FEDERAL RESERVE BANK *of* ATLANTA

Model Comparison Using the Hansen-Jagannathan Distance

Raymond Kan and Cesare Robotti

Working Paper 2007-4
February 2007

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA | WORKING PAPER SERIES

Model Comparison Using the Hansen-Jagannathan Distance

Raymond Kan and Cesare Robotti

Working Paper 2007-4
February 2007

Abstract: Although it is of interest to empirical researchers to test whether or not a particular asset-pricing model is true, a more useful task 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 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 are nested or nonnested. 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 the same data as in Hodrick and Zhang (2001), we show that the commonly used returns and factors are, for the most part, too noisy to conclude that one model is superior to the other models in terms of Hansen-Jagannathan distance. 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.

JEL classification: G12

Key words: Hansen-Jagannathan distance, asset-pricing models, model misspecification, risk premia

---

The authors thank Yaxuan Qi, Jay Shanken, Guofu Zhou, seminar participants at the National University of Ireland, and participants at the 2006 All Georgia Finance Conference for helpful discussions and comments. Kan 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. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, 416-978-4291, 416-978-5433 (fax), kan@chass.utoronto.ca, or Cesare Robotti, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA 30305, 404-498-8543, 404-498-8810 (fax), cesare.robotti@atl.frb.org.

Model Comparison Using the Hansen-Jagannathan Distance

Asset pricing models are, at best, an approximation 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. The existing tests proposed by Hansen, Heaton, and Luttmer (1995), Jagannathan and Wang (1996), and Hansen and Jagannathan (1997) only allow us to test whether a given model has a particular HJ-distance value, but do not allow us to test whether or not two competing models have the same HJ-distance.[1] Because the p-values from this kind of tests 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 difficult to justify because the difference between the sample HJ-distances of competing models might not be statistically significant. The first methodological contribution of this paper consists in the proposal of a new methodology to formally test whether or not two competing linear asset pricing models have the same HJ-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. We provide the asymptotic distribution of our test statistic under general distributional assumptions as well as for the special case in which returns

[1] The 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.

and factors are multivariate elliptically distributed. The results for the multivariate elliptical case enable us to gain further intuition on the important determinants of the asymptotic distribution of the test statistic.

In addition to model comparisons, researchers are also interested in whether or not a particular 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 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 SDF parameters that are applicable to both correctly specified and misspecified models. When factors and returns are multivariate elliptically distributed, we are able to show analytically that the standard errors under potentially misspecified models are always bigger than the standard errors that assume the model is correctly specified. We call the difference between the asymptotic variances of the SDF parameter estimates under correctly specified and misspecified models the misspecification adjustment term and show that the magnitude of this term depends on, among other things, the correlations between the factors and the 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 factors are macroeconomic variables.

After describing the econometric methodology, we provide an in-depth empirical analysis to demonstrate the relevance of our new test. We focus on the empirical performance of several unconditional and conditional asset pricing models using the same dataset as in Hodrick and Zhang (2001). 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 correct), 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 no evidence that conditional and intertemporal CAPM-type specifications such as the Campbell (1996), Cochrane (1996), and Jagannathan and Wang (1996) models outperform the Fama-French three and five factor models in terms of HJ-distance.

The rest of the paper is organized as follows. Section I 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 II introduces tests of equalities of squared HJ-distances for two competing models and provides the asymptotic distributions of their sample counterparts for different scenarios. Section III presents the empirical analysis. The final section summarizes our findings and the Appendix contains proofs of all propositions.

I. Asymptotic Analysis Under Potentially Misspecified Models

A. Pricing Errors and HJ-distance

Let y be a proposed SDF with mean μ_y and let 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

$$e \equiv E[Ry] - 1_N = 0_N, \tag{1}$$

where 1_N is an N-vector of ones and 0_N is an N-vector of zeros.[2] 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

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

[2] We 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., $q = 0_N$).

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

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

where $U = E[RR']$ is the second moment matrix of R.

Hansen and Jagannathan (1997) provide two nice interpretations of the HJ-distance. The first interpretation is that the HJ-distance measures the minimum distance between the proposed SDF and the set of correct SDFs ($\mathcal{M}$),

$$\delta = \min_{m \in \mathcal{M}} \|m - y\|, \tag{4}$$

where $\|X\| = E[X^2]^{\frac{1}{2}}$ is the standard L^2 norm. The second interpretation is that it represents the maximum pricing error of a portfolio of R that has a unit second moment. Define ξ as the random payoff of a portfolio. Hansen and Jagannathan (1997) show that

$$\delta = \max_{\|\xi\|=1} |\pi(\xi) - \pi^y(\xi)|, \tag{5}$$

where $\pi(\xi)$ and $\pi^y(\xi)$ are the prices of ξ assigned by the true and the proposed SDFs, respectively.

In this paper, we focus on linear asset pricing models because they are the most popular models in the empirical asset pricing literature. However, with some additional efforts, our analysis could also be extended to the case of nonlinear models. A linear factor asset pricing model suggests that y is a linear function of K systematic factors f

$$y(\lambda_0, \lambda_1) = \lambda_0 + \lambda_1' f = \lambda' x, \tag{6}$$

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

$$\mu = E[Y] \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \tag{7}$$

$$V = \mathrm{Var}[Y] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{8}$$

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

$$e(\lambda) = E[Ry] - 1_N = E[Rx'\lambda] - 1_N = D\lambda - 1_N, \tag{9}$$

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

$$\delta^2 = \min_{\lambda}(D\lambda - 1_N)'V_{22}^{-1}(D\lambda - 1_N) = 1_N'V_{22}^{-1}1_N - 1_N'V_{22}^{-1}D(D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}1_N. \tag{10}$$

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 as

$$\lambda_{HJ} = (D'V_{22}^{-1}D)^{-1}(D'V_{22}^{-1}1_N). \tag{11}$$

In the subsequent analysis, we drop the subscript from λ_{HJ} for brevity reasons. 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.

B. 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 as

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \tag{12}$$

$$\hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \tag{13}$$

The sample squared HJ-distance and the SDF parameter estimates are simply the sample counterparts of (10) and (11)

$$\hat{\delta}^2 = 1_N'\hat{V}_{22}^{-1}1_N - 1_N'\hat{V}_{22}^{-1}\hat{D}(\hat{D}'\hat{V}_{22}^{-1}\hat{D})^{-1}\hat{D}'\hat{V}_{22}^{-1}1_N, \tag{14}$$

$$\hat{\lambda} = (\hat{D}'\hat{V}_{22}^{-1}\hat{D})^{-1}(\hat{D}'\hat{V}_{22}^{-1}1_N), \tag{15}$$

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 $\delta = 0$

$$T\hat{\delta}^2 \overset{A}{\sim} \sum_{i=1}^{N-K-1} \xi_i x_i, \tag{16}$$

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

$$S^{\frac{1}{2}} V_{22}^{-1} S^{\frac{1}{2}} - S^{\frac{1}{2}} V_{22}^{-1} D (D' V_{22}^{-1} D)^{-1} D' V_{22}^{-1} S^{\frac{1}{2}}, \tag{17}$$

where S is the asymptotic covariance matrix of

$$\bar{g}_T(\lambda) = \frac{1}{T} \sum_{t=1}^{T} (R_t x_t' \lambda - 1_N). \tag{18}$$

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 $\delta \neq 0$

$$\sqrt{T}(\hat{\delta}^2 - \delta^2) \overset{A}{\sim} N(0, v), \tag{19}$$

$$\sqrt{T}(\hat{\delta} - \delta) \overset{A}{\sim} N\left(0, \frac{v}{4\delta^2}\right), \tag{20}$$

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

$$q_t = y_t^2 - (y_t - \eta' R_t)^2 - 2\eta' 1_N - \delta^2 = 2\eta' R_t y_t - (\eta' R_t)^2 - 2\eta' 1_N - \delta^2, \tag{21}$$

with $\eta = U^{-1} e$. Under the linear factor model, 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

$$q_t = 2u_t y_t - u_t^2 + \delta^2, \tag{22}$$

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

In conducting statistical tests, we need a consistent estimate of $\text{Avar}[\hat{\delta}^2]$. This can be accomplished by replacing q_t with

$$\hat{q}_t = 2\hat{u}_t \hat{y}_t - \hat{u}_t^2 + \hat{\delta}^2, \tag{23}$$

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 $\hat{e} = \hat{D}\hat{\lambda} - 1_N$. For example, if q_t is uncorrelated over time, then we have $\text{Avar}[\hat{\delta}^2] = E[q_t^2]$, and its consistent estimator is given by

$$\widehat{\text{Avar}}[\hat{\delta}^2] = \frac{1}{T} \sum_{t=1}^{T} \hat{q}_t^2, \tag{24}$$

which is convenient to compute. When q_t is autocorrelated, one can use Newey and West's (1987) method to obtain a consistent estimator of Avar$[\hat{\delta}^2]$. For example, if q_t has an MA(m) structure, then a consistent estimator of Avar$[\hat{\delta}^2]$ is given by

$$\widehat{\mathrm{Avar}}[\hat{\delta}^2] = \frac{1}{T}\sum_{t=1}^{T}\hat{q}_t^2 + \frac{2}{T}\sum_{k=1}^{m}\left(1-\frac{k}{m+1}\right)\sum_{t=1}^{T-k}\hat{q}_t\hat{q}_{t+k}. \tag{25}$$

With additional assumptions, we can further simplify the asymptotic distribution of $\hat{\delta}^2$. Lemma 1 presents the asymptotic distribution of $\hat{\delta}^2$ under the correctly specified and misspecified models when Y_t is i.i.d. multivariate elliptically distributed. The expression for the correctly specified model is available in Kan and Zhou (2004) but the expression for the misspecified model is new.

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.*[3] *Let $\mu_y = E[y_t] = \lambda_0 + \lambda_1'\mu_1$ and $\sigma_y^2 =$ Var$[y_t] = \lambda_1'V_{11}\lambda_1$. When $\delta = 0$*

$$T\hat{\delta}^2 \overset{A}{\sim} [\mu_y^2 + (1+\kappa)\sigma_y^2]\chi^2_{N-K-1}. \tag{27}$$

When $\delta \neq 0$

$$\sqrt{T}(\hat{\delta}^2 - \delta^2) \overset{A}{\sim} N(0, 4[\mu_y^2 + (1+\kappa)\sigma_y^2]\delta^2 + (2+3\kappa)\delta^4). \tag{28}$$

The results in Lemma 1 show that when $\delta \neq 0$, the asymptotic variance of $\hat{\delta}^2$ increases with δ^2 and with μ_y^2 and σ_y^2. Therefore, it is not entirely clear that a specification test of $H_0 : \delta = 0$ has more power to reject a model with large HJ-distance than to reject a model with small HJ-distance. In addition, Lemma 1 shows that the asymptotic variance of the sample HJ-distance increases with the kurtosis parameter κ. This is hardly surprising since the fatter the tails of the returns, the more likely it is that there will be outliers in the sample covariance matrix which, in turn, make the sample HJ-distance more volatile.

[3] The multivariate kurtosis parameter of Y_t is defined as

$$\kappa = \frac{E[((Y_t-\mu)'V^{-1}(Y_t-\mu))^2]}{(N+K)(N+K+2)} - 1, \tag{26}$$

which is the same as the univariate kurtosis parameter for the case of multivariate elliptical distribution.

C. Asymptotic Distribution of the SDF Parameter Estimates Under Potentially Misspecified Models

In many empirical studies, interest lies in the point estimates of the SDF parameter λ. A statistically significant $\hat{\lambda}$ associated with a given factor is often interpreted as evidence that the factor is priced. However, in 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 present an analysis of the asymptotic distribution of $\hat{\lambda}$ under potentially misspecified models. Our analysis closely follows those of Hall and Inoue (2003) and Kan and Robotti (2006).[4]

Proposition 1: *Under a potentially misspecified model*

$$\sqrt{T}(\hat{\lambda} - \lambda) \stackrel{A}{\sim} N(0_{K+1}, V(\hat{\lambda})), \tag{29}$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{30}$$

with

$$h_t = HD'V_{22}^{-1}R_t y_t - H[D'V_{22}^{-1}(R_t - \mu_2) - x_t]u_t - \lambda, \tag{31}$$

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*

$$h_t = HD'V_{22}^{-1}R_t y_t - \lambda. \tag{32}$$

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 (31) instead of the sample counterpart of (32). This is because the latter is only valid when the model is correctly specified whereas the former is valid for both correctly specified and misspecified models.

With additional assumptions, we can further simplify the expression of $V(\hat{\lambda})$. In Lemma 2, we present the asymptotic variances of $\hat{\lambda}$ when Y_t is i.i.d. multivariate elliptically distributed.

[4] It should be noted that Hansen, Heaton, and Luttmer (1995, Appendix C) also presents 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.

Lemma 2 *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 $\hat{\lambda}$ is given by*

$$V(\hat{\lambda}) = [\mu_y^2 + (1+\kappa)\sigma_y^2]H + \begin{bmatrix} \sigma_y^2 - \mu_y^2 + \lambda_0^2 + 2\kappa(\mu_1'\lambda_1) & (\lambda_0 - 2\kappa\mu_1'\lambda_1)\lambda_1' \\ (\lambda_0 - 2\kappa\mu_1'\lambda_1)\lambda_1 & (1+2\kappa)\lambda_1\lambda_1' \end{bmatrix}$$
$$+ \delta^2 H \left([1 + (1+\kappa)\mu_2' V_{22}^{-1}\mu_2] \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix} \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix}' + \begin{bmatrix} 0 & 0_K' \\ 0_K & (1+\kappa)(V_{11} - V_{12}V_{22}^{-1}V_{21}) \end{bmatrix} \right) H, \tag{33}$$

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

When the model is correctly specified, $\delta = 0$ and

$$V(\hat{\lambda}) = [\mu_y^2 + (1+\kappa)\sigma_y^2]H + \begin{bmatrix} \sigma_y^2 - \mu_y^2 + \lambda_0^2 + 2\kappa(\mu_1'\lambda_1) & (\lambda_0 - 2\kappa\mu_1'\lambda_1)\lambda_1' \\ (\lambda_0 - 2\kappa\mu_1'\lambda_1)\lambda_1 & (1+2\kappa)\lambda_1\lambda_1' \end{bmatrix}. \tag{34}$$

Let P be an $N \times (N-1)$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}}\mu_2$. Applying the partitioned matrix inverse formula to H, one can verify that its lower right $K \times K$ submatrix is $(B'B)^{-1}$, where $B = P'V_{22}^{-\frac{1}{2}}V_{21}$. With this, it is straightforward to show that the asymptotic variance of $\hat{\lambda}_1$ is given by

$$\begin{aligned} V(\hat{\lambda}_1) &= [\mu_y^2 + (1+\kappa)\sigma_y^2](B'B)^{-1} + (1+2\kappa)\lambda_1\lambda_1' \\ &\quad + \delta^2(B'B)^{-1}[(1+\kappa)(V_{11} - B'B) + \nu\nu'](B'B)^{-1}, \end{aligned} \tag{35}$$

where $\nu = V_{12}V_{22}^{-1}\mu_2/(\mu'V_{22}^{-1}\mu_2)$.

Note that the last term in $V(\hat{\lambda}_1)$ only exists when the model is misspecified. Then, it is natural to define the matrix

$$\delta^2(B'B)^{-1}[(1+\kappa)(V_{11} - B'B) + \nu\nu'](B'B)^{-1} \tag{36}$$

as the misspecification adjustment term. The adjustment term is positive definite because

$$V_{11} - B'B = V_{11} - V_{12}V_{22}^{-\frac{1}{2}}PP'V_{22}^{-\frac{1}{2}}V_{21} = V_{11} - V_{12}V_{22}^{-1}V_{21} + \frac{V_{12}V_{22}^{-1}\mu_2\mu_2'V_{22}^{-1}V_{21}}{\mu_2'V_{22}^{-1}\mu_2} \tag{37}$$

is a positive definite matrix. As expected, the adjustment is positively related to the squared HJ-distance δ^2, suggesting that the degree of model misspecification plays an important role in determining the magnitude of the adjustment. The adjustment is also positively related to κ which suggests that the fatter the tails of the returns, the larger the adjustment. The final determinants of the adjustment are related to $(B'B)^{-1}$ and $V_{11} - B'B$. To understand what these two matrices represent, consider the normalized returns $R_t^* = P'V_{22}^{-\frac{1}{2}}R_t$. These normalized returns have the properties of $E[R_t^*] = 0_{N-1}$ and and $\mathrm{Var}[R_t^*] = I_{N-1}$. Projecting the factors on a constant term and R_t^* in the following multivariate regression

$$f_t = c_0 + c_1 R_t^* + \varepsilon_t, \tag{38}$$

it is easy to verify that $c_1 = V_{12}V_{22}^{-\frac{1}{2}}P$. Letting $f_t^* = c_1 R_t^* = V_{12}V_{22}^{-\frac{1}{2}}PP'V_{22}^{-\frac{1}{2}}R_t$ be the mimicking portfolios of f_t, we can verify that $E[f_t^*] = 0_K$ and $\mathrm{Var}[f_t^*] = B'B$. With this, it follows that $(B'B)^{-1} = \mathrm{Var}[f_t^*]^{-1}$ and $V_{11} - B'B = \mathrm{Var}[f_t] - \mathrm{Var}[f_t^*]$.

The magnitudes of these terms heavily depend on how well the factors can be explained by the returns. When the factors are portfolio returns, we expect $\mathrm{Var}[f_t^*]^{-1}$ and $\mathrm{Var}[f_t] - \mathrm{Var}[f_t^*]$ to be small. However, when the factors are macroeconomic factors, they may have very low correlations with the returns and $\mathrm{Var}[f_t^*]$ can be very small. In those cases, the magnitude of the misspecification adjustment term can be very large and model misspecification can have a serious impact on the standard errors of $\hat{\lambda}_1$. 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.

II. Tests of Equality of the Squared HJ-Distances of Two Models

When testing the equality of the squared HJ-distances of two competing linear SDFs, we need to consider two separate cases: nested models and non-nested models. In addition, the two models can either be both correctly specified or both misspecified.[5]

A. Nested Models

Let $f = [f_1', \ f_2']'$, where f_1 is $K_1 \times 1$ and f_2 is $K_2 \times 1$, and $K = K_1 + K_2$. For the SDF of model 1, we assume that it is linear in f_1

$$y_1 = \eta_0 + \eta_1' f_1 = \eta' A_1 x, \tag{39}$$

[5] Under the null hypothesis that the HJ-distances of the two models are equal, we do not need to consider the case that one model is correctly specified while the other model is misspecified. Our analysis is similar in spirit to the Vuong's (1989) and Rivers and Vuong's (2002) model selection methodology using the likelihood ratio test statistic.

where $x = [1,\ f_1']'$, $A_1 = [I_{K_1+1},\ \mathrm{O}_{(K_1+1)\times K_2}]$ and $\eta = [\eta_0,\ \eta_1']'$ is the (K_1+1)-vector of the SDF parameters of model 1. For the SDF of model 2, we assume that it is linear in f

$$y_2 = \lambda_0 + \lambda_1' f = \lambda' x, \tag{40}$$

where λ is the $(K+1)$-vector of the SDF parameters of model 2. Therefore, model 2 nests model 1 as a special case. The squared HJ-distances for the two models are

$$\delta_1^2 = 1_N' V_{22}^{-1} 1_N - 1_N' V_{22}^{-1} D A_1' (A_1 D' V_{22}^{-1} D A_1')^{-1} A_1 D' V_{22}^{-1} 1_N, \tag{41}$$

$$\delta_2^2 = 1_N' V_{22}^{-1} 1_N - 1_N' V_{22}^{-1} D (D' V_{22}^{-1} D)^{-1} D' V_{22}^{-1} 1_N. \tag{42}$$

As model 2 nests model 1 as a special case, we must have $\delta_1^2 \geq \delta_2^2$. Lemma 3 shows that, when the two models have the same HJ-distance, there are some restrictions on the SDF parameters of model 2.

Lemma 3 *Partition $\lambda = (D'V_{22}^{-1}D)^{-1}(D'V_{22}^{-1}1_N)$ as $[\lambda^{(1)\prime},\ \lambda^{(2)\prime}]'$, where $\lambda^{(2)}$ is a K_2-vector of the SDF parameters associated with f_2. Then $\delta_1^2 = \delta_2^2$ if and only if $\lambda^{(2)} = 0_{K_2}$.*

Note that Lemma 3 does not require the models to be correctly specified. It is applicable even when the models are misspecified. In order to test the equality of HJ-distances of the two models, Lemma 3 suggests that one can simply perform a test of $H_0 : \lambda^{(2)} = 0_{K_2}$ in model 2. Suppose $\hat{V}(\hat{\lambda}^{(2)})$ is a consistent estimate of the asymptotic variance of $\hat{\lambda}^{(2)}$. Then, under the null hypothesis of $H_0 : \lambda^{(2)} = 0_{K_2}$

$$T \hat{\lambda}^{(2)\prime} \hat{V}(\hat{\lambda}^{(2)})^{-1} \hat{\lambda}^{(2)} \stackrel{A}{\sim} \chi^2_{K_2}, \tag{43}$$

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 $\hat{\lambda}$ which assumes that model 2 is correctly specified. Instead, we need to rely on the misspecification robust standard errors of $\hat{\lambda}$ based on (31) to perform the test of $H_0 : \lambda^{(2)} = 0_{K_2}$.

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: *Let $A_2 = [\mathrm{O}_{K_2\times(K_1+1)},\ I_{K_2}]$. Under the null hypothesis of $H_0 : \delta_1^2 = \delta_2^2$*

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

where x_i's are independent χ^2_1 random variables and ξ_i's are the eigenvalues of $(A_2HA_2')^{-1}V(\hat{\lambda}^{(2)})$, with $V(\hat{\lambda}^{(2)})$ being the asymptotic variance of $\hat{\lambda}^{(2)}$.

Again, it should be emphasized that the misspecification robust version of $V(\hat{\lambda}^{(2)})$ should be used to conduct the test of $H_0 : \delta_1^2 = \delta_2^2$. In actual testing, we replace ξ_i by its sample counterpart $\hat{\xi}_i$, where the $\hat{\xi}_i$'s are the eigenvalues of

$$(A_2\hat{H}A_2')^{-1}\hat{V}(\hat{\lambda}^{(2)}), \tag{45}$$

and $\hat{H}$ and $\hat{V}(\hat{\lambda}^{(2)})$ are the sample counterparts of H and $V(\hat{\lambda}^{(2)})$, respectively.

When Y_t is multivariate elliptically distributed, we can further simplify the test of $H_0 : \delta_1^2 = \delta_2^2$. Lemma 4 summarizes our results.

Lemma 4 *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. When $\delta_1^2 = \delta_2^2 = \delta^2$, then $E[y_{1t}] = E[y_{2t}] = \lambda_0 + \lambda_1'\mu_1 \equiv \mu_y$ and $\mathrm{Var}[y_{1t}] = \mathrm{Var}[y_{2t}] = \lambda_1'V_{11}\lambda_1 \equiv \sigma_y^2$, and ξ_i's are the eigenvalues of*

$$\begin{aligned}&[\mu_y^2 + (1+\kappa)\sigma_y^2]I_{K_2} + \delta^2(A_2HA_2')^{-1}\\ &\times A_2H\left([1+(1+\kappa)\mu_2'V_{22}^{-1}\mu_2]\begin{bmatrix}1\\ \mu_1\end{bmatrix}\begin{bmatrix}1\\ \mu_1\end{bmatrix}' + \begin{bmatrix}0 & 0_K'\\ 0_K & (1+\kappa)(V_{11}-V_{12}V_{22}^{-1}V_{21})\end{bmatrix}\right)HA_2'.\end{aligned} \tag{46}$$

For the special case that both models are correctly specified, then

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \overset{A}{\sim} [\mu_y^2 + (1+\kappa)\sigma_y^2]\chi^2_{K_2}. \tag{47}$$

Since the eigenvalues ξ_i's under misspecified models are all greater than $\mu_y^2 + (1+\kappa)\sigma_y^2$ (the value of ξ_i's when the models are correctly specified), model misspecification creates additional sampling variation in $\hat{\delta}_1^2 - \hat{\delta}_2^2$. Without taking into account potential model misspecification, one might mistakenly reject $H_0 : \delta_1^2 = \delta_2^2$.

B. Non-nested Models

Let $f = [f_1',\ f_2',\ f_3']'$, where f_i is $K_i \times 1$ and $K = K_1 + K_2 + K_3$. Let $x_1 = [1,\ f_1',\ f_2']'$ and $x_2 = [1,\ f_2',\ f_3']'$. We assume that the SDF of model 1 is linear in x_1 and is given by

$$y_1 = \eta_0 + \eta_1'[f_1',\ f_2']' = \eta'x_1, \tag{48}$$

whereas the SDF of model 2 is linear in x_2 and is given by

$$y_2 = \lambda_0 + \lambda_1'[f_2',\ f_3']' = \lambda' x_2. \tag{49}$$

Note that $K_2 = 0$ implies that the two models do not have common factors. 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. The pricing errors of the two models are

$$\begin{aligned} e_1 &= D_1(D_1'V_{22}^{-1}D_1)^{-1}D_1'V_{22}^{-1}1_N - 1_N, && (50)\\ e_2 &= D_2(D_2'V_{22}^{-1}D_2)^{-1}D_2'V_{22}^{-1}1_N - 1_N, && (51)\end{aligned}$$

and the squared HJ-distances of the two models are given by

$$\begin{aligned} \delta_1^2 &= 1_N'V_{22}^{-1}1_N - 1_N'V_{22}^{-1}D_1(D_1'V_{22}^{-1}D_1)^{-1}D_1'V_{22}^{-1}1_N, && (52)\\ \delta_2^2 &= 1_N'V_{22}^{-1}1_N - 1_N'V_{22}^{-1}D_2(D_2'V_{22}^{-1}D_2)^{-1}D_2'V_{22}^{-1}1_N. && (53)\end{aligned}$$

For non-nested models, there are two cases to consider: (i) both models are correctly specified, and (ii) both models are misspecified. We take up these two cases in turn.

B.1. Both Models are Correctly Specified

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 the following population moment conditions

$$E[g_t(\theta)] = E\begin{bmatrix} g_{1t}(\eta) \\ g_{2t}(\lambda) \end{bmatrix} = E\begin{bmatrix} R_t x_{1t}'\eta - 1_N \\ R_t x_{2t}'\lambda - 1_N \end{bmatrix} = 0_{2N}, \tag{54}$$

where $\theta = [\eta',\ \lambda']'$. The sample moment conditions are then given by

$$\bar{g}_T(\theta) = \begin{bmatrix} \bar{g}_{1T}(\eta) \\ \bar{g}_{2T}(\lambda) \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T R_t x_{1t}'\eta - 1_N \\ \frac{1}{T}\sum_{t=1}^T R_t x_{2t}'\lambda - 1_N \end{bmatrix} = \begin{bmatrix} \hat{D}_1\eta - 1_N \\ \hat{D}_2\lambda - 1_N \end{bmatrix}. \tag{55}$$

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

$$A_T\bar{g}_T(\theta) = 0_{2N}, \tag{56}$$

where

$$A_T = \begin{bmatrix} \hat{D}_1'\hat{V}_{22}^{-1} & \mathrm{O}_{(K_1+K_2+1)\times N} \\ \mathrm{O}_{(K_2+K_3+1)\times N} & \hat{D}_2'\hat{V}_{22}^{-1} \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} D_1'V_{22}^{-1} & \mathrm{O}_{(K_1+K_2+1)\times N} \\ \mathrm{O}_{(K_2+K_3+1)\times N} & D_2'V_{22}^{-1} \end{bmatrix} \equiv A. \tag{57}$$

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

$$G_T(\theta) = \begin{bmatrix} \hat{D}_1 & \mathrm{O}_{N\times(K_2+K_3+1)} \\ \mathrm{O}_{N\times(K_1+K_2+1)} & \hat{D}_2 \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} D_1 & \mathrm{O}_{N\times(K_2+K_3+1)} \\ \mathrm{O}_{N\times(K_1+K_2+1)} & D_2 \end{bmatrix} \equiv G. \quad (58)$$

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

$$\sqrt{T}(\hat{\theta} - \theta) \overset{A}{\sim} N(0_{K+K_2+2}, (AG)^{-1}ASA'(G'A')^{-1}), \quad (59)$$

where

$$S = \sum_{j=-\infty}^{\infty} E[g_t(\theta)g_{t+j}(\theta)'] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad (60)$$

and S_{ij} is an $N \times N$ submatrix of S. The asymptotic distribution of $\bar{g}_T(\hat{\theta})$ is given by

$$\sqrt{T}\bar{g}_T(\hat{\theta}) \sim N(0_{2N}, [I_{2N} - G(AG)^{-1}A]S[I_{2N} - G(AG)^{-1}A]'). \quad (61)$$

After simplification, we can write

$$\sqrt{T} \begin{bmatrix} \bar{g}_{1T}(\hat{\eta}) \\ \bar{g}_{2T}(\hat{\lambda}) \end{bmatrix} \sim N\left(0_{2N}, \begin{bmatrix} G_1S_{11}G_1' & G_1S_{12}G_2' \\ G_2S_{21}G_1' & G_2S_{22}G_2' \end{bmatrix}\right), \quad (62)$$

where $G_1 = I_N - D_1(D_1'V_{22}^{-1}D_1)^{-1}D_1'V_{22}^{-1}$ and $G_2 = I_N - D_2(D_2'V_{22}^{-1}D_2)^{-1}D_2'V_{22}^{-1}$.

Proposition 3 presents the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ when both models are correctly specified.

Proposition 3: *Denote $n_1 = N - K_1 - K_2 - 1$ and $n_2 = N - K_2 - K_3 - 1$. Let P_1 be an $N \times n_1$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}}D_1$. Similarly, let P_2 be an $N \times n_2$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}}D_2$. When $\delta_1^2 = \delta_2^2 = 0$, then*

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \overset{A}{\sim} \sum_{i=1}^{n_1+n_2} \xi_i x_i, \quad (63)$$

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

$$\begin{bmatrix} P_1'V_{22}^{-\frac{1}{2}}S_{11}V_{22}^{-\frac{1}{2}}P_1 & P_1'V_{22}^{-\frac{1}{2}}S_{12}V_{22}^{-\frac{1}{2}}P_2 \\ -P_2'V_{22}^{-\frac{1}{2}}S_{21}V_{22}^{-\frac{1}{2}}P_1 & -P_2'V_{22}^{-\frac{1}{2}}S_{22}V_{22}^{-\frac{1}{2}}P_2 \end{bmatrix}. \quad (64)$$

Note that the ξ_i's are not all positive because $\hat{\delta}_1^2 - \hat{\delta}_2^2$ can take positive or negative values. Therefore, for the non-nested model case, we must perform a a two-tailed test of $H_0 : \delta_1^2 = \delta_2^2$ instead of a one-tailed test as in the nested models case.

When Y_t is multivariate elliptically distributed, we can further simplify the ξ_i's. The results are given in the Lemma 5.

Lemma 5 *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 $A_1 = [I_{K_1+K_2} \ \mathrm{O}_{(K_1+K_2)\times K_3}]$ and $A_2 = [\mathrm{O}_{(K_2+K_3)\times K_1}, \ I_{K_2+K_3}]$, the first two moments of the two SDFs can be obtained as*

$$
\begin{aligned}
\mu_{y_1} &= E[y_{1t}] = \eta_0 + \eta_1' A_1 \mu_1, && (65)\\
\mu_{y_2} &= E[y_{2t}] = \lambda_0 + \lambda_1' A_2 \mu_1, && (66)\\
\sigma_{y_1}^2 &= \mathrm{Var}[y_{1t}] = \eta_1' A_1 V_{11} A_1' \eta_1, && (67)\\
\sigma_{y_2}^2 &= \mathrm{Var}[y_{2t}] = \lambda_1' A_2 V_{11} A_2' \lambda_1, && (68)\\
\sigma_{y_1,y_2} &= \mathrm{Cov}[y_{1t}, y_{2t}] = \eta_1' A_1 V_{11} A_2' \lambda_1, && (69)
\end{aligned}
$$

and ξ_i's are the eigenvalues of

$$
\begin{bmatrix} [\mu_{y_1}^2 + (1+\kappa)\sigma_{y_1}^2] I_{n_1} & [\mu_{y_1}\mu_{y_2} + (1+\kappa)\sigma_{y_1,y_2}] P_1' P_2 \\ -[\mu_{y_1}\mu_{y_2} + (1+\kappa)\sigma_{y_1,y_2}] P_2' P_1 & -[\mu_{y_2}^2 + (1+\kappa)\sigma_{y_2}^2] I_{n_2} \end{bmatrix}. \tag{70}
$$

As an example, we consider a case with $f_{3t} = f_{1t} + \epsilon_t$, where ϵ_t is a zero-mean measurement error uncorrelated with f_{1t}, f_{2t} and R_t. Under this setup, we have $n_1 = n_2$ and model 2 is effectively the same as model 1 except that some of its factors are more noisy. Since $\mathrm{Cov}[R_t, f_{1t}'] = \mathrm{Cov}[R_t, f_{3t}']$, it is straightforward to show that $\mu_{y_1} = \mu_{y_2}$, $\sigma_{y_1}^2 = \sigma_{y_1,y_2} < \sigma_{y_2}^2$ and $P_2 = P_1$. It follows that ξ_i's are the eigenvalues of the matrix

$$
\begin{bmatrix} \mu_{y_1}^2 + (1+\kappa)\sigma_{y_1}^2 & \mu_{y_1}^2 + (1+\kappa)\sigma_{y_1}^2 \\ -\mu_{y_1}^2 - (1+\kappa)\sigma_{y_1}^2 & -\mu_{y_1}^2 - (1+\kappa)\sigma_{y_2}^2 \end{bmatrix} \otimes I_{n_1}. \tag{71}
$$

It can be verified that this matrix has n_1 pairs of eigenvalues of

$$
\zeta_1, \zeta_2 = \frac{-(1+\kappa)(\sigma_{y_2}^2 - \sigma_{y_1}^2) \pm \sqrt{(1+\kappa)(\sigma_{y_2}^2 - \sigma_{y_1}^2)[4\mu_{y_1}^2 + (1+\kappa)(3\sigma_{y_1}^2 + \sigma_{y_2}^2)]}}{2}, \tag{72}
$$

so we have

$$
T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \stackrel{A}{\sim} \zeta_1 \chi_{n_1}^2 + \zeta_2 \chi_{n_1}^2. \tag{73}
$$

Since $E[\zeta_1\chi^2_{n_1} + \zeta_2\chi^2_{n_1}] = -n_1(1+\kappa)(\sigma^2_{y_2} - \sigma^2_{y_1}) < 0$, we expect the model with noisier factors to have a larger sample HJ-distance.

B.2. Both Models are Misspecified

When two non-nested models are both misspecified, the asymptotic distribution of $\hat\delta_1^2 - \hat\delta_2^2$ is given in Proposition 4.

Proposition 4: *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 $\delta_2 \neq 0$, then

$$\sqrt{T}(\hat\delta_1^2 - \hat\delta_2^2 - (\delta_1^2 - \delta_2^2)) \overset{A}{\sim} N(0, v_d), \tag{74}$$

where

$$v_d = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]. \tag{75}$$

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

$$\sqrt{T}(\hat\delta_1^2 - \hat\delta_2^2) \overset{A}{\sim} N(0, v_d) \tag{76}$$

and d_t can be simplified to

$$d_t = 2u_{1t}y_{1t} - u_{1t}^2 - 2u_{2t}y_{2t} + u_{2t}^2. \tag{77}$$

When Y_t is multivariate elliptically distributed, we can further simplify the asymptotic distribution of $\hat\delta_1^2 - \hat\delta_2^2$. The results are given in Lemma 6.

Lemma 6 *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. The asymptotic variance of $\hat\delta_1^2 - \hat\delta_2^2$ is given by*

$$\begin{aligned} v_d &= 4(\mu_{y_1}^2\delta_1^2 + \mu_{y_2}^2\delta_2^2 - 2\mu_{y_1}\mu_{y_2}\delta_{12}) \\ &\quad + 4(1+\kappa)(\sigma_{y_1}^2\delta_1^2 + \sigma_{y_2}^2\delta_2^2 - 2\sigma_{y_1,y_2}\delta_{12} - \delta_1^2\delta_2^2 + \delta_{12}^2) + (2+3\kappa)(\delta_1^2 - \delta_2^2)^2, \end{aligned} \tag{78}$$

where $\delta_{12} = e_1' V_{22}^{-1} e_2$. Under the null hypothesis of $H_0 : \delta_1^2 = \delta_2^2 = \delta^2 \neq 0$, v_d can be simplified to

$$v_d = 4(\mu_{y_1}^2\delta^2 + \mu_{y_2}^2\delta^2 - 2\mu_{y_1}\mu_{y_2}\delta_{12}) + 4(1+\kappa)(\sigma_{y_1}^2\delta^2 + \sigma_{y_2}^2\delta^2 - 2\sigma_{y_1,y_2}\delta_{12} - \delta^4 + \delta_{12}^2). \tag{79}$$

III. Empirical Analysis

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

A. Data and Asset Pricing Models

We use the same data as in Hodrick and Zhang (2001). For monthly models, the data cover the period from 1952/1 to 1997/12 (552 monthly observations). The only exception is the consumption CAPM, for which we have monthly data starting in 1959/2 (467 monthly observations). For quarterly models, the data cover the period from 1953 Q1 to 1997 Q4 (180 quarterly observations). The asset returns are the returns on the 25 Fama-French size and book-to-market portfolios in excess of the T-bill rate plus the gross return on the T-bill. Monthly excess returns are obtained by subtracting the one-month T-bill return from the returns on the 25 Fama-French portfolios. Quarterly excess returns are obtained by compounding the monthly Fama-French returns to a quarterly frequency and subtracting the three-month T-bill return. Following Hodrick and Zhang (2001), we consider unconditional as well as conditional models. For conditional monthly models, the conditioning variables are the lagged values of the cyclical component of the industrial production index (Lag IP) and a January dummy (JAN). For conditional quarterly models, the conditioning variables are the lagged values of the cyclical component of real GNP (Lag GNP), the lagged values of the consumption-wealth ratio (Lag CAY) of Lettau and Ludvigson (2001), and a January dummy (JAN) that takes on the value of one for the first quarter and the value of zero for all other quarters.[6]

For monthly models, we consider six different empirical specifications. The first model is the CAPM, which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{vw} r_t^{vw}, \tag{80}$$

where r_t^{vw} is the excess return on the CRSP value-weighted index. The second model is a linearized

[6]We thank Robert Hodrick and Xiaoyan Zhang for sharing their data with us. We also thank Martin Lettau and Kenneth French for making the rest of the data available through their web sites.

consumption CAPM (C–CAPM), which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{cg} r_t^{cg}, \tag{81}$$

where r_t^{cg} is the growth rate in real nondurables consumption. The third model (JW) is the conditional CAPM of Jagannathan and Wang (1996), which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{vw} r_t^{vw} + \lambda_{prem} r_{t-1}^{prem} + \lambda_{lab} r_t^{lab}, \tag{82}$$

where r_t^{vw} is the excess return on the CRSP value-weighted index,[7] r_{t-1}^{prem} is the lagged yield spread between low and high-grade corporate bonds, and r_t^{lab} is the growth rate in per capita income. The fourth model (CAMP) is a linearized version of Campbell's (1996) intertemporal capital asset pricing model, which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{rvw} r_t^{rvw} + \lambda_{clab} r_t^{clab} + \lambda_{div} r_t^{div} + \lambda_{rtb} r_t^{rtb} + \lambda_{trm} r_t^{trm}, \tag{83}$$

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^{div} is the dividend yield on the CRSP value-weighted market portfolio, r_t^{rtb} is the difference between the one-month T-bill rate and its one-year backward moving average, and r_t^{trm} is the yield spread between long and short-term government bonds. The fifth model (FF3) is the Fama-French (1993) three-factor model, which assumes that the SDF is

$$y_t = \lambda_0 + \lambda_{vw} r_t^{vw} + \lambda_{smb} r_t^{smb} + \lambda_{hml} r_t^{hml}, \tag{84}$$

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

$$y_t = \lambda_0 + \lambda_{vw} r_t^{vw} + \lambda_{smb} r_t^{smb} + \lambda_{hml} r_t^{hml} + \lambda_{term} r_t^{term} + \lambda_{def} r_t^{def}, \tag{85}$$

where r_t^{term} is the yield spread between a thirty-year bond and the one-month T-bill, and r_t^{def} is the yield spread between low and high-grade corporate bonds (same series as in the JW model but not lagged).

[7]Jagannathan and Wang (1996) actually use the gross return on the CRSP value-weighted index and not the excess return. However, we use the excess return to be consistent with Hodrick and Zhang (2001).

To form monthly 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.[8] 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), Campbell (1996), and Ferson and Harvey (1999), among others.

For quarterly models, we consider seven empirical specifications: the six models described above and, in addition, the production based asset pricing model (COCH) of Cochrane (1996). The corresponding SDF is

$$y_t = \lambda_0 + \lambda_{gnr} r_t^{gnr} + \lambda_{gr} r_t^{gr}, \tag{86}$$

where r_t^{gnr} is the growth rate on real nonresidential investment and r_t^{gr} is the growth rate on real residential investment. Quarterly conditional models are formed by scaling the factors of the quarterly unconditional models by a constant and either Lag GNP, Lag CAY or JAN.

B. 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 squared HJ-distances of two models.

B.1. Summary of the Models

Table I provides a summary of the different monthly and quarterly asset pricing models. The results are largely identical to the ones reported in Table 3 of Hodrick and Zhang (2001). The estimates of the HJ-distance are denoted with $\hat{\delta}$. The p-value of the test of $H_0 : \delta = 0$ from equation (16) is $p(\delta = 0)$. The standard error of the sample HJ-distance from equation (20) computed under

[8] Although the JW model is already an unconditional version of a conditional model, we follow Hodrick and Zhang (2001) and scale its factors by a constant and either Lag IP or JAN, which implies a total of eight factors in the model SDF.

the alternative hypothesis that $\delta \neq 0$ is se($\hat{\delta}$).[9] The 95% confidence interval of δ based on se($\hat{\delta}$) is CI(δ). No. of par. denotes the number of parameters in each asset pricing model.

Table I about here

Starting from the monthly unconditional asset pricing models in Panel A, most of the models are rejected by the data at the 5% level. This provides compelling evidence to incorporate model misspecification into our statistical analysis. The CAMP and FF5 models have the lowest HJ-distances and are the only ones that pass the test of $H_0 : \delta = 0$ at the 5% level. However, an examination of the 95% confidence intervals of δ for the CAMP and FF5 models indicates that the HJ-distances of these two models are far from zero. The reason behind the different outcomes provided by the specification tests and confidence intervals analyses is that the p-value of $H_0 : \delta = 0$ is computed under the hypothesis that the model is correctly specified, while the confidence interval of δ uses a standard error that is only valid when the model is misspecified. Since the asymptotic distributions of $\hat{\delta}$ under correctly specified and misspecified models are different, the conclusions that we obtain from the two types of analyses can also be different. In addition, the confidence intervals of δ for different models significantly overlap each other, possibly suggesting that, after accounting for sampling variability, it might be difficult to detect substantial differences in the HJ-distances of competing models.

When scaling the factors by either Lag IP or JAN in Panels B and C, the estimates of the HJ-distances of the conditional models are smaller than the corresponding estimates of the unconditional models. The smaller HJ-distances of the conditional models can be due to two reasons: (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 so the conditional models are better able to fit the data. In the scaled factor case, more 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 factors by JAN, the C-CAPM, JW, CAMP, FF3 and FF5 models are not rejected by

[9]The se($\hat{\delta}$)'s are computed assuming no serial correlation. A separate set of results (available upon request) considers a 12-lag and a 4-lag Newey-West (1987) adjustment for monthly and quarterly models, respectively. Overall, accounting for serial correlation in the data makes the standard errors of $\hat{\delta}$ slightly bigger. In addition, the lag adjustments generally deliver higher p-values for testing $H_0 : \delta = 0$, thus making the models harder to reject.

the data at the 5% level, as shown in Panel C. Similar to the unconditional models, an inspection of the confidence intervals of the HJ-distances suggests that the HJ-distances of all models are far from zero. In addition, the confidence intervals of δ for the different models significantly overlap each other.

When considering quarterly models, all the unconditional models are rejected by the data, as shown in Panel D. Panel E shows that scaling factors by Lag GNP makes it more difficult to reject the models. Specifically, the CAMP, COCH, FF3 and FF5 models pass the HJ-distance test at the 5% level. In contrast, when scaling the factors by Lag CAY in Panel F, only the CAMP model is not rejected by the data. When scaling the factors by JAN in Panel G, we cannot reject the JW, CAMP, COCH and FF5 models using the sample HJ-distance. Similar to the monthly case, scaling the factors of quarterly models by different instruments results in consistently lower HJ-distances than the ones that we observe in the unconditional models case. Nevertheless, the 95% confidence intervals of δ indicate that the HJ-distances of the competing models are far from zero. In addition, the confidence intervals of δ for the different models significantly overlap each other. Consistent with the monthly case, the confidence intervals analysis for quarterly models suggests that, after accounting for sampling variability, there might not be substantial differences in the HJ-distances of competing asset pricing models.

As we mentioned above, going from unconditional to conditional models always delivers smaller sample HJ-distances. Hence, 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, as shown in Panels A through G. The larger standard errors reflect the additional noise brought into the model by the instruments. A direct implication is that competing models may become even more difficult to distinguish once conditioning information is introduced into the models. 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 would be tempted to conclude that the CAMP model should be viewed as the preferred SDF because: (i) the model overall passes the HJ-distance test; and (ii) the model produces HJ-distance estimates that are consistently among the lowest. However, neither the sample HJ-distance nor its p-value allow us to formally compare models. In the subsequent empirical analysis, we will conduct our tests of equality to investigate whether a specific asset pricing model outperforms the others.

B.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 section, 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 fact, this proves to be the case.

In Table II, we focus on the SDF parameter estimates, $\hat{\lambda}$, of unconditional monthly and quarterly models. We report $\hat{\lambda}$ and associated t-ratios under correctly specified and potentially misspecified models.[10] In computing t-ratios under correctly specified models, we use the sample counterparts of (32), while in computing t-ratios under potential model misspecification, we use the sample counterparts of (31). 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 monthly CAPM in Panel A of Table II. The t-ratios on $\hat{\lambda}_{vw}$ for correctly specified and potentially misspecified models are almost identical (-3.31 and -3.32, respectively). The same type

[10]The 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 and a 4-lag Newey-West (1987) adjustment for monthly and quarterly models, respectively. Overall, accounting for serial correlation in the data makes the standard errors of $\hat{\lambda}$ bigger.

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 -1.90 to a t-ratio of -1.08 and, for the FF5 model, we go from a t-ratio on $\hat{\lambda}_{term}$ of 2.67 to a t-ratio of 1.70. For the quarterly unconditional models in Panel B, we see a similar pattern. For example, for the COCH model, we go from a t-ratio on $\hat{\lambda}_{gr}$ of -2.00 to a t-ratio of -1.33. To summarize, we find that for non-traded factors, all the t-ratios under potentially misspecified models are smaller than the t-ratios under correctly specified models. Hence, ignoring model misspecification can lead to the erroneous conclusion that certain factors are priced.

Table II 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 Wald test we focus on is the test of the hypothesis that the parameters associated with the scaled factors are jointly equal to zero. Given Lemma 3 above, 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 III, we report the Wald test statistics under correct specification (cs) and potential misspecification (m) for monthly conditional models in Panels A and B, and for quarterly conditional models in Panels C through E. Once again, we find that ignoring potential model misspecification makes a substantial difference in terms of the p-values of the Wald tests. For monthly models, with the exception of the JW model with factors scaled by JAN, we cannot reject the null hypothesis that the parameters of the scaled factors are all equal to zero at the 5% level, when using misspecification robust Wald tests. Similarly, for quarterly models, with the exception of the CAPM model with factors scaled by JAN, we cannot reject the null hypothesis that the parameters of the scaled factors are all zero at the 5% level, when using misspecification robust Wald tests.[11] Therefore, although conditional models always deliver lower sample HJ-distances than unconditional models, we do not find much statistical evidence to conclude that conditional models are better than unconditional models in terms of HJ-distance after we account for potential model misspecification.

[11]The p-values of the Wald tests are computed assuming no serial correlation. A separate set of results (available upon request) considers a 12-lag and a 4-lag Newey-West (1987) adjustment for monthly and quarterly models, respectively. Overall, accounting for serial correlation in the data makes the p-values even larger.

Table III 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 is taken into account. For example, when scaling monthly consumption growth by JAN, we go from a t-ratio of -2.42 under correctly specified models to a t-ratio of -1.65 under potentially misspecified models. In the Fama-French (1993) three-factor model, when scaling the monthly *smb* factor with JAN, we go from a t-ratio of -2.18 to a t-ratio of -1.24. Finally, in the COCH model, when scaling the *gr* factor with Lag GNP, we go from a t-ratio of -2.26 to a t-ratio of -1.64.

To summarize, accounting for model misspecification can often make a qualitative difference in terms of 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 correct, potential model misspecification should be accounted for when computing the standard errors of the estimates of SDF parameters.

B.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 3 to conduct the tests of equality of HJ-distances.[12] 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

[12] Results obtained using Lemma 3 (not reported in the paper) are largely consistent with the ones shown in the tables.

as misspecified models. For non-nested models, the asymptotic distribution of our test statistic depends on whether the competing models are correctly specified or misspecified. Therefore, in the non-nested case, we need to take a stand in order to conduct the tests of equality of HJ-distances. We decide to present our empirical results under the assumption that the competing models are misspecified because we believe that this is the more realistic scenario. In Tables IV–VI, 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 Table IV, we report differences between the squared sample HJ-distances of two models and the associated p-values (in parentheses).[13]

Table IV about here

In Panels A and D, we compare monthly and quarterly models with unscaled factors; in Panels B, C, E, F and G, we compare monthly and quarterly models with factors scaled by the same conditioning variable. For monthly models with unscaled factors, we observe that the CAPM and the JW models are outperformed by the CAMP, FF3 and FF5 models, while the C-CAPM is outperformed by the FF3 and FF5 models. However, when we consider models with scaled factors, no model clearly outperforms the others since all the p-values are greater than 0.05. For quarterly unconditional models, the CAPM and the COCH models are outperformed by the FF3 and FF5 models. When we scale factors by Lag GNP, the FF3 model outperforms the CAPM. When we scale factors by Lag CAY, the FF3 model outperforms the COCH model. When we scale factors by JAN, no model significantly outperforms the others. Out of 129 pairwise tests of equality, only in 14 cases we find differences between models that are statistically significant at the 5% level. The only models that seem to underperform in a few circumstances are the CAPM, the C-CAPM, the JW, and the COCH models. In addition, we find no evidence that intertemporal CAPM-type specifications such as the Campbell (1996) model outperform the Fama-French three and five factor models.

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 I. However, it is inappropriate to conclude that the instruments actually help to reduce the

[13]Note that in the case of non-nested models, the p-values are two-tailed p-values.

pricing errors without performing a formal comparison of the unconditional models vs. the conditional models. In Table V, we report the results from testing the equality of HJ-distances between conditional and unconditional models.

Table V about here

Panels A and B are for monthly models, while Panels C through E are for quarterly models. 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, we cannot find statistically significant differences in HJ-distances between the conditional version and the unconditional version of the model. For comparisons across models, we see a pattern which is similar to the one that we observe in Table IV. Namely, for monthly models, the unconditional C-CAPM model is outperformed by the conditional CAMP, FF3, and FF5 models when scaling by Lag IP, and by the FF3, and FF5 when scaling by JAN. The unconditional JW model is outperformed by the conditional CAMP, FF3, and FF5 models. However, the unconditional CAPM is now only outperformed by the conditional CAMP model, indicating that the instruments add noise to the data, thus making it harder to detect significant differences between the HJ-distances of two competing models. For quarterly models, similarly to Table IV, we find some evidence of underperformance of the unconditional C-CAPM and COCH models. In synthesis, out of 219 model comparisons, we find that only in 19 cases the differences between models are statistically significant at the 5% level. Once again, the data are generally too noisy for us to conclude that one model clearly outperforms the others.

Finally, in Table VI, we compare conditional models with factors scaled by one instrument with conditional models that use a different instrument. The reason behind this type of exercise is that different conditional models might capture different characteristics of the economy and that the type of scaling might affect their absolute and relative performances.

Table VI about here

For monthly models, we find that the performances of all the competing models cannot be distinguished. For quarterly models, we find some evidence of underperformance of the CAPM and the COCH models. In synthesis, out of 183 model comparisons, we find that only in 5 cases the differences between models are statistically significant at the 5% level.[14] 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.

IV. Conclusion

In this paper, we propose a new methodology to test whether or not two competing linear asset pricing models have the same HJ-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. We provide the asymptotic distribution of our test statistic under general assumptions as well as under the assumption that factors and returns are jointly elliptically distributed. The multivariate elliptical case allows us to gain further intuition on the important determinants of the asymptotic distribution of the test statistic.

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 multivariate elliptical case, we are able to show analytically that the standard errors under potentially misspecified models are always bigger than the standard errors that assume that the model is correctly specified. In addition, we show that the misspecification adjustment depends on, among other things, the correlation between the factor and the returns of 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 using the same data as in Hodrick and Zhang (2001). 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 the others in terms

[14] All the p-values in Tables IV–VI are computed assuming no serial correlation. A separate set of results (available upon request) considers a 12-lag and a 4-lag Newey-West (1987) adjustment for monthly and quarterly models, respectively. Overall, accounting for serial correlation in the data makes the p-values of the test statistics larger and the differences between models even harder to detect.

of HJ-distance. Specifically, we find no evidence that conditional and intertemporal CAPM-type specifications outperform the Fama-French (1993) three and five-factor models in terms of HJ-distance. Our results appear to be robust to the horizon considered and to factor scaling.

While we do not find 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 (like OLS and GLS R^2 or sum of squares 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 importantly, 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

We first present some expressions for the mixed moments of multivariate elliptical distributions, which will be used repeatedly in the Appendix.

Claim: Suppose (X_i, X_j, X_k, X_l) follow a multivariate elliptical distribution with multivariate kurtosis parameter κ. Denoting $\mu_i = E[X_i]$ and $\sigma_{ij} = \mathrm{Cov}[X_i, X_j]$, we have

$$\begin{aligned}
E[X_iX_j] &= \sigma_{ij} + \mu_i\mu_j, && \text{(A1)}\\
E[X_iX_jX_k] &= \mu_i\mu_j\mu_k + \mu_i\sigma_{jk} + \mu_j\sigma_{ik} + \mu_k\sigma_{ij}, && \text{(A2)}\\
E[X_iX_jX_kX_l] &= (1+\kappa)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) + \mu_i\mu_j\mu_k\mu_l \\
&\quad + \sigma_{ij}\mu_k\mu_l + \sigma_{ik}\mu_j\mu_l + \sigma_{il}\mu_j\mu_k + \sigma_{jk}\mu_i\mu_l + \sigma_{jl}\mu_i\mu_k + \sigma_{kl}\mu_i\mu_j. && \text{(A3)}
\end{aligned}$$

Proof: (A1) follows from the definition of covariance. For (A2) and (A3), Lemma 2 of Maruyama and Seo (2003) shows that

$$\begin{aligned}
E[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)] &= 0, && \text{(A4)}\\
E[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)] &= (1+\kappa)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}). && \text{(A5)}
\end{aligned}$$

Using (A1) and (A4), we obtain (A2). For the product moment of $X_iX_jX_kX_l$, we use (A2) and (A4) to write

$$\begin{aligned}
& E[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)] \\
=\ & E[X_i(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)] \\
=\ & E[X_iX_j(X_k-\mu_k)(X_l-\mu_l)] - \mu_jE[X_i(X_k-\mu_k)(X_l-\mu_l)] \\
=\ & E[X_iX_jX_kX_l] - \mu_kE[X_iX_jX_l] - \mu_lE[X_iX_jX_k] + \mu_k\mu_l(\sigma_{ij} + \mu_i\mu_j) \\
& - \mu_jE[(X_i-\mu_i)(X_k-\mu_k)(X_l-\mu_l)] - \mu_i\mu_jE[(X_k-\mu_k)(X_l-\mu_l)] \\
=\ & E[X_iX_jX_kX_l] - \mu_i\mu_j\mu_k\mu_l - \sigma_{ij}\mu_k\mu_l - \sigma_{ik}\mu_j\mu_l - \sigma_{il}\mu_j\mu_k \\
& - \sigma_{jk}\mu_i\mu_l - \sigma_{jl}\mu_i\mu_k - \sigma_{kl}\mu_i\mu_j. \qquad \text{(A6)}
\end{aligned}$$

Using this equation and (A5), we obtain (A3). This completes the proof.

Proof of Lemma 1: The result for $\delta = 0$ is in Kan and Zhou (2004). Therefore, we only provide the proof for the $\delta \neq 0$ case. Since Y_t is multivariate elliptically distributed, $u_t = e'V_{22}^{-1}R_t$ and $y_t =$

$\mu_y + \lambda_1'(f_t - \mu_1)$ are bivariate elliptically distributed because both of them are linear combinations of the elements of Y_t. Using the properties of multivariate elliptical distributions (see Muirhead, 1982, p.41) and the fact that $e'V_{22}^{-1}\mu_2 = 0$, we have $E[u_t] = 0$, $E[u_t^2] = e'V_{22}^{-1}e = \delta^2$, $E[u_t^3] = 0$, $E[u_t^4] = 3(1+\kappa)E[u_t^2]^2 = 3(1+\kappa)\delta^4$, $E[y_t] = \mu_y$, $E[y_t^2] = \mu_y^2 + \sigma_y^2$, where κ is the kurtosis parameter of the elliptical distribution. In addition, using the identity $D'V_{22}^{-1}e = 0_{K+1}$, we have

$$E[u_t y_t] = E[e'V_{22}^{-1}R_t x_t'\lambda] = e'V_{22}^{-1}D\lambda = 0. \tag{A7}$$

Therefore, u_t and y_t are uncorrelated. Applying (A3), we obtain $E[u_t^2 y_t^2] = (1+\kappa)\delta^2\sigma_y^2 + \delta^2\mu_y^2$ and $E[u_t^3 y_t] = 0$. Using these moments of u_t and y_t, we have

$$E[q_t^2] \;=\; 4[\mu_y^2 + (1+\kappa)\sigma_y^2]\delta^2 + (2+3\kappa)\delta^4. \tag{A8}$$

This completes the proof.

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

$$\phi = \begin{bmatrix} \mu \\ \mathrm{vec}(V) \end{bmatrix}, \qquad \hat{\phi} = \begin{bmatrix} \hat{\mu} \\ \mathrm{vec}(\hat{V}) \end{bmatrix}. \tag{A9}$$

Under some standard regularity conditions, we can assume[15]

$$\sqrt{T}(\hat{\phi} - \phi) \overset{A}{\sim} N(0_{(N+K)\times(N+K+1)}, S_0). \tag{A10}$$

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

$$r_t(\phi) = \begin{bmatrix} Y_t - \mu \\ \mathrm{vec}((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}. \tag{A11}$$

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

$$S_0 = \sum_{j=-\infty}^{\infty} E[r_t(\phi) r_{t+j}(\phi)']. \tag{A12}$$

[15]Note 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 $[\hat{\mu}',\ \mathrm{vech}(\hat{V})']'$, but the results are the same under both specifications.

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

$$\sqrt{T}(\hat{\lambda}-\lambda) \quad \overset{A}{\sim} \quad N\left(0, \left[\frac{\partial\lambda}{\partial\phi'}\right]' S_0 \left[\frac{\partial\lambda}{\partial\phi'}\right]\right). \tag{A13}$$

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

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

$$\frac{\partial\lambda}{\partial\mu_1'} = -\left[1,\ 0_K'\right]'\lambda_1', \tag{A14}$$

$$\frac{\partial\lambda}{\partial\mu_2'} = -H\left[1,\ \mu_1'\right]' e'V_{22}^{-1} - HD'V_{22}^{-1}\mu_y, \tag{A15}$$

$$\begin{aligned}\frac{\partial\lambda}{\partial \mathrm{vec}(V)'} &= \left[H[0_K,\ I_K]',\ \mathrm{O}_{(K+1)\times N}\right]\otimes\left[0_K',\ -e'V_{22}^{-1}\right] \\ &\quad + \left[-\lambda_1',\ e'V_{22}^{-1}\right]\otimes\left[\mathrm{O}_{(K+1)\times K},\ HD'V_{22}^{-1}\right].\end{aligned} \tag{A16}$$

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

$$\frac{\partial d}{\partial\mu_1'} = \begin{bmatrix}\mathrm{O}_{N\times K}\\ I_K\otimes\mu_2\end{bmatrix} = \begin{bmatrix}0_K'\\ I_K\end{bmatrix}\otimes\mu_2, \tag{A17}$$

$$\frac{\partial d}{\partial\mu_2'} = \begin{bmatrix}I_N\\ \mu_1\otimes I_N\end{bmatrix} = \begin{bmatrix}1\\ \mu_1\end{bmatrix}\otimes I_N, \tag{A18}$$

$$\frac{\partial d}{\partial \mathrm{vec}(V)'} = [[0_K,\ I_K]',\ \mathrm{O}_{(K+1)\times N}]\otimes[\mathrm{O}_{N\times K},\ I_N]. \tag{A19}$$

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

$$\frac{\partial \mathrm{vec}(D')}{\partial d'} = \frac{\partial K_{N,K+1}d}{\partial d'} = K_{N,K+1}, \tag{A20}$$

$$\begin{aligned}\frac{\partial \mathrm{vec}(D'V_{22}^{-1}D)}{\partial d'} &= (D'V_{22}^{-1}\otimes I_{K+1})\frac{\partial \mathrm{vec}(D')}{\partial d'} + (I_{K+1}\otimes D'V_{22}^{-1})\frac{\partial d}{\partial d'} \\ &= (D'V_{22}^{-1}\otimes I_{K+1})K_{N,K+1} + (I_{K+1}\otimes D'V_{22}^{-1}) \\ &= (I_{(K+1)^2} + K_{K+1})(I_{K+1}\otimes D'V_{22}^{-1}),\end{aligned} \tag{A21}$$

$$\frac{\partial \mathrm{vec}((D'V_{22}^{-1}D)^{-1})}{\partial \mathrm{vec}(D'V_{22}^{-1}D)'} = -(D'V_{22}^{-1}D)^{-1}\otimes(D'V_{22}^{-1}D)^{-1}, \tag{A22}$$

$$\frac{\partial \mathrm{vec}((D'V_{22}^{-1}D)^{-1})}{\partial d'} = -(I_{(K+1)^2} + K_{K+1})[(D'V_{22}^{-1}D)^{-1}\otimes(D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}], \tag{A23}$$

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

$$
\begin{aligned}
&- (D \otimes I_{K+1})(I_{(K+1)^2} + K_{K+1})[(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}] \\
= \;& K_{N,K+1}[(D'V_{22}^{-1}D)^{-1} \otimes I_N] - D(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1} \\
&- K_{N,K+1}[(D'V_{22}^{-1}D)^{-1} \otimes D(D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}] \\
= \;& K_{N,K+1}[(D'V_{22}^{-1}D)^{-1} \otimes [I_N - D(D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}]] \\
&- D(D'V_{22}^{-1}D)^{-1} \otimes (D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}. \qquad \text{(A24)}
\end{aligned}
$$

Therefore,

$$
\frac{\partial \lambda}{\partial d'} = (1_N' V_{22}^{-1} \otimes I_{K+1}) \frac{\partial \mathrm{vec}(HD')}{\partial d'} = -H \otimes e'V_{22}^{-1} - \lambda' \otimes HD'V_{22}^{-1}. \qquad \text{(A25)}
$$

It follows that

$$
\begin{aligned}
\frac{\partial \lambda}{\partial \mu_1'} &= \frac{\partial \lambda}{\partial d'} \frac{\partial d}{\partial \mu_1'} = -\lambda_1' \otimes HD'V_{22}^{-1}\mu_2 = -[1,\ 0_K']'\lambda_1', \qquad \text{(A26)} \\
\frac{\partial \lambda}{\partial \mu_2'} &= \frac{\partial \lambda}{\partial d'} \frac{\partial d}{\partial \mu_2'} = -H[1,\ \mu_1']'e'V_{22}^{-1} - HD'V_{22}^{-1}\mu_y. \qquad \text{(A27)}
\end{aligned}
$$

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

$$
\frac{\partial \lambda}{\partial \mathrm{vec}(V)'} = (1_N' V_{22}^{-1} D \otimes I_{K+1}) \frac{\partial \mathrm{vec}(H)}{\partial \mathrm{vec}(V)'} + (1_N' V_{22}^{-1} \otimes H) \frac{\partial \mathrm{vec}(D')}{\partial \mathrm{vec}(V)'} + (1_N' \otimes HD') \frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'}. \qquad \text{(A28)}
$$

The last two terms are given by

$$
\begin{aligned}
(1_N' V_{22}^{-1} \otimes H) \frac{\partial \mathrm{vec}(D')}{\partial \mathrm{vec}(V)'} &= [H\,[0_K,\ I_K]',\mathrm{O}_{(K+1)\times N}] \otimes [0_K',\ 1_N' V_{22}^{-1}], \qquad \text{(A29)} \\
(1_N' \otimes HD') \frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'} &= -[0_K',\ 1_N' V_{22}^{-1}] \otimes [\mathrm{O}_{(K+1)\times K},\ HD'V_{22}^{-1}]. \qquad \text{(A30)}
\end{aligned}
$$

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

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

$$
\begin{aligned}
&+[[0_K,\ I_K]',\ \mathrm{O}_{(K+1)\times N}] \otimes [\mathrm{O}_{(K+1)\times K},\ D'V_{22}^{-1}]\} \\
=\ & [H\,[0_K,\ I_K]',\ \mathrm{O}_{(K+1)\times N}] \otimes [0_K',\ -\lambda' D'V_{22}^{-1}] + [0_K',\ \lambda' D'V_{22}^{-1}] \otimes [\mathrm{O}_{(K+1)\times K},\ HD'V_{22}^{-1}] \\
& -[\lambda_1',\ 0_N'] \otimes [\mathrm{O}_{(K+1)\times K},\ HD'V_{22}^{-1}]. \qquad \text{(A31)}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial \lambda}{\partial \mathrm{vec}(V)'} \ =\ & \left[H[0_K,\ I_K]',\ \mathrm{O}_{(K+1)\times N}\right] \otimes \left[0_K',\ -e'V_{22}^{-1}\right] \\
& + \left[-\lambda_1',\ e'V_{22}^{-1}\right] \otimes \left[\mathrm{O}_{(K+1)\times K},\ HD'V_{22}^{-1}\right]. \qquad \text{(A32)}
\end{aligned}
$$

This completes the proof of the claim.

Using the expression of $\partial\lambda/\partial\phi'$, we can simplify the asymptotic variance of $\hat{\lambda}$ to

$$
V(\hat{\lambda}) \ =\ \sum_{j=-\infty}^{\infty} E[h_t(\phi) h_{t+j}(\phi)'], \qquad \text{(A33)}
$$

where

$$
\begin{aligned}
h_t(\phi) \ =\ & \frac{\partial \lambda}{\partial \phi'} r_t(\phi) \\
=\ & -\left[1,\ 0_K'\right]' \lambda_1'(f_t - \mu_1) - (H\left[1,\ \mu_1'\right]' e'V_{22}^{-1} + \mu_y HD'V_{22}^{-1})(R_t - \mu_2) \\
& + \mathrm{vec}\left([0_K',\ -e'V_{22}^{-1}][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} [0_K,\ I_K]H \\ \mathrm{O}_{N\times(K+1)} \end{bmatrix} \right) \\
& + \mathrm{vec}\left([\mathrm{O}_{(K+1)\times K},\ HD'V_{22}^{-1}][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} -\lambda_1 \\ V_{22}^{-1} e \end{bmatrix} \right) \\
=\ & \begin{bmatrix} -\lambda_1'(f_t - \mu_1) \\ 0_K \end{bmatrix} - H \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix} u_t - HD'V_{22}^{-1}(R_t - \mu_2)\mu_y \\
& - H[0_K,\ I_K]'(f_t - \mu_1)u_t - HD'V_{22}^{-1}(R_t - \mu_2)(f_t - \mu_1)'\lambda_1 + HD'V_{22}^{-1}(R_t - \mu_2)u_t \\
& + HD'V_{22}^{-1}V_{21}\lambda_1 + H[0_K,\ I_K]'V_{12}V_{22}^{-1}e - HD'V_{22}^{-1}e \\
=\ & -HD'V_{22}^{-1}(R_t - \mu_2)(y_t - u_t) - Hx_t u_t - \begin{bmatrix} y_t \\ 0_K \end{bmatrix} + \lambda \\
=\ & -HD'V_{22}^{-1}R_t y_t + H[D'V_{22}^{-1}(R_t - \mu_2) - x_t]u_t + \lambda. \qquad \text{(A34)}
\end{aligned}
$$

Equation (A34) follows from the fact that $HD'V_{22}^{-1}V_{21}\lambda_1 = [-\mu_1'\lambda_1,\ \lambda_1']'$ and $HD'V_{22}^{-1}\mu_2 = [1,\ 0_K']'$. 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

$$
h_t(\phi) = -HD'V_{22}^{-1}R_t y_t + \lambda. \qquad \text{(A35)}
$$

This completes the proof.

Proof of Lemma 2: Let $q_t = HD'V_{22}^{-1}(R_t - \mu_2)$, $w_t = D'V_{22}^{-1}(R_t - \mu_2) - x_t$ 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

$$HD'V_{22}^{-1}\mu_2 y_t = HD'V_{22}^{-1}D \begin{bmatrix} 1 \\ 0_K \end{bmatrix} y_t = \begin{bmatrix} y_t \\ 0_K \end{bmatrix}, \tag{A36}$$

we can write

$$h_t = q_t y_t - Hw_t u_t + z_t. \tag{A37}$$

It is straightforward to obtain $E[q_t] = 0_{K+1}$, $E[w_t] = -[1,\ \mu_1']'$, $E[z_t] = [\lambda_1'\mu_1,\ -\lambda_1']'$, $\mathrm{Var}[q_t] = H$,

$$\mathrm{Var}[w_t] = (\mu_2' V_{22}^{-1}\mu_2) \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix} \begin{bmatrix} 1 \\ \mu_1 \end{bmatrix}' + \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11} - V_{12}V_{22}^{-1}V_{21} \end{bmatrix}, \tag{A38}$$

$$\mathrm{Var}[z_t] = \begin{bmatrix} \sigma_y^2 & 0_K' \\ 0_K & \mathrm{O}_{K\times K} \end{bmatrix}. \tag{A39}$$

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 (A2) and (A3), we obtain

$$E[q_t w_t' y_t u_t] = \mathrm{O}_{(K+1)\times(K+1)}, \tag{A40}$$

$$E[w_t z_t' u_t] = \mathrm{O}_{(K+1)\times(K+1)}, \tag{A41}$$

$$E[w_t w_t' u_t^2] = \delta^2(E[w_t]E[w_t]' + (1+\kappa)\mathrm{Var}[w_t]), \tag{A42}$$

$$E[q_t q_t' y_t^2] = [\mu_y^2 + (1+\kappa)\sigma_y^2]H + 2(1+\kappa) \begin{bmatrix} -\mu_1'\lambda_1 \\ \lambda_1 \end{bmatrix} \begin{bmatrix} -\mu_1'\lambda_1 \\ \lambda_1 \end{bmatrix}', \tag{A43}$$

$$E[q_t z_t' y_t] = \begin{bmatrix} -\mu_1'\lambda_1 \\ \lambda_1 \end{bmatrix} \begin{bmatrix} \mu_y + \mu_1'\lambda_1 \\ -\lambda_1 \end{bmatrix}'. \tag{A44}$$

Using (A40) and (A41), we can write

$$V(\hat{\lambda}) = E[h_t h_t'] = E[q_t q_t' y_t^2] + E[q_t z_t' y_t] + E[z_t q_t' y_t] + E[z_t z_t'] + HE[w_t w_t' u_t^2]H. \tag{A45}$$

Substituting (A42)–(A44) in (A45) and after simplification, we obtain our expression of $V(\hat{\lambda})$. This completes the proof.

Proof of Lemma 3: Partition $B = D'V_{22}^{-1}D$ as

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \tag{A46}$$

where B_{11} is $(K_1+1)\times(K_1+1)$ and B_{22} is $K_2\times K_2$. We can write the difference of the squared HJ-distances of the two models as

$$\begin{aligned}
\delta_1^2-\delta_2^2 &= 1_N'V_{22}^{-1}DHD'V_{22}^{-1}1_N - 1_N'V_{22}^{-1}D\begin{bmatrix} B_{11}^{-1} & \mathrm{O}_{(K_1+1)\times K_2} \\ \mathrm{O}_{K_2\times(K_1+1)} & \mathrm{O}_{K_2\times K_2}\end{bmatrix}D'V_{22}^{-1}1_N \\
&= \lambda'B\lambda-\lambda'B\begin{bmatrix} B_{11}^{-1} & \mathrm{O}_{(K_1+1)\times K_2} \\ \mathrm{O}_{K_2\times(K_1+1)} & \mathrm{O}_{K_2\times K_2}\end{bmatrix}B\lambda \\
&= \lambda'B\lambda-\lambda'\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{21}B_{11}^{-1}B_{12}\end{bmatrix}\lambda \\
&= \lambda^{(2)\prime}(B_{22}-B_{21}B_{11}^{-1}B_{12})\lambda^{(2)}. \qquad \text{(A47)}
\end{aligned}$$

As D is assumed to be of full column rank, $B_{22}-B_{21}B_{11}^{-1}B_{12}$ is a positive definite matrix. Therefore, $\delta_1^2=\delta_2^2$ if and only if $\lambda^{(2)}=0_{K_2}$. This completes the proof.

Proof of Proposition 2: Let $z=\sqrt{T}V(\hat\lambda^{(2)})^{-\frac{1}{2}}\hat\lambda^{(2)}\overset{A}{\sim}N(0_{K_2},I_{K_2})$. From the proof of Lemma 3 and the fact that $A_2HA_2'=A_2B^{-1}A_2'=(B_{22}-B_{21}B_{11}^{-1}B_{12})^{-1}$, we can write

$$T(\hat\delta_1^2-\hat\delta_2^2)=z'V(\hat\lambda^{(2)})^{\frac{1}{2}}(B_{22}-B_{21}B_{11}^{-1}B_{12})V(\hat\lambda^{(2)})^{\frac{1}{2}}z=z'V(\hat\lambda^{(2)})^{\frac{1}{2}}(A_2HA_2')^{-1}V(\hat\lambda^{(2)})^{\frac{1}{2}}z. \qquad \text{(A48)}$$

Let $Q\Xi Q'$ be the eigenvalue decomposition of $V(\hat\lambda^{(2)})^{\frac{1}{2}}(A_2HA_2')^{-1}V(\hat\lambda^{(2)})^{\frac{1}{2}}$, where $\Xi=\mathrm{Diag}(\xi_1,\ldots,\xi_{K_2})$ is a diagonal matrix of the eigenvalues of $V(\hat\lambda^{(2)})^{\frac{1}{2}}(A_2HA_2')^{-1}V(\hat\lambda^{(2)})^{\frac{1}{2}}$, or equivalently the eigenvalues of $(A_2HA_2')^{-1}V(\hat\lambda^{(2)})$, and Q is a matrix of the corresponding eigenvectors. Writing $\tilde z=Q'z\overset{A}{\sim}N(0_{K_2},I_{K_2})$, we have

$$T(\hat\delta_1^2-\hat\delta_2^2)=\tilde z'\Xi\tilde z=\sum_{i=1}^{K_2}\xi_i\tilde z_i^2, \qquad \text{(A49)}$$

where $\tilde z_i^2\overset{A}{\sim}\chi_1^2$, $i=1,\ldots,K_2$, and they are asymptotically independent of each other. This completes the proof.

Proof of Lemma 4: When Y_t is i.i.d. multivariate elliptically distributed, we use (33) to obtain the asymptotic variance of $\hat\lambda^{(2)}$ as

$$\begin{aligned}
& A_2V(\hat\lambda)A_2' \\
=\ & [\mu_y^2+(1+\kappa)\sigma_y^2]A_2HA_2'+(1+2\kappa)\lambda^{(2)}\lambda^{(2)\prime}+\delta^2A_2H \\
\times & \left([1+(1+\kappa)\mu_2'V_{22}^{-1}\mu_2]\begin{bmatrix}1\\ \mu_1\end{bmatrix}\begin{bmatrix}1\\ \mu_1\end{bmatrix}'+\begin{bmatrix}0 & 0_K' \\ 0_K & (1+\kappa)(V_{11}-V_{12}V_{22}^{-1}V_{21})\end{bmatrix}\right)HA_2'. \qquad \text{(A50)}
\end{aligned}$$

Premultiplying (A50) by $(A_2 H A_2')^{-1}$ and using the fact that $A_2\lambda = \lambda^{(2)} = 0_{K_2}$ under the null hypothesis, we obtain (46). This completes the proof.

Proof of Proposition 3: Let

$$z = \sqrt{T}\begin{bmatrix} V_{22}^{-\frac{1}{2}}\bar{g}_{1T}(\hat{\eta}) \\ V_{22}^{-\frac{1}{2}}\bar{g}_{2T}(\hat{\lambda}) \end{bmatrix} \overset{A}{\sim} N(0_{2N}, V_z), \tag{A51}$$

where

$$V_z = \begin{bmatrix} V_{22}^{-\frac{1}{2}}G_1 S_{11} G_1' V_{22}^{-\frac{1}{2}} & V_{22}^{-\frac{1}{2}}G_1 S_{12} G_2' V_{22}^{-\frac{1}{2}} \\ V_{22}^{-\frac{1}{2}}G_2 S_{21} G_1' V_{22}^{-\frac{1}{2}} & V_{22}^{-\frac{1}{2}}G_2 S_{22} G_2' V_{22}^{-\frac{1}{2}} \end{bmatrix}. \tag{A52}$$

Note that $V_{22}^{-\frac{1}{2}}G_1$ and $V_{22}^{-\frac{1}{2}}G_1$ can be written as

$$\begin{aligned} V_{22}^{-\frac{1}{2}}G_1 &= V_{22}^{-\frac{1}{2}}[I_N - D_1(D_1'V_{22}^{-1}D_1)^{-1}D_1'V_{22}^{-1}] \\ &= [I_N - V_{22}^{-\frac{1}{2}}D_1(D_1'V_{22}^{-1}D_1)^{-1}D_1'V_{22}^{-\frac{1}{2}}]V_{22}^{-\frac{1}{2}} \\ &= P_1P_1'V_{22}^{-\frac{1}{2}}, \qquad\qquad (A53) \\ V_{22}^{-\frac{1}{2}}G_2 &= P_2P_2'V_{22}^{-\frac{1}{2}}. \qquad\qquad (A54) \end{aligned}$$

With these identities, we can write V_z as

$$\begin{bmatrix} P_1 & \mathrm{O}_{N\times n_2} \\ \mathrm{O}_{N\times n_1} & P_2 \end{bmatrix}\begin{bmatrix} P_1'V_{22}^{-\frac{1}{2}}S_{11}V_{22}^{-\frac{1}{2}}P_1 & P_1'V_{22}^{-\frac{1}{2}}S_{12}V_{22}^{-\frac{1}{2}}P_2 \\ P_2'V_{22}^{-\frac{1}{2}}S_{21}V_{22}^{-\frac{1}{2}}P_1 & P_2'V_{22}^{-\frac{1}{2}}S_{22}V_{22}^{-\frac{1}{2}}P_2 \end{bmatrix}\begin{bmatrix} P_1' & \mathrm{O}_{n_1\times N} \\ \mathrm{O}_{n_2\times N} & P_2' \end{bmatrix}. \tag{A55}$$

By defining $\tilde{z} = V_z^{-\frac{1}{2}}z \overset{A}{\sim} N(0_{2N}, I_{2N})$, we can write

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) = z'\begin{bmatrix} I_N & \mathrm{O}_{N\times N} \\ \mathrm{O}_{N\times N} & -I_N \end{bmatrix}z = \tilde{z}'V_z^{\frac{1}{2}}\begin{bmatrix} I_N & \mathrm{O}_{N\times N} \\ \mathrm{O}_{N\times N} & -I_N \end{bmatrix}V_z^{\frac{1}{2}}\tilde{z} = \sum_{i=1}^{n_1+n_2}\xi_i\tilde{z}_i^2, \tag{A56}$$

where $\tilde{z}_i^2 \overset{A}{\sim} \chi_1^2$, $i = 1, \ldots, K_2$, are asymptotically independent of each other and the ξ_i's are the nonzero eigenvalues of

$$\begin{aligned} &\begin{bmatrix} I_N & \mathrm{O}_{N\times N} \\ \mathrm{O}_{N\times N} & -I_N \end{bmatrix}V_z \\ = &\begin{bmatrix} P_1 & \mathrm{O}_{N\times n_2} \\ \mathrm{O}_{N\times n_1} & P_2 \end{bmatrix}\begin{bmatrix} P_1'V_{22}^{-\frac{1}{2}}S_{11}V_{22}^{-\frac{1}{2}}P_1 & P_1'V_{22}^{-\frac{1}{2}}S_{12}V_{22}^{-\frac{1}{2}}P_2 \\ -P_2'V_{22}^{-\frac{1}{2}}S_{21}V_{22}^{-\frac{1}{2}}P_1 & -P_2'V_{22}^{-\frac{1}{2}}S_{22}V_{22}^{-\frac{1}{2}}P_2 \end{bmatrix}\begin{bmatrix} P_1' & \mathrm{O}_{n_1\times N} \\ \mathrm{O}_{n_2\times N} & P_2' \end{bmatrix}, \end{aligned} \tag{A57}$$

or equivalently the eigenvalues of (64). This completes the proof.

Proof of Lemma 5: Let $r_{1t} = P_1' V_{22}^{-\frac{1}{2}} R_t$ and $r_{2t} = P_2' V_{22}^{-\frac{1}{2}} R_t$. We can write

$$
\begin{aligned}
P_1' V_{22}^{-\frac{1}{2}} g_{1t} &= P_1' V_{22}^{-\frac{1}{2}} (R_t y_{1t} - 1_N) = r_{1t} y_{1t} - P_1' V_{22}^{-\frac{1}{2}} D_1 \eta = r_{1t} y_{1t}, &\text{(A58)}\\
P_2' V_{22}^{-\frac{1}{2}} g_{2t} &= P_2' V_{22}^{-\frac{1}{2}} (R_t y_{2t} - 1_N) = r_{2t} y_{2t} - P_2' V_{22}^{-\frac{1}{2}} D_2 \lambda = r_{2t} y_{2t}. &\text{(A59)}
\end{aligned}
$$

It is straightforward to show that $E[r_{1t}] = 0_{n_1}$, $E[r_{2t}] = 0_{n_2}$, $\mathrm{Var}[r_{1t}] = I_{n_1}$, $\mathrm{Var}[r_{2t}] = I_{n_2}$, $E[r_{1t} r_{2t}'] = P_1' P_2$, $E[r_{1t} y_{1t}] = E[r_{1t} y_{2t}] = 0_{n_1}$, $E[r_{2t} y_{1t}] = E[r_{2t} y_{2t}] = 0_{n_2}$. With these moments, we can apply (A3) to obtain

$$
\begin{aligned}
P_1' V_{22}^{-\frac{1}{2}} S_{11} V_{22}^{-\frac{1}{2}} P_1 &= E[P_1' V_{22}^{-\frac{1}{2}} g_{1t} g_{1t}' V_{22}^{-\frac{1}{2}} P_1] = E[r_{1t} r_{1t}' y_{1t}^2] = [\mu_{y_1}^2 + (1+\kappa)\sigma_{y_1}^2] I_{n_1}, &\text{(A60)}\\
P_1' V_{22}^{-\frac{1}{2}} S_{12} V_{22}^{-\frac{1}{2}} P_2 &= E[r_{1t} r_{2t}' y_{1t} y_{2t}] = [\mu_{y_1}\mu_{y_2} + (1+\kappa)\sigma_{y_1,y_2}] P_1' P_2, &\text{(A61)}\\
P_2' V_{22}^{-\frac{1}{2}} S_{22} V_{22}^{-\frac{1}{2}} P_2 &= E[r_{2t} r_{2t}' y_{2t}^2] = [\mu_{y_2}^2 + (1+\kappa)\sigma_{y_2}^2] I_{n_2}. &\text{(A62)}
\end{aligned}
$$

This completes the proof.

Proof of Proposition 4: We first present an 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

$$
\frac{\partial \delta^2}{\partial \phi} = \begin{bmatrix} 2\mu_y \\ 2\lambda_1 \\ -V_{22}^{-1} e \end{bmatrix} \otimes \begin{bmatrix} 0_K \\ V_{22}^{-1} e \end{bmatrix}. \qquad \text{(A63)}
$$

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

$$
\begin{aligned}
\frac{\partial \delta^2}{\partial \mu_1} &= 2\lambda_1 \mu_2' V_{22}^{-1} e = 0_K, &\text{(A64)}\\
\frac{\partial \delta^2}{\partial \mu_2} &= 2\mu_y V_{22}^{-1} e. &\text{(A65)}
\end{aligned}
$$

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

$$
\frac{\partial \delta^2}{\partial \mathrm{vec}(V)'} = \frac{\partial e' V_{22}^{-1} e}{\partial \mathrm{vec}(V)'} = 2e' V_{22}^{-1} \frac{\partial e}{\partial \mathrm{vec}(V)'} + (e' \otimes e') \frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'}. \qquad \text{(A66)}
$$

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

$$
\begin{aligned}
2e' V_{22}^{-1} \frac{\partial e}{\partial \mathrm{vec}(V)'} &= 2e' V_{22}^{-1} \frac{\partial D\lambda}{\partial \mathrm{vec}(V)'}\\
&= 2e' V_{22}^{-1} \left[(\lambda' \otimes I_N) \frac{\partial \mathrm{vec}(D)}{\partial \mathrm{vec}(V)'} + D \frac{\partial \lambda}{\partial \mathrm{vec}(V)'} \right]\\
&= 2e' V_{22}^{-1} \left[(\lambda' \otimes I_N) \frac{\partial \mathrm{vec}(D)}{\partial \mathrm{vec}(V)'} \right]. &\text{(A67)}
\end{aligned}
$$

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

$$\begin{aligned} 2e'V_{22}^{-1}\frac{\partial e}{\partial \mathrm{vec}(V)'} &= 2e'V_{22}^{-1}(\lambda'\otimes I_N)\left(\begin{bmatrix} 0_K' & 0_N' \\ I_K & \mathrm{O}_{K\times N}\end{bmatrix}\otimes[\mathrm{O}_{N\times K},\ I_N]\right) \\ &= [2\lambda_1',\ 0_N']\otimes[0_K',\ e'V_{22}^{-1}]. \end{aligned} \tag{A68}$$

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

$$\begin{aligned} (e'\otimes e')\frac{\partial\mathrm{vec}(V_{22}^{-1})}{\partial\mathrm{vec}(V)'} &= (e'\otimes e')\frac{\partial\mathrm{vec}(V_{22}^{-1})}{\partial\mathrm{vec}(V_{22})'}\frac{\partial\mathrm{vec}(V_{22})}{\partial\mathrm{vec}(V)'} \\ &= -(e'\otimes e')(V_{22}^{-1}\otimes V_{22}^{-1})([\mathrm{O}_{N\times K},\ I_N]\otimes[\mathrm{O}_{N\times K},\ I_N]) \\ &= -[0_K',\ e'V_{22}^{-1}]\otimes[0_K',\ e'V_{22}^{-1}]. \end{aligned} \tag{A69}$$

Combining these two terms, we have

$$\frac{\partial\delta^2}{\partial\mathrm{vec}(V)} = \begin{bmatrix} 2\lambda_1 \\ -V_{22}^{-1}e\end{bmatrix}\otimes\begin{bmatrix} 0_K \\ V_{22}^{-1}e\end{bmatrix}. \tag{A70}$$

This completes the proof of the claim.

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

$$\begin{aligned} q_t(\phi) &= \left[\frac{\partial\delta^2}{\partial\phi}\right]' r_t(\phi) \\ &= 2\mu_y e'V_{22}^{-1}(R_t-\mu_2) + \\ &\quad \left(\begin{bmatrix} 2\lambda_1 \\ -V_{22}^{-1}e\end{bmatrix}'\otimes\begin{bmatrix} 0_K \\ V_{22}^{-1}e\end{bmatrix}'\right)\mathrm{vec}((Y_t-\mu)(Y_t-\mu)'-V) \\ &= 2\mu_y e'V_{22}^{-1}(R_t-\mu_2) + \mathrm{vec}\left(\begin{bmatrix} 0_K \\ V_{22}^{-1}e\end{bmatrix}'((Y_t-\mu)(Y_t-\mu)'-V)\begin{bmatrix} 2\lambda_1 \\ -V_{22}^{-1}e\end{bmatrix}\right) \\ &= 2\mu_y e'V_{22}^{-1}(R_t-\mu_2) + \\ &\quad e'V_{22}^{-1}(R_t-\mu_2)\left[2\lambda_1'(f_t-\mu_1) - e'V_{22}^{-1}(R_t-\mu_2)\right] + e'V_{22}^{-1}e - 2e'V_{22}^{-1}V_{21}\lambda_1 \\ &= 2u_t y_t - u_t^2 + \delta^2 - 2e'V_{22}^{-1}V_{21}\lambda_1, \end{aligned} \tag{A71}$$

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

$$q_t(\phi) = 2u_t y_t - u_t^2 + \delta^2. \tag{A72}$$

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

$$q_{1t}(\phi) = \left[\frac{\partial \delta_1^2}{\partial \phi}\right]' r_t(\phi) = 2u_{1t}y_{1t} - u_{1t}^2 + \delta_1^2, \tag{A73}$$

$$q_{2t}(\phi) = \left[\frac{\partial \delta_2^2}{\partial \phi}\right]' r_t(\phi) = 2u_{2t}y_{2t} - u_{2t}^2 + \delta_2^2. \tag{A74}$$

Now, using the delta method and equations (A9)–(A12), the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ when both models are misspecified is given by

$$\sqrt{T}(\hat{\delta}_1^2 - \hat{\delta}_2^2 - (\delta_1^2 - \delta_2^2)) \stackrel{A}{\sim} N\left(0, \left[\frac{\partial(\delta_1^2 - \delta_2^2)}{\partial \phi}\right]' S_0 \left[\frac{\partial(\delta_1^2 - \delta_2^2)}{\partial \phi}\right]\right). \tag{A75}$$

With the analytical expressions of $q_{1t}(\phi)$ and $q_{2t}(\phi)$, the asymptotic variance of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ can be written as

$$v_d = \text{Avar}[\hat{\delta}_1^2 - \hat{\delta}_2^2] = \sum_{j=-\infty}^{\infty} E[d_t(\phi) d_{t+j}(\phi)], \tag{A76}$$

where

$$d_t(\phi) = \left(\frac{\partial \delta_1^2}{\partial \phi} - \frac{\partial \delta_2^2}{\partial \phi}\right)' r_t(\phi) = q_{1t}(\phi) - q_{2t}(\phi). \tag{A77}$$

This completes the proof.

Proof of Lemma 6: When d_t is uncorrelated over time, we have

$$v_d = E[d_t^2] = E[q_{1t}^2] + E[q_{2t}^2] - 2E[q_{1t}q_{2t}]. \tag{A78}$$

When Y_t is i.i.d. multivariate elliptically distributed, $E[q_{1t}^2]$ and $E[q_{2t}^2]$ can be obtained using the proof in Lemma 1 and they are given by

$$E[q_{1t}^2] = 4[\mu_{y_1}^2 + (1+\kappa)\sigma_{y_1}^2]\delta_1^2 + (2+3\kappa)\delta_1^4, \tag{A79}$$

$$E[q_{2t}^2] = 4[\mu_{y_2}^2 + (1+\kappa)\sigma_{y_2}^2]\delta_2^2 + (2+3\kappa)\delta_2^4. \tag{A80}$$

Following the proof in Lemma 1, we can show that $E[u_{1t}] = E[u_{2t}] = 0$, $E[u_{1t}^2] = \delta_1^2$, $E[u_{2t}^2] = \delta_2^2$, $E[u_{1t}u_{2t}] = e_1' V_{22}^{-1} e_2 = \delta_{12}$, $E[u_{1t}y_{1t}] = E[u_{2t}y_{2t}] = 0$. Using the fact that $e_1 = D_1\eta - 1_N$, $e_2 = D_2\lambda - 1_N$, $D_1' V_{22}^{-1} e_1 = 0_{n_1}$ and $D_2' V_{22}^{-1} e_2 = 0_{n_2}$, we can show that

$$E[u_{1t}y_{2t}] = e_1' V_{22}^{-1} D_2 \lambda = e_1' V_{22}^{-1}(e_2 - e_1 + D_1\eta) = \delta_{12} - \delta_1^2, \tag{A81}$$

$$E[u_{2t}y_{1t}] = e_2' V_{22}^{-1} D_1 \eta = e_2' V_{22}^{-1}(e_1 - e_2 + D_2\lambda) = \delta_{12} - \delta_2^2. \tag{A82}$$

With these moments available, we can apply (A3) to show that

$$
\begin{aligned}
E[u_{1t}u_{2t}y_{1t}y_{2t}] &= [\mu_{y_1}\mu_{y_2} + (1+\kappa)\sigma_{y_1,y_2}]\delta_{12} + (1+\kappa)(\delta_{12} - \delta_1^2)(\delta_{12} - \delta_2^2), && \text{(A83)} \\
E[u_{1t}^2 u_{2t}y_{2t}] &= 2(1+\kappa)(\delta_{12} - \delta_1^2)\delta_{12}, && \text{(A84)} \\
E[u_{2t}^2 u_{1t}y_{1t}] &= 2(1+\kappa)(\delta_{12} - \delta_2^2)\delta_{12}, && \text{(A85)} \\
E[u_{1t}^2 u_{2t}^2] &= (1+\kappa)(\delta_1^2\delta_2^2 + 2\delta_{12}^2). && \text{(A86)}
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E[q_{1t}q_{2t}] &= 4E[u_{1t}u_{2t}y_{1t}y_{2t}] - 2E[u_{1t}^2u_{2t}y_{2t}] - 2E[u_{2t}^2u_{1t}y_{1t}] + E[u_{1t}^2u_{2t}^2] - \delta_1^2\delta_2^2 \\
&= 4[\mu_{y_1}\mu_{y_2} + (1+\kappa)\sigma_{y_1,y_2}]\delta_{12} + (2+3\kappa)\delta_1^2\delta_2^2 + 2(1+\kappa)(\delta_1^2\delta_2^2 - \delta_{12}^2), && \text{(A87)} \\
E[d_t^2] &= E[q_{1t}^2] + E[q_{2t}^2] - 2E[q_{1t}q_{2t}] \\
&= 4[\mu_{y_1}^2 + (1+\kappa)\sigma_{y_1}^2]\delta_1^2 + 4[\mu_{y_2}^2 + (1+\kappa)\sigma_{y_2}^2]\delta_2^2 - 8[\mu_{y_1}\mu_{y_2} + (1+\kappa)\sigma_{y_1,y_2}]\delta_{12} \\
&\quad + (2+3\kappa)(\delta_1^2 - \delta_2^2)^2 - 4(1+\kappa)(\delta_1^2\delta_2^2 - \delta_{12}^2) \\
&= 4(\mu_{y_1}^2\delta_1^2 + \mu_{y_2}^2\delta_2^2 - 2\mu_{y_1}\mu_{y_2}\delta_{12}) \\
&\quad + 4(1+\kappa)(\sigma_{y_1}^2\delta_1^2 + \sigma_{y_2}^2\delta_2^2 - 2\sigma_{y_1,y_2}\delta_{12} - \delta_1^2\delta_2^2 + \delta_{12}^2) + (2+3\kappa)(\delta_1^2 - \delta_2^2)^2. && \text{(A88)}
\end{aligned}
$$

This completes the proof.

References

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

Chen, Xiaohong, and Sydney C. Ludvigson, 2004, Land of addicts? An empirical investigation of habit-based asset pricing models, working paper, New York University.

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

Dittmar, Robert F., 2002, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *Journal of Finance* 57, 369–403.

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

Farnsworth, Heber, Wayne E. Ferson, David Jackson, and Steven Todd, 2002, Performance evaluation with stochastic discount factors, *Journal of Business* 75, 473–503.

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

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross section of stock returns, *Journal of Finance* 54, 1325–1360.

Hall, Alastair R., and Atsushi Inoue, 2003, The large sample behaviour of the generalized method of moments estimator in misspecified models, *Journal of Econometrics* 114, 361–394.

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

Hansen, Lars P., John Heaton, and Erzo G. J. Luttmer, 1995, Econometric evaluation of asset pricing models, *Review of Financial Studies* 8, 237–274.

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

Hodrick, Robert J., and Xiaoyan Zhang, 2001, Evaluating the specification errors of asset pricing models, *Journal of Financial Economics* 62, 327–376.

Jagannathan, Ravi, Keiichi Kubota, and Hitoshi Takehara, 1998, Relationship between labor-income risk and average return: Empirical evidence from the Japanese stock market, *Journal of Business* 71, 319–348.

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

Kan, Raymond, and Cesare Robotti, 2006, Specification tests of asset pricing models using excess returns, working paper, University of Toronto and Federal Reserve Bank of Atlanta.

Kan, Raymond, and Guofu Zhou, 2004, Hansen-Jagannathan distance: Geometry and exact distribution, working paper, University of Toronto and Washington University in St. Louis.

Lettau, Martin, and Sidney Ludvigson, 2001, Consumption, aggregate wealth and expected stock returns, *Journal of Finance* 56, 815–849.

Magnus, Jan R., and Heinz Neudecker, 1999, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edition, (Wiley, New York).

Maruyama, Yosihito, and Takashi Seo, 2003, Estimation of moment parameter in elliptical distributions, *Journal of the Japan Statistical Society* 33, 215–229.

Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory* (Wiley, New York).

Newey, Whitney, and Kenneth West, 1987, A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Rivers, Douglas, and Quang H. Vuong, 2002, Model selection tests for nonlinear dynamic models, *Econometrics Journal* 5, 1–39.

Vuong, Quang H., 1989, Likelihood ratio tests for model selection and non nested hypotheses, *Econometrica* 57, 307–333.

Table I
Summary of the Models

The table presents a summary of six monthly and seven quarterly asset pricing models. The monthly 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 quarterly models include the Cochrane (1996, COCH) investment model in addition to the previous models. The asset returns are the returns on the 25 Fama-French portfolios in excess of the T-bill rate and the gross T-bill return. Monthly data are from 1952/1 to 1997/12. Quarterly data are from 1953 Q1 to 1997 Q4. IP is the cyclical element in the industrial production index. GNP is the cyclical element in real GNP. CAY is the consumption to wealth ratio from Lettau and Ludvigson (2001). JAN is a dummy variable with a value of one for January (monthly models) or first quarter (quarterly models) 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 of δ based on $se(\hat{\delta})$.

Panel A: Monthly Models (Unscaled Factors)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
$\hat{\delta}$	0.390	0.429	0.387	0.298	0.322	0.287
$p(\delta = 0)$	0.000	0.000	0.000	0.318	0.000	0.073
se($\hat{\delta}$)	0.043	0.051	0.044	0.062	0.045	0.052
2.5% CI(δ)	0.305	0.329	0.300	0.176	0.234	0.185
97.5% CI(δ)	0.474	0.530	0.474	0.419	0.411	0.388
No. of par.	2	2	4	6	4	6

Panel B: Monthly Models (Factors Scaled by Lag IP)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
$\hat{\delta}$	0.353	0.389	0.314	0.256	0.296	0.270
$p(\delta = 0)$	0.017	0.038	0.062	0.566	0.009	0.074
se($\hat{\delta}$)	0.061	0.062	0.059	0.075	0.054	0.058
2.5% CI(δ)	0.234	0.268	0.198	0.109	0.191	0.157
97.5% CI(δ)	0.472	0.510	0.430	0.404	0.401	0.384
No. of par.	4	4	8	12	8	12

Panel C: Monthly Models (Factors Scaled by JAN)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
$\hat{\delta}$	0.366	0.366	0.272	0.286	0.285	0.229
$p(\delta = 0)$	0.000	0.055	0.672	0.103	0.093	0.638
se($\hat{\delta}$)	0.051	0.072	0.081	0.060	0.051	0.070
2.5% CI(δ)	0.267	0.226	0.113	0.167	0.185	0.092
97.5% CI(δ)	0.465	0.507	0.432	0.404	0.385	0.366
No. of par.	4	4	8	12	8	12

Panel D: Quarterly Models (Unscaled Factors)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
$\hat{\delta}$	0.620	0.618	0.604	0.550	0.625	0.537	0.516
$p(\delta = 0)$	0.000	0.001	0.001	0.015	0.000	0.001	0.019
se($\hat{\delta}$)	0.079	0.089	0.090	0.090	0.082	0.084	0.093
2.5% CI(δ)	0.465	0.443	0.428	0.374	0.465	0.372	0.333
97.5% CI(δ)	0.775	0.792	0.781	0.726	0.784	0.701	0.699
No. of par.	2	2	4	6	3	4	6

Panel E: Quarterly Models (Factors Scaled by Lag GNP)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
$\hat{\delta}$	0.599	0.612	0.561	0.503	0.558	0.449	0.427
$p(\delta = 0)$	0.001	0.000	0.002	0.141	0.107	0.547	0.373
se($\hat{\delta}$)	0.090	0.082	0.086	0.114	0.115	0.102	0.105
2.5% CI(δ)	0.423	0.451	0.393	0.279	0.333	0.249	0.222
97.5% CI(δ)	0.775	0.772	0.730	0.727	0.783	0.650	0.633
No. of par.	4	4	8	12	6	8	12

Panel F: Quarterly Models (Factors Scaled by Lag CAY)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
$\hat{\delta}$	0.613	0.607	0.591	0.515	0.623	0.531	0.500
$p(\delta = 0)$	0.000	0.000	0.002	0.096	0.000	0.001	0.011
se($\hat{\delta}$)	0.082	0.088	0.097	0.103	0.083	0.085	0.099
2.5% CI(δ)	0.453	0.434	0.401	0.312	0.459	0.366	0.307
97.5% CI(δ)	0.773	0.780	0.781	0.717	0.786	0.697	0.694
No. of par.	4	4	8	12	6	8	12

Panel G: Quarterly Models (Factors Scaled by JAN)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
$\hat{\delta}$	0.563	0.581	0.486	0.375	0.509	0.508	0.402
$p(\delta = 0)$	0.001	0.000	0.774	0.980	0.434	0.005	0.759
se($\hat{\delta}$)	0.085	0.086	0.136	0.165	0.113	0.082	0.121
2.5% CI(δ)	0.396	0.412	0.219	0.051	0.288	0.348	0.165
97.5% CI(δ)	0.730	0.750	0.753	0.699	0.731	0.668	0.639
No. of par.	4	4	8	12	6	8	12

Table II
Estimates and t-ratios of Parameters in Various Stochastic Discount Factor Models under Correctly Specified and Misspecified Models: Unscaled Factors

The table presents the estimation results of monthly and quarterly asset pricing models with unscaled factors. The asset returns are the returns on the 25 Fama-French portfolios in excess of the T-bill rate and the gross T-bill return. Monthly data are from 1952/1 to 1997/12. Quarterly data are from 1953 Q1 to 1997 Q4. We report parameter estimates $\hat{\lambda}$, t-ratios under correctly specified models (t-ratio$_{cs}$) and model misspecification robust t-ratios (t-ratio$_m$).

Panel A: Monthly Models

	CAPM		**C-CAPM**	
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_0$	$\hat{\lambda}_{cg}$
Estimate	1.02	−3.77	1.09	−46.02
t-ratio$_{cs}$	75.14	−3.31	20.15	−1.90
t-ratio$_m$	47.56	−3.32	12.08	−1.08

	JW				**CAMP**					
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\lambda}_0$	$\hat{\lambda}_{rvw}$	$\hat{\lambda}_{clab}$	$\hat{\lambda}_{div}$	$\hat{\lambda}_{rtb}$	$\hat{\lambda}_{trm}$
Estimate	0.78	−3.12	−2.91	52.54	−1.06	1.26	13.44	65.79	93.61	−70.14
t-ratio$_{cs}$	1.74	−2.26	−0.07	1.06	−0.82	0.45	0.33	1.94	0.22	−2.51
t-ratio$_m$	0.81	−1.73	−0.03	0.54	−0.64	0.45	0.25	1.50	0.17	−2.03

	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	−5.36	−1.04	−9.97	−0.33	−2.98	−4.58	−10.01	33.62	−65.79
t-ratio$_{cs}$	48.37	−4.42	−0.61	−5.29	−0.57	−1.59	−1.73	−3.91	2.67	−0.96
t-ratio$_m$	40.83	−4.43	−0.61	−5.26	−0.37	−1.36	−1.49	−3.86	1.70	−0.68

Panel B: Quarterly Models

	CAPM		C-CAPM		JW			
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_0$	$\hat{\lambda}_{cg}$	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$
Estimate	1.02	−2.43	1.38	−67.00	0.61	−0.44	−59.99	63.71
t-ratio$_{cs}$	34.69	−2.21	7.63	−2.39	0.79	−0.24	−1.14	1.38
t-ratio$_m$	18.55	−2.22	4.84	−1.46	0.33	−0.13	−0.65	0.58

	CAMP						COCH		
	$\hat{\lambda}_0$	$\hat{\lambda}_{rvw}$	$\hat{\lambda}_{clab}$	$\hat{\lambda}_{div}$	$\hat{\lambda}_{rtb}$	$\hat{\lambda}_{trm}$	$\hat{\lambda}_0$	$\hat{\lambda}_{gnr}$	$\hat{\lambda}_{gr}$
Estimate	0.23	−0.02	10.44	27.77	−21.81	−55.52	0.93	8.99	−7.02
t-ratio$_{cs}$	0.22	−0.01	0.64	1.05	−0.08	−2.54	10.30	1.19	−2.00
t-ratio$_m$	0.15	−0.01	0.42	0.70	−0.06	−1.88	5.40	0.67	−1.33

	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	1.11	−3.53	−0.60	−6.79	1.21	−4.89	−0.60	−6.13	−20.38	122.01
t-ratio$_{cs}$	21.75	−2.86	−0.41	−4.03	2.35	−2.94	−0.37	−3.41	−1.78	1.56
t-ratio$_m$	16.94	−2.89	−0.41	−4.04	1.54	−2.56	−0.36	−3.31	−1.16	0.97

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

The table presents Wald tests that the SDF parameters of the scaled factors are jointly equal to zero. The asset returns are the returns on the 25 Fama-French portfolios in excess of the T-bill rate and the gross T-bill return. Monthly data are from 1952/1 to 1997/12. Quarterly data are from 1953 Q1 to 1997 Q4. 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.

Panel A: Monthly Models (Factors Scaled by Lag IP)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
Wald(cs)	8.64	7.52	12.35	4.58	5.84	2.80
p-value	(0.013)	(0.023)	(0.015)	(0.599)	(0.212)	(0.834)
Wald(m)	2.84	4.71	7.54	2.12	2.66	1.05
p-value	(0.241)	(0.095)	(0.110)	(0.908)	(0.616)	(0.984)

Panel B: Monthly Models (Factors Scaled by JAN)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
Wald(cs)	6.22	6.08	11.05	1.57	6.67	4.23
p-value	(0.045)	(0.048)	(0.026)	(0.954)	(0.154)	(0.645)
Wald(m)	5.80	3.19	9.64	0.94	3.76	3.67
p-value	(0.055)	(0.203)	(0.047)	(0.988)	(0.439)	(0.721)

Panel C: Quarterly Models (Factors Scaled by Lag GNP)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
Wald(cs)	3.23	1.33	6.22	3.16	5.41	5.37	6.47
p-value	(0.199)	(0.514)	(0.183)	(0.788)	(0.144)	(0.252)	(0.373)
Wald(m)	0.80	0.28	3.22	1.86	2.88	3.85	4.90
p-value	(0.670)	(0.867)	(0.522)	(0.932)	(0.410)	(0.427)	(0.556)

Panel D: Quarterly Models (Factors Scaled by Lag CAY)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
Wald(cs)	1.72	2.19	1.80	3.25	0.34	0.94	2.23
p-value	(0.422)	(0.335)	(0.773)	(0.777)	(0.952)	(0.918)	(0.897)
Wald(m)	0.83	0.64	0.95	1.82	0.10	0.30	0.94
p-value	(0.661)	(0.725)	(0.918)	(0.936)	(0.992)	(0.990)	(0.988)

Panel E: Quarterly Models (Factors Scaled by JAN)

Model	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
Wald(cs)	17.77	9.12	8.00	5.64	10.06	4.45	6.46
p-value	(0.000)	(0.010)	(0.092)	(0.464)	(0.018)	(0.348)	(0.374)
Wald(m)	11.23	4.12	5.88	4.15	5.06	2.45	4.41
p-value	(0.004)	(0.128)	(0.209)	(0.656)	(0.167)	(0.654)	(0.622)

Table IV
Tests of Equality of Squared HJ-Distances

The table presents pairwise tests of equality of the squared HJ-distances of the monthly and quarterly asset pricing models with unscaled and scaled factors. The asset returns are the returns on the 25 Fama-French portfolios in excess of the T-bill rate and the gross T-bill return. Monthly data are from 1952/1 to 1997/12. Quarterly data are from 1953 Q1 to 1997 Q4. The scaling variables are Lag IP and JAN for monthly models and Lag GNP, Lag CAY and JAN for quarterly models. 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.

Panel A: Monthly Models (Unscaled Factors)

Unscaled	Unscaled: **C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
CAPM	−0.007	0.002	0.063	0.048	0.070
	(0.707)	(0.870)	(**0.035**)	(**0.000**)	(**0.007**)
C-CAPM		0.015	0.057	0.057	0.095
		(0.546)	(0.124)	(**0.025**)	(**0.015**)
JW			0.061	0.046	0.068
			(**0.048**)	(**0.033**)	(**0.018**)
CAMP				−0.015	0.006
				(0.605)	(0.842)
FF3					0.022
					(0.167)

Panel B: Monthly Models (Factors Scaled by Lag IP)

Lag IP	Lag IP: **C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
CAPM	0.010	0.026	0.059	0.037	0.052
	(0.783)	(0.560)	(0.103)	(0.170)	(0.732)
C-CAPM		0.032	0.095	0.029	0.073
		(0.476)	(0.074)	(0.387)	(0.154)
JW			0.033	0.011	0.026
			(0.415)	(0.732)	(0.396)
CAMP				−0.022	−0.007
				(0.488)	(0.855)
FF3					0.015
					(0.891)

Panel C: Monthly Models (Factors Scaled by JAN)

	JAN				
JAN	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
CAPM	0.021 (0.621)	0.060 (0.221)	0.052 (0.091)	0.053 (0.132)	0.081 (0.358)
C-CAPM		0.038 (0.558)	0.020 (0.708)	0.029 (0.535)	0.071 (0.200)
JW			−0.008 (0.862)	−0.007 (0.884)	0.022 (0.617)
CAMP				0.000 (0.989)	0.029 (0.453)
FF3					0.029 (0.600)

Panel D: Quarterly Models (Unscaled Factors)

	Unscaled					
Unscaled	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	0.003 (0.961)	0.019 (0.737)	0.082 (0.156)	−0.006 (0.854)	0.096 (**0.000**)	0.117 (**0.044**)
C-CAPM		0.016 (0.839)	0.079 (0.331)	−0.009 (0.875)	0.093 (0.142)	0.115 (0.095)
JW			0.063 (0.370)	−0.025 (0.627)	0.077 (0.270)	0.099 (0.231)
CAMP				−0.088 (0.182)	0.014 (0.817)	0.036 (0.603)
COCH					0.102 (**0.032**)	0.124 (**0.049**)
FF3						0.021 (0.517)

Panel E: Quarterly Models (Factors Scaled by Lag GNP)

	Lag GNP					
Lag GNP	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	−0.015 (0.854)	0.044 (0.822)	0.106 (0.304)	0.048 (0.649)	0.157 (**0.038**)	0.176 (0.367)
C-CAPM		0.059 (0.427)	0.121 (0.256)	0.063 (0.552)	0.172 (0.086)	0.191 (0.053)
JW			0.062 (0.496)	0.004 (0.971)	0.113 (0.241)	0.133 (0.124)
CAMP				−0.058 (0.576)	0.051 (0.631)	0.071 (0.489)
COCH					0.109 (0.389)	0.129 (0.303)
FF3						0.019 (0.945)

Panel F: Quarterly Models (Factors Scaled by Lag CAY)

	Lag CAY					
Lag CAY	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	0.007 (0.895)	0.026 (0.919)	0.110 (0.148)	−0.012 (0.761)	0.093 (0.127)	0.125 (0.622)
C-CAPM		0.019 (0.793)	0.103 (0.248)	−0.019 (0.734)	0.086 (0.176)	0.118 (0.133)
JW			0.084 (0.307)	−0.038 (0.490)	0.067 (0.364)	0.099 (0.303)
CAMP				−0.122 (0.129)	−0.017 (0.832)	0.015 (0.875)
COCH					0.105 (**0.043**)	0.137 (0.069)
FF3						0.032 (0.885)

Panel G: Quarterly Models (Factors Scaled by JAN)

	JAN					
JAN	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	−0.021 (0.551)	0.081 (0.737)	0.176 (0.157)	0.057 (0.494)	0.059 (0.371)	0.155 (0.546)
C-CAPM		0.101 (0.372)	0.197 (0.151)	0.078 (0.328)	0.079 (0.172)	0.176 (0.096)
JW			0.096 (0.508)	−0.023 (0.867)	−0.022 (0.848)	0.074 (0.618)
CAMP				−0.119 (0.459)	−0.118 (0.335)	−0.021 (0.883)
COCH					0.001 (0.987)	0.098 (0.417)
FF3						0.096 (0.428)

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

The table compares the performance of monthly and quarterly asset pricing models with unscaled factors with the performance of the corresponding models with scaled factors. The asset returns are the returns on the 25 Fama-French portfolios in excess of the T-bill rate and the gross T-bill return. Monthly data are from 1952/1 to 1997/12. Quarterly data are from 1953 Q1 to 1997 Q4. The scaling variables are Lag IP and JAN for monthly models and Lag GNP, Lag CAY and JAN for quarterly models. 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.

Panel A: Monthly Models (Unscaled vs. Scaled by Lag IP)

	Lag IP					
Unscaled	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
CAPM	0.027	0.026	0.053	0.086	0.064	0.079
	(0.347)	(0.469)	(0.451)	(**0.017**)	(0.133)	(0.615)
C-CAPM	0.023	0.033	0.065	0.128	0.062	0.106
	(0.509)	(0.199)	(0.083)	(**0.009**)	(**0.031**)	(**0.021**)
JW	0.025	0.018	0.051	0.084	0.062	0.077
	(0.448)	(0.650)	(0.306)	(**0.026**)	(**0.030**)	(**0.018**)
CAMP	−0.036	−0.024	−0.010	0.023	0.001	0.016
	(0.316)	(0.598)	(0.784)	(0.905)	(0.981)	(0.646)
FF3	−0.021	−0.024	0.005	0.038	0.016	0.031
	(0.525)	(0.517)	(0.873)	(0.253)	(0.736)	(0.937)
FF5	−0.042	−0.062	−0.017	0.016	−0.006	0.009
	(0.285)	(0.168)	(0.652)	(0.691)	(0.836)	(0.994)

Panel B: Monthly Models (Unscaled vs. Scaled by JAN)

	JAN					
Unscaled	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
CAPM	0.018	0.043	0.078	0.070	0.071	0.099
	(0.325)	(0.343)	(0.286)	(**0.019**)	(0.172)	(0.375)
C-CAPM	0.028	0.050	0.088	0.070	0.079	0.121
	(0.322)	(0.152)	(0.092)	(0.063)	(**0.035**)	(**0.010**)
JW	0.016	0.035	0.076	0.068	0.069	0.097
	(0.492)	(0.467)	(0.167)	(**0.029**)	(**0.020**)	(**0.009**)
CAMP	−0.045	−0.007	0.014	0.007	0.007	0.036
	(0.186)	(0.897)	(0.741)	(0.994)	(0.836)	(0.361)
FF3	−0.030	−0.007	0.030	0.022	0.023	0.051
	(0.173)	(0.880)	(0.484)	(0.422)	(0.614)	(0.705)
FF5	−0.052	−0.045	0.008	0.000	0.001	0.030
	(0.112)	(0.395)	(0.854)	(0.987)	(0.973)	(0.813)

Panel C: Quarterly Models (Unscaled vs. Scaled by Lag GNP)

	Lag GNP						
Unscaled	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	0.025	0.010	0.069	0.131	0.073	0.182	0.201
	(0.677)	(0.850)	(0.865)	(0.158)	(0.441)	(0.147)	(0.471)
C-CAPM	0.022	0.007	0.066	0.128	0.070	0.179	0.199
	(0.781)	(0.871)	(0.392)	(0.229)	(0.491)	(0.079)	(**0.048**)
JW	0.006	−0.009	0.050	0.112	0.054	0.163	0.183
	(0.930)	(0.905)	(0.758)	(0.270)	(0.607)	(0.112)	(0.066)
CAMP	−0.057	−0.072	−0.013	0.049	−0.009	0.100	0.120
	(0.486)	(0.353)	(0.852)	(0.951)	(0.937)	(0.267)	(0.154)
COCH	0.031	0.016	0.075	0.137	0.079	0.188	0.208
	(0.634)	(0.748)	(0.219)	(0.145)	(0.398)	(**0.048**)	(**0.028**)
FF3	−0.071	−0.086	−0.027	0.035	−0.023	0.086	0.105
	(0.313)	(0.129)	(0.683)	(0.706)	(0.811)	(0.399)	(0.786)
FF5	−0.092	−0.108	−0.049	0.013	−0.045	0.065	0.084
	(0.244)	(0.108)	(0.491)	(0.886)	(0.660)	(0.410)	(0.712)

Panel D: Quarterly Models (Unscaled vs. Scaled by Lag CAY)

	Lag CAY						
Unscaled	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	0.009	0.016	0.035	0.119	−0.004	0.102	0.134
	(0.782)	(0.766)	(0.963)	(0.116)	(0.919)	(0.336)	(0.726)
C-CAPM	0.006	0.013	0.032	0.116	−0.006	0.099	0.131
	(0.922)	(0.740)	(0.678)	(0.232)	(0.911)	(0.131)	(0.109)
JW	−0.010	−0.003	0.016	0.100	−0.022	0.083	0.115
	(0.838)	(0.966)	(0.948)	(0.206)	(0.668)	(0.253)	(0.207)
CAMP	−0.073	−0.066	−0.047	0.037	−0.085	0.020	0.052
	(0.243)	(0.387)	(0.548)	(0.971)	(0.200)	(0.758)	(0.503)
COCH	0.015	0.022	0.041	0.125	0.003	0.108	0.140
	(0.699)	(0.686)	(0.453)	(0.114)	(0.993)	(**0.033**)	(0.060)
FF3	−0.087	−0.080	−0.061	0.023	−0.100	0.006	0.038
	(0.050)	(0.179)	(0.382)	(0.769)	(0.052)	(0.989)	(0.986)
FF5	−0.109	−0.102	−0.083	0.001	−0.121	−0.016	0.016
	(0.068)	(0.125)	(0.323)	(0.987)	(0.064)	(0.701)	(0.994)

Panel E: Quarterly Models (Unscaled vs. Scaled by JAN)

	JAN						
Unscaled	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	0.067	0.047	0.148	0.244	0.124	0.126	0.222
	(0.142)	(0.383)	(0.637)	(0.076)	(0.158)	(0.204)	(0.429)
C-CAPM	0.065	0.044	0.145	0.241	0.122	0.123	0.220
	(0.322)	(0.218)	(0.251)	(0.101)	(0.206)	(0.115)	(0.072)
JW	0.048	0.028	0.129	0.225	0.106	0.107	0.203
	(0.502)	(0.726)	(0.429)	(0.120)	(0.334)	(0.186)	(0.099)
CAMP	−0.014	−0.035	0.066	0.162	0.043	0.044	0.141
	(0.842)	(0.659)	(0.580)	(0.658)	(0.694)	(0.530)	(0.204)
COCH	0.074	0.053	0.154	0.250	0.131	0.132	0.228
	(0.156)	(0.328)	(0.156)	(0.078)	(0.124)	(**0.037**)	(**0.039**)
FF3	−0.029	−0.049	0.052	0.148	0.028	0.030	0.126
	(0.563)	(0.372)	(0.649)	(0.258)	(0.748)	(0.799)	(0.720)
FF5	−0.050	−0.071	0.030	0.126	0.007	0.008	0.105
	(0.426)	(0.280)	(0.807)	(0.368)	(0.941)	(0.886)	(0.647)

Table VI
Tests of Equality of Squared HJ-Distances: Factors Scaled by Different Conditioning Variables

The table compares the performance of monthly and quarterly conditional asset pricing models with factors scaled by different conditioning variables. The asset returns are the returns on the 25 Fama-French portfolios in excess of the T-bill rate and the gross T-bill return. Monthly data are from 1952/1 to 1997/12. Quarterly data are from 1953 Q1 to 1997 Q4. The scaling variables are Lag IP and JAN for monthly models and Lag GNP, Lag CAY and JAN for quarterly models. 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 of potential model misspecification.

Panel A: Monthly Models (Scaled by Lag IP vs. Scaled by JAN)

	JAN					
Lag IP	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**FF3**	**FF5**
CAPM	−0.009	0.027	0.050	0.043	0.043	0.072
	(0.805)	(0.612)	(0.323)	(0.253)	(0.263)	(0.102)
C-CAPM	−0.004	0.017	0.055	0.037	0.046	0.088
	(0.917)	(0.719)	(0.356)	(0.429)	(0.309)	(0.093)
JW	−0.035	−0.015	0.025	0.017	0.018	0.046
	(0.353)	(0.770)	(0.598)	(0.655)	(0.627)	(0.226)
CAMP	−0.068	−0.078	−0.008	−0.016	−0.015	0.013
	(0.076)	(0.196)	(0.863)	(0.638)	(0.695)	(0.772)
FF3	−0.046	−0.012	0.014	0.006	0.007	0.035
	(0.102)	(0.788)	(0.761)	(0.837)	(0.811)	(0.308)
FF5	−0.061	−0.056	−0.001	−0.009	−0.008	0.021
	(0.086)	(0.334)	(0.981)	(0.796)	(0.792)	(0.449)

Panel B: Quarterly Models (Scaled by Lag GNP vs. Scaled by Lag CAY)

	Lag CAY						
Lag GNP	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	−0.017	−0.009	0.009	0.094	−0.029	0.077	0.109
	(0.788)	(0.913)	(0.891)	(0.319)	(0.664)	(0.265)	(0.211)
C-CAPM	−0.001	0.006	0.025	0.109	−0.013	0.092	0.124
	(0.980)	(0.858)	(0.741)	(0.229)	(0.799)	(0.126)	(0.120)
JW	−0.060	−0.053	−0.034	0.050	−0.072	0.033	0.065
	(0.286)	(0.463)	(0.596)	(0.536)	(0.257)	(0.622)	(0.388)
CAMP	−0.122	−0.115	−0.096	−0.012	−0.134	−0.029	0.003
	(0.199)	(0.282)	(0.377)	(0.903)	(0.154)	(0.758)	(0.974)
COCH	−0.064	−0.057	−0.038	0.046	−0.076	0.029	0.061
	(0.509)	(0.588)	(0.719)	(0.709)	(0.368)	(0.765)	(0.574)
FF3	−0.173	−0.166	−0.148	−0.063	−0.186	−0.080	−0.048
	(0.056)	(0.108)	(0.148)	(0.547)	(0.055)	(0.292)	(0.578)
FF5	−0.193	−0.186	−0.167	−0.083	−0.205	−0.100	−0.067
	(**0.034**)	(0.063)	(0.099)	(0.406)	(**0.032**)	(0.198)	(0.404)

Panel C: Quarterly Models (Scaled by Lag GNP vs. Scaled by JAN)

	JAN						
Lag GNP	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	0.042	0.021	0.123	0.219	0.099	0.101	0.197
	(0.572)	(0.787)	(0.340)	(0.139)	(0.297)	(0.207)	(0.088)
C-CAPM	0.057	0.037	0.138	0.234	0.115	0.116	0.212
	(0.358)	(0.489)	(0.259)	(0.098)	(0.244)	(0.096)	(0.069)
JW	−0.002	−0.022	0.079	0.175	0.056	0.057	0.153
	(0.983)	(0.774)	(0.496)	(0.219)	(0.580)	(0.469)	(0.179)
CAMP	−0.063	−0.084	0.017	0.113	−0.006	−0.005	0.092
	(0.509)	(0.401)	(0.895)	(0.423)	(0.961)	(0.962)	(0.439)
COCH	−0.005	−0.026	0.075	0.171	0.052	0.053	0.150
	(0.954)	(0.773)	(0.569)	(0.270)	(0.689)	(0.609)	(0.243)
FF3	−0.115	−0.136	−0.034	0.062	−0.058	−0.056	0.040
	(0.223)	(0.174)	(0.802)	(0.665)	(0.593)	(0.482)	(0.702)
FF5	−0.134	−0.155	−0.053	0.042	−0.077	−0.075	0.021
	(0.142)	(0.114)	(0.691)	(0.764)	(0.491)	(0.358)	(0.839)

Panel D: Quarterly Models (Scaled by Lag CAY vs. Scaled by JAN)

	JAN						
Lag CAY	**CAPM**	**C-CAPM**	**JW**	**CAMP**	**COCH**	**FF3**	**FF5**
CAPM	0.059	0.038	0.139	0.235	0.116	0.117	0.214
	(0.200)	(0.493)	(0.203)	(0.089)	(0.202)	(**0.036**)	(**0.046**)
C-CAPM	0.052	0.031	0.132	0.228	0.109	0.110	0.207
	(0.439)	(0.583)	(0.291)	(0.123)	(0.263)	(0.132)	(0.077)
JW	0.033	0.012	0.113	0.209	0.090	0.091	0.188
	(0.651)	(0.881)	(0.251)	(0.159)	(0.402)	(0.265)	(0.143)
CAMP	−0.052	−0.072	0.029	0.125	0.006	0.007	0.103
	(0.558)	(0.460)	(0.813)	(0.342)	(0.963)	(0.933)	(0.397)
COCH	0.071	0.050	0.152	0.247	0.128	0.130	0.226
	(0.193)	(0.378)	(0.165)	(0.084)	(0.201)	(0.051)	(**0.047**)
FF3	−0.034	−0.055	0.046	0.142	0.023	0.024	0.121
	(0.550)	(0.369)	(0.697)	(0.292)	(0.801)	(0.601)	(0.224)
FF5	−0.067	−0.087	0.014	0.110	−0.009	−0.008	0.088
	(0.394)	(0.274)	(0.915)	(0.463)	(0.925)	(0.914)	(0.369)