat{\delta})$'s are computed assuming no serial correlation. A separate set of results (available upon request) considers a 12-lag Newey-West (1987) adjustment. Overall, accounting for serial correlation in the data makes the standard errors of $\hat{\delta}$ and the p -values for testing $H_0 : \delta = 0$ slightly higher.

¹⁵In the statistics literature, it is not uncommon to see dramatic changes in the asymptotic distribution of parameter estimates and sample test statistics moving from one true value to another. For example, in the unit-root literature, the asymptotic distribution of the AR(1) parameter estimate substantially changes when the true parameter is near or on the unit boundary.

it might be difficult to detect substantial differences in the HJ-distances of competing models.

In Panels B and C, we report the estimation results when we scale the factors by either *Lag IP* or *JAN*, respectively. In both cases, the estimates of the HJ-distances of the conditional models are smaller than the corresponding estimates of the unconditional models. There can be two reasons for the smaller HJ-distances of the conditional models: (i) the conditioning information reduces the pricing errors by allowing the prices of risk to vary with the business cycle; and (ii) the use of conditioning information effectively doubles the number of factors and parameters making the conditional models better able to fit the data. In the scaled factor case, some models pass the HJ-distance test. Specifically, when we scale the factors by *Lag IP*, the JW, CAMP, and FF5 models are not rejected by the data at the 5% level, as shown in Panel B. When scaling the factors by *JAN*, the JW and the CAMP models are not rejected by the data at the 5% level, as shown in Panel C. However, an inspection of the confidence intervals for the HJ-distances suggests that the HJ-distances of all models are far from zero. In addition, the confidence intervals for δ of different models significantly overlap with each other.

By observing that conditional models always deliver smaller sample HJ-distances than the unconditional models, one might be tempted to conclude that conditional models perform better than their unconditional counterparts. However, there are two issues to be aware of when considering conditional models. The first effect of scaling is that the standard errors of $\hat{\delta}$ become larger. The larger standard errors reflect the additional noise brought into the model by the instruments. A direct implication is that it may be hard to distinguish conditional models from their unconditional counterparts. The formal model comparison tests discussed below will confirm this intuition. The second effect of scaling is that the number of factors becomes large relative to the number of assets. When K is large relative to N , Kan and Zhou (2004) argue that using asymptotic results might not be entirely appropriate and derive the finite sample distribution of $\hat{\delta}$ under the null and the alternative hypotheses for the case in which factors and returns are jointly normally distributed.

From this preliminary analysis, one may conclude that no model consistently outperforms the others because (i) different models pass the HJ-distance test depending on the scaling; and (ii) the confidence intervals for the HJ-distances of different models significantly overlap with each other. However, neither the p -values of the sample HJ-distances nor the confidence intervals analysis allow us to compare models formally. In the subsequent empirical analysis, we conduct our tests

of equality of HJ-distances to investigate whether a specific asset pricing model outperforms the others.

3.2.2 Properties of the SDF Parameter Estimates under Correctly Specified and Potentially Misspecified Models

Before turning to model comparison, we empirically investigate whether model misspecification substantially affects the properties of the SDF parameter estimates. Statistically significant SDF parameter estimates are often interpreted as evidence that the underlying factors are priced sources of risk. All existing studies test whether or not a factor is priced by using a standard error that assumes that the model is correctly specified. As we argued in the introduction, it is difficult to justify this practice when estimating the SDF parameters for many different models because some (if not all) of the models are bound to be misspecified. In this subsection, we empirically investigate whether using an asymptotic variance that is robust to model misspecification instead of an asymptotic variance that assumes a correctly specified model could lead us to different conclusions in terms of a factor being priced or not.

In Table 2, we focus on the SDF parameter estimates, $\hat{\lambda}$, of unconditional models. For each model, we report $\hat{\lambda}$ and associated t -ratios under correctly specified and potentially misspecified models.¹⁶ In computing t -ratios under correctly specified models, we use the sample counterpart of Equation (28), while in computing t -ratios under potential model misspecification, we use the sample counterpart of Equation (27). Consistent with our theoretical results, we find that the t -ratios under correctly specified and potentially misspecified models are about the same for factors that are traded, while they largely differ for factors that are not traded, such as macroeconomic factors. Consider, for example, the CAPM results. The t -ratios on $\hat{\lambda}_{vw}$ for correctly specified and potentially misspecified models are practically identical. The same type of conclusion emerges from an inspection of the FF3 model. However, when we consider models with non-traded factors, the picture substantially changes. For example, for the C-CAPM, we go from a t -ratio on $\hat{\lambda}_{cg}$ of -2.97 to a t -ratio of -1.91 and, for the JW model, we go from a t -ratio on $\hat{\lambda}_{lab}$ of 2.90 to a t -ratio of 1.40 . To summarize, we find that for non-traded factors, all the t -ratios under potentially misspecified

¹⁶The t -ratios are computed by assuming that the errors have no serial correlation. A separate set of results (available upon request) considers a 12-lag Newey-West (1987) adjustment. Overall, accounting for serial correlation in the data makes the standard errors of $\hat{\lambda}$ bigger.

models are smaller (in absolute value) than the t -ratios under correctly specified models. Hence, ignoring model misspecification can lead to the erroneous conclusion that certain factors are priced.

Table 2 about here

For many of the conditional models there are a lot of parameters. Instead of reporting all the parameter estimates, we explore the impact of potential model misspecification on the Wald tests of joint significance of the parameters. The null hypothesis of the Wald test is that the parameters associated with the scaled factors are jointly equal to zero. Given the results in Lemma 2, this Wald test is also a test of $H_0 : \delta_1^2 = \delta_2^2$, where model 1 is the unconditional model, which is nested by model 2, the conditional model. In Table 3, we report the Wald test statistics under correct specification (cs) and potential misspecification (m) for various conditional models. Panels A and B contain the results when we scale the factors by *Lag IP* and *JAN*, respectively. Once again, we find that ignoring potential model misspecification makes a substantial difference in terms of the p -values of the Wald tests. When using the traditional Wald test that assumes the models are correctly specified, we can reject the null that the parameters of the scaled factors are jointly equal to zero for several conditional models (see, for example, the C-CAPM and the JW model). However, when we account for potential model misspecification, the p -values of the Wald tests substantially increase, and we can no longer reject the null hypothesis that the conditional C-CAPM is just as good as its unconditional version.¹⁷ Therefore, although conditional models always deliver lower sample HJ-distances than unconditional models, we do not find strong statistical evidence to conclude that conditional models are better than unconditional models in terms of HJ-distance after we account for potential model misspecification.

Table 3 about here

Although not reported (results are available upon request), we also compute the t -ratios of the estimates of the conditional models under both correctly specified and potentially misspecified models. We find that most of the scaled factors have very low correlations with returns. As a result, many of the scaled factors are no longer statistically significant once potential model misspecification

¹⁷The p -values of the Wald tests are computed assuming no serial correlation. A separate set of results (available upon request) considers a 12-lag Newey-West adjustment. Overall, accounting for serial correlation in the data makes the p -values even larger.

is taken into account. For example, for the C-CAPM with the consumption growth scaled by JAN , we go from a t -ratio of -2.19 under correctly specified models to a t -ratio of -1.41 under potentially misspecified models. For the CAMP model, we go from a t -ratio of 2.25 to a t -ratio of 1.73 for the trm factor that is scaled with JAN .

To summarize, accounting for model misspecification can often make a qualitative difference in determining whether or not a factor is priced, especially when the factor has low correlation with asset returns. This would typically be the case when the factor is a macroeconomic factor, or when the factor is scaled by an instrument. Unless one is certain that a model is correctly specified, potential model misspecification should be accounted for when computing the standard errors of the estimates of SDF parameters.

3.2.3 Tests of Equality of the HJ-Distances of Two Models

In this subsection, we empirically investigate whether competing asset pricing models exhibit significantly different sample HJ-distances. Failure to find significant differences across models would imply that the commonly used returns and factors are too noisy for us to conclude that one model is clearly superior to the others. In the theoretical section of the paper, we show that the asymptotic distribution of our test statistic, the difference between the sample squared HJ-distances of two models, depends on whether the competing models are correctly specified or misspecified and on whether they are nested or non-nested. For nested models, we use Proposition 2 instead of Lemma 2 to conduct the tests of equality of HJ-distances.¹⁸ For nested models, we report our results using the misspecification robust version of $\hat{V}(\hat{\lambda}_2)$ because it is applicable to correctly specified as well as misspecified models. For non-nested models, we use the asymptotic normal distribution in Proposition 6 to compute the p -value of the test statistic.

In Table 4, we report pairwise tests of equality of squared HJ-distances for different models, some of them being nested models and others being non-nested models. In Panel A, we provide pairwise comparisons of different unconditional models. In Panels B and C, we compare different conditional models when they are scaled by $Lag IP$ and JAN , respectively. In each panel, we report the differences between the squared sample HJ-distances of different pairs of models and the

¹⁸Results obtained using Lemma 2 (not reported in the paper) are largely consistent with the ones shown in the tables.

associated p -values (in parentheses).¹⁹

Table 4 about here

When comparing unconditional models, we observe that the CAPM is outperformed by the FF3 and FF5. However, we find no evidence that intertemporal CAPM-type specifications such as the Campbell (1996) or the Jagannathan and Wang (1996) models outperform the unconditional CAPM. When considering conditional models with scaled factors, we find it equally hard to distinguish the performances of different models. Only the conditional JW model outperforms the conditional CAPM when the factors are scaled by *Lag IP* (see Panel B). All the other models are indistinguishable from each other in terms of HJ-distance.

Next, we investigate whether conditional models perform substantially better than unconditional models. The reason behind this type of exercise is that the HJ-distances of the conditional models are always lower than the HJ-distances of their unconditional counterparts, as shown in Table 1. However, it may be premature to conclude that the instruments actually help to reduce the pricing errors without performing a formal comparison of the unconditional models versus the conditional models. In addition, we also investigate whether conditional models scaled by one instrument are better than conditional models scaled by another instrument. This exercise is also of interest because different conditional models might capture different characteristics of the economy, and the type of scaling might affect their absolute and relative performances.

In Table 5, we report the results from testing the equality of HJ-distances between conditional and unconditional models. Panels A and B compare the unconditional models with conditional models that are scaled by *Lag IP* and *JAN*, respectively. Panel C compares two sets of conditional models, with one set scaled by *Lag IP* and the other set scaled by *JAN*. The first noticeable pattern is that the p -values along the main diagonal of each panel are not significant at the 5% level. This suggests that for a given model, overall we cannot find statistically significant differences in HJ-distances between the different conditional versions and the unconditional version of the model. These findings indicate that the instruments add noise to the data, making it hard to detect significant differences between the different conditional versions and the unconditional version of a given model.

¹⁹Note that in the case of non-nested models, the reported p -values are two-tailed p -values.

Table 5 about here

Across different model specifications, we find that the unconditional CAPM is outperformed by the conditional JW model when scaling by *Lag IP*. When scaling by *JAN*, we find that the unconditional CAPM is outperformed by the conditional JW and CAMP models. The unconditional C-CAPM is always outperformed by the conditional FF5 model. When scaling by *JAN*, we also find the unconditional C-CAPM to be outperformed by the CAMP and FF3 models. In addition, we find that there are a few cases in which some unconditional models perform better than some conditional models. For example, the unconditional FF3 and FF5 models in Panel A perform better than the conditional CAPM. Finally, Panel C shows that there are a few cases in which the equality of HJ-distances is rejected at the 5% level. We find that the conditional CAPM scaled by *Lag IP* is dominated by the conditional JW, CAMP, FF3, and FF5 models scaled by *JAN*. Overall, our econometric analysis suggests that once instruments are used, there is too much noise in the data for us to conclude that one conditional model clearly outperforms the others.

In synthesis, out of 153 pairwise model comparisons in Tables 4 and 5, we find only 16 cases in which the differences in sample HJ-distances between models are statistically significant at the 5% level. Note that all the p -values in Tables 4 and 5 are computed assuming no serial correlation. When we consider a 12-lag Newey-West (1987) adjustment (results are available upon request), we find that most of the p -values of the test statistics become larger and that the differences between models are even harder to detect. Out of 153 pairwise model comparisons in Tables 4 and 5, there are now only five cases where the differences in sample HJ-distances between models are statistically significant at the 5% level. These low rejection rates suggest that the data are generally too noisy for us to conclude that one model clearly outperforms the others.

3.3 Robustness Analysis

Lewellen, Nagel, and Shanken (2006) note that since the returns on the Fama-French size and book-to-market ranked portfolios exhibit a strong factor structure, it is relatively easy for a factor model to produce a mechanically good fit of the expected returns on these portfolios. Following their suggestion, we perform an analysis by adding 30 industry portfolios to our test assets to help us differentiate competing asset pricing models. For this larger set of test assets, it is also difficult

to find statistically significant difference in the HJ-distances of different models. Out of the 153 pairwise model comparisons, we find only 22 cases in which the differences in sample HJ-distances between models are statistically significant at the 5% level.²⁰ Therefore, even with a much larger set of test assets, the data are still too noisy for us to conclude that one model clearly outperforms the others.

4. Conclusion

In this paper, we propose a methodology to test whether or not two competing linear asset pricing models have the same HJ-distance. Under general distributional assumptions, we present the asymptotic distribution of the difference between the sample squared HJ-distances of two models. We show that the asymptotic distribution of this difference depends on whether the competing models are correctly specified or misspecified, and on whether the competing models are nested or non-nested.

In addition, we contribute to the existing literature by proposing a simple methodology for computing the standard errors of the estimated SDF parameters that are robust to model misspecification. For the case in which returns and factors are multivariate elliptically distributed, we are able to show analytically that the standard errors under misspecified models are always bigger than the standard errors that assume the model is correctly specified. Moreover, we show that the misspecification adjustment depends on, among other things, the correlation between the factor and the returns on the test assets. This adjustment can be very large when the underlying factor is poorly mimicked by asset returns. A nice feature of our misspecification robust standard errors is that they can be used whether the model is correctly specified or misspecified.

We conduct our empirical analysis on a variety of asset pricing models that have been proposed in the literature. We find that many of the non-traded factors in several intertemporal CAPM-type specifications are no longer priced when potential model misspecification is taken into account. On the contrary, the statistical significance of the traded factors is not greatly affected when we use our misspecification robust standard errors. In addition, we find that the commonly used returns and factors are, for the most part, too noisy for us to conclude that one model outperforms

²⁰To conserve space, we do not report the estimation results using this larger set of test assets, but the results are available upon request.

the others in terms of HJ-distance. Specifically, there is little evidence that conditional and intertemporal CAPM-type specifications outperform even the simple unconditional CAPM in terms of HJ-distance.

While we do not find many statistically significant differences between the HJ-distances of the scaled factor models and the unscaled factor models, this does not necessarily mean that the conditional models do not perform better than the unconditional models. The sample HJ-distances of competing models may be very noisy and have little power in differentiating good models from bad models. However, explicitly accounting for the uncertainty associated with the difference between the sample HJ-distances of two competing models is still better than simply relying on the point estimates of the HJ-distances. Moreover, it is not clear that other measures of model misspecification (such as the ordinary least squares (OLS) R^2 or other aggregate measures of pricing errors) would allow us to overcome this problem. As aggregates of sample pricing errors, these other measures can be just as noisy as the sample HJ-distance, and more important, they may not be economically as meaningful as the HJ-distance.

Our analysis could be extended in a number of ways. For instance, our methodology could be modified to accommodate nonlinear stochastic discount factors. In addition, testing the equality of HJ-distances of more than two models is, in principle, feasible. Future research should also address the small sample properties of the test statistics proposed in this paper. Finally, our analysis can also be used to develop tests of equality of other measures of model misspecification.

Appendix

Proof of Proposition 1: Note that $\hat{\lambda}$ is a smooth function of $\hat{\mu}$ and \hat{V} . Therefore, once we have the asymptotic distribution of $\hat{\mu}$ and \hat{V} , we can use the delta method to obtain the asymptotic distribution of $\hat{\lambda}$. Let

$$\phi = \begin{bmatrix} \mu \\ \text{vec}(V) \end{bmatrix}, \quad \hat{\phi} = \begin{bmatrix} \hat{\mu} \\ \text{vec}(\hat{V}) \end{bmatrix}. \quad (\text{A1})$$

Under some standard regularity conditions, we can assume²¹

$$\sqrt{T}(\hat{\phi} - \phi) \overset{A}{\approx} N(0_{(N+K) \times (N+K+1)}, S_0). \quad (\text{A2})$$

We first note that $\hat{\mu}$ and \hat{V} can be written as the GMM estimator that uses the moment conditions $E[r_t(\phi)] = 0_{(N+K)(N+K+1)}$, where

$$r_t(\phi) = \begin{bmatrix} Y_t - \mu \\ \text{vec}((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}. \quad (\text{A3})$$

Since this is an exactly identified system of moment conditions, it is straightforward to verify that the asymptotic variance of $\sqrt{T}(\hat{\phi} - \phi)$ is given by

$$S_0 = \sum_{j=-\infty}^{\infty} E[r_t(\phi)r_{t+j}(\phi)']. \quad (\text{A4})$$

Using the delta method, the asymptotic distribution of $\hat{\lambda}$ under the misspecified model is given by

$$\sqrt{T}(\hat{\lambda} - \lambda) \overset{A}{\approx} N\left(0_{K+1}, \left[\frac{\partial \lambda}{\partial \phi'} \right]' S_0 \left[\frac{\partial \lambda}{\partial \phi'} \right] \right). \quad (\text{A5})$$

The expression of $\partial \lambda / \partial \phi'$ is presented next.

Claim: Let $e = D\lambda - 1_N$. We have

$$\frac{\partial \lambda}{\partial \mu'_1} = -[1, 0'_K]' \lambda'_1, \quad (\text{A6})$$

$$\frac{\partial \lambda}{\partial \mu'_2} = -H [1, \mu'_1]' e' V_{22}^{-1} - HD' V_{22}^{-1} \mu_y, \quad (\text{A7})$$

$$\begin{aligned} \frac{\partial \lambda}{\partial \text{vec}(V)'} &= [H[0_K, I_K]', 0_{(K+1) \times N}] \otimes [0'_K, -e' V_{22}^{-1}] \\ &\quad + [-\lambda'_1, e' V_{22}^{-1}] \otimes [0_{(K+1) \times K}, HD' V_{22}^{-1}]. \end{aligned} \quad (\text{A8})$$

²¹Note that S_0 is a singular matrix as \hat{V} is symmetric, so there are redundant elements in $\hat{\phi}$. We could have written $\hat{\phi}$ as $[\hat{\mu}', \text{vech}(\hat{V})']'$, but the results are the same under both specifications.

Proof: Let $d = \text{vec}(D)$. It is straightforward to show that

$$\frac{\partial d}{\partial \mu'_1} = \begin{bmatrix} 0_{N \times K} \\ I_K \otimes \mu_2 \end{bmatrix} = \begin{bmatrix} 0'_K \\ I_K \end{bmatrix} \otimes \mu_2, \quad (\text{A9})$$

$$\frac{\partial d}{\partial \mu'_2} = \begin{bmatrix} I_N \\ \mu_1 \otimes I_N \end{bmatrix} = \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix} \otimes I_N, \quad (\text{A10})$$

$$\frac{\partial d}{\partial \text{vec}(V)'} = [[0_K, I_K]', 0_{(K+1) \times N}] \otimes [0_{N \times K}, I_N]. \quad (\text{A11})$$

Define $K_{m,n}$ as a commutation matrix (see, e.g., Magnus and Neudecker (1999)) such that $K_{m,n} \text{vec}(A) = \text{vec}(A')$ where A is an $m \times n$ matrix. In addition, we denote $K_{n,n}$ by K_n . Note that

$$\frac{\partial \text{vec}(D')}{\partial d'} = \frac{\partial K_{N,K+1} d}{\partial d'} = K_{N,K+1}, \quad (\text{A12})$$

$$\begin{aligned} \frac{\partial \text{vec}(D'V_{22}^{-1}D)}{\partial d'} &= (D'V_{22}^{-1} \otimes I_{K+1}) \frac{\partial \text{vec}(D')}{\partial d'} + (I_{K+1} \otimes D'V_{22}^{-1}) \frac{\partial d}{\partial d'} \\ &= (D'V_{22}^{-1} \otimes I_{K+1}) K_{N,K+1} + (I_{K+1} \otimes D'V_{22}^{-1}) \\ &= (I_{(K+1)^2} + K_{K+1})(I_{K+1} \otimes D'V_{22}^{-1}), \end{aligned} \quad (\text{A13})$$

$$\frac{\partial \text{vec}((D'V_{22}^{-1}D)^{-1})}{\partial \text{vec}(D'V_{22}^{-1}D)'} = -(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1}, \quad (\text{A14})$$

$$\frac{\partial \text{vec}((D'V_{22}^{-1}D)^{-1})}{\partial d'} = -(I_{(K+1)^2} + K_{K+1})[(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1} D'V_{22}^{-1}], \quad (\text{A15})$$

$$\begin{aligned} \frac{\partial \text{vec}((D'V_{22}^{-1}D)^{-1}D')}{\partial d'} &= [I_N \otimes (D'V_{22}^{-1}D)^{-1}] \frac{\partial \text{vec}(D')}{\partial d'} + (D \otimes I_{K+1}) \frac{\partial \text{vec}((D'V_{22}^{-1}D)^{-1})}{\partial d'} \\ &= [I_N \otimes (D'V_{22}^{-1}D)^{-1}] K_{N,K+1} \\ &\quad - (D \otimes I_{K+1})(I_{(K+1)^2} + K_{K+1})[(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1} D'V_{22}^{-1}] \\ &= K_{N,K+1}[(D'V_{22}^{-1}D)^{-1} \otimes I_N] - D(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1} D'V_{22}^{-1} \\ &\quad - K_{N,K+1}[(D'V_{22}^{-1}D)^{-1} \otimes D(D'V_{22}^{-1}D)^{-1} D'V_{22}^{-1}] \\ &= K_{N,K+1}[(D'V_{22}^{-1}D)^{-1} \otimes [I_N - D(D'V_{22}^{-1}D)^{-1} D'V_{22}^{-1}]] \\ &\quad - D(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1} D'V_{22}^{-1}. \end{aligned} \quad (\text{A16})$$

Therefore

$$\frac{\partial \lambda}{\partial d'} = (1'_N V_{22}^{-1} \otimes I_{K+1}) \frac{\partial \text{vec}(HD')}{\partial d'} = -H \otimes e'V_{22}^{-1} - \lambda' \otimes HD'V_{22}^{-1}. \quad (\text{A17})$$

It follows that

$$\frac{\partial \lambda}{\partial \mu'_1} = \frac{\partial \lambda}{\partial d'} \frac{\partial d}{\partial \mu'_1} = -\lambda'_1 \otimes HD'V_{22}^{-1} \mu_2 = -[1, 0'_K]' \lambda'_1, \quad (\text{A18})$$

$$\frac{\partial \lambda}{\partial \mu'_2} = \frac{\partial \lambda}{\partial d'} \frac{\partial d}{\partial \mu'_2} = -H[1, \mu'_1]' e'V_{22}^{-1} - HD'V_{22}^{-1} \mu_y, \quad (\text{A19})$$

where the last equality in Equation (A18) follows from the fact that $HD'V_{22}^{-1}\mu_2 = HD'V_{22}^{-1}D[1, 0'_K]' = [1, 0'_K]'$.

For the derivative of λ with respect to $\text{vec}(V)$, we use the product rule to obtain

$$\frac{\partial \lambda}{\partial \text{vec}(V)'} = (1'_N V_{22}^{-1} D \otimes I_{K+1}) \frac{\partial \text{vec}(H)}{\partial \text{vec}(V)'} + (1'_N V_{22}^{-1} \otimes H) \frac{\partial \text{vec}(D')}{\partial \text{vec}(V)'} + (1'_N \otimes HD') \frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'}. \quad (\text{A20})$$

The last two terms are given by

$$(1'_N V_{22}^{-1} \otimes H) \frac{\partial \text{vec}(D')}{\partial \text{vec}(V)'} = [H [0_K, I_K]', 0_{(K+1) \times N}] \otimes [0'_K, 1'_N V_{22}^{-1}], \quad (\text{A21})$$

$$(1'_N \otimes HD') \frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'} = -[0'_K, 1'_N V_{22}^{-1}] \otimes [0_{(K+1) \times K}, HD'V_{22}^{-1}]. \quad (\text{A22})$$

For the first term, we use the chain rule to obtain

$$\begin{aligned} & (1'_N V_{22}^{-1} D \otimes I_{K+1}) \frac{\partial \text{vec}(H)}{\partial \text{vec}(V)'} \\ = & (1'_N V_{22}^{-1} D \otimes I_{K+1}) \frac{\partial \text{vec}((D'V_{22}^{-1}D)^{-1})}{\partial \text{vec}(D'V_{22}^{-1}D)'} \frac{\partial \text{vec}(D'V_{22}^{-1}D)}{\partial \text{vec}(V)'} \\ = & -(1'_N V_{22}^{-1} D \otimes I_{K+1})(H \otimes H) \left[(D'V_{22}^{-1} \otimes I_{K+1}) \frac{\partial \text{vec}(D')}{\partial \text{vec}(V)'} \right. \\ & \left. + (D' \otimes D') \frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'} + (I_{K+1} \otimes D'V_{22}^{-1}) \frac{\partial \text{vec}(D)}{\partial \text{vec}(V)'} \right] \\ = & -(\lambda' \otimes H) \{ [[0_{(K+1) \times K}, D'V_{22}^{-1}] \otimes [[0_K, I_K]', 0_{(K+1) \times N}]] K_{N+K} \\ & - [0_{(K+1) \times K}, D'V_{22}^{-1}] \otimes [0_{(K+1) \times K}, D'V_{22}^{-1}] \\ & + [[0_K, I_K]', 0_{(K+1) \times N}] \otimes [0_{(K+1) \times K}, D'V_{22}^{-1}] \} \\ = & [H [0_K, I_K]', 0_{(K+1) \times N}] \otimes [0'_K, -\lambda' D'V_{22}^{-1}] + [0'_K, \lambda' D'V_{22}^{-1}] \otimes [0_{(K+1) \times K}, HD'V_{22}^{-1}] \\ & - [\lambda'_1, 0'_N] \otimes [0_{(K+1) \times K}, HD'V_{22}^{-1}]. \end{aligned} \quad (\text{A23})$$

Combining the three terms and using the identity $e = D\lambda - 1_N$, we have

$$\begin{aligned} \frac{\partial \lambda}{\partial \text{vec}(V)'} &= [H [0_K, I_K]', 0_{(K+1) \times N}] \otimes [0'_K, -e'V_{22}^{-1}] \\ &+ [-\lambda'_1, e'V_{22}^{-1}] \otimes [0_{(K+1) \times K}, HD'V_{22}^{-1}]. \end{aligned} \quad (\text{A24})$$

This completes the proof of the claim.

Using the expression of $\partial \lambda / \partial \phi'$, we can simplify the asymptotic variance of $\sqrt{T}(\hat{\lambda} - \lambda)$ to

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[h_t(\phi)h_{t+j}(\phi)'], \quad (\text{A25})$$

where

$$\begin{aligned}
h_t(\phi) &= \frac{\partial \lambda}{\partial \phi'} r_t(\phi) \\
&= -[1, 0'_K]' \lambda'_1(f_t - \mu_1) - (H [1, \mu'_1]' e' V_{22}^{-1} + \mu_y H D' V_{22}^{-1})(R_t - \mu_2) \\
&\quad + \text{vec} \left([0'_K, -e' V_{22}^{-1}] [(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} [0_K, I_K] H \\ 0_{N \times (K+1)} \end{bmatrix} \right) \\
&\quad + \text{vec} \left([0_{(K+1) \times K}, H D' V_{22}^{-1}] [(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} -\lambda_1 \\ V_{22}^{-1} e \end{bmatrix} \right) \\
&= \begin{bmatrix} -\lambda'_1(f_t - \mu_1) \\ 0_K \end{bmatrix} - H \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix} u_t - H D' V_{22}^{-1} (R_t - \mu_2) \mu_y \\
&\quad - H [0_K, I_K]' (f_t - \mu_1) u_t - H D' V_{22}^{-1} (R_t - \mu_2) (f_t - \mu_1)' \lambda_1 + H D' V_{22}^{-1} (R_t - \mu_2) u_t \\
&\quad + H [0_K, I_K]' V_{12} V_{22}^{-1} e + H D' V_{22}^{-1} V_{21} \lambda_1 - H D' V_{22}^{-1} e \\
&= -H D' V_{22}^{-1} (R_t - \mu_2) (y_t - u_t) - H x_t u_t - \begin{bmatrix} y_t \\ 0_K \end{bmatrix} + \lambda \\
&= -H D' V_{22}^{-1} R_t y_t + H [D' V_{22}^{-1} (R_t - \mu_2) - x_t] u_t + \lambda. \tag{A26}
\end{aligned}$$

Equation (A26) follows from the fact that $H D' V_{22}^{-1} V_{21} \lambda_1 = [-\mu'_1 \lambda_1, \lambda'_1]'$ and $H D' V_{22}^{-1} \mu_2 = [1, 0'_K]'$. In addition, the first-order condition of $D' V_{22}^{-1} e = 0_{K+1}$ implies that $\mu'_2 V_{22}^{-1} e = 0$ and $V_{12} V_{22}^{-1} e = 0_K$. Note that when the model is correctly specified, we have $e = 0_N$ and $u_t = 0$. In this case, we have

$$h_t(\phi) = -H D' V_{22}^{-1} R_t y_t + \lambda. \tag{A27}$$

This completes the proof.

Proof of Lemma 1: Let $q_t = H D' V_{22}^{-1} (R_t - \mu_2)$, $w_t = D' V_{22}^{-1} (R_t - \mu_2) - x_t$ and $z_t = [\lambda'_1 f_t, -\lambda'_1]'$. Since q_t , w_t , and z_t are linear functions of R_t and f_t , they are also jointly elliptically distributed.

Using the identity

$$H D' V_{22}^{-1} \mu_2 y_t = H D' V_{22}^{-1} D \begin{bmatrix} 1 \\ 0_K \end{bmatrix} y_t = \begin{bmatrix} y_t \\ 0_K \end{bmatrix}, \tag{A28}$$

we can write

$$h_t = -q_t y_t + H w_t u_t - z_t. \tag{A29}$$

It is straightforward to obtain $E[q_t] = 0_{K+1}$, $E[w_t] = -[1, \mu'_1]'$, $E[z_t] = [\lambda'_1 \mu_1, -\lambda'_1]'$, $\text{Var}[q_t] = H$, and

$$\text{Var}[w_t] = (\mu'_2 V_{22}^{-1} \mu_2) \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix} \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix}' + \begin{bmatrix} 0 & 0'_K \\ 0_K & V_{11} - V_{12} V_{22}^{-1} V_{21} \end{bmatrix}, \tag{A30}$$

$$\text{Var}[z_t] = \begin{bmatrix} \sigma_y^2 & 0'_K \\ 0_K & 0_{K \times K} \end{bmatrix}. \quad (\text{A31})$$

In addition, using the identity $D'V_{22}^{-1}e = 0_{K+1}$, we can obtain the following joint moments $E[q_t u_t] = 0_{K+1}$, $E[q_t y_t] = [-\mu'_1 \lambda_1, \lambda'_1]'$, $E[w_t u_t] = 0_{K+1}$, $E[z_t u_t] = 0_{K+1}$, $E[u_t y_t] = 0$. Using these moments and applying Lemma 2 of Maruyama and Seo (2003), we obtain

$$E[q_t w'_t y_t u_t] = 0_{(K+1) \times (K+1)}, \quad (\text{A32})$$

$$E[w_t z'_t u_t] = 0_{(K+1) \times (K+1)}, \quad (\text{A33})$$

$$E[w_t w'_t u_t^2] = \delta^2(E[w_t]E[w_t]' + (1 + \kappa)\text{Var}[w_t]), \quad (\text{A34})$$

$$E[q_t q'_t y_t^2] = [\mu_y^2 + (1 + \kappa)\sigma_y^2]H + 2(1 + \kappa) \begin{bmatrix} -\mu'_1 \lambda_1 \\ \lambda_1 \end{bmatrix} \begin{bmatrix} -\mu'_1 \lambda_1 \\ \lambda_1 \end{bmatrix}', \quad (\text{A35})$$

$$E[q_t z'_t y_t] = \begin{bmatrix} -\mu'_1 \lambda_1 \\ \lambda_1 \end{bmatrix} \begin{bmatrix} \mu_y + \mu'_1 \lambda_1 \\ -\lambda_1 \end{bmatrix}'. \quad (\text{A36})$$

Using Equations (A32) and (A33), we can write

$$V(\hat{\lambda}) = E[h_t h'_t] = E[q_t q'_t y_t^2] + E[q_t z'_t y_t] + E[z_t q'_t y_t] + E[z_t z'_t] + H E[w_t w'_t u_t^2] H. \quad (\text{A37})$$

Substituting Equations (A34)–(A36) in Equation (A37) and after simplification, we obtain our expression of $V(\hat{\lambda})$. This completes the proof.

Proof of Lemmas 2 and 3: For nested models, $\delta_1^2 = \delta_2^2$ holds if and only if $y_1 = y_2$. Since we can view Lemma 2 as a special case of Lemma 3 when $K_2 = 0$, we provide the proof of Lemma 3 here. Given that $y_1 = y_2$ if and only if $\eta_1 = \lambda_1$, $\eta_2 = 0_{K_2}$, and $\lambda_2 = 0_{K_3}$, it suffices to show that $\eta_2 = 0_{K_2}$ and $\lambda_2 = 0_{K_3}$ imply $\eta_1 = \lambda_1$. Premultiplying both sides of Equation (32) by $D'_1 V_{22}^{-1} D_1$, we obtain

$$\begin{bmatrix} D'_{1a} V_{22}^{-1} D_{1a} & D'_{1a} V_{22}^{-1} D_{1b} \\ D'_{1b} V_{22}^{-1} D_{1a} & D'_{1b} V_{22}^{-1} D_{1b} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} D'_{1a} V_{22}^{-1} 1_N \\ D'_{1b} V_{22}^{-1} 1_N \end{bmatrix}, \quad (\text{A38})$$

where D_{1a} is the first $K_1 + 1$ columns of D_1 and D_{1b} is the last K_2 columns of D_1 . The first block of this equation gives us

$$D'_{1a} V_{22}^{-1} D_{1a} \eta_1 + D'_{1a} V_{22}^{-1} D_{1b} \eta_2 = D'_{1a} V_{22}^{-1} 1_N. \quad (\text{A39})$$

When $\eta_2 = 0_{K_2}$, we have

$$\eta_1 = (D'_{1a} V_{22}^{-1} D_{1a})^{-1} D'_{1a} V_{22}^{-1} 1_N. \quad (\text{A40})$$

Similarly, premultiplying both sides of Equation (33) by $D_2' V_{22}^{-1} D_2$, when $\lambda_2 = 0_{K_3}$ we have

$$\lambda_1 = (D_{2a}' V_{22}^{-1} D_{2a})^{-1} D_{2a}' V_{22}^{-1} 1_N, \quad (\text{A41})$$

where D_{2a} is the first $K_1 + 1$ columns of D_2 . Since D_{1a} and D_{2a} are both equal to $E[R_t[1, f_{1t}]]$, we have $\eta_1 = \lambda_1$. This completes the proof.

Proof of Propositions 2 and 3: Since Proposition 2 is a special case of Proposition 3 when $K_2 = 0$, we prove only Proposition 3 here. We first provide a simplified expression for $\delta_1^2 - \delta_2^2$. Consider a model 0 that is linear in $x_0 = [1, f_1']'$. Let D_{1a} and D_{1b} be the first $K_1 + 1$ and the last K_2 columns of D_1 , respectively. The difference between the squared HJ-distances of models 0 and 1 is given by

$$\begin{aligned} \delta_0^2 - \delta_1^2 &= 1_N' V_{22}^{-1} D_1 H_1 D_1' V_{22}^{-1} 1_N - 1_N' V_{22}^{-1} D_1 \begin{bmatrix} (D_{1a}' V_{22}^{-1} D_{1a})^{-1} & 0_{(K_1+1) \times K_2} \\ 0_{K_2 \times (K_1+1)} & 0_{K_2 \times K_2} \end{bmatrix} D_1' V_{22}^{-1} 1_N \\ &= \eta'(D_1' V_{22}^{-1} D_1) \eta - \eta'(D_1' V_{22}^{-1} D_1) \begin{bmatrix} (D_{1a}' V_{22}^{-1} D_{1a})^{-1} & 0_{(K_1+1) \times K_2} \\ 0_{K_2 \times (K_1+1)} & 0_{K_2 \times K_2} \end{bmatrix} (D_1' V_{22}^{-1} D_1) \eta \\ &= \eta'(D_1' V_{22}^{-1} D_1) \eta - \eta' \begin{bmatrix} D_{1a}' V_{22}^{-1} D_{1a} & D_{1a}' V_{22}^{-1} D_{1b} \\ D_{1b}' V_{22}^{-1} D_{1a} & (D_{1b}' V_{22}^{-1} D_{1a})(D_{1a}' V_{22}^{-1} D_{1a})^{-1} (D_{1a}' V_{22}^{-1} D_{1b}) \end{bmatrix} \eta \\ &= \eta_2' [D_{1b}' V_{22}^{-1} D_{1b} - (D_{1b}' V_{22}^{-1} D_{1a})(D_{1a}' V_{22}^{-1} D_{1a})^{-1} (D_{1a}' V_{22}^{-1} D_{1b})] \eta_2 \\ &= \eta_2' H_{1,22}^{-1} \eta_2. \end{aligned} \quad (\text{A42})$$

Similarly, we have

$$\delta_0^2 - \delta_2^2 = \lambda_2' H_{2,22}^{-1} \lambda_2. \quad (\text{A43})$$

Subtracting Equation (A42) from Equation (A43), we obtain the following simple expression of $\delta_1^2 - \delta_2^2$:

$$\delta_1^2 - \delta_2^2 = -\eta_2' H_{1,22}^{-1} \eta_2 + \lambda_2' H_{2,22}^{-1} \lambda_2 = \psi' \begin{bmatrix} -H_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & H_{2,22}^{-1} \end{bmatrix} \psi. \quad (\text{A44})$$

Since this equation also holds for its sample counterpart, we can write

$$\hat{\delta}_1^2 - \hat{\delta}_2^2 = \hat{\psi}' \begin{bmatrix} -\hat{H}_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{H}_{2,22}^{-1} \end{bmatrix} \hat{\psi}. \quad (\text{A45})$$

Under the null hypothesis $H_0 : \psi = 0_{K_2+K_3}$, we have $z = \sqrt{T} V(\hat{\psi})^{-\frac{1}{2}} \hat{\psi} \stackrel{A}{\sim} N(0_{K_2+K_3}, I_{K_2+K_3})$ and

$$V(\hat{\psi})^{\frac{1}{2}} \begin{bmatrix} -\hat{H}_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{H}_{2,22}^{-1} \end{bmatrix} V(\hat{\psi})^{\frac{1}{2}} \xrightarrow{\text{a.s.}} V(\hat{\psi})^{\frac{1}{2}} \begin{bmatrix} -H_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & H_{2,22}^{-1} \end{bmatrix} V(\hat{\psi})^{\frac{1}{2}}. \quad (\text{A46})$$

It follows that

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \xrightarrow{d} z'V(\hat{\psi})^{\frac{1}{2}} \begin{bmatrix} -H_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & H_{2,22}^{-1} \end{bmatrix} V(\hat{\psi})^{\frac{1}{2}} z. \quad (\text{A47})$$

Let $Q\Xi Q'$ be the eigenvalue decomposition of the matrix in the middle, where $\Xi = \text{Diag}(\xi_1, \dots, \xi_{K_2+K_3})$ is a diagonal matrix of the eigenvalues of the matrix in the middle, or equivalently the eigenvalues of Equation (45), and Q is a matrix of the corresponding eigenvectors. Writing $\tilde{z} = Q'z \stackrel{A}{\sim} N(0_{K_2+K_3}, I_{K_2+K_3})$, we have

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \xrightarrow{d} \tilde{z}'\Xi\tilde{z} = \sum_{i=1}^{K_2+K_3} \xi_i \tilde{z}_i^2, \quad (\text{A48})$$

where $\tilde{z}_i^2 \stackrel{A}{\sim} \chi_1^2$, $i = 1, \dots, K_2 + K_3$, and they are independent of each other. This completes the proof.

Proof of Propositions 4 and 5: In order to obtain the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ for correctly specified models, we employ the Generalized Method of Moments (GMM) of Hansen (1982). When both models are correctly specified, we have $E[g_t(\theta)] = 0_{2N}$. The sample moment conditions are then given by

$$\bar{g}_T(\theta) = \begin{bmatrix} \bar{g}_{1T}(\eta) \\ \bar{g}_{2T}(\lambda) \end{bmatrix} = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T R_t x'_{1t} \eta - 1_N \\ \frac{1}{T} \sum_{t=1}^T R_t x'_{2t} \lambda - 1_N \end{bmatrix} = \begin{bmatrix} \hat{D}_1 \eta - 1_N \\ \hat{D}_2 \lambda - 1_N \end{bmatrix}. \quad (\text{A49})$$

The sample estimator of θ can be written as the solution to the following conditions:

$$A_T \bar{g}_T(\theta) = 0_{2K_1+K_2+K_3+2}, \quad (\text{A50})$$

where

$$A_T = \begin{bmatrix} \hat{D}'_1 \hat{V}_{22}^{-1} & 0_{(K_1+K_2+1) \times N} \\ 0_{(K_1+K_3+1) \times N} & \hat{D}'_2 \hat{V}_{22}^{-1} \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} D'_1 V_{22}^{-1} & 0_{(K_1+K_2+1) \times N} \\ 0_{(K_1+K_3+1) \times N} & D'_2 V_{22}^{-1} \end{bmatrix} \equiv A. \quad (\text{A51})$$

We define the derivative of the sample moment conditions with respect to the parameters as

$$G_T(\theta) = \begin{bmatrix} \hat{D}_1 & 0_{N \times (K_1+K_3+1)} \\ 0_{N \times (K_1+K_2+1)} & \hat{D}_2 \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} D_1 & 0_{N \times (K_1+K_3+1)} \\ 0_{N \times (K_1+K_2+1)} & D_2 \end{bmatrix} \equiv G. \quad (\text{A52})$$

Under joint stationarity and ergodicity assumptions on factors and returns and assuming that their fourth moments exist, the asymptotic distribution of $\bar{g}_T(\hat{\theta})$ is given by

$$\sqrt{T} \bar{g}_T(\hat{\theta}) \stackrel{A}{\sim} N(0_{2N}, [I_{2N} - G(AG)^{-1}A]S[I_{2N} - G(AG)^{-1}A]'). \quad (\text{A53})$$

After simplification, we can write

$$\sqrt{T} \begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \end{bmatrix} = \sqrt{T} \begin{bmatrix} \bar{g}_{1T}(\hat{\eta}) \\ \bar{g}_{2T}(\hat{\lambda}) \end{bmatrix} \stackrel{A}{\sim} N \left(0_{2N}, \begin{bmatrix} Q_1 S_{11} Q_1' & Q_1 S_{12} Q_2' \\ Q_2 S_{21} Q_1' & Q_2 S_{22} Q_2' \end{bmatrix} \right), \quad (\text{A54})$$

where

$$\begin{aligned} Q_1 &= I_N - D_1 (D_1' V_{22}^{-1} D_1)^{-1} D_1' V_{22}^{-1} \\ &= V_{22}^{\frac{1}{2}} [I_N - V_{22}^{-\frac{1}{2}} D_1 (D_1' V_{22}^{-1} D_1)^{-1} D_1' V_{22}^{-\frac{1}{2}}] V_{22}^{-\frac{1}{2}} \\ &= V_{22}^{\frac{1}{2}} P_1 P_1' V_{22}^{-\frac{1}{2}}, \end{aligned} \quad (\text{A55})$$

$$Q_2 = V_{22}^{\frac{1}{2}} P_2 P_2' V_{22}^{-\frac{1}{2}}. \quad (\text{A56})$$

Let

$$z = \sqrt{T} \begin{bmatrix} \hat{P}_1' \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_1 \\ \hat{P}_2' \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_2 \end{bmatrix} \stackrel{A}{\sim} N(0_{n_1+n_2}, V_z), \quad (\text{A57})$$

where

$$V_z = \begin{bmatrix} P_1' V_{22}^{-\frac{1}{2}} S_{11} V_{22}^{-\frac{1}{2}} P_1 & P_1' V_{22}^{-\frac{1}{2}} S_{12} V_{22}^{-\frac{1}{2}} P_2 \\ P_2' V_{22}^{-\frac{1}{2}} S_{21} V_{22}^{-\frac{1}{2}} P_1 & P_2' V_{22}^{-\frac{1}{2}} S_{22} V_{22}^{-\frac{1}{2}} P_2 \end{bmatrix}, \quad (\text{A58})$$

we can then write the test statistic in Equation (48) as $z' \hat{V}_z^{-1} z$. Since $\hat{V}_z \xrightarrow{\text{a.s.}} V_z$, we have $z' \hat{V}_z^{-1} z \stackrel{A}{\sim} \chi_{n_1+n_2}^2$.

Using the fact that $\hat{D}_1' \hat{V}_{22}^{-1} \hat{e}_1 = 0_{K_1+K_2+1}$, we can write

$$\begin{aligned} T \hat{\delta}_1^2 &= T \hat{e}_1' \hat{V}_{22}^{-\frac{1}{2}} [P_1 \hat{P}_1' + \hat{V}_{22}^{-\frac{1}{2}} \hat{D}_1 (\hat{D}_1' \hat{V}_{22}^{-1} \hat{D}_1)^{-1} \hat{D}_1' \hat{V}_{22}^{-\frac{1}{2}}] \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_1 \\ &= T \hat{e}_1' \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_1 \hat{P}_1' \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_1 = z_1' z_1, \end{aligned} \quad (\text{A59})$$

where z_1 is the first n_1 elements of z . Similarly, $T \hat{\delta}_2^2 = z_2' z_2$, where z_2 is the last n_2 elements of z .

Using these expressions and letting $Q \Xi Q'$ be the eigenvalue decomposition of

$$V_z^{\frac{1}{2}} \begin{bmatrix} I_{n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & -I_{n_2} \end{bmatrix} V_z^{\frac{1}{2}}, \quad (\text{A60})$$

where $\Xi = \text{Diag}(\xi_1, \dots, \xi_{n_1+n_2})$ is a diagonal matrix of the eigenvalues of Equation (A60), or equivalently the eigenvalues of Equation (50), and Q is a matrix of the corresponding eigenvectors. Writing $\tilde{z} = Q' V_z^{-\frac{1}{2}} z \stackrel{A}{\sim} N(0_{n_1+n_2}, I_{n_1+n_2})$, we have

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) = z' \begin{bmatrix} I_{n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & -I_{n_2} \end{bmatrix} z = z' V_z^{-\frac{1}{2}} Q \Xi Q' V_z^{-\frac{1}{2}} z = \sum_{i=1}^{n_1+n_2} \xi_i \tilde{z}_i^2, \quad (\text{A61})$$

where $\tilde{z}_i^2 \stackrel{A}{\sim} \chi_1^2$, $i = 1, \dots, n_1 + n_2$, and they are asymptotically independent of each other. This completes the proof.

Proof of Proposition 6: We first present the expression of $\partial\delta^2/\partial\phi$ for a general linear SDF model.

Claim: Let $\lambda = (D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}1_N$ and $e = D\lambda - 1_N$. We have

$$\frac{\partial\delta^2}{\partial\phi} = \begin{bmatrix} 2\mu_y \\ 2\lambda_1 \\ -V_{22}^{-1}e \end{bmatrix} \otimes \begin{bmatrix} 0_K \\ V_{22}^{-1}e \end{bmatrix}. \quad (\text{A62})$$

Proof: Note that $D'V_{22}^{-1}e = 0_{K+1}$ implies $\mu_2'V_{22}^{-1}e = 0$. Then, it is easy to show that

$$\frac{\partial\delta^2}{\partial\mu_1} = 2\lambda_1\mu_2'V_{22}^{-1}e = 0_K, \quad (\text{A63})$$

$$\frac{\partial\delta^2}{\partial\mu_2} = 2\mu_y V_{22}^{-1}e. \quad (\text{A64})$$

For the derivative of δ^2 with respect to $\text{vec}(V)$, we write $\delta^2 = e'V_{22}^{-1}e$ and use the product rule to obtain

$$\frac{\partial\delta^2}{\partial\text{vec}(V)'} = \frac{\partial e'V_{22}^{-1}e}{\partial\text{vec}(V)'} = 2e'V_{22}^{-1} \frac{\partial e}{\partial\text{vec}(V)'} + (e' \otimes e') \frac{\partial\text{vec}(V_{22}^{-1})}{\partial\text{vec}(V)'}. \quad (\text{A65})$$

For the first term, we use the product rule and the fact that $D'V_{22}^{-1}e = 0_{K+1}$ to obtain

$$\begin{aligned} 2e'V_{22}^{-1} \frac{\partial e}{\partial\text{vec}(V)'} &= 2e'V_{22}^{-1} \frac{\partial D\lambda}{\partial\text{vec}(V)'} \\ &= 2e'V_{22}^{-1} \left[(\lambda' \otimes I_N) \frac{\partial\text{vec}(D)}{\partial\text{vec}(V)'} + D \frac{\partial\lambda}{\partial\text{vec}(V)'} \right] \\ &= 2e'V_{22}^{-1} \left[(\lambda' \otimes I_N) \frac{\partial\text{vec}(D)}{\partial\text{vec}(V)'} \right]. \end{aligned} \quad (\text{A66})$$

Writing $D = [\mu_2, [0_{N \times K}, I_N]V[I_K, 0_{K \times N}]' + \mu_2\mu_1']$, we can simplify the first term to

$$\begin{aligned} 2e'V_{22}^{-1} \frac{\partial e}{\partial\text{vec}(V)'} &= 2e'V_{22}^{-1}(\lambda' \otimes I_N) \left(\begin{bmatrix} 0'_K & 0'_N \\ I_K & 0_{K \times N} \end{bmatrix} \otimes [0_{N \times K}, I_N] \right) \\ &= [2\lambda'_1, 0'_N] \otimes [0'_K, e'V_{22}^{-1}]. \end{aligned} \quad (\text{A67})$$

For the second term, we use the fact that for a nonsingular matrix A , we have $\partial\text{vec}(A^{-1})/\partial\text{vec}(A)' = -(A^{-1} \otimes A^{-1})'$. Using this identity and the chain rule, we have

$$\begin{aligned} (e' \otimes e') \frac{\partial\text{vec}(V_{22}^{-1})}{\partial\text{vec}(V)'} &= (e' \otimes e') \frac{\partial\text{vec}(V_{22}^{-1})}{\partial\text{vec}(V_{22})'} \frac{\partial\text{vec}(V_{22})}{\partial\text{vec}(V)'} \\ &= -(e' \otimes e')(V_{22}^{-1} \otimes V_{22}^{-1})([0_{N \times K}, I_N] \otimes [0_{N \times K}, I_N]) \\ &= -[0'_K, e'V_{22}^{-1}] \otimes [0'_K, e'V_{22}^{-1}]. \end{aligned} \quad (\text{A68})$$

Combining these two terms, we have

$$\frac{\partial \delta^2}{\partial \text{vec}(V)} = \begin{bmatrix} 2\lambda_1 \\ -V_{22}^{-1}e \end{bmatrix} \otimes \begin{bmatrix} 0_K \\ V_{22}^{-1}e \end{bmatrix}. \quad (\text{A69})$$

This completes the proof of the claim.

With the analytical expression of $\partial \delta^2 / \partial \phi$ available, we can show that

$$\begin{aligned} q_t(\phi) &= \left[\frac{\partial \delta^2}{\partial \phi} \right]' r_t(\phi) \\ &= 2\mu_y e' V_{22}^{-1} (R_t - \mu_2) + \\ &\quad \left(\begin{bmatrix} 2\lambda_1 \\ -V_{22}^{-1}e \end{bmatrix}' \otimes \begin{bmatrix} 0_K \\ V_{22}^{-1}e \end{bmatrix}' \right) \text{vec}((Y_t - \mu)(Y_t - \mu)' - V) \\ &= 2\mu_y e' V_{22}^{-1} (R_t - \mu_2) + \text{vec} \left(\begin{bmatrix} 0_K \\ V_{22}^{-1}e \end{bmatrix}' ((Y_t - \mu)(Y_t - \mu)' - V) \begin{bmatrix} 2\lambda_1 \\ -V_{22}^{-1}e \end{bmatrix} \right) \\ &= 2\mu_y e' V_{22}^{-1} (R_t - \mu_2) + \\ &\quad e' V_{22}^{-1} (R_t - \mu_2) [2\lambda_1' (f_t - \mu_1) - e' V_{22}^{-1} (R_t - \mu_2)] + e' V_{22}^{-1} e - 2e' V_{22}^{-1} V_{21} \lambda_1 \\ &= 2u_t y_t - u_t^2 + \delta^2 - 2e' V_{22}^{-1} V_{21} \lambda_1, \end{aligned} \quad (\text{A70})$$

by denoting $u_t = e' V_{22}^{-1} R_t$ and $y_t = \lambda_0 + \lambda_1' f_t$. Using the identity $e' V_{22}^{-1} D = 0'_{K+1}$, which implies that $e' V_{22}^{-1} V_{21} = -e' V_{22}^{-1} \mu_2 \mu_1' = 0'_K$, we can further simplify $q_t(\phi)$ to

$$q_t(\phi) = 2u_t y_t - u_t^2 + \delta^2. \quad (\text{A71})$$

Applying a similar derivation for models 1 and 2, we get

$$q_{1t}(\phi) = \left[\frac{\partial \delta_1^2}{\partial \phi} \right]' r_t(\phi) = 2u_{1t} y_{1t} - u_{1t}^2 + \delta_1^2, \quad (\text{A72})$$

$$q_{2t}(\phi) = \left[\frac{\partial \delta_2^2}{\partial \phi} \right]' r_t(\phi) = 2u_{2t} y_{2t} - u_{2t}^2 + \delta_2^2. \quad (\text{A73})$$

Now, using the delta method and Equations (A1)–(A4), the asymptotic distribution of $\hat{\delta}_1^2 - \delta_1^2$ when both models are misspecified is given by

$$\sqrt{T}(\hat{\delta}_1^2 - \delta_1^2 - (\delta_1^2 - \delta_2^2)) \stackrel{A}{\sim} N \left(0, \left[\frac{\partial(\delta_1^2 - \delta_2^2)}{\partial \phi} \right]' S_0 \left[\frac{\partial(\delta_1^2 - \delta_2^2)}{\partial \phi} \right] \right). \quad (\text{A74})$$

With the analytical expressions of $q_{1t}(\phi)$ and $q_{2t}(\phi)$, the asymptotic variance of $\sqrt{T}(\hat{\delta}_1^2 - \delta_1^2)$ can be written as

$$v_d = \sum_{j=-\infty}^{\infty} E[d_t(\phi) d_{t+j}(\phi)], \quad (\text{A75})$$

where

$$d_t(\phi) = \left(\frac{\partial \delta_1^2}{\partial \phi} - \frac{\partial \delta_2^2}{\partial \phi} \right)' r_t(\phi) = q_{1t}(\phi) - q_{2t}(\phi). \quad (\text{A76})$$

This completes the proof.

References

- Andrews, D. W. K. (1991). “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation.” *Econometrica* 59:817–858.
- Bentler, P. M., and M. Berkane (1986). “Greatest Lower Bound to the Elliptical Theory Kurtosis Parameter.” *Biometrika* 73:240–241.
- Campbell, J. Y. (1996). “Understanding Risk and Return.” *Journal of Political Economy* 104:298–345.
- Campbell, J. Y., and J. H. Cochrane (2000). “Explaining the Poor Performance of Consumption-based Asset Pricing Models.” *Journal of Finance* 55:2863–2878.
- Casella, G., and R. L. Berger (1990). *Statistical Inference*. Belmont, CA: Duxbury Press.
- Chen, X., and S. C. Ludvigson (2004). “Land of Addicts? An Empirical Investigation of Habit-based Asset Pricing Models.” Working paper, New York University.
- Dittmar, R. F. (2002). “Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns.” *Journal of Finance* 57:369–403.
- Fama, E. F., and K. R. French (1993). “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics* 33:3–56.
- Farnsworth, H., W. E. Ferson, D. Jackson, and S. Todd (2002). “Performance Evaluation with Stochastic Discount Factors.” *Journal of Business* 75:473–503.
- Ferson, W. E., and C. R. Harvey (1991). “The Variation of Economic Risk Premiums.” *Journal of Political Economy* 99:385–415.
- Ferson, W. E., and C. R. Harvey (1999). “Conditioning Variables and the Cross Section of Stock Returns.” *Journal of Finance* 54:1325–1360.
- Gallant, A. R., and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford, UK: Basil Blackwell.

- Golden, R. M. (2003). “Discrepancy Risk Model Selection Test Theory for Comparing Possibly Misspecified or Nonnested Models.” *Psychometrika* 68:229–249.
- Hall, A. R., and A. Inoue (2003). “The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models.” *Journal of Econometrics* 114:361–394.
- Hansen, L. P. (1982). “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica* 50:1029–1054.
- Hansen, L. P., J. Heaton, and E. G. J. Luttmer (1995). “Econometric Evaluation of Asset Pricing Models.” *Review of Financial Studies* 8:237–274.
- Hansen, L. P., and R. Jagannathan (1997). “Assessing Specification Errors in Stochastic Discount Factor Models.” *Journal of Finance* 52:557–590.
- Hodrick, R. J., and E. Prescott (1997). “Postwar U.S. Business Cycles: An Empirical Investigation.” *Journal of Money, Credit and Banking* 29:1–16.
- Hodrick, R. J., and X. Zhang (2001). “Evaluating the Specification Errors of Asset Pricing Models.” *Journal of Financial Economics* 62:327–376.
- Hou, K., and R. Kimmel (2006). “On the Estimation of Risk Premia in Linear Factor Models.” Working paper, Ohio State University.
- Jagannathan, R., K. Kubota, and H. Takehara (1998). “Relationship between Labor-Income Risk and Average Return: Empirical Evidence from the Japanese Stock Market.” *Journal of Business* 71:319–348.
- Jagannathan, R., and Z. Wang (1996). “The Conditional CAPM and the Cross-Section of Expected Returns.” *Journal of Finance* 51:3–53.
- Kan, R., and C. Robotti (2007). “The Exact Distribution of the Hansen-Jagannathan Bound.” Working paper, University of Toronto and Federal Reserve Bank of Atlanta.
- Kan, R., and C. Robotti (2008). “Specification Tests of Asset Pricing Models Using Excess Returns.” *Journal of Empirical Finance* 15:816–838.

- Kan, R., and G. Zhou (2004). “Hansen-Jagannathan Distance: Geometry and Exact Distribution.” Working paper, University of Toronto and Washington University in St. Louis.
- Lettau, M., and S. Ludvigson (2001). “Consumption, Aggregate Wealth and Expected Stock Returns.” *Journal of Finance* 56:815–849.
- Lewellen, J. W., S. Nagel, and J. Shanken (2006). “A Skeptical Appraisal of Asset-Pricing Tests.” Working paper, Dartmouth College.
- Magnus, J., and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York, NY: Wiley.
- Maruyama, Y., and T. Seo (2003). “Estimation of Moment Parameter in Elliptical Distributions.” *Journal of the Japan Statistical Society* 33:215–229.
- Newey, W., and K. West (1987). “A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica* 55:703–708.
- Rivers, D., and Q. H. Vuong (2002). “Model Selection Tests for Nonlinear Dynamic Models.” *Econometrics Journal* 5:1–39.
- Vuong, Q. H. (1989). “Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses.” *Econometrica* 57:307–333.
- White, H. (1994). *Estimation, Inference and Specification Analysis*. New York, NY: Cambridge University Press.
- Wang, Z., and X. Zhang (2005). “Empirical Evaluation of Asset Pricing Models: Arbitrage and Pricing Errors over Contingent Claims.” Working paper, Federal Reserve Bank of New York and Cornell University.

Table 1
Summary of the Models

Panel A: Unscaled Factors

Model	CAPM	C-CAPM	JW	CAMP	FF3	FF5
$\hat{\delta}$	0.423	0.435	0.395	0.331	0.360	0.356
$p(\delta = 0)$	0.000	0.000	0.000	0.030	0.000	0.000
$se(\hat{\delta})$	0.039	0.050	0.049	0.055	0.041	0.041
2.5% CI(δ)	0.353	0.349	0.311	0.241	0.289	0.285
97.5% CI(δ)	0.509	0.548	0.505	0.459	0.452	0.448
<i>No. of par.</i>	2	2	4	6	4	6

Panel B: Factors Scaled by *Lag IP*

Model	CAPM	C-CAPM	JW	CAMP	FF3	FF5
$\hat{\delta}$	0.420	0.400	0.280	0.308	0.358	0.277
$p(\delta = 0)$	0.000	0.009	0.567	0.075	0.000	0.589
$se(\hat{\delta})$	0.041	0.056	0.066	0.068	0.042	0.080
2.5% CI(δ)	0.348	0.306	0.180	0.204	0.285	0.161
97.5% CI(δ)	0.509	0.528	0.441	0.472	0.453	0.477
<i>No. of par.</i>	4	4	8	12	8	12

Panel C: Factors Scaled by *JAN*

Model	CAPM	C-CAPM	JW	CAMP	FF3	FF5
$\hat{\delta}$	0.397	0.387	0.305	0.234	0.346	0.337
$p(\delta = 0)$	0.000	0.012	0.241	0.487	0.000	0.000
$se(\hat{\delta})$	0.047	0.059	0.074	0.063	0.046	0.044
2.5% CI(δ)	0.316	0.290	0.194	0.141	0.268	0.262
97.5% CI(δ)	0.504	0.522	0.486	0.393	0.448	0.437
<i>No. of par.</i>	4	4	8	12	8	12

The table presents a summary of six asset pricing models. The models include the market CAPM (CAPM), the consumption CAPM (C-CAPM), the conditional CAPM of Jagannathan and Wang (1996, JW), the Campbell (1996) five-factor model (CAMP), the Fama-French (1993) three-factor model (FF3) and the Fama-French (1993) five-factor model (FF5). The models are estimated using the monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate and the gross one-month T-bill return. Most of the data are from January 1952 to December 2006, but the data for the C-CAPM model start in February 1959, and the data for the CAMP model cover only the period February 1952 to December 1990. The scaling variables are *Lag IP* and *JAN*, where *IP* is the cyclical element in the industrial production index and *JAN* is a dummy variable with a value of one for January and zero otherwise. $\hat{\delta}$ is the sample HJ-distance. $p(\delta = 0)$ is the p -value for the test of $H_0 : \delta = 0$. $se(\hat{\delta})$ is the standard error of the sample HJ-distance under the alternative. CI(δ) is the 95% confidence interval for δ based on the statistical method. *No. of par.* is the number of parameters.

Table 2

Estimates and t -ratios of Parameters in Various SDF Models under Correctly Specified and Misspecified Models: Unscaled Factors

	CAPM		C-CAPM	
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_0$	$\hat{\lambda}_{cg}$
Estimate	1.02	-3.38	1.16	-72.90
t -ratio _{cs}	89.68	-3.32	17.90	-2.97
t -ratio _m	89.67	-3.32	12.65	-1.91

	JW				CAMP					
	$\hat{\lambda}_0$	$\hat{\lambda}_{jvw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\lambda}_0$	$\hat{\lambda}_{rvw}$	$\hat{\lambda}_{clab}$	$\hat{\lambda}_{div}$	$\hat{\lambda}_{rtb}$	$\hat{\lambda}_{trm}$
Estimate	0.92	-1.79	-71.44	162.63	1.01	0.15	-1.04	42.97	23.17	19.60
t -ratio _{cs}	1.51	-1.46	-1.33	2.90	27.74	0.89	-2.28	0.86	2.40	1.70
t -ratio _m	0.70	-1.27	-0.65	1.40	26.57	0.55	-1.44	0.54	1.79	1.20

	FF3				FF5					
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\lambda}_{term}$	$\hat{\lambda}_{def}$
Estimate	1.07	-4.97	-2.46	-8.88	1.07	-2.58	-2.74	-7.36	-11.19	-20.92
t -ratio _{cs}	52.11	-4.37	-1.68	-5.60	46.91	-1.24	-1.41	-3.61	-1.43	-0.85
t -ratio _m	52.20	-4.37	-1.68	-5.61	46.25	-0.64	-0.91	-2.06	-0.73	-0.35

The table presents the estimation results of six asset pricing models with unscaled factors. The models are estimated using the monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate and the gross one-month T-bill return. Most of the data are from January 1952 to December 2006, but the data for the C-CAPM model start in February 1959, and the data for the CAMP model cover only the period February 1952 to December 1990. We report parameter estimates $\hat{\lambda}$, t -ratios under correctly specified models (t -ratio_{cs}), and model misspecification robust t -ratios (t -ratio_m).

Table 3
Wald Tests of SDF Parameters of Conditional Models under Correct Specification and Potential Misspecification

Panel A: Factors Scaled by *Lag IP*

Model	CAPM	C-CAPM	JW	CAMP	FF3	FF5
Wald(<i>cs</i>)	1.33	7.47	14.96	3.69	1.04	7.65
<i>p</i> -value	(0.513)	(0.024)	(0.005)	(0.718)	(0.904)	(0.265)
Wald(<i>m</i>)	0.27	3.88	10.72	1.69	0.21	4.24
<i>p</i> -value	(0.874)	(0.144)	(0.030)	(0.946)	(0.995)	(0.644)

Panel B: Factors Scaled by *JAN*

Model	CAPM	C-CAPM	JW	CAMP	FF3	FF5
Wald(<i>cs</i>)	7.64	6.51	14.22	9.43	3.47	4.13
<i>p</i> -value	(0.022)	(0.039)	(0.007)	(0.151)	(0.483)	(0.659)
Wald(<i>m</i>)	7.04	4.62	12.91	6.50	1.64	1.75
<i>p</i> -value	(0.030)	(0.099)	(0.012)	(0.370)	(0.802)	(0.941)

The table presents Wald tests that the SDF parameters of the scaled factors are jointly equal to zero. The models are estimated using the monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate and the gross one-month T-bill return. Most of the data are from January 1952 to December 2006, but the data for the C-CAPM model start in February 1959, and the data for the CAMP model cover only the period February 1952 to December 1990. The scaling variables are *Lag IP* and *JAN*, where *IP* is the cyclical element in the industrial production index and *JAN* is a dummy variable with a value of one for January and zero otherwise. We report the Wald-test statistic under correctly specified (*cs*) and potentially misspecified (*m*) models. The *p*-values of the Wald tests are shown in parentheses.

Table 4
Tests of Equality of Squared HJ-Distances

Panel A: Unscaled Factors

Unscaled	Unscaled				
	C-CAPM	JW	CAMP	FF3	FF5
CAPM	0.008 (0.700)	0.023 (0.349)	0.039 (0.124)	0.049 (0.000)	0.052 (0.047)
C-CAPM		0.012 (0.710)	0.028 (0.449)	0.044 (0.087)	0.046 (0.087)
JW			0.015 (0.632)	0.026 (0.393)	0.029 (0.313)
CAMP				-0.001 (0.970)	0.009 (0.717)
FF3					0.003 (0.797)

Panel B: Factors Scaled by *Lag IP*

<i>Lag IP</i>	<i>Lag IP</i>				
	C-CAPM	JW	CAMP	FF3	FF5
CAPM	0.032 (0.270)	0.098 (0.013)	0.044 (0.207)	0.048 (0.072)	0.100 (0.419)
C-CAPM		0.041 (0.378)	0.024 (0.603)	0.028 (0.419)	0.073 (0.100)
JW			-0.017 (0.690)	-0.050 (0.190)	0.002 (0.969)
CAMP				-0.013 (0.678)	0.025 (0.618)
FF3					0.051 (0.503)

Table 4 (Continued)
Tests of Equality of Squared HJ-Distances

Panel C: Factors Scaled by *JAN*

<i>JAN</i>	<i>JAN</i>				
	C-CAPM	JW	CAMP	FF3	FF5
CAPM	0.018 (0.601)	0.065 (0.095)	0.069 (0.105)	0.038 (0.177)	0.044 (0.680)
C-CAPM		0.030 (0.598)	0.029 (0.460)	0.023 (0.446)	0.033 (0.373)
JW			0.025 (0.571)	-0.027 (0.535)	-0.020 (0.625)
CAMP				-0.027 (0.402)	-0.021 (0.515)
FF3					0.006 (0.961)

The table presents pairwise tests of equality of the squared HJ-distances of six different asset pricing models with unscaled and scaled factors. The models are estimated using the monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate and the gross one-month T-bill return. Most of the data are from January 1952 to December 2006, but the data for the C-CAPM model start in February 1959, and the data for the CAMP model cover only the period February 1952 to December 1990. The scaling variables are *Lag IP* and *JAN*, where *IP* is the cyclical element in the industrial production index and *JAN* is a dummy variable with a value of one for January and zero otherwise. We report the difference between the sample squared HJ-distances of the models in row *i* and column *j*, $\hat{\delta}_i^2 - \hat{\delta}_j^2$, and the associated *p*-value (in parentheses) for the test of $H_0 : \delta_i^2 = \delta_j^2$. The *p*-values are computed under the assumption that the models are potentially misspecified.

Table 5
Tests of Equality of Squared HJ-Distances: Unconditional vs. Conditional Models

Panel A: Unscaled Factors vs. Factors Scaled by *Lag IP*

Unscaled	<i>Lag IP</i>					
	CAPM	C-CAPM	JW	CAMP	FF3	FF5
CAPM	0.002 (0.906)	0.038 (0.246)	0.100 (0.010)	0.053 (0.136)	0.050 (0.197)	0.102 (0.535)
C-CAPM	-0.002 (0.930)	0.029 (0.234)	0.070 (0.099)	0.056 (0.201)	0.058 (0.076)	0.102 (0.014)
JW	-0.020 (0.461)	0.018 (0.676)	0.077 (0.090)	0.030 (0.500)	0.028 (0.384)	0.079 (0.110)
CAMP	-0.029 (0.310)	0.005 (0.913)	0.031 (0.384)	0.014 (0.966)	0.001 (0.954)	0.039 (0.417)
FF3	-0.046 (0.015)	-0.015 (0.675)	0.051 (0.176)	0.015 (0.644)	0.002 (0.993)	0.053 (0.814)
FF5	-0.050 (0.017)	-0.017 (0.645)	0.048 (0.201)	0.005 (0.873)	-0.002 (0.901)	0.050 (0.707)

Panel B: Unscaled Factors vs. Factors Scaled by *JAN*

Unscaled	<i>JAN</i>					
	CAPM	C-CAPM	JW	CAMP	FF3	FF5
CAPM	0.021 (0.235)	0.048 (0.172)	0.086 (0.033)	0.094 (0.006)	0.059 (0.157)	0.065 (0.567)
C-CAPM	0.022 (0.480)	0.040 (0.176)	0.070 (0.174)	0.086 (0.042)	0.063 (0.035)	0.073 (0.038)
JW	-0.002 (0.953)	0.028 (0.523)	0.063 (0.339)	0.070 (0.083)	0.036 (0.281)	0.042 (0.205)
CAMP	-0.015 (0.707)	0.029 (0.515)	0.029 (0.494)	0.054 (0.432)	0.028 (0.363)	0.033 (0.293)
FF3	-0.028 (0.222)	-0.004 (0.892)	0.037 (0.351)	0.055 (0.085)	0.010 (0.843)	0.016 (0.981)
FF5	-0.031 (0.201)	-0.006 (0.851)	0.034 (0.382)	0.046 (0.157)	0.007 (0.713)	0.013 (0.961)

Table 5 (Continued)
Tests of Equality of Squared HJ-Distances: Unconditional vs. Conditional Models

Panel C: Factors Scaled by *Lag IP* vs. Factors Scaled by *JAN*

<i>Lag IP</i>	<i>JAN</i>					
	CAPM	C-CAPM	JW	CAMP	FF3	FF5
CAPM	0.018 (0.442)	0.042 (0.262)	0.083 (0.043)	0.084 (0.033)	0.057 (0.018)	0.063 (0.011)
C-CAPM	-0.007 (0.861)	0.010 (0.792)	0.041 (0.466)	0.054 (0.231)	0.034 (0.366)	0.043 (0.300)
JW	-0.079 (0.069)	-0.030 (0.521)	-0.014 (0.777)	0.024 (0.531)	-0.041 (0.274)	-0.035 (0.357)
CAMP	-0.029 (0.509)	0.001 (0.990)	0.015 (0.765)	0.040 (0.326)	0.013 (0.719)	0.019 (0.624)
FF3	-0.030 (0.236)	-0.018 (0.648)	0.035 (0.378)	0.053 (0.101)	0.008 (0.658)	0.015 (0.446)
FF5	-0.081 (0.089)	-0.063 (0.158)	-0.016 (0.768)	0.015 (0.752)	-0.043 (0.313)	-0.037 (0.399)

The table compares the performance of six asset pricing models with unscaled factors with the performance of the corresponding models with scaled factors. The models are estimated using the monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate and the gross one-month T-bill return. Most of the data are from January 1952 to December 2006, but the data for the C-CAPM model start in February 1959, and the data for the CAMP model cover only the period February 1952 to December 1990. The scaling variables are *Lag IP* and *JAN*, where *IP* is the cyclical element in the industrial production index and *JAN* is a dummy variable with a value of one for January and zero otherwise. We report the difference between the sample squared HJ-distances of the models in row i and column j , $\hat{\delta}_i^2 - \hat{\delta}_j^2$, and the associated p -value (in parentheses) for the test of $H_0 : \delta_i^2 = \delta_j^2$. The p -values are computed under the assumption that the models are potentially misspecified.