Chi-squared tests for evaluation and comparison of asset pricing models

Nikolay Gospodinov[a], Raymond Kan[b], and Cesare Robotti[c]

[a]*Corresponding author, Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec H3G 1M8, Canada, and CIREQ, Phone: (514) 848-2424, Fax: (514) 848-4536, E-mail:* `nikolay.gospodinov@concordia.ca`

[b]*Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario M5S 3E6, Canada, E-mail:* `kan@chass.utoronto.ca`

[c]*Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St., Atlanta, GA 30309, United States, and EDHEC Risk Institute, France, E-mail:* `cesare.robotti@atl.frb.org`

Abstract

This paper presents a general statistical framework for estimation, testing and comparison of asset pricing models using the unconstrained distance measure of Hansen and Jagannathan (1997). The limiting results cover both linear and nonlinear models that could be correctly specified or misspecified. We propose modified versions of the existing model selection tests and new pivotal specification and model comparison tests with improved finite-sample properties. In addition, we provide formal tests of multiple model comparison. The excellent size and power properties of the proposed tests are demonstrated using simulated data from linear and nonlinear asset pricing models.

JEL Classification: C12, C13, G12

Keywords: Asset pricing models; Hansen-Jagannathan distance; Model selection; Model misspecification

1. Introduction

It is common for financial economists to view all asset pricing models only as approximations to reality. Although these models are likely to be misspecified, it is still useful to empirically evaluate the degree of misspecification and their relative pricing performance using actual data. In their seminal paper, Hansen and Jagannathan (1997, HJ hereafter) propose measures of model misspecification that are now routinely used for parameter estimation, specification testing and comparison of competing asset pricing models. The unconstrained (constrained) HJ-distance measures the distance between the stochastic discount factor (SDF) of a proposed model and the set of (nonnegative) admissible stochastic discount factors. But despite the recent advances in developing the appropriate econometric theory for comparing asset pricing models based on the HJ-distance, a general statistical procedure for model selection in this context is still incomplete. As a result, many researchers are still ranking alternative models by comparing their corresponding sample HJ-distances without any use of a formal statistical criterion that takes into account the sampling and model misspecification uncertainty. In this paper, we provide a comprehensive statistical framework for estimation, evaluation and comparison of linear and nonlinear (potentially misspecified) asset pricing models based on the unconstrained HJ-distance. Given some unappealing theoretical properties of the constrained HJ-distance (Gospodinov, Kan and Robotti, 2011), we do not consider explicitly the sample constrained HJ-distance but the generality of our analytical framework allows us to easily extend the main results for the unconstrained HJ-distance that we derive in this paper to its constrained analog (a detailed econometric analysis of the sample constrained HJ-distance is available from the authors upon request). Our framework could also be used to study the statistical properties of other measures of model misspecification.

The econometric methodology for using the unconstrained HJ-distance as a specification test for linear and nonlinear models is developed by Hansen, Heaton and Luttmer (1995), Jagannathan and Wang (1996) and Parker and Julliard (2005). Kan and Robotti (2009) provide a statistical procedure for comparing linear asset pricing models based on the unconstrained HJ-distance. Furthermore, Kan and Robotti (2009) propose standard errors for the SDF parameter estimates that are valid for misspecified models. Almeida and Garcia (2012) consider estimation and inference in SDF models based on more general minimum discrepancy measures of model misspecification. The objective of this paper is to provide a unifying framework for improved statistical inference, specification testing and (pairwise and multiple) model comparison based on the sample HJ-distances of competing linear and nonlinear asset pricing models.

Our main contributions can be summarized as follows. First, we propose a new Lagrange multiplier test for correct model specification. This new specification test is asymptotically chi-squared distributed and enjoys improved finite-sample properties compared to the specification test based on the HJ-distance. Second, we derive the non-degenerate joint asymptotic distribution of the parameters and the Lagrange multipliers which are not always asymptotically normally distributed.[1] Third, we improve upon the model selection testing procedures in the existing literature. This is achieved by incorporating the appropriate null hypotheses which leads to simpler model comparison tests that require the estimation of far fewer parameters than the existing testing procedures. While the practice of not imposing the null hypotheses in constructing the test statistics can be justified based on asymptotic arguments, it produces the undesirable outcome of comparing test statistics that are positive by construction (as in the nested model case discussed in Section 3) to distributions that can take on negative values. Our modifications are new to the literature on model selection tests and lead to substantial power improvements in situations with many test assets (moment conditions). Importantly, the proposed tests can be easily adapted to other setups including the quasi-likelihood framework of Vuong (1989). Fourth, we propose pivotal (asymptotically chi-squared distributed) versions of the model comparison tests that are easier to implement and analyze than their weighted chi-squared counterparts. The chi-squared tests appear to possess excellent finite-sample properties and their improved power proves to be particularly important in situations where they are used as pre-tests in sequential testing procedures for non-nested models. Fifth, we develop tests for multiple model comparison as well as fast numerical algorithms for computing their asymptotic p-values.[2] Finally, we investigate the finite-sample performance of the proposed inference procedures using simulated data from some popular linear and nonlinear asset pricing models.

The rest of the paper is organized as follows. Section 2 introduces the population and sample HJ-distance problems. It also presents the basic assumptions and the asymptotic properties of the sample HJ-distance and its corresponding estimators. Section 3 develops our pairwise and

[1] A more complete analysis of this problem is presented in Gospodinov, Kan and Robotti (2012).

[2] The Matlab codes for implementing all the statistical tests and procedures discussed in the paper are available upon request.

multiple model comparison tests based on the sample HJ-distances. Section 4 studies the finite-sample properties of our testing procedures using Monte Carlo simulation experiments. Section 5 concludes. Proofs are collected in the Appendix and some additional theoretical and simulation results are provided in an online appendix available on the authors' websites.

The paper adopts the following notation. Let $\overset{A}{\sim}$ stand for "asymptotically distributed as," χ^2_p signify a chi-squared random variable with p degrees of freedom, $|w| = (w'w)^{\frac{1}{2}}$ denote the Euclidean norm of a vector w and $||A|| = \sqrt{\text{tr}(A'A)}$ be the Euclidean or Frobenius norm of a matrix A, where $\text{tr}(\cdot)$ is the trace operator. Finally, let $Z = (Z_1, \ldots, Z_s)'$ be a vector of s independent standard normal random variables, and let $\xi = (\xi_1, \ldots, \xi_s)'$ be a vector of s real numbers. Then, $F_s(\xi) = \sum_{i=1}^{s} \xi_i Z_i^2$ denotes a random variable which is distributed as a weighted sum of s independent chi-squared random variables with one degree of freedom.

2. Estimation and model evaluation based on the HJ-distance

2.1. Population HJ-distance

Let x_t denote a vector of payoffs of n test assets at the end of period t and q_{t-1} be the corresponding costs of these n assets at the end of period $t-1$ with $E[q_{t-1}] \neq 0_n$.[3] This setup can accommodate both gross and excess returns on test assets as well as payoffs of trading strategies that are based on time-varying information. In addition, we assume that $U = E[x_t x_t']$ is nonsingular so that none of the test assets is redundant.

Let m_t represent an admissible SDF at time t and let $\mathcal{M}$ be the set of all admissible SDFs. An SDF m_t is admissible if it prices the test assets correctly, i.e.,[4]

$$E[x_t m_t] = E[q_{t-1}]. \tag{1}$$

Suppose that $y_t(\gamma)$ is a candidate SDF at time t that depends on a k-vector of unknown parameters $\gamma \in \Gamma$, where Γ is the parameter space of γ. An asset pricing model is correctly specified if there exists a $\gamma \in \Gamma$ such that $y_t(\gamma) \in \mathcal{M}$. The model is misspecified if $y_t(\gamma) \notin \mathcal{M}$ for all $\gamma \in \Gamma$.

[3]When $E[q_{t-1}] = 0_n$, the mean of the SDF cannot be identified and researchers have to choose some normalization of the SDF (see, for example, Kan and Robotti, 2008).

[4]Strictly speaking, the set of admissible SDFs should be defined in terms of conditional expectations. In this paper, we use an unconditional version of the fundamental pricing equation. This, in principle, could be justified by incorporating conditioning information through scaled payoffs (see, for example, Section 8.1 in Cochrane, 2005).

When the asset pricing model is misspecified, we are interested in measuring the degree of model misspecification. HJ suggest using

$$\delta = \min_{\gamma \in \Gamma} \min_{m_t \in \mathcal{M}} \left(E[(y_t(\gamma) - m_t)^2] \right)^{\frac{1}{2}} \tag{2}$$

as a misspecification measure of $y_t(\gamma)$. We refer to δ as the HJ-distance measure.

Instead of solving the above primal problem to obtain δ, HJ suggest that it is sometimes more convenient to solve the following dual problem:

$$\delta^2 = \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} E[y_t(\gamma)^2 - (y_t(\gamma) - \lambda' x_t)^2 - 2\lambda' q_{t-1}], \tag{3}$$

where λ is an n-vector of Lagrange multipliers.

Let $\theta = [\gamma' \, , \lambda']'$ and denote by $\theta^* = [\gamma^{*\prime} \, , \lambda^{*\prime}]'$ the pseudo-true value that solves the population dual problem in (3):

$$\theta^* = \arg \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} E[\phi_t(\theta)], \tag{4}$$

where $\phi_t(\theta) \equiv y_t(\gamma)^2 - m_t(\theta)^2 - 2\lambda' q_{t-1}$ and $m_t(\theta) \equiv y_t(\gamma) - \lambda' x_t$. Note that $y_t(\gamma^*)$ prices the n test assets correctly if the vector of pricing errors is zero, i.e.,

$$e(\gamma^*) = E[x_t y_t(\gamma^*) - q_{t-1}] = 0_n. \tag{5}$$

In this case, $y_t(\gamma^*) \in \mathcal{M}$, $\lambda^* = 0_n$ and we refer to γ^* as the true value.[5]

By rearranging the dual problem in (3), it is easy to show that $\lambda^* = U^{-1} e(\gamma^*)$ and

$$\delta^2 = e(\gamma^*)' U^{-1} e(\gamma^*). \tag{6}$$

While the quadratic form in the pricing errors in (6) has been widely used in the empirical finance literature for parameter estimation, model evaluation and comparison, the potential usefulness of the information regarding model specification contained in the Lagrange multipliers has been largely ignored. In this paper, we explicitly exploit this information to develop a Lagrange multiplier model specification test.

[5] The optimization problem in (4) bears strong resemblance to the structure of the Euclidean likelihood problem defined as $\min_\gamma \max_\lambda E[\mathbf{h}(\lambda' e(\gamma))]$ with $\mathbf{h}(\varsigma) = -\frac{1}{2}\varsigma^2 - \varsigma$. Other choices of $\mathbf{h}(\varsigma)$ give rise to some popular members of the class of generalized empirical likelihood (GEL) estimators. See Almeida and Garcia (2012) for further discussion of the class of GEL estimators in the context of asset pricing models. While the analysis in this paper can be easily extended to GEL estimators, we choose to present our main results for the HJ-distance measure given its popularity in empirical asset pricing, nice economic (maximum pricing error) interpretation and computational simplicity (closed-form solution for the Lagrange multipliers).

2.2. Sample estimators and assumptions

Since the population HJ-distance of a model and its associated parameters are unobservable, they have to be estimated from the data. The estimator of θ^* in (4) is obtained as the solution to the sample dual problem

$$\hat{\theta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\lambda} \end{bmatrix} = \arg\min_{\gamma\in\Gamma} \max_{\lambda\in\Re^n} \frac{1}{T}\sum_{t=1}^{T}\phi_t(\theta). \tag{7}$$

Alternatively, let $e_t(\gamma) = x_t y_t(\gamma) - q_{t-1}$, $e_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T} e_t(\gamma)$ and $\hat{U} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t'$. Then, the estimator $\hat{\theta} = (\hat{\gamma}', \hat{\lambda}')'$ can be obtained sequentially as

$$\hat{\gamma} = \arg\min_{\gamma\in\Gamma}\ e_T(\gamma)'\hat{U}^{-1}e_T(\gamma), \tag{8}$$

and $\hat{\lambda} = \hat{U}^{-1}e_T(\hat{\gamma})$.

In the following analysis, we appeal to the empirical process theory to derive the limiting behavior of the estimators and test statistics under correctly specified and misspecified models. The main regularity conditions for the consistency and the asymptotic distribution theory are listed below. They include restrictions on the dependence of the data, identification conditions for the pseudo-true values and some standard assumptions for deriving the limiting distributions.

ASSUMPTION A. *Assume that (i) $\phi_t(\theta)$ is m-dependent,* (ii) *the parameter space Θ is compact, (iii) $\phi_t(\theta)$ is continuous in $\theta \in \Theta$ almost surely, (iv) $|\phi_t(\theta_1) - \phi_t(\theta_2)| \leq A_t |\theta_1 - \theta_2|\ \forall\, \theta_1, \theta_2 \in \Theta$, where A_t is a bounded random variable that satisfies $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E[|A_t|^{2+\omega}] < \infty$ for some $\omega > 0$, (v) $\sup_{\theta\in\Theta} E[|\phi_t(\theta)|^{2+\omega}] < \infty$ for some $\omega > 0$, (vi) the population dual problem (4) has a unique solution θ^* which is in the interior of Θ.*

Assumptions A(i)–A(v) ensure the stochastic equicontinuity of $\phi_t(\theta)$ (see Andrews, 1994 and Stock and Wright, 2000) and imply that

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\phi_t(\theta) - E[\phi_t(\theta)]\right| \xrightarrow{p} 0. \tag{9}$$

The m-dependence can be relaxed although results for empirical processes with more general dependence structure are still limited (see, for instance, Andrews, 1993). Assumption A(vi) is an identification condition that ensures the uniqueness of the pseudo-true value θ^*. The uniform convergence in (9) and Assumption A(vi) are sufficient for establishing the consistency of $\hat{\theta}$.

Let

$$H \equiv \begin{bmatrix} H_{\gamma\gamma} & H'_{\lambda\gamma} \\ H_{\lambda\gamma} & H_{\lambda\lambda} \end{bmatrix} = \frac{\partial^2 E[\phi_t(\theta^*)]}{\partial\theta\partial\theta'} \tag{10}$$

and

$$M \equiv \begin{bmatrix} M_{\gamma\gamma} & M'_{\lambda\gamma} \\ M_{\lambda\gamma} & M_{\lambda\lambda} \end{bmatrix} = \lim_{T\to\infty} \operatorname{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\phi_t(\theta^*)}{\partial\theta}\right]. \tag{11}$$

The next assumption provides conditions for the existence and uniform convergence of the limiting matrices in (10) and (11).

Assumption B. *Let $\mathcal{N}(\theta^*)$ be a neighborhood of θ^*. Assume that (i) $E[\phi_t(\theta)]$ is twice continuously differentiable in θ for $\theta \in \mathcal{N}(\theta^*)$, (ii) $\sup_{\theta\in\mathcal{N}(\theta^*)}\left\|\frac{\partial^2 E[\phi_t(\theta)]}{\partial\theta\partial\theta'}\right\| < \infty$ and H is of full rank, (iii) M is a finite positive definite matrix when $\delta > 0$, or $M_{\lambda\lambda}$ is a finite positive definite matrix when $\delta = 0$.*

Following Andrews (1994), let $h_t(\theta) = \partial\phi_t(\theta)/\partial\theta$ and define the empirical process $\sqrt{T}\bar{v}_T(\theta)$, where

$$\bar{v}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} v_t(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T}(h_t(\theta) - E[h_t(\theta)]). \tag{12}$$

Assumption C below ensures that $\sqrt{T}\bar{v}_T(\theta)$ obeys the central limit theorem.

Assumption C. *Assume that $v_t(\theta)$ satisfies the conditions: (i) $|v_t(\theta_1) - v_t(\theta_2)| \leq B_t|\theta_1 - \theta_2|$ $\forall$ $\theta_1, \theta_2 \in \Theta$, where B_t is a bounded random variable that satisfies $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E[|B_t|^{2+\omega}] < \infty$ for some $\omega > 0$, (ii) $\sup_{\theta\in\Theta} E[|v_t(\theta)|^{2+\omega}] < \infty$ for some $\omega > 0$.*

It proves useful for our subsequent analysis to provide explicit expressions for the partitioned matrices in (10) and (11). Let

$$C = E\left[u_t\frac{\partial^2 y_t(\gamma^*)}{\partial\gamma\partial\gamma'}\right], \tag{13}$$

$$D = E\left[x_t\frac{\partial y_t(\gamma^*)}{\partial\gamma'}\right], \tag{14}$$

$$S = \sum_{j=-\infty}^{\infty} E\left[e_t(\gamma^*)e_{t+j}(\gamma^*)'\right], \tag{15}$$

where $u_t = e(\gamma^*)'U^{-1}x_t$. Using the fact that

$$\frac{\partial\phi_t(\theta^*)}{\partial\gamma} = 2[y_t(\gamma^*) - m_t(\theta^*)]\frac{\partial y_t(\gamma^*)}{\partial\gamma}, \tag{16}$$

$$\frac{\partial\phi_t(\theta^*)}{\partial\lambda} = 2[x_t m_t(\theta^*) - q_{t-1}], \tag{17}$$

and under Assumptions A, B and C, we can write

$$
\begin{aligned}
H_{\gamma\gamma} &= 2E\left[(y_t(\gamma^*) - m_t(\theta^*))\frac{\partial^2 y_t(\gamma^*)}{\partial\gamma\partial\gamma'}\right] = 2C, && (18)\\
H_{\lambda\gamma} &= 2E\left[x_t\frac{\partial y_t(\gamma^*)}{\partial\gamma'}\right] = 2D, && (19)\\
H_{\lambda\lambda} &= -2E\left[x_t x_t'\right] \equiv -2U, && (20)
\end{aligned}
$$

and

$$
M_{\lambda\lambda} = 4\sum_{j=-\infty}^{\infty} E\left[(x_t m_t(\theta^*) - q_{t-1})(x_{t+j}m_{t+j}(\theta^*) - q_{t+j-1})'\right]. \tag{21}
$$

If the model is correctly specified, we have $\lambda^* = 0_n$ and $y_t(\gamma^*) = m_t(\theta^*)$. Then, it follows that $H_{\gamma\gamma} = 0_{k\times k}$ and $M_{\lambda\lambda} = 4S$. Furthermore, we have $\partial\phi_t(\theta^*)/\partial\gamma = 0_k$ which yields $M_{\gamma\gamma} = 0_{k\times k}$ and $M_{\lambda\gamma} = 0_{n\times k}$. This is the reason why Assumption B(iii) requires only $M_{\lambda\lambda}$, and not M, to be positive definite when $\delta = 0$.

2.3. Asymptotic results

Let P denote an $n\times(n-k)$ orthonormal matrix whose columns are orthogonal to $U^{-\frac{1}{2}}D$ and

$$
\Pi = \begin{bmatrix} I_k & 0_{k\times n} \\ 0_{(n-k)\times k} & P'U^{\frac{1}{2}} \end{bmatrix}. \tag{22}
$$

The following lemma establishes the asymptotic normality of the estimates of the SDF parameters and of the Lagrange multipliers for misspecified and correctly specified models.

Lemma 1. *Under Assumptions A, B and C,*

(a) *if* $\delta > 0$,

$$
\sqrt{T}(\hat\theta - \theta^*) \overset{A}{\sim} N(0_{n+k}, \Sigma), \tag{23}
$$

where $\Sigma = \sum_{j=-\infty}^{\infty} E[l_t l_{t+j}']$ *with* $l_t = [l_{1t}',\ l_{2t}']'$ *given by*

$$
\begin{aligned}
l_{1t} &= (C + D'U^{-1}D)^{-1}\left[D'U^{-1}e_t(\gamma^*) + \left\{\frac{\partial y_t(\gamma^*)}{\partial\gamma} - D'U^{-1}x_t\right\}u_t\right], && (24)\\
l_{2t} &= U^{-1}[Dl_{1t} - e_t(\gamma^*) + x_t u_t]. && (25)
\end{aligned}
$$

(b) *if* $\delta = 0$,

$$
\sqrt{T}\Pi(\hat\theta - \theta^*) \overset{A}{\sim} N(0_n, \tilde\Sigma), \tag{26}
$$

where $\tilde{\Sigma} = \sum_{j=-\infty}^{\infty} E[\tilde{l}_t \tilde{l}'_{t+j}]$ *with* $\tilde{l}_t = [\tilde{l}'_{1t},\ \tilde{l}'_{2t}]'$ *given by*

$$\tilde{l}_{1t} = (D'U^{-1}D)^{-1}D'U^{-1}e_t(\gamma^*), \tag{27}$$

$$\tilde{l}_{2t} = -P'U^{-\frac{1}{2}}e_t(\gamma^*). \tag{28}$$

The covariance matrices Σ and $\tilde{\Sigma}$ in Lemma 1 can be consistently estimated using a nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator (see, for example, Newey and West, 1987) based on the sample analogs of (24)–(25) and (27)–(28). Tests of parameter restrictions based on the Wald or distance metric statistics can be easily developed from the results in Lemma 1.

While the estimator $\hat{\gamma}$ is asymptotically normally distributed under both the null and alternative hypotheses, the asymptotic distribution of some linear combinations of $\hat{\lambda}$ is not always normal when $\delta = 0$. To illustrate this, note that when $\delta = 0$, the expression for l_{2t} in (25) simplifies to

$$l_{2t} = [U^{-1}D(D'U^{-1}D)^{-1}D' - I_n]U^{-1}e_t(\gamma^*). \tag{29}$$

Since $D'l_{2t} = 0_k$, the asymptotic covariance matrix of $\sqrt{T}\hat{\lambda}$ is singular when $\delta = 0$. This implies that for a nonzero vector α in the span of the column space of D, $\sqrt{T}\alpha'\hat{\lambda}$ is not asymptotically normal because $\alpha' l_{2t} = 0$.[6] More generally, Gospodinov, Kan and Robotti (2012) show that when α is in the span of the column space of D, then

$$T\alpha'\hat{\lambda} \xrightarrow{d} -z_1' z_2, \tag{30}$$

where z_1 and z_2 are jointly normally distributed vectors of random variables.

The possible breakdown in the asymptotic normality of $\sqrt{T}\hat{\lambda}$ is the reason why in Lemma 1 we report the asymptotic distribution of $\sqrt{T}P'U^{\frac{1}{2}}\hat{\lambda}$ which always has a non-degenerate asymptotic normal distribution. It is also interesting to note that premultiplying $\hat{\lambda}$ by $P'U^{\frac{1}{2}}$ is similar in spirit to the decomposition of Sowell (1996) in which the n-vector of normalized population moment conditions $U^{-\frac{1}{2}}e_t(\gamma^*)$ is decomposed into k identifying restrictions used for the estimation of γ that characterize the space of identifying restrictions and $(n-k)$ over-identifying restrictions that characterize the space of over-identifying restrictions. This type of decomposition provides the

[6] Hansen, Heaton and Luttmer (1995, Proposition 4.1) present the asymptotic distribution of the estimated Lagrange multipliers when the SDF does not have parameters. In this case, $\sqrt{T}\hat{\lambda}$ has a non-degenerate asymptotic normal distribution even when $\delta = 0$.

basis for establishing the limiting distribution of the test for over-identifying restrictions. Next, we use the asymptotic result for $\sqrt{T}P'U^{\frac{1}{2}}\hat{\lambda}$ in part (b) of Lemma 1 to develop a Lagrange multiplier (LM) test for model specification.

Theorem 1. *Let $\hat{S}$ be a nonparametric HAC estimator of S and $\hat{P}$ be an orthonormal matrix whose columns are orthogonal to $\hat{U}^{-\frac{1}{2}}\hat{D}$ with $\hat{D} = \frac{1}{T}\sum_{t=1}^{T}\left[x_t\frac{\partial y_t(\hat{\gamma})}{\partial \gamma'}\right]$. Define the LM statistic as*

$$LM_{\hat{\lambda}} \equiv T\hat{\lambda}'\hat{U}^{\frac{1}{2}}\hat{P}\left(\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\right)^{-1}\hat{P}'\hat{U}^{\frac{1}{2}}\hat{\lambda}. \tag{31}$$

Then, under $H_0 : \delta = 0$ and Assumptions A, B and C,

$$LM_{\hat{\lambda}} \overset{A}{\sim} \chi^2_{n-k}. \tag{32}$$

Since $\delta = 0$ if and only if $\lambda = 0_n$, the LM test in Theorem 1 provides an alternative model specification test that is based on the distance of the Lagrange multipliers from zero.[7] Similar arguments can be used for developing an asymptotically equivalent specification test on the model's pricing errors. The existing specification test in the literature is based on measuring the distance of the sample squared HJ-distance, $\hat{\delta}^2 = e_T(\hat{\gamma})'\hat{U}^{-1}e_T(\hat{\gamma})$, from zero and is asymptotically distributed as a weighted sum of independent chi-squared random variables with one degree of freedom (Jagannathan and Wang, 1996; Parker and Julliard, 2005). Unlike the LM test in Theorem 1, the HJ-distance test is asymptotically non-pivotal and it tends to overreject substantially under the null when the number of test assets n is large relative to the number of time series observations T (see Ahn and Gadarowski, 2004).

3. Model selection tests

In this section, we first present model selection tests of two competing models. Our analysis is similar in spirit to the model selection methodology of Vuong (1989), Rivers and Vuong (2002), Golden (2003), Marcellino and Rossi (2008), and Li, Xu and Zhang (2010), but we provide several improvements upon the results available in the literature.[8] First, since for nested models the

[7] A similar test is used by Newey (1985) for generic GMM problems and by Smith (1997) and Imbens, Spady and Johnson (1998) in the context of GEL estimation of moment condition models.

[8] Kitamura (2000) and Chen, Hong and Shum (2007), among others, develop test procedures for comparing misspecified models within the generalized empirical likelihood framework. Almeida and Garcia (2012) provide a discussion on how these tests can be adapted to asset pricing model selection.

HJ-distance of the nesting model is always smaller than the HJ-distance of the nested model, the difference between the sample HJ-distances of two nested models should be compared with a distribution that only takes on positive values. However, the existing tests do not impose this restriction and are expected to exhibit finite-sample distortions and loss of power. In contrast, we take into account the nested model structure and develop model comparison tests with this desirable property. Second, we develop chi-squared versions of the model comparison tests for nested and non-nested models that are easier to implement than the weighted chi-squared tests. In addition to model selection tests of two competing models, we provide multiple model comparison tests that allow us to compare a benchmark model with a set of alternative models in terms of their HJ-distances.

3.1. Pairwise model comparison

Define models $\mathcal{F} = \{y_t^{\mathcal{F}}(\gamma_{\mathcal{F}});\ \gamma_{\mathcal{F}} \in \Gamma_{\mathcal{F}}\}$ and $\mathcal{G} = \{y_t^{\mathcal{G}}(\gamma_{\mathcal{G}});\ \gamma_{\mathcal{G}} \in \Gamma_{\mathcal{G}}\}$, where $\gamma_{\mathcal{F}}$ and $\gamma_{\mathcal{G}}$ are k_1 and k_2 parameter vectors, respectively, and $\Gamma_{\mathcal{F}}$ and $\Gamma_{\mathcal{G}}$ denote their corresponding parameter spaces. The population squared HJ-distances for models $\mathcal{F}$ and $\mathcal{G}$ are given by

$$\delta_{\mathcal{F}}^2 = \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{F}}} E[\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}})], \tag{33}$$

$$\delta_{\mathcal{G}}^2 = \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} E[\phi_t^{\mathcal{G}}(\theta_{\mathcal{G}})], \tag{34}$$

where $\lambda_{\mathcal{F}}$ and $\lambda_{\mathcal{G}}$ are the vectors of Lagrange multipliers for models $\mathcal{F}$ and $\mathcal{G}$, respectively, $\theta_{\mathcal{F}} = [\gamma_{\mathcal{F}}',\ \lambda_{\mathcal{F}}']'$, $\theta_{\mathcal{G}} = [\gamma_{\mathcal{G}}',\ \lambda_{\mathcal{G}}']'$, $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}) \equiv y_t^{\mathcal{F}}(\gamma_{\mathcal{F}})^2 - [m_t^{\mathcal{F}}(\theta_{\mathcal{F}})]^2 - 2\lambda_{\mathcal{F}}' q_{t-1}$, $\phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}) \equiv y_t^{\mathcal{G}}(\gamma_{\mathcal{G}})^2 - [m_t^{\mathcal{G}}(\theta_{\mathcal{G}})]^2 - 2\lambda_{\mathcal{G}}' q_{t-1}$, $m_t^{\mathcal{F}}(\theta_{\mathcal{F}}) \equiv y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}) - \lambda_{\mathcal{F}}' x_t$, and $m_t^{\mathcal{G}}(\theta_{\mathcal{G}}) \equiv y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}) - \lambda_{\mathcal{G}}' x_t$. Denote by $\theta_{\mathcal{F}}^* = [\gamma_{\mathcal{F}}^{*\prime},\ \lambda_{\mathcal{F}}^{*\prime}]'$ and $\theta_{\mathcal{G}}^* = [\gamma_{\mathcal{G}}^{*\prime},\ \lambda_{\mathcal{G}}^{*\prime}]'$ the pseudo-true parameters of models $\mathcal{F}$ and $\mathcal{G}$, respectively. For nested models, we have $\mathcal{F} \subset \mathcal{G}$ or $\mathcal{G} \subset \mathcal{F}$. On the other hand, if $\mathcal{F} \not\subset \mathcal{G}$, and $\mathcal{G} \not\subset \mathcal{F}$, we refer to $\mathcal{F}$ and $\mathcal{G}$ as non-nested models. Non-nested models can be further decomposed into strictly non-nested (if $\mathcal{F} \cap \mathcal{G} = \emptyset$) and overlapping (if $\mathcal{F} \cap \mathcal{G} \neq \emptyset$).

A simple way of testing $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$ is suggested by Hansen, Heaton and Luttmer (1995, pp. 255–256) who establish that the difference between the sample squared HJ-distances of models $\mathcal{F}$ and $\mathcal{G}$ under $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$ is asymptotically normally distributed:

$$\sqrt{T}(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} N(0, \sigma_d^2), \tag{35}$$

where $\sigma_d^2 = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]$ and $d_t = \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) - \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$.

It is important to emphasize that the result in (35) holds only if $\sigma_d^2 \neq 0$. To determine whether the use of the normal test in (35) is appropriate, one could do a pre-test of $H_0 : \sigma_d^2 = 0$ (see, for example, Rivers and Vuong, 2002, Golden, 2003 and Marcellino and Rossi, 2008). Alternatively, since $\sigma_d^2 = 0$ if and only if $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$, one could do a pre-test of $H_0 : \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$. This is the approach that we pursue in this paper.[9] There are two possible reasons for $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$: (i) the two SDFs are equal, i.e., $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, or (ii) the two SDFs are different but correctly specified, so that $\delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$, which implies $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*) = 0$.

For nested models, the test in (35) should not be performed because under $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$, we must have $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. The reason is that, in general, the larger model has a smaller HJ-distance and the only case in which the two models can have the same HJ-distance is when $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. Therefore, we should only perform a test of $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ for nested models. For overlapping models, it is possible that either $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ or $\delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$, so we need to conduct two pre-tests before using the test in (35). Finally, for strictly non-nested models, we cannot have $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ and, as a result, we only have to test $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ before using the test in (35). We discuss the nested and non-nested cases in the following two subsections.

3.1.1. Nested models

For nested models, σ_d^2 is zero by construction under the null of equal HJ-distances. Therefore, the normal test in (35) cannot be used. In addition, for nested models, $\delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$ if and only if $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, so we can simply test $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$.

Without loss of generality, we assume $\mathcal{F} \subset \mathcal{G}$. Suppose that the null hypothesis $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ can be written as a parametric restriction of the form $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$ for model $\mathcal{G}$ against $H_1 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) \neq 0_{k_2-k_1}$, where $\psi(\cdot)$ is a twice continuously differentiable function in its argument. Define

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma_{\mathcal{G}}'} \tag{36}$$

as a $(k_2 - k_1) \times k_2$ derivative matrix of the parametric restrictions $\psi_{\mathcal{G}}$. For many models of interest,

[9] Other inference procedures such as subsampling and m-out-of-n ($m < n$) bootstrap could potentially deal with the degeneracy of the asymptotic distribution that occurs at $\sigma_d^2 = 0$ and provide correct inference. To the best of our knowledge, the asymptotic validity of these procedures in our context has not been established in the literature. Also, since both of these resampling procedures reduce the number of effective time series observations per moment condition, it is not clear whether they can provide any finite-sample size and power improvements given the excellent finite-sample properties of our asymptotic tests reported in Section 4 below.

$y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}})$ when a subset of the parameters of model $\mathcal{G}$ is equal to zero (or a constant vector c). In this case, we can rearrange the parameters such that $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_1)\times k_1},\ I_{k_2-k_1}]\gamma_{\mathcal{G}} - c$. Then, $\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_1)\times k_1},\ I_{k_2-k_1}]$ is a selector matrix that picks only the part of the parameter vector $\gamma_{\mathcal{G}}$ that is not contained in model $\mathcal{F}$. Also, let $\Sigma_{\hat{\gamma}_{\mathcal{G}}}$ be the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{G}}$ given by the upper left block of Σ in part (a) of Lemma 1, $\Psi_*^{\mathcal{G}} \equiv \Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, and $\tilde{H}_{\mathcal{G}} = (C^{\mathcal{G}} + D^{\mathcal{G}\prime}U^{-1}D^{\mathcal{G}})^{-1}$, where the matrices C, D, and U are defined in Section 2. Finally, define the Wald test statistic

$$Wald_{\hat{\psi}_{\mathcal{G}}} = T\hat{\psi}_{\mathcal{G}}'(\hat{\Psi}^{\mathcal{G}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{G}}}\hat{\Psi}^{\mathcal{G}\prime})^{-1}\hat{\psi}_{\mathcal{G}}, \tag{37}$$

where $\hat{\psi}_{\mathcal{G}} = \psi_{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})$, $\hat{\Psi}^{\mathcal{G}} = \Psi^{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})$, and $\hat{\Sigma}_{\hat{\gamma}_{\mathcal{G}}}$ is a consistent estimator of $\Sigma_{\hat{\gamma}_{\mathcal{G}}}$.

Theorem 2 below presents the asymptotic distribution of $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ and the Wald test under $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$.

Theorem 2. *Suppose that Assumptions A, B and C hold and $\mathcal{F} \subset \mathcal{G}$. Then, under $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$,*

(a)

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_2-k_1}(\xi), \tag{38}$$

where the ξ_i's are the eigenvalues of the matrix

$$(\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1}\Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}, \tag{39}$$

(b)

$$Wald_{\hat{\psi}_{\mathcal{G}}} \overset{A}{\sim} \chi^2_{k_2-k_1}. \tag{40}$$

Part (a) of Theorem 2 shows that, under $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, the limiting distribution of $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ is a linear combination of $k_2 - k_1$ chi-squared random variables with one degree of freedom and weights (estimated from the sample version of the matrix (39)) that are guaranteed to be positive. The Wald test in part (b) of Theorem 2 offers an alternative way of testing the equality of two nested SDFs by testing directly $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$. This Wald test is asymptotically pivotal and is easier to implement than the test in part (a).

3.1.2. Non-nested models

We first consider the test of $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. Since for strictly non-nested models $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*)$ cannot be equal to $y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ by definition ($\mathcal{F} \cap \mathcal{G} = \emptyset$), we focus only on overlapping models. It is well known that for linear models, the equality of the SDFs implies some restrictions on the parameter vectors (see, for example, Lien and Vuong, 1987 and Kan and Robotti, 2009). Similar restrictions can also be obtained for nonlinear models. Let $y_t^{\mathcal{H}}(\gamma_{\mathcal{H}})$ be the SDF of model $\mathcal{H}$, where $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ and $\gamma_{\mathcal{H}}$ is a k_3-vector. Therefore, $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ implies $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ and $y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$. Suppose that $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ and $y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ can be written as a parametric restriction of the form $H_0 : \psi_{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_3}$, where $\psi_{\mathcal{F}}(\cdot)$ and $\psi_{\mathcal{G}}(\cdot)$ are some twice continuously differentiable functions of their arguments. Let

$$\Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}) = \frac{\partial \psi_{\mathcal{F}}(\gamma_{\mathcal{F}})}{\partial \gamma_{\mathcal{F}}'} \tag{41}$$

and

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma_{\mathcal{G}}'} \tag{42}$$

be $(k_1 - k_3) \times k_1$ and $(k_2 - k_3) \times k_2$ derivative matrices of the parametric restrictions $\psi_{\mathcal{F}}$ and $\psi_{\mathcal{G}}$, respectively. In many situations, $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ and $y_t^{\mathcal{F}}(\gamma_{\mathcal{G}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ implies that a subset of parameters of models $\mathcal{F}$ and $\mathcal{G}$ are equal to zero. For such cases, we can arrange the parameters so that $\Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}) = [0_{(k_1-k_3)\times k_3},\ I_{k_1-k_3}]$ and $\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_3)\times k_3},\ I_{k_2-k_3}]$. Let $\Sigma_{\hat{\gamma}_{\mathcal{FG}}}$ be the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{FG}} = [\hat{\gamma}_{\mathcal{F}}',\ \hat{\gamma}_{\mathcal{G}}']'$, $\tilde{H}_{\mathcal{F}} = (C^{\mathcal{F}} + D^{\mathcal{F}\prime}U^{-1}D^{\mathcal{F}})^{-1}$, $\tilde{H}_{\mathcal{G}} = (C^{\mathcal{G}} + D^{\mathcal{G}\prime}U^{-1}D^{\mathcal{G}})^{-1}$, $\Psi_*^{\mathcal{F}} = \Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}^*)$, $\Psi_*^{\mathcal{G}} = \Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ and

$$\Psi_*^{\mathcal{FG}} \equiv \begin{bmatrix} \Psi_*^{\mathcal{F}} & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi_*^{\mathcal{G}} \end{bmatrix}. \tag{43}$$

Define the Wald test statistic

$$Wald_{\hat{\psi}_{\mathcal{FG}}} = T\hat{\psi}_{\mathcal{FG}}'(\hat{\Psi}^{\mathcal{FG}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}\hat{\Psi}^{\mathcal{FG}\prime})^{-1}\hat{\psi}_{\mathcal{FG}}, \tag{44}$$

where $\hat{\psi}_{\mathcal{FG}} = [\psi_{\mathcal{F}}(\hat{\gamma}_{\mathcal{F}})',\ \psi_{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})']'$,

$$\hat{\Psi}^{\mathcal{FG}} = \begin{bmatrix} \Psi^{\mathcal{F}}(\hat{\gamma}_{\mathcal{F}}) & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi^{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}}) \end{bmatrix}, \tag{45}$$

and $\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}$ is a consistent estimator of $\Sigma_{\hat{\gamma}_{\mathcal{FG}}}$.

The next theorem establishes the asymptotic distribution of $T(\hat{\delta}^2_{\mathcal{F}}-\hat{\delta}^2_{\mathcal{G}})$ and the Wald test under the null hypothesis $H_0 : \psi_{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_3}$.

Theorem 3. *Suppose that $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \not\subset \mathcal{G}$, $\mathcal{G} \not\subset \mathcal{F}$, and Assumptions A, B and C hold. Then, under $H_0 : \psi_{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_3}$,*

(a)

$$T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}) \overset{A}{\sim} F_{k_1+k_2-2k_3}(\xi), \tag{46}$$

where the ξ_i's are the eigenvalues of the matrix

$$\begin{bmatrix} -(\Psi^{\mathcal{F}}_* \tilde{H}_{\mathcal{F}} \Psi^{\mathcal{F}\prime}_*)^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi^{\mathcal{G}}_* \tilde{H}_{\mathcal{G}} \Psi^{\mathcal{G}\prime}_*)^{-1} \end{bmatrix} \Psi^{\mathcal{F}\mathcal{G}}_* \Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}} \Psi^{\mathcal{F}\mathcal{G}\prime}_*, \tag{47}$$

(b)

$$Wald_{\hat{\psi}_{\mathcal{F}\mathcal{G}}} \overset{A}{\sim} \chi^2_{k_1+k_2-2k_3}. \tag{48}$$

Unlike the case of nested models, the eigenvalues in part (a) of Theorem 3 are not always positive because $\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}$ can take on both positive and negative values. As a result, we need to perform a two-sided test of $H_0 : y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})$. Similarly to the nested case, an alternative way of testing the equality of two overlapping SDFs is to directly test the constraints $\psi_{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_3}$ using the asymptotically pivotal Wald test in part (b) of Theorem 3.

For overlapping and strictly non-nested models, the variance σ^2_d can be zero when both models are correctly specified.[10] Asymptotic weighted chi-squared and joint LM specification tests of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ are proposed by Kan and Robotti (2009) for linear models. The extension of these tests to nonlinear models is provided in an online appendix (see also Hall and Pelletier, 2011).

In summary, our proposed sequential testing procedure of equality of the squared HJ-distances of two overlapping models is the following. First, we need to test whether the SDFs of the two models are equal using the tests in Theorem 3. Since the test in part (a) of Theorem 3 will not be consistent against the alternative $H_1 : y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) \neq y^{\mathcal{H}}_t(\gamma^*_{\mathcal{H}})$ when both models are correctly specified,

[10]In a likelihood framework (see Vuong, 1989), two strictly non-nested models cannot be both correctly specified. However, in our context, a correctly specified model is defined in terms of moment conditions and it is possible for two strictly non-nested models to be both correctly specified. See Kan and Robotti (2009) and Hall and Pelletier (2011) for further discussion of this point. Similarly, two overlapping SDFs can also be both correctly specified. Examples for these situations are available from the authors upon request.

our recommendation is to use the Wald test of $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. If the null is rejected, we need to proceed with testing if the two models are both correctly specified. Finally, if we still reject, we can then perform the normal test in (35). The significance level of this procedure, as a test of $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$, is asymptotically bounded above by $\max[\alpha_1, \alpha_2, \alpha_3]$, where α_1, α_2, and α_3 are the asymptotic significance levels used in these three tests.[11]

The results in Theorems 2 and 3 offer substantial advantages over the inference procedures available in the literature. Imposing the parametric restrictions that directly arise from the structure of the models and the appropriate null hypotheses results in a drastic reduction of the number of weights that are used to compute the critical values of the tests. More specifically, the number of eigenvalues in the weighted chi-squared distribution is reduced from $2n + k_1 + k_2$ to $k_2 - k_1$ for nested and to $k_1 + k_2 - 2k_3$ for overlapping models. This proves to be particularly advantageous when the number of test assets n is large. The reduced dimensions of the matrices in part (a) of Theorems 2 and 3 are expected to lead to improved finite-sample behavior of the model selection tests.

3.2. Multiple model comparison

Thus far, we have considered pairwise model comparison. However, when multiple models are involved, pairwise model comparison may not determine unambiguously the best performing model. In this subsection, we develop formal multiple model comparison tests for non-nested and nested models. Our non-nested model comparison test is a multivariate inequality test based on the results of Wolak (1987, 1989).[12] Suppose we have $p + 1$ models. We are interested in testing the null hypothesis that the benchmark model, model 1 (we could think of model 1 as model $\mathcal{F}$ in the pairwise model comparison subsection), performs at least as well as the other p models. Let δ_i^2 denote the population squared HJ-distance of model i and let $\rho \equiv (\rho_2, \ldots, \rho_{p+1})$, where $\rho_i \equiv \delta_1^2 - \delta_i^2$. Therefore, the null hypothesis is $H_0 : \rho \leq 0_p$ while the alternative is $H_1 : \rho \in \Re^p$.

The test is based on the sample counterpart, $\hat{\rho} \equiv (\hat{\rho}_2, \ldots, \hat{\rho}_{p+1})$, where $\hat{\rho}_i \equiv \hat{\delta}_1^2 - \hat{\delta}_i^2$. Assume

[11] In an online appendix, we also provide unrestricted and restricted (as in Theorems 2 and 3) versions of the test of $H_0 : \sigma_d^2 = 0$.

[12] Kan, Robotti and Shanken (2012) adapt the multivariate inequality test of Wolak (1987, 1989) to compare the performance of alternative asset pricing models in a two-pass cross-sectional regression framework.

that

$$\sqrt{T}(\hat{\rho} - \rho) \overset{A}{\sim} N(0_p, \Omega_{\hat{\rho}}). \tag{49}$$

As in Section 3.1, sufficient conditions for asymptotic normality are: i) $\delta_i^2 > 0$, and ii) the SDFs of the different models are distinct.[13] Let $\tilde{\rho}$ be the optimal solution in the following quadratic programming problem:

$$\min_{\rho} (\hat{\rho} - \rho)' \hat{\Omega}_{\hat{\rho}}^{-1} (\hat{\rho} - \rho) \quad \text{s.t.} \quad \rho \leq 0_p, \tag{50}$$

where $\hat{\Omega}_{\hat{\rho}}$ is a consistent estimator of $\Omega_{\hat{\rho}}$. The likelihood ratio-type test of the null hypothesis is

$$LR = T(\hat{\rho} - \tilde{\rho})' \hat{\Omega}_{\hat{\rho}}^{-1} (\hat{\rho} - \tilde{\rho}). \tag{51}$$

Since the null hypothesis is composite, to construct a test with the desired size, we require the distribution of LR under the least favorable value of ρ, which is $\rho = 0_p$. Under this value, LR follows a "chi-bar-squared distribution,"

$$LR \overset{A}{\sim} \sum_{i=0}^{p} w_{p-i}(\Omega_{\hat{\rho}}) X_i, \tag{52}$$

where the X_i's are independent χ^2 random variables with i degrees of freedom, χ_0^2 is simply defined as the constant zero and the weights w_i sum up to one.[14] We use this procedure to obtain asymptotically valid p-values.[15]

Before using the multivariate inequality test to compare a benchmark model with a set of alternative models, we remove those alternative models that are nested by the benchmark model since, by construction, $\rho_i \leq 0$ in this case. If any of the remaining alternatives is nested by another alternative model, we remove the "nested" model since the δ^2 of the nesting model will be at least as small. Finally, we also eliminate from consideration any alternative models that nest the

[13] Note that a pre-test of equality of SDFs can be easily developed also for multiple models by generalizing the chi-squared test in (48) to the $p > 1$ case.

[14] An explicit formula for the weights $w_i(\Omega_{\hat{\rho}})$ is given in Kudo (1963) and the computational details are provided in the Appendix. Although the Monte Carlo simulations in the next section show that our asymptotic approximation works well in experiments with realistic sample sizes, researchers could also use bootstrap methods to obtain the p-value for the proposed test statistic. See also Andrews and Soares (2010), among others, for various subsampling, m-out-of-n bootstrap and "plug-in-asymptotic" procedures for parameter inference in models defined by moment inequalities.

[15] There are alternatives to the multivariate inequality test described above. Under the assumption in (49), White (2000) and Hansen (2005) provide reality check tests that can be used to compare the performance of multiple models. Computing p-values for their tests, however, requires either Monte Carlo simulation or bootstrap methods and can be very time consuming. See Chen and Ludvigson (2009) for an application of the tests of White (2000) and Hansen (2005) to the study of the HJ-distances of competing asset pricing models.

benchmark because the asymptotic normality assumption on $\hat{\rho}_i$ does not hold under the null of $\rho_i = 0$.

Since the multivariate inequality test described above is not applicable when the benchmark is nested by some alternative models, a different multiple model comparison test is needed in this situation. When the alternative models nesting the benchmark are nested within each other, we remove the "nested" models since the δ^2 of the nesting model will be at least as small. In this scenario, one could simply use the pairwise model comparison techniques developed in Section 3.1. The situation, however, becomes more complicated when the alternative models exhibit an overlapping structure.

Suppose that the benchmark (with k_1 parameter vector γ_1) is nested by model i (with k_i parameter vector γ_i, $i = 2, \ldots, p+1$). Similar to the setup of Section 3.1, suppose that $y_t^1(\gamma_1^*) = y_t^i(\gamma_i^*)$ can be written as a parametric restriction of the form $\psi_i(\gamma_i^*) = 0_{k_i - k_1}$, where $\psi_i(\cdot)$ is a twice continuously differentiable function in its argument. The null hypothesis for multiple model comparison can therefore be formulated as $H_0 : \psi_2(\gamma_2^*) = 0_{k_2-k_1}, \ldots, \psi_{p+1}(\gamma_{p+1}^*) = 0_{k_{p+1}-k_1}$. Having derived the asymptotic distribution of $\hat{\gamma}_i$ in Lemma 1, we can use the delta method to obtain the asymptotic distribution of $\hat{\psi}_i = \psi_i(\hat{\gamma}_i)$. Specifically, let

$$\psi = \begin{pmatrix} \psi_2(\gamma_2^*) \\ \vdots \\ \psi_{p+1}(\gamma_{p+1}^*) \end{pmatrix} \tag{53}$$

and denote by $\hat{\psi}$ a consistent estimator of ψ. Also, let $\Sigma_{\hat{\psi}}$ be the asymptotic covariance matrix of $\hat{\psi}$ with rank l under the null hypothesis and $\hat{\Sigma}_{\hat{\psi}}$ denote its consistent estimator. Then, we have

$$Wald_{\hat{\psi}} = T\hat{\psi}'\hat{\Sigma}_{\hat{\psi}}^{+}\hat{\psi} \overset{A}{\sim} \chi_l^2, \tag{54}$$

where $\hat{\Sigma}_{\hat{\psi}}^{+}$ is the generalized inverse of $\hat{\Sigma}_{\hat{\psi}}$. To perform this test, we need to determine the rank of $\Sigma_{\hat{\psi}}$ under the null hypothesis. For linear SDFs, l is simply the number of distinct factors in the set of alternative models minus the number of factors in the benchmark model.[16] For nonlinear SDFs, determining the rank of $\Sigma_{\hat{\psi}}$ under H_0 depends on the particular overlapping structure of the nesting models which needs to be analyzed on a case-by-case basis.

[16]For example, suppose we have three linear models with factors $[1, f_{1t}]'$, $[1, f_{1t}, f_{2t}, f_{3t}]'$ and $[1, f_{1t}, f_{2t}, f_{4t}]'$, respectively. Note that the first model is the nested (benchmark) model and the last two models are the alternative models nesting the benchmark. The number of distinct factors in the set of alternative models is five ($1, f_{1t}, f_{2t}, f_{3t}$ and f_{4t}) and $l = 3$ is obtained by subtracting the number of factors in the benchmark model (1 and f_{1t}). This procedure is used in Section 4.2 below.

To summarize, if the benchmark model is nested by some competing models, one should separate the set of competing models into two subsets. The first subset includes competing models that nest the benchmark. To test whether the benchmark performs as well as the models in this subset, one can use the chi-squared nested multiple model comparison test described above. The second subset includes competing models that do not nest the benchmark. For this second subset, we can use the test in (52). If we perform each test at a significance level of $\alpha/2$ and fail to reject the null hypothesis in both tests, then, by the Bonferroni inequality, the size of the joint test will be less than or equal to α.

4. Monte Carlo simulations

In this section, we undertake a Monte Carlo experiment to explore the small-sample properties of the test statistics discussed in the theoretical part of the paper.[17] We consider both linear and nonlinear asset pricing models. To make our simulations more realistic, we calibrate the parameters by using almost 50 years, 1952:2–2000:4, of U.S. quarterly gross returns on the three-month T-bill and the well-known 25 Fama-French size and book-to-market portfolios ($n = 26$). The time-series sample size is $T = 120, 240, 360, 480$ and 600. These choices of T reflect sample sizes that are typically encountered in empirical work. We compare actual rejection rates over 100,000 iterations to the nominal 10%, 5% and 1% levels of our tests. A detailed description of the various simulation designs can be found in the Appendix.

In the linear case, the SDF takes the form $y_t(\gamma) = \gamma'\tilde{f}_t$ with $\gamma = [\gamma_0,\ \gamma_1']'$ and $\tilde{f}_t = [1,\ f_t']'$, where f_t is a $(k-1)$-vector of risk factors at time t. In our simulations, we consider the following linear models: the consumption capital asset pricing model (CCAPM) with the log consumption growth rate of non-durable goods ($\triangle c_{ndur}$) as a risk factor; the durable consumption CAPM of Yogo (YOGO, 2006) with the excess market return (r_{mkt}), $\triangle c_{ndur}$ and the log consumption growth rate of durable goods ($\triangle c_{dur}$) as risk factors; the three-factor model of Fama and French (FF3, 1993) with r_{mkt}, the return difference between portfolios of small and large stocks (r_{smb}) and the return difference between portfolios of high and low book-to-market ratios (r_{hml}) as risk factors; the conditional consumption CAPM of Lettau and Ludvigson (LL, 2001) with $\triangle c_{ndur}$, the lagged

[17] In an online appendix available on the authors' websites, we also analyze the size and power properties of the model selection tests for nested and non-nested models that have been proposed in the literature. This investigation suggests that our tests are more powerful than the existing ones in realistic simulation settings.

consumption-wealth ratio (cay) and the interaction term between $\triangle c_{ndur}$ and cay as risk factors.

In the nonlinear case, we assume that the SDF takes the form $y_t(\gamma) = \exp(\gamma' \tilde{f}_t)$. Many popular asset pricing models can be cast in this log-linear framework. More specifically, we consider the nonlinear version of Yogo's (2006) model and the external habit model (EHM) of Abel (1990), both of which nest the nonlinear CCAPM model. Let C_{ndur} denote consumption of non-durable goods, C_{dur} be consumption of durable goods and $U(C_{ndur}, C_{dur}) = C_{ndur}^{1-\phi} C_{dur}^{\phi}$ denote the Cobb-Douglas intraperiod utility function, where $\phi \in [0,1]$ is the budget share of durable consumption. For Yogo's (2006) model, the SDF is given by

$$y_t^{YOGO}(\alpha, \beta, \rho, \phi) = \beta^{\frac{1-\rho}{1-\alpha}} \left(\frac{C_{ndur,t}}{C_{ndur,t-1}} \right)^{-\alpha\left(\frac{1-\rho}{1-\alpha}\right)} \left(\frac{C_{dur,t}/C_{ndur,t}}{C_{dur,t-1}/C_{ndur,t-1}} \right)^{\phi(1-\rho)} R_{mkt,t}^{\frac{\alpha-\rho}{1-\alpha}}, \tag{55}$$

where R_{mkt} is the gross market return, β is the discount rate, $\rho > 0$ is the coefficient of relative risk aversion and $1/\alpha \geq 0$ is the elasticity of intertemporal substitution. Taking logarithms yields

$$\begin{aligned} \ln(y_t^{YOGO}(\alpha, \beta, \rho, \phi)) \quad = \quad & \frac{1-\rho}{1-\alpha} \ln(\beta) - \frac{(1-\rho)(\alpha(1-\phi)+\phi)}{1-\alpha} \triangle c_{ndur,t} \\ & + \phi(1-\rho) \triangle c_{dur,t} + \frac{\alpha-\rho}{1-\alpha} \ln(R_{mkt,t}), \end{aligned} \tag{56}$$

where $\triangle c_{ndur}$ and $\triangle c_{dur}$ are defined above. When $\phi = 0$, we have the classical non-expected (Epstein-Zin) utility model. By imposing the additional restriction $\alpha = \rho$, we obtain the standard expected utility model (nonlinear CCAPM). Similarly, the SDF for EHM with time-separability parameter $\tau \geq 0$ is given by

$$y_t^{EHM}(\beta, \rho, \tau) = \beta \left(\frac{C_{ndur,t}}{C_{ndur,t-1}} \right)^{-\rho} \left(\frac{C_{ndur,t-1}}{C_{ndur,t-2}} \right)^{\tau(\rho-1)} \tag{57}$$

or

$$\ln(y_t^{EHM}(\beta, \rho, \tau)) = \ln(\beta) - \rho \triangle c_{ndur,t} + \tau(\rho - 1) \triangle c_{ndur,t-1}. \tag{58}$$

When $\tau = 0$, EHM reduces to the nonlinear CCAPM. Also note that nonlinear YOGO and EHM are overlapping models with the restrictions $[\phi,\ \alpha - \rho,\ \tau]' = 0_3$ rendering the two SDFs equal. The last log-linear SDF considered in our simulations is the nonlinear CAPM of Brown and Gibbons (1985) with $\ln(R_{mkt})$ as a risk factor, which is nested by nonlinear YOGO.

In our simulations, we make the following joint distributional assumption on the factors and returns. For linear models, we assume that the factors and the gross returns on the test assets are i.i.d. multivariate normally distributed.[18] For nonlinear models, we assume that the factors and the continuously compounded returns on the test assets are i.i.d. multivariate normally distributed.[19]

4.1. Parameter estimates and model specification tests

In this subsection, we investigate the size properties of the SDF parameter estimates and the size and power properties of the model specification tests. The data are simulated using the linear and nonlinear specifications of Yogo's (2006) model.

We start by analyzing the finite-sample properties of the SDF parameter estimates under model misspecification. One way to summarize the sampling behavior of the SDF parameter estimates and their corresponding asymptotic approximations is to focus on the rejection rates of the t-tests of $H_0 : \gamma_i = 0$. In the simulations, the expected returns are chosen such that the SDF parameter associated with a given factor is equal to zero.[20] The t-tests are constructed using the asymptotic covariance matrices in Lemma 1 and are compared against the critical values from a standard normal distribution. We refer to the t-tests based on (23) and (24) as *t-tests under potentially misspecified models.* For comparison, we also report results using the traditional standard errors derived under correctly specified models based on the asymptotic covariance matrix in (26) and (27). We refer to the corresponding t-tests as *t-tests under correctly specified models.* The reason for investigating the finite-sample performances of the t-tests under correctly specified models in a simulation setup where the model fails to hold exactly is that researchers typically rely on these t-tests in drawing inferences on the SDF parameters even when a model is strongly rejected by the data.

Panels A and B of Table 1 present the empirical size of both t-tests of the null hypothesis

[18] Since asset return distributions often exhibit fat tails, we also draw factors and returns from a multivariate t-distribution with 8 degrees of freedom. The t-distribution results are very similar to the ones under normality and can be found in an online appendix on the authors' websites. The only noteworthy difference is a slight increase in the empirical size of the tests in the t-distribution case.

[19] As pointed out by a referee, endogenously derived returns from many equilibrium models are often not normally distributed. This is a scenario in which our simulation results may not provide a very accurate assessment of the properties of the proposed tests.

[20] For nonlinear YOGO, the implied HJ-distances from choosing the expected returns such that $\gamma_i = 0$ are equal to 0.6403 (when setting the SDF parameter associated with $\ln(R_{mkt})$ equal to zero), 0.5892 (when setting the SDF parameter associated with Δc_{ndur} equal to zero) and 0.5877 (when setting the SDF parameter associated with Δc_{dur} equal to zero). As a basis for comparison, note that the HJ-distance for the nonlinear YOGO model when no parameter restrictions are imposed is 0.6357. For linear YOGO, the implied HJ-distances from choosing the expected returns such that $\gamma_i = 0$ are always equal to 0.6514 which is also the value of the HJ-distance for the unrestricted linear YOGO model.

$H_0 : \gamma_i = 0$ for linear YOGO, while Panels C and D of the same table report the empirical size of both t-tests of the null hypothesis $H_0 : \gamma_i = 0$ for nonlinear YOGO (to preserve space, we do not report simulation results for the t-ratios associated with the SDF intercept terms). Panels A and C are for the t -tests under potentially misspecified models, while Panels B and D are for the t-tests under correctly specified models.

Table 1 about here

The results in Table 1 reveal that the finite-sample performance of these two tests differs considerably. The t-test under potentially misspecified models is well-behaved and its empirical size is always close to the nominal level.[21] On the other hand, the t-test under correctly specified models tends to overreject substantially. For example, the t-test on the durable consumption parameter in nonlinear YOGO rejects the null hypothesis 33% of the time at the 5% nominal level for $T = 600$. Interestingly, the presence of non-traded factors in the YOGO specifications also leads to significant size distortions of the t-test on the traded factor. Finally, the performance of the t-tests under correctly specified models deteriorates as T increases.

This difference in behavior between the two t-tests warrants some explanation. In the case of linear SDFs, Kan and Robotti (2009) prove that when factors and returns are multivariate elliptically distributed, the standard errors under potentially misspecified models are always bigger than the standard errors constructed under the assumption that the model is correctly specified. They show that the magnitude of the misspecification adjustment term, that reflects the difference between the asymptotic variances of the SDF parameter estimates under correctly specified and misspecified models, depends on, among other things, the degree of model misspecification (as measured by the HJ-distance measure) and the correlations of the factors with the returns. The misspecification adjustment term can be huge when the underlying factor is poorly mimicked by asset returns – a situation that typically arises when some of the factors are macroeconomic variables as, for example, in linear YOGO. Therefore, when the model is misspecified and the factors are poorly spanned by the returns, the t-test under correctly specified models can lead to the erroneous conclusion that certain factors are priced. Our simulation evidence further demonstrates that the t-test under correctly specified models can be seriously oversized for both linear and log-linear SDFs

[21] We should note that the t-test under potentially misspecified models maintains its good size properties even when the data are generated under correctly specified models (results are not reported to conserve space).

and that researchers should exercise caution when using it to determine whether a risk factor is priced. Another related issue is the deterioration in the size properties of the t-test under correctly specified models as T increases. This is likely to be a symptom of the fact that some non-traded factors such as $\triangle c_{ndur}$ and $\triangle c_{dur}$ are almost uncorrelated with the returns. For further discussion, we refer the reader to Kan and Zhang (1999) who show that when the model is misspecified and a factor is "useless," i.e., independent of the returns, increasing the sample size also increases the severity of the overrejection problem. For these reasons, we strongly recommend using the t-test under potentially misspecified models in factor pricing.

We now turn our attention to the model specification tests. In particular, we assess the finite-sample performance of the conventional HJ-distance test and LM test in Theorem 1. To examine size, the return means are set such that the model holds exactly, i.e., $\delta = 0$. To examine power, the return means are chosen based on the means estimated from the data, which implies that the HJ-distances for linear and nonlinear YOGO are 0.6514 and 0.6357, respectively. The empirical size and power of the two tests are presented in Panels A and B of Table 2.

Table 2 about here

The overrejections of the HJ-distance test have already been documented in the literature (see, for example, Ahn and Gadarowski, 2004). The overrejection problem is particularly severe when the number of assets is large relative to the number of time series observations. Our results confirm the overrejections of the HJ-distance test across the linear and nonlinear YOGO specifications. For example, for nonlinear YOGO, the empirical size of the HJ-distance test is 17.3% at the 5% significance level for $T = 120$ and approaches the nominal level of the test as T increases (6.5% at the 5% significance level for $T = 600$). The LM test in Theorem 1 has excellent size properties, being only slightly oversized for $T = 120$. The improved size of the LM test is accompanied by a power performance that is very similar to the one of the HJ-distance test which overrejects under the null.[22]

[22] All tables in this section report actual power since computing size-adjusted power seems infeasible for several of our tests. Their null distributions depend on many nuisance parameters and the simulation of their exact distributions is complicated by the fact that those nuisance parameters are in general not known.

4.2. Model selection tests for nested models

In Table 3, we investigate the size and power properties of pairwise and multiple nested model comparison tests. Panels A and B are for linear SDFs, while Panels C and D are for log-linear SDFs.

Table 3 about here

For linear SDFs, CCAPM represents our benchmark model. For pairwise model comparison, we consider CCAPM nested by YOGO, while for multiple model comparison we consider CCAPM nested by YOGO and LL. In the log-linear SDF case, nonlinear CCAPM represents our benchmark model. For pairwise model comparison, we consider nonlinear CCAPM nested by nonlinear YOGO, while for multiple model comparison we consider nonlinear CCAPM nested by nonlinear YOGO and EHM.

The tests under investigation are the weighted chi-squared test and the Wald test in Theorem 2, as well as the chi-squared multiple model comparison test in (54). To analyze the finite-sample behavior of the pairwise model comparison tests under the null of equality of squared HJ-distances, we choose the return means such that the SDF parameters associated with factors that are not in the benchmark model are zero and all the models are misspecified. The implied HJ-distances are 0.6514 and 0.6414 in the linear and log-linear cases, respectively. To analyze power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for linear (nonlinear) CCAPM and linear (nonlinear) YOGO are 0.6768 (0.6630) and 0.6514 (0.6357), respectively. Turning to multiple model comparison, the size of the chi-squared test is evaluated by choosing the return means such that the SDF parameters associated with factors that are not in the benchmark model are zero and all the models are misspecified. The implied HJ-distances are 0.6410 and 0.6251 in the linear and log-linear cases, respectively. To evaluate power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for linear (nonlinear) CCAPM, linear (nonlinear) YOGO, LL and EHM are 0.6768 (0.6630), 0.6514 (0.6357), 0.6561 and 0.6363, respectively.

Panels A and C of Table 3 show that the Wald test in Theorem 2 has very good size properties and high power (despite the small differences in HJ-distances between models). The weighted chi-squared tests in Theorem 2 is a bit conservative under the null (especially in the nonlinear case) and overall exhibits lower power than the Wald test.

For multiple model comparison, the size and power of the chi-squared test in Panels B and D of Table 3 are impressive. This simulation evidence is very encouraging for the use of this new test in empirical work.

4.3. Model selection tests for non-nested models

The case of non-nested (overlapping) models is arguably the most important case in practice since many empirical asset pricing specifications contain a constant term and different systematic factors.

Starting with pairwise model comparison, we evaluate the finite-sample behavior of the pre-tests of equality of SDFs in Theorem 3. In the linear SDF case (Table 4.A), the simulated data are generated using FF3 and YOGO. To evaluate size, we choose the return means such that the SDF parameters associated with the non-overlapping factors in FF3 and YOGO are zero and the two models are misspecified, which implies that the population HJ-distance is 0.5733. To analyze power, the return means are chosen based on the means estimated from the data which implies that the population HJ-distances for FF3 and YOGO are 0.5822 and 0.6514, respectively. For log-linear SDFs (Table 4.C), we consider EHM and nonlinear YOGO. Similarly to the linear case, we study size by setting the SDF parameters associated with the non-overlapping factors in EHM and nonlinear YOGO equal to zero, while keeping both models misspecified (the implied HJ-distance is 0.6251 in this scenario). In the power experiments, the return means are again chosen based on the means estimated from the data which implies that the population HJ-distances for EHM and nonlinear YOGO are 0.6363 and 0.6357, respectively.

Table 4 about here

Panels A and C show that the Wald test in Theorem 3 has excellent size. The weighted chi-squared test also enjoys good size properties but is a bit conservative under the null in the nonlinear SDF case. Both tests exhibit very good power in the linear setting, but the Wald test outperforms the weighted chi-squared test when it comes to log-linear SDFs. The high power of the Wald test appears to be particularly important given the fact that this test serves only as a preliminary step in establishing whether two (or more) models have equal pricing performance.

If the null hypotheses of SDF equality and correct specification of the two models are rejected, then the researcher can proceed with the normal test in (35). In the size computations for linear SDFs (Panel B), the data are simulated from two misspecified three-factor models. The two models have r_{smb} and r_{hml} as common factors. The third factor in each model is created by adding a normally distributed error to r_{mkt}. The error term in each model has a mean of zero and a variance of 20% of the variance of r_{mkt}. The two error terms are independent of each other as well as of the returns and the factors. This implies that the population HJ-distances of the two models are both equal to 0.5822. For power evaluation, the two overlapping models are FF3 and YOGO with population HJ-distances of 0.5822 and 0.6514, respectively. For log-linear SDFs (Panel D), the data are simulated from two misspecified one-factor models, where the factor in each model is equal to $\ln(R_{mkt})$ plus a normally distributed error generated as in the linear case (the only difference being that the error term has a variance of 20% of the variance of $\ln(R_{mkt})$). This implies that the population HJ-distances of the two models are both equal to 0.6377. For power evaluation, the two non-nested models are nonlinear CCAPM and FF3 with population HJ-distances of 0.6630 and 0.5822, respectively. Panels B and D ($p = 1$ case) show that the size properties of the normal test are very good even for small T and that the empirical power quickly approaches 1 as T increases.

Finally, we extend the simulation setup described in the previous paragraph to multiple model comparison and employ the LR test in (52). In the linear SDF case, we consider 3 three-factor models. In the size comparison, we add another model with r_{smb} and r_{hml} as common factors and a non-overlapping part given by r_{mkt} plus a normally distributed error generated as in the pairwise model comparison case. This guarantees that the population HJ-distances of the three models are all equal to 0.5822. For power evaluation, we consider LL (the benchmark model) in addition to YOGO and FF3 (the population HJ-distances of the three models are 0.6561, 0.6514 and 0.5822, respectively). In the log-linear SDF case, we study size by adding another one-factor model, where the factor is equal to $\ln(R_{mkt})$ plus a normally distributed error generated as in pairwise model comparison case (the population HJ-distances of the three models are all equal to 0.6377). To study power, we consider nonlinear CCAPM (the benchmark model) in addition to nonlinear CAPM and FF3 (the population HJ-distances of the three models are 0.6630, 0.6377 and 0.5822, respectively). Panels B and D ($p = 2$ case) reveal the very good finite-sample properties of the LR test for comparing multiple asset pricing models.

Overall, our simulation results suggest that the tests developed in this paper should be fairly reliable for the sample sizes typically encountered in empirical work. In addition, the proposed Wald tests exhibit higher power than their weighted chi-squared counterparts.

5. Concluding remarks

This paper develops a general statistical framework for evaluation and comparison of possibly misspecified asset pricing models using the unconstrained HJ-distance. We derive new versions of the weighted chi-squared specification and model comparison tests that are computationally efficient and possess improved finite-sample properties compared to the existing tests in the literature. We also propose new pivotal (asymptotically chi-squared distributed) specification and model selection tests. Finally, we develop computationally attractive tests for multiple model comparison. The excellent size and power properties of the proposed tests are demonstrated using simulated data from popular linear and nonlinear asset pricing models. The simulation results clearly suggest that the standard tests for model specification and selection as well as the typical practice of conducting inference on the SDF parameters under the assumption of correctly specified models could be highly misleading in various realistic setups. One of the main findings that emerges from our analysis is that properly incorporating the uncertainty arising from model misspecification, as well as imposing the additional restrictions implied by the structure of the models, leads to substantially improved inference.

Although our simulation results are encouraging, the small-sample properties of the test statistics proposed in this paper should be explored further. In addition, it is of interest to compare the empirical performance of various linear and nonlinear asset pricing models using the testing procedures developed in this paper. Finally, some recent research (for example, Nagel and Singleton, 2011) has emphasized the importance of incorporating conditioning information in the estimation and testing of asset pricing models. Extending the results on model comparison to conditional moment restrictions is a promising area for future research.

Acknowledgments

Gospodinov gratefully acknowledges financial support from FQRSC, IFM2, and SSHRC. Kan gratefully acknowledges financial support from the National Bank Financial of Canada, SSHRC, and the Center for Financial Innovation and Stability at the Federal Reserve Bank of Atlanta. We would like to thank the Editor (Ron Gallant), an Associate Editor, three anonymous referees, Esther Eiling, Wayne Ferson, Jonathan Fletcher, Eric Jondeau, Serena Ng, Denis Pelletier, B. Ravikumar, Sergei Sarkissian, Zhaogong Song, Jonathan Wright, Chu Zhang, Guofu Zhou, seminar participants at Carleton University, EDHEC Business School, Emory University, HEC Lausanne, University of Montreal, and participants at the 2010 IFM2 Symposium on Mathematical Finance, the 2010 Meetings of the Society for Nonlinear Dynamics and Econometrics, the 2010 NBER–NSF Time Series Conference, the 2011 Fifth CIREQ Time Series Conference, and the 2012 Northern Finance Association Conference for helpful discussions and comments. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Appendix

A. Preliminary lemma

We first present a preliminary lemma that develops an expansion of the sample HJ-distance which will be used in the proofs of the subsequent lemmas and theorems.

Lemma A.1. *Under Assumptions A, B and C,*

$$\hat{\delta}^2 - \delta^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\theta^*) - E[\phi_t(\theta^*)]\right) - \frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right). \tag{A.1}$$

Proof. We start by expanding $E[\phi_t(\hat{\theta})]$ about θ^*. Rearranging and noting that $\partial E[\phi_t(\theta^*)]/\partial\theta = E[\partial\phi_t(\theta^*)/\partial\theta] = 0_{n+k}$ (using Assumption C and the definition of θ^*), we obtain

$$\frac{1}{T}\sum_{t=1}^{T}\phi_t(\hat{\theta}) = E[\phi_t(\theta^*)] + \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\right) + \frac{1}{2}(\hat{\theta}-\theta^*)'\frac{\partial^2 E[\phi_t(\tilde{\theta})]}{\partial\theta\partial\theta'}(\hat{\theta}-\theta^*), \tag{A.2}$$

where $\tilde{\theta}$ is an intermediate point between $\hat{\theta}$ and θ^*. Furthermore, a mean value expansion of $\frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\right)$ about θ^* yields (Pollard, 1982)

$$\frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\theta^*) - E[\phi_t(\theta^*)]\right) + \bar{v}_T(\theta^*)'(\hat{\theta}-\theta^*) + o_p\left(\frac{1}{T}\right) \tag{A.3}$$

using the definition of $\bar{v}_T(\theta)$ in (12). Plugging (A.3) into (A.2) and from the consistency of $\hat{\theta}$ and Assumption B(ii), we obtain

$$\hat{\delta}^2 - \delta^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\theta^*) - E[\phi_t(\theta^*)]\right) + \bar{v}_T(\theta^*)'(\hat{\theta}-\theta^*) + \frac{1}{2}(\hat{\theta}-\theta^*)'H(\hat{\theta}-\theta^*) + o_p\left(\frac{1}{T}\right). \tag{A.4}$$

Let $\bar{h}_T^*(\theta) = \frac{1}{T}\sum_{t=1}^{T}E[h_t(\theta)]$. A mean value expansion of $\bar{h}_T^*(\theta^*)$ about $\hat{\theta}$ gives

$$0_{n+k} = \sqrt{T}\bar{h}_T^*(\theta^*) = \sqrt{T}\bar{h}_T^*(\hat{\theta}) - \frac{\partial\bar{h}_T^*(\check{\theta})}{\partial\theta}\sqrt{T}(\hat{\theta}-\theta^*), \tag{A.5}$$

where $\check{\theta}$ is another intermediate point on the line segment joining $\hat{\theta}$ and θ^*. From Assumption B(ii) and the consistency of $\hat{\theta}$, we have

$$\sqrt{T}(\hat{\theta}-\theta^*) = H^{-1}\sqrt{T}\bar{h}_T^*(\hat{\theta}) + o_p(1). \tag{A.6}$$

From the first order condition of $\frac{1}{T}\sum_{t=1}^{T} h_t(\hat{\theta}) = 0_{n+k}$, it follows that

$$\sqrt{T}\bar{v}_T(\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(h_t(\hat{\theta}) - E[h_t(\hat{\theta})]\right) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E[h_t(\hat{\theta})] = -\sqrt{T}\bar{h}_T^*(\hat{\theta}). \tag{A.7}$$

This allows us to rewrite $\sqrt{T}\bar{h}_T^*(\hat{\theta})$ as

$$\sqrt{T}\bar{h}_T^*(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\hat{\theta}) = \sqrt{T}[\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})] - \sqrt{T}\bar{v}_T(\theta^*). \tag{A.8}$$

By the consistency of $\hat{\theta}$, $P[|\hat{\theta} - \theta^*| > \omega] \to 0$ for any arbitrarily small $\omega > 0$. Then,

$$\sqrt{T}|\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})| \leq \sup_{\theta\in\Theta:|\theta-\theta^*|\leq\omega} \sqrt{T}\left|\bar{v}_T(\theta^*) - \bar{v}_T(\theta)\right|. \tag{A.9}$$

From the stochastic equicontinuity of the empirical process $\sqrt{T}\bar{v}_T(\cdot)$,

$$\sup_{\theta\in\Theta:|\theta-\theta^*|\leq\omega} \sqrt{T}\left|\bar{v}_T(\theta^*) - \bar{v}_T(\theta)\right| \xrightarrow{p} 0. \tag{A.10}$$

Therefore, we have $\sqrt{T}[\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})] = o_p(1)$ and

$$\sqrt{T}\bar{h}_T^*(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\theta^*) + o_p(1). \tag{A.11}$$

Substituting (A.11) into (A.6) yields

$$\sqrt{T}(\hat{\theta} - \theta^*) = -H^{-1}\sqrt{T}\bar{v}_T(\theta^*) + o_p(1). \tag{A.12}$$

Thus, after plugging (A.12) in (A.4), we obtain

$$\hat{\delta}^2 - \delta^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\theta^*) - E[\phi_t(\theta^*)]\right) - \frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right). \tag{A.13}$$

This completes the proof. ■

B. Proofs

Proof of Lemma 1. (a) For $\delta > 0$ and under Assumptions A, B and C,

$$\sqrt{T}\bar{v}_T(\theta^*) \overset{A}{\sim} N(0_{n+k}, M). \tag{A.14}$$

Then, combining (A.12) and (A.14), we obtain

$$\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{n+k}, H^{-1}MH^{-1}). \tag{A.15}$$

To derive an explicit expression for the asymptotic covariance matrix of $\hat{\theta}$, we write

$$H^{-1}MH^{-1} = \sum_{j=-\infty}^{\infty} E[l_t l'_{t+j}], \tag{A.16}$$

where

$$l_t \equiv \begin{bmatrix} l_{1t} \\ l_{2t} \end{bmatrix} = H^{-1}\frac{\partial \phi_t(\theta^*)}{\partial \theta}. \tag{A.17}$$

From the definition of H in (10), we can use the partitioned matrix inverse formula to obtain

$$H^{-1} = \begin{bmatrix} 2C & 2D' \\ 2D & -2U \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} \tilde{H} & \tilde{H}D'U^{-1} \\ U^{-1}D\tilde{H} & -U^{-1} + U^{-1}D\tilde{H}D'U^{-1} \end{bmatrix}, \tag{A.18}$$

where $\tilde{H} = (C + D'U^{-1}D)^{-1}$. Using (A.18), (16) and (17), we can express l_{1t} and l_{2t} as

$$\begin{aligned} l_{1t} &= (C + D'U^{-1}D)^{-1}\left[(y_t(\gamma^*) - m_t(\theta^*))\frac{\partial y_t(\gamma^*)}{\partial \gamma} + D'U^{-1}[x_t m_t(\theta^*) - q_{t-1}]\right], && \text{(A.19)} \\ l_{2t} &= U^{-1}[Dl_{1t} - x_t m_t(\theta^*) + q_{t-1}]. && \text{(A.20)} \end{aligned}$$

Using the definition of $m_t(\theta^*)$ and rearranging the terms delivers the desired result. This completes the proof of part (a).

(b) When $\delta = 0$, $C = 0_{k\times k}$ and $m_t(\theta^*) = y_t(\gamma^*)$. Therefore, l_{1t} and l_{2t} simplify to

$$\begin{aligned} l_{1t} &= \tilde{l}_{1t} = (D'U^{-1}D)^{-1}D'U^{-1}e_t(\gamma^*), && \text{(A.21)} \\ l_{2t} &= U^{-1}[Dl_{1t} - e_t(\gamma^*)]. && \text{(A.22)} \end{aligned}$$

Premultiplying l_{2t} by $P'U^{\frac{1}{2}}$ yields $\tilde{l}_{2t} = -P'U^{-\frac{1}{2}}e_t(\gamma^*)$. This completes the proof of part (b). ■

Proof of Theorem 1. From part (b) of Lemma 1, we have

$$\sqrt{T}P'U^{\frac{1}{2}}\hat{\lambda} \overset{A}{\sim} N(0_{n-k}, P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P) \tag{A.23}$$

when $\delta = 0$, or equivalently

$$\sqrt{T}(P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P)^{-\frac{1}{2}}P'U^{\frac{1}{2}}\hat{\lambda} \overset{A}{\sim} N(0_{n-k}, I_{n-k}). \tag{A.24}$$

Then, under Assumptions A, B and C,

$$\begin{aligned} LM_{\hat{\lambda}} &= T\hat{\lambda}'\hat{U}^{\frac{1}{2}}\hat{P}\left(\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\right)^{-1}\hat{P}'\hat{U}^{\frac{1}{2}}\hat{\lambda} \\ &= T\hat{\lambda}'U^{\frac{1}{2}}P\left(P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P\right)^{-1}P'U^{\frac{1}{2}}\hat{\lambda} + o_p(1) \\ &\overset{A}{\sim} \chi^2_{n-k}. && \text{(A.25)} \end{aligned}$$

This completes the proof. ■

Proof of Theorem 2. (a) Since $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ under the null, it follows that $\lambda_{\mathcal{F}}^* = \lambda_{\mathcal{G}}^*$ and $m_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = m_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$ which implies that $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$. Using these identities, we have

$$\frac{\partial \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)}{\partial \lambda_{\mathcal{F}}} = 2[x_t m_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) - q_{t-1}] = 2[x_t m_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*) - q_{t-1}] = \frac{\partial \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)}{\partial \lambda_{\mathcal{G}}} \tag{A.26}$$

and

$$\bar{v}_{2,T}^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \bar{v}_{2,T}^{\mathcal{G}}(\theta_{\mathcal{G}}^*). \tag{A.27}$$

It is convenient to express the null hypothesis $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$ as a functional dependence

$$H_0 : \gamma_{\mathcal{G}}^* = g(\gamma_{\mathcal{F}}^*), \tag{A.28}$$

where $g(\cdot)$ is a twice continuously differentiable function from $\Gamma_{\mathcal{F}}$ to $\Gamma_{\mathcal{G}}$ (see Gallant, 1987 and Vuong, 1989).[23] Denote by

$$G(\gamma_{\mathcal{F}}) = \frac{\partial g(\gamma_{\mathcal{F}})}{\partial \gamma_{\mathcal{F}}'} \tag{A.29}$$

the $k_2 \times k_1$ matrix of derivatives of $g(\gamma_{\mathcal{F}})$ with respect to $\gamma_{\mathcal{F}}$. Gallant (1987, p. 241) shows that

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)G(\gamma_{\mathcal{F}}^*) = \Psi^{\mathcal{G}}(g(\gamma_{\mathcal{F}}^*))G(\gamma_{\mathcal{F}}^*) = 0_{(k_2-k_1)\times k_1}. \tag{A.30}$$

Define the matrices

$$\mathbb{S} = [\Psi_*^{\mathcal{G}},\ 0_{(k_2-k_1)\times n}], \qquad \mathbb{Q} = \begin{bmatrix} G(\gamma_{\mathcal{F}}^*) & 0_{k_2\times n} \\ 0_{n\times k_1} & I_n \end{bmatrix} \tag{A.31}$$

and note that $\mathbb{S}\mathbb{Q} = 0_{(k_2-k_1)\times(n+k_1)}$. Then, using (A.27) and (A.28), it follows that (see Lemma B in Vuong, 1989)

$$\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \mathbb{Q}'\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \tag{A.32}$$

and

$$H_{\mathcal{F}} = \mathbb{Q}'H_{\mathcal{G}}\mathbb{Q}. \tag{A.33}$$

By Lemma A.1 and the fact that $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$ under the null, we obtain

$$\begin{aligned}
& T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \\
= \;& -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)'H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
= \;& -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'\mathbb{Q}(\mathbb{Q}'H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
= \;& \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-\frac{1}{2}}\left[I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\ \mathbb{Q}(\mathbb{Q}'H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'H^{\frac{1}{2}}_{\mathcal{G}}\right]H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1).
\end{aligned} \tag{A.34}$$

[23] Gallant (1987, Section 3.6) provides a discussion of these two alternative representations of the null hypothesis.

Using $\mathbb{S}\mathbb{Q} = 0_{(k_2-k_1)\times(n+k_1)}$, it can be shown that (see pp. 241–242 in Gallant, 1987)

$$I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\mathbb{Q}(\mathbb{Q}'H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'H_{\mathcal{G}}^{\frac{1}{2}} = H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}. \tag{A.35}$$

Substituting (A.35) into (A.34) yields

$$\begin{aligned} T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) &= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-\frac{1}{2}}[H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}]H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\ &= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \end{aligned} \tag{A.36}$$

Furthermore, invoking

$$\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \overset{A}{\sim} N\left(0_{n+k_2}, M_{\mathcal{G}}\right), \tag{A.37}$$

we have

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} \frac{1}{2}z'\left[M_{\mathcal{G}}^{\frac{1}{2}}H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}^{\frac{1}{2}}\right]z, \tag{A.38}$$

where $z \sim N(0_{n+k_2}, I_{n+k_2})$. Denote by $\Sigma_{\hat{\theta}_{\mathcal{G}}}$ the asymptotic covariance matrix of $\hat{\theta}_{\mathcal{G}}$ given in part (a) of Lemma 1. Since the eigenvalues of the matrix $\frac{1}{2}M_{\mathcal{G}}^{\frac{1}{2}}H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}^{\frac{1}{2}}$ are the same as the eigenvalues of the matrix

$$\frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}' = \frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}\Sigma_{\hat{\theta}_{\mathcal{G}}}\mathbb{S}' = (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1}\Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}, \tag{A.39}$$

we conclude that

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_2-k_1}(\xi), \tag{A.40}$$

where the ξ_i's are the eigenvalues of the matrix in (A.39). Since $A = \Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime}$ and $B = \Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}$ are two symmetric positive definite matrices, $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ is also symmetric positive definite with positive eigenvalues. Furthermore, because $A^{-1}B$ and $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ share the same eigenvalues, the eigenvalues of $A^{-1}B$ are also positive. This completes the proof of part (a).

(b) Note that, under the null and using the delta method,

$$\sqrt{T}\hat{\psi}_{\mathcal{G}} \overset{A}{\sim} N(0_{k_2-k_1}, \Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}). \tag{A.41}$$

Substituting consistent estimators for $\Psi_*^{\mathcal{G}}$ and $\Sigma_{\hat{\gamma}_{\mathcal{G}}}$ and constructing the Wald test delivers the desired result. This completes the proof of part (b). ■

Proof of Theorem 3. (a) Define the following matrices

$$\mathbb{S}_{\mathcal{F}} = \left[\Psi_*^{\mathcal{F}},\ 0_{(k_1-k_3)\times n}\right], \qquad \mathbb{S}_{\mathcal{G}} = \left[\Psi_*^{\mathcal{G}},\ 0_{(k_2-k_3)\times n}\right]. \tag{A.42}$$

Since $\mathcal{H} \subset \mathcal{F}$ and $\mathcal{H} \subset \mathcal{G}$, we can use the results from the proof of part (a) of Theorem 2 to obtain

$$T(\hat{\delta}^2_{\mathcal{H}} - \hat{\delta}^2_{\mathcal{F}}) = \frac{1}{2}\sqrt{T}\bar{v}^{\mathcal{F}}_T(\theta^*_{\mathcal{F}})'\mathcal{A}_{\mathcal{F}}\sqrt{T}\bar{v}^{\mathcal{F}}_T(\theta^*_{\mathcal{F}}) + o_p(1) \tag{A.43}$$

and

$$T(\hat{\delta}^2_{\mathcal{H}} - \hat{\delta}^2_{\mathcal{G}}) = \frac{1}{2}\sqrt{T}\bar{v}^{\mathcal{G}}_T(\theta^*_{\mathcal{G}})'\mathcal{A}_{\mathcal{G}}\sqrt{T}\bar{v}^{\mathcal{G}}_T(\theta^*_{\mathcal{G}}) + o_p(1), \tag{A.44}$$

where $\mathcal{A}_{\mathcal{F}} = H^{-1}_{\mathcal{F}}\mathbb{S}'_{\mathcal{F}}(\mathbb{S}_{\mathcal{F}}H^{-1}_{\mathcal{F}}\mathbb{S}'_{\mathcal{F}})^{-1}\mathbb{S}_{\mathcal{F}}H^{-1}_{\mathcal{F}}$ and $\mathcal{A}_{\mathcal{G}} = H^{-1}_{\mathcal{G}}\mathbb{S}'_{\mathcal{G}}(\mathbb{S}_{\mathcal{G}}H^{-1}_{\mathcal{G}}\mathbb{S}'_{\mathcal{G}})^{-1}\mathbb{S}_{\mathcal{G}}H^{-1}_{\mathcal{G}}$. Taking the difference yields

$$T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}) = -\frac{1}{2}\sqrt{T}\bar{v}^{\mathcal{F}}_T(\theta^*_{\mathcal{F}})'\mathcal{A}_{\mathcal{F}}\sqrt{T}\bar{v}^{\mathcal{F}}_T(\theta^*_{\mathcal{F}}) + \frac{1}{2}\sqrt{T}\bar{v}^{\mathcal{G}}_T(\theta^*_{\mathcal{G}})'\mathcal{A}_{\mathcal{G}}\sqrt{T}\bar{v}^{\mathcal{G}}_T(\theta^*_{\mathcal{G}}) + o_p(1). \tag{A.45}$$

From Assumptions A, B and C, the joint empirical process $\sqrt{T}[\bar{v}^{\mathcal{F}}_T(\theta^*_{\mathcal{F}})'\,,\bar{v}^{\mathcal{G}}_T(\theta^*_{\mathcal{G}})']'$ converges to a Gaussian process:

$$\begin{bmatrix} \sqrt{T}\bar{v}^{\mathcal{F}}_T(\theta^*_{\mathcal{F}}) \\ \sqrt{T}\bar{v}^{\mathcal{G}}_T(\theta^*_{\mathcal{G}}) \end{bmatrix} \overset{A}{\sim} N\left(0_{2n+k_1+k_2}, \mathbb{M}\right), \tag{A.46}$$

where

$$\mathbb{M} = \begin{bmatrix} M_{\mathcal{F}} & M_{\mathcal{F}\mathcal{G}} \\ M_{\mathcal{G}\mathcal{F}} & M_{\mathcal{G}} \end{bmatrix} = \lim_{T\to\infty} \mathrm{Var}\begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial \phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}})}{\partial \theta_{\mathcal{F}}} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial \phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}})}{\partial \theta_{\mathcal{G}}} \end{bmatrix}. \tag{A.47}$$

Hence,

$$T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}) \overset{A}{\sim} \frac{1}{2}z'\left[\mathbb{M}^{\frac{1}{2}}\begin{pmatrix} -\mathcal{A}_{\mathcal{F}} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & \mathcal{A}_{\mathcal{G}} \end{pmatrix}\mathbb{M}^{\frac{1}{2}}\right]z, \tag{A.48}$$

where $z \sim N(0_{2n+k_1+k_2}, I_{2n+k_1+k_2})$.

Then, using the fact that AB and BA share the same nonzero eigenvalues, the matrix in the square brackets in (A.48) has the same nonzero eigenvalues as the matrix

$$\begin{aligned} &\frac{1}{2}\begin{bmatrix} -(\mathbb{S}_{\mathcal{F}}H^{-1}_{\mathcal{F}}\mathbb{S}'_{\mathcal{F}})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\mathbb{S}_{\mathcal{G}}H^{-1}_{\mathcal{G}}\mathbb{S}'_{\mathcal{G}})^{-1} \end{bmatrix} \\ &\times \begin{bmatrix} \mathbb{S}_{\mathcal{F}}H^{-1}_{\mathcal{F}} & 0_{(k_1-k_3)\times(n+k_2)} \\ 0_{(k_2-k_3)\times(n+k_1)} & \mathbb{S}_{\mathcal{G}}H^{-1}_{\mathcal{G}} \end{bmatrix}\mathbb{M}\begin{bmatrix} H^{-1}_{\mathcal{F}}\mathbb{S}'_{\mathcal{F}} & 0_{(n+k_1)\times(k_2-k_3)} \\ 0_{(n+k_2)\times(k_1-k_3)} & H^{-1}_{\mathcal{G}}\mathbb{S}'_{\mathcal{G}} \end{bmatrix}. \end{aligned} \tag{A.49}$$

Using the fact that $\mathbb{S}_{\mathcal{F}}H^{-1}_{\mathcal{F}}\mathbb{S}'_{\mathcal{F}} = \frac{1}{2}\Psi^{\mathcal{F}}_*\tilde{H}_{\mathcal{F}}\Psi^{\mathcal{F}\prime}_*$, $\mathbb{S}_{\mathcal{G}}H^{-1}_{\mathcal{G}}\mathbb{S}'_{\mathcal{G}} = \frac{1}{2}\Psi^{\mathcal{G}}_*\tilde{H}_{\mathcal{G}}\Psi^{\mathcal{G}\prime}_*$ and

$$\Sigma_{\hat{\theta}_{\mathcal{F}\mathcal{G}}} = \begin{bmatrix} H^{-1}_{\mathcal{F}} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H^{-1}_{\mathcal{G}} \end{bmatrix}\mathbb{M}\begin{bmatrix} H^{-1}_{\mathcal{F}} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H^{-1}_{\mathcal{G}} \end{bmatrix} \tag{A.50}$$

is the asymptotic covariance matrix of $[\hat{\theta}'_{\mathcal{F}},\ \hat{\theta}'_{\mathcal{G}}]'$, the matrix in (A.49) can be written as

$$\begin{bmatrix} -(\Psi^{\mathcal{F}}_*\tilde{H}_{\mathcal{F}}\Psi^{\mathcal{F}\prime}_*)^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi^{\mathcal{G}}_*\tilde{H}_{\mathcal{G}}\Psi^{\mathcal{G}\prime}_*)^{-1} \end{bmatrix}\Psi^{\mathcal{F}\mathcal{G}}_*\Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}}\Psi^{\mathcal{F}\mathcal{G}\prime}_*. \tag{A.51}$$

Therefore,

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_1+k_2-2k_3}(\xi), \tag{A.52}$$

where the ξ_i's are the eigenvalues of the matrix in (A.51). This completes the proof of part (a).

(b) By the delta method,

$$\sqrt{T}\hat{\psi}_{\mathcal{FG}} \overset{A}{\sim} N(0_{k_1+k_2-2k_3}, \Psi_*^{\mathcal{FG}} \Sigma_{\hat{\gamma}_{\mathcal{FG}}} \Psi_*^{\mathcal{FG}\prime}). \tag{A.53}$$

Using consistent estimators of $\Psi_*^{\mathcal{FG}}$ and $\Sigma_{\hat{\gamma}_{\mathcal{FG}}}$ for constructing

$$Wald_{\hat{\psi}_{\mathcal{FG}}} = T\hat{\psi}_{\mathcal{FG}}'(\hat{\Psi}^{\mathcal{FG}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}\hat{\Psi}^{\mathcal{FG}\prime})^{-1}\hat{\psi}_{\mathcal{FG}}, \tag{A.54}$$

we obtain immediately

$$Wald_{\hat{\psi}_{\mathcal{FG}}} \overset{A}{\sim} \chi^2_{k_1+k_2-2k_3}. \tag{A.55}$$

This completes the proof of part (b). ■

C. Computation of the p-values for the multivariate inequality test

The biggest hurdle in determining the p-value of LR in (52) is the computation of the weights. For a given $p \times p$ covariance matrix $\Omega = (\omega_{ij})$, the expressions for the weights $w_i(\Omega)$, $i = 0, \ldots, p$, are given in Kudo (1963). The weights depend on Ω through the correlation coefficients $\rho_{ij} = \omega_{ij}/(\omega_i\omega_j)$. When $p = 1$, $w_0 = w_1 = 1/2$. For $p = 2$, $w_0 = 1/2 - w_2$, $w_1 = 1/2$ and $w_2 = 1/4 + \arcsin(\rho_{12})/(2\pi)$. When $p = 3$, $w_0 = 1/2 - w_2$, $w_1 = 1/2 - w_3$,

$$w_2 = \frac{3}{8} + \frac{\arcsin(\rho_{12\cdot3}) + \arcsin(\rho_{13\cdot2}) + \arcsin(\rho_{23\cdot1})}{4\pi}, \tag{A.56}$$

and

$$w_3 = \frac{1}{8} + \frac{\arcsin(\rho_{12}) + \arcsin(\rho_{13}) + \arcsin(\rho_{23})}{4\pi}, \tag{A.57}$$

where

$$\rho_{ij\cdot k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{[(1-\rho_{ik}^2)(1-\rho_{jk}^2)]^{\frac{1}{2}}}. \tag{A.58}$$

For $p > 3$, the computation of the weights is more complicated. Following Kudo (1963), let $\mathcal{P} = \{1, \ldots, p\}$. There are 2^p subsets of $\mathcal{P}$, which are indexed by M. Let $n(M)$ be the number of elements in M and M' be the complement of M relative to $\mathcal{P}$. Define Ω_M as the submatrix of Ω

that consists of the rows and columns in the set M, $\Omega_{M'}$ as the submatrix of Ω that consists of the rows and columns in the set M', $\Omega_{M,M'}$ the submatrix of Ω with rows corresponding to the elements in M and columns corresponding to the elements in M' ($\Omega_{M',M}$ is similarly defined), and $\Omega_{M\cdot M'} = \Omega_M - \Omega_{M,M'}\Omega_{M'}^{-1}\Omega_{M',M}$. Kudo (1963) shows that

$$w_i(\Omega) = \sum_{M:\, n(M)=i} P(\Omega_{M'}^{-1})P(\Omega_{M\cdot M'}), \tag{A.59}$$

where $P(A)$ is the probability for a multivariate normal distribution with zero mean and covariance matrix A to have all positive elements. In the above equation, we use the convention that $P[\Omega_{\emptyset\cdot\mathcal{P}}] = 1$ and $P[\Omega_{\emptyset}^{-1}] = 1$. Using (A.59), we have $w_0(\Omega) = P(\Omega^{-1})$ and $w_p(\Omega) = P(\Omega)$.

Researchers have typically used a Monte Carlo approach to compute the positive orthant probability $P(A)$. However, the Monte Carlo approach is not efficient because it requires a large number of simulations to achieve the accuracy of a few digits, even when p is relatively small.

We overcome this problem by using a formula for the positive orthant probability due to Childs (1967) and Sun (1988a). Let $R = (r_{ij})$ be the correlation matrix corresponding to A. Childs (1967) and Sun (1988a) show that

$$\begin{aligned} P_{2k}(A) &= \frac{1}{2^{2k}} + \frac{1}{2^{2k-1}\pi} \sum_{1\le i<j\le 2k} \arcsin(r_{ij}) \\ &\quad + \sum_{j=2}^{k} \frac{1}{2^{2k-j}\pi^j} \sum_{1\le i_1<\cdots<i_{2j}\le 2k} I_{2j}\left(R_{(i_1,\ldots,i_{2j})}\right), \end{aligned} \tag{A.60}$$

$$\begin{aligned} P_{2k+1}(A) &= \frac{1}{2^{2k+1}} + \frac{1}{2^{2k}\pi} \sum_{1\le i<j\le 2k+1} \arcsin(r_{ij}) \\ &\quad + \sum_{j=2}^{k} \frac{1}{2^{2k+1-j}\pi^j} \sum_{1\le i_1<\cdots<i_{2j}\le 2k+1} I_{2j}\left(R_{(i_1,\ldots,i_{2j})}\right), \end{aligned} \tag{A.61}$$

where $R_{(i_1,\ldots,i_{2j})}$ denotes the submatrix consisting of the $(i_1,\ldots,i_{2j})$-th rows and columns of R, and

$$I_{2j}(\Lambda) = \frac{(-1)^j}{(2\pi)^j} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{2j} \frac{1}{\omega_i}\right) \exp\left(-\frac{\omega'\Lambda\omega}{2}\right) \mathrm{d}\omega_1 \cdots \mathrm{d}\omega_{2j}, \tag{A.62}$$

where Λ is a $2j \times 2j$ covariance matrix and $\omega = (\omega_1,\ldots,\omega_{2j})'$. Sun (1988a) provides a recursive relation for $I_{2j}(\Lambda)$ that allows us to obtain I_{2j} starting from I_2. Sun's formula enables us to compute the $2j$-th order multivariate integral I_{2j} using a $(j-1)$-th order multivariate integral, which can be obtained numerically using the Gauss-Legendre quadrature method. Sun (1988b) provides a

Fortran subroutine to compute $P(A)$ for $p \leq 9$. We improve on Sun's program and are able to accurately compute $P(A)$ and hence $w_i(\Omega)$ for $p \leq 11$.

D. Description of the simulation designs for linear and nonlinear models

Since generating data from correctly specified and misspecified models (nested and non-nested) that are calibrated to actual data is nontrivial, we present the details of our simulation designs. Let R_t be an n-vector of payoffs and f_t be a $(k-1)$-vector of risk factors. For linear models, let $Y_t = [f_t',\ R_t']'$. For log-linear models, let $Y_t = [f_t',\ r_t']'$, where $r_t = \ln(R_t)$. In addition, let

$$\mu = E[Y_t] = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{A.63}$$

and

$$V = \mathrm{Var}[Y_t] = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{A.64}$$

Let $\hat{\mu}_2$ and $\hat{V}$ denote the sample estimates obtained from actual data. In all our simulations, the covariance matrix of the factors and returns, V, is chosen based on the covariance matrix estimated from the data, i.e., $V = \hat{V}$. In addition, since the results are invariant to the mean of the factors, without loss of generality we set $\mu_1 = 0_{k-1}$ in all simulation designs. In this appendix, we discuss how to choose μ_2 and the SDF parameters γ, such that the model is correctly specified or misspecified and the γ's satisfy certain restrictions (for example, $\gamma_i = 0$).

Linear models

Correctly specified models. For a given $\gamma = [\gamma_0,\ \gamma_1']'$, the pricing errors of a linear SDF $y_t(\gamma) = \gamma_0 + \gamma_1' f_t$ are given by

$$e = E[R_t y_t(\gamma)] - 1_n = \mu_2\gamma_0 + V_{21}\gamma_1 - 1_n. \tag{A.65}$$

The vector of pricing errors can be re-written as

$$e = \gamma_0 \left(\mu_2 + V_{21}\frac{\gamma_1}{\gamma_0} - 1_n\frac{1}{\gamma_0} \right) = \frac{\mu_2 - V_{21}\eta_1 - 1_n\eta_0}{\eta_0}, \tag{A.66}$$

where $\eta_0 = 1/\gamma_0$ and $\eta_1 = -\gamma_1/\gamma_0$. From this equation, we can see that in order for the model to be correctly specified, we need to set $\mu_2 = X\eta$, where $X = [1_n,\ V_{21}]$ and $\eta = [\eta_0,\ \eta_1']'$.

When we do not need to impose restrictions on the parameter vector γ (as in the size experiment in Table 2.A), η is chosen such that the implied μ_2 is as close as possible to $\hat{\mu}_2$, i.e.,

$$\min_{\eta}(\hat{\mu}_2 - X\eta)' V_{22}^{-1}(\hat{\mu}_2 - X\eta). \tag{A.67}$$

The solution to this minimization problem can be easily obtained as

$$\eta^* = (X' V_{22}^{-1} X)^{-1}(X' V_{22}^{-1}\hat{\mu}_2). \tag{A.68}$$

The solution η^* is then used to obtain $\mu_2 = X\eta^*$ and back out the original parameters γ^* as

$$\gamma_1^* = -\frac{\eta_1^*}{\eta_0^*}, \qquad \gamma_0^* = \frac{1}{\eta_0^*}. \tag{A.69}$$

Under correctly specified models, interest may lie in evaluating the statistical significance of individual coefficients as well as imposing the null hypothesis of SDF equality in nested and overlapping model comparison tests. These are situations in which we need to choose η (or, equivalently, γ) subject to certain constraints, say $g(\eta) = 0_m$, by solving the following problem:

$$\min_{\eta}(\hat{\mu}_2 - X\eta)' V_{22}^{-1}(\hat{\mu}_2 - X\eta) \quad \text{s.t.} \quad g(\eta) = 0_m. \tag{A.70}$$

In the case of zero restrictions on the parameters, let $\eta_1 = [\eta_{1a}', \ \eta_{1b}']'$, where η_{1a} is $k_1 \times 1$ and η_{1b} is $k_2 \times 1$, and $k_1 + k_2 = k - 1$. Without loss of generality, we set $\eta_{1b} = 0_{k_2}$. This can be easily accomplished by choosing $\tilde{\eta}^* = [\tilde{\eta}_c^{*\prime}, \ 0_{k_2}']'$, where

$$\tilde{\eta}_c^* = (X_c' V_{22}^{-1} X_c)^{-1} X_c' V_{22}^{-1}\hat{\mu}_2, \tag{A.71}$$

and $X_c = [1_n, \ V_{21}^a]$ with V_{21}^a being the first k_1 columns of V_{21}. As before, $\mu_2 = X\tilde{\eta}^*$ and the values of the original parameters γ^* can be recovered from $\tilde{\eta}^*$.

Potentially misspecified models. We now discuss the issue of how to set μ_2 such that the model is misspecified but the pseudo-true SDF parameters are still equal to a given value of γ. Let $D = [\mu_2, \ V_{21}]$. From the definition of pseudo-true SDF parameters, we have

$$\gamma = (D' V_{22}^{-1} D)^{-1} D' V_{22}^{-1} 1_n. \tag{A.72}$$

Multiplying both sides of the equation above by $D' V_{22}^{-1} D$, we obtain the following first order conditions:

$$\mu_2' V_{22}^{-1}(\mu_2\gamma_0 + V_{21}\gamma_1 - 1_n) \;=\; 0, \tag{A.73}$$

$$V_{12} V_{22}^{-1}(\mu_2\gamma_0 + V_{21}\gamma_1 - 1_n) \;=\; 0_{k-1}. \tag{A.74}$$

Let z be an n-vector of constants and

$$\mu_2 = \frac{1_n - V_{21}\gamma_1 + z}{\gamma_0} = (z + 1_n)\eta_0 + V_{21}\eta_1 = X\eta + z\eta_0. \tag{A.75}$$

Then, the above first order conditions imply that z must satisfy

$$V_{12}V_{22}^{-1}z = 0_{k-1}, \tag{A.76}$$

$$z'V_{22}^{-1}1_n = -z'V_{22}^{-1}z. \tag{A.77}$$

There are many possible vectors z that can satisfy these conditions. In order to have a misspecified model with pseudo-true parameters γ, a convenient choice of z is $\hat{e}$, where

$$\hat{e} = \hat{D}(\hat{D}'V_{22}^{-1}\hat{D})^{-1}(\hat{D}'V_{22}^{-1}1_n) - 1_n, \tag{A.78}$$

and $\hat{D} = [\hat{\mu}_2,\ V_{21}]$. Note that $\hat{e}$ has the following properties: (1) $\hat{D}'V_{22}^{-1}\hat{e} = 0_k$, which implies $V_{12}V_{22}^{-1}\hat{e} = 0_{k-1}$, and (2) $\hat{e}'V_{22}^{-1}1_n = -\hat{e}'V_{22}^{-1}\hat{e}$. In the simulations (Tables 1.A and 1.B, Tables 3.A and 3.B (size experiments), Table 4.A (size experiments)), we use this choice of z and set the mean of the returns for misspecified models as $\mu_2 = X\tilde{\eta}^* + \hat{e}\tilde{\eta}_0^*$, where $\tilde{\eta}^*$ is obtained from (A.71).

Log-linear models

Suppose that $y_t(\gamma) = \exp(\gamma_0 + \gamma_1' f_t)$ and

$$\begin{bmatrix} f_t \\ r_t \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}\right). \tag{A.79}$$

Again, the goal is to set μ_2 such that the model is correctly specified or misspecified with a desired parameter vector γ.

Correctly specified models. For a given $\gamma = [\gamma_0,\ \gamma_1']'$, using the properties of the log-normal distribution, we can obtain the pricing errors of the log-linear model as

$$\begin{aligned} e &= E[R_t y_t(\gamma)] - 1_n = E[\exp(r_t + \gamma_0 + \gamma_1' f_t)] - 1_n \\ &= \exp\left(\gamma_0 + \mu_2 + \frac{\gamma_1' V_{11}\gamma_1}{2} + V_{21}\gamma_1 + \frac{v_r}{2}\right) - 1_n, \end{aligned} \tag{A.80}$$

where $v_r = \mathrm{Diag}(V_{22})$.

For convenience, let $\tilde{\mu}_2 = \mu_2 + (v_r/2)$ and $\tilde{\gamma}_0 = \gamma_0 + (\gamma_1' V_{11}\gamma_1)/2$. We can then write the vector of pricing errors as

$$e = \exp\left(\tilde{\mu}_2 + 1_n\tilde{\gamma}_0 + V_{21}\gamma_1\right) - 1_n = \exp\left(\tilde{\mu}_2 + X\tilde{\gamma}\right) - 1_n, \tag{A.81}$$

where $X = [1_n,\ V_{21}]$ and $\tilde{\gamma} = [\tilde{\gamma}_0,\ \gamma_1']'$.

From the above equation, we can see that a model is correctly specified if and only if $\tilde{\mu}_2 = -X\tilde{\gamma}$, or equivalently

$$\mu_2 = -\frac{v_r}{2} - X\tilde{\gamma}. \tag{A.82}$$

Therefore, for any given value of $\tilde{\gamma}$, we can set μ_2 as in (A.82) and the model will be correctly specified.

When we do not need to impose restrictions on the parameter vector $\tilde{\gamma}$ (as in the size experiment in Table 2.B), $\tilde{\gamma}$ is determined so that the implied μ_2 is as close as possible to $\hat{\mu}_2$, i.e.,

$$\min_{\tilde{\gamma}}(\hat{\mu}_2 - \mu_2)'V_{22}^{-1}(\hat{\mu}_2 - \mu_2) = \min_{\tilde{\gamma}}\left(\hat{\mu}_2 + \frac{v_r}{2} + X\tilde{\gamma}\right)' V_{22}^{-1}\left(\hat{\mu}_2 + \frac{v_r}{2} + X\tilde{\gamma}\right). \tag{A.83}$$

The solution to this minimization problem is given by

$$\tilde{\gamma}^* = -(X'V_{22}^{-1}X)^{-1}X'V_{22}^{-1}\left(\hat{\mu}_2 + \frac{v_r}{2}\right). \tag{A.84}$$

As in the linear case, we may need to choose $\tilde{\gamma}$ subject to certain constraints, say $g(\tilde{\gamma}) = 0_m$. In these situations, we choose $\tilde{\gamma}$ by solving the following problem:

$$\min_{\tilde{\gamma}}\left(\hat{\mu}_2 + \frac{v_r}{2} + X\tilde{\gamma}\right)' V_{22}^{-1}\left(\hat{\mu}_2 + \frac{v_r}{2} + X\tilde{\gamma}\right) \quad \text{s.t.} \quad g(\tilde{\gamma}) = 0_m. \tag{A.85}$$

In the case of zero restrictions on the parameters, let $\gamma_1 = [\gamma_{1a}',\ \gamma_{1b}']'$, where γ_{1a} is $k_1 \times 1$, γ_{1b} is $k_2 \times 1$, and $k_1 + k_2 = k - 1$. Without loss of generality, we set $\gamma_{1b} = 0_{k_2}$. This can be accomplished by choosing $\tilde{\gamma}^* = [\tilde{\gamma}_c^{*\prime},\ 0_{k_2}']'$, where

$$\tilde{\gamma}_c^* = -(X_c'V_{22}^{-1}X_c)^{-1}X_c'V_{22}^{-1}\left(\hat{\mu}_2 + \frac{v_r}{2}\right), \tag{A.86}$$

and $X_c = [1_n,\ V_{21}^a]$ with V_{21}^a being the first k_1 columns of V_{21}. We can then set $\mu_2 = -\frac{v_r}{2} - X\tilde{\gamma}^*$.

Potentially misspecified models. Setting μ_2 such that the model is misspecified and the pseudo-true SDF parameters are equal to $\tilde{\gamma}$ proves to be a more challenging task. Without loss of generality, we assume

$$\tilde{\mu}_2 = -X\tilde{\gamma} + z, \tag{A.87}$$

where z is an n-vector of constants. When the pseudo-true SDF parameters are equal to $\tilde{\gamma}$, there are k constraints on z, given by the following first order conditions:

$$D'U^{-1}e = 0_k, \tag{A.88}$$

where $U = E[R_tR_t']$, $e = E[R_ty_t(\tilde{\gamma})] - 1_n$, and

$$D = E\left[R_t\frac{\partial y_t(\tilde{\gamma})}{\partial\tilde{\gamma}'}\right]. \tag{A.89}$$

Note that with the reparametrization from γ to $\tilde{\gamma}$, $y_t(\tilde{\gamma})$ is given by

$$y_t(\tilde{\gamma}) = \exp\left(\tilde{\gamma}_0 - \frac{\tilde{\gamma}_1'V_{11}\tilde{\gamma}_1}{2} + \tilde{\gamma}_1'f_t\right). \tag{A.90}$$

We first derive explicit expressions for e, U and D. For e, we have

$$\begin{aligned} e &= E[R_ty_t(\tilde{\gamma})] - 1_n \\ &= E\left[\exp\left(r_t + \tilde{\gamma}_0 - \frac{\tilde{\gamma}_1'V_{11}\tilde{\gamma}_1}{2} + \tilde{\gamma}_1'f_t\right)\right] - 1_n \\ &= \exp(z) - 1_n \\ &= \tilde{z} - 1_n, \end{aligned} \tag{A.91}$$

by denoting $\tilde{z} = \exp(z)$. To obtain U, we need to compute $E[R_{it}R_{jt}]$. Note that

$$E[\exp(r_{it} + r_{jt})] = \exp\left(\mu_i + \mu_j + \frac{v_{ii}}{2} + \frac{v_{jj}}{2} + v_{ij}\right) = E[R_{it}]E[R_{jt}]\exp(v_{ij}), \tag{A.92}$$

where $v_{ii} = \mathrm{Var}[r_{it}]$ and $v_{ij} = \mathrm{Cov}[r_{it}, r_{jt}]$. With this result, we have

$$U = E[R_tR_t'] = E[R_t]E[R_t]' \circ \exp(V_{22}), \tag{A.93}$$

where $\circ$ is the Hadamard product and

$$E[R_t] = \exp(\tilde{\mu}_2) = \exp(-X\tilde{\gamma}) \circ \tilde{z}. \tag{A.94}$$

We now turn our attention to D and derive the expression for its first column. It is straightforward to show that

$$E\left[R_t\frac{\partial y_t(\tilde{\gamma})}{\partial\tilde{\gamma}_0}\right] = E[R_ty_t(\tilde{\gamma})] = 1_n + e = \tilde{z}. \tag{A.95}$$

To derive the other columns of D, we need to compute the expectation of

$$R_t\frac{\partial y_t(\tilde{\gamma})}{\partial\tilde{\gamma}_1'} = R_ty_t(\tilde{\gamma})(-\tilde{\gamma}_1'V_{11} + f_t'). \tag{A.96}$$

We appeal to Stein's lemma, which states that if (u, v) are bivariate normally distributed, we have

$$\mathrm{Cov}[u, g(v)] = \sigma_{uv} E[g'(v)] \tag{A.97}$$

when $E[g'(v)]$ exists. Using Stein's lemma, we can easily obtain

$$\begin{aligned} E\left[R_t \frac{\partial y_t(\tilde{\gamma})}{\partial \tilde{\gamma}_1'}\right] &= E\left[e^{r_t} y_t(\tilde{\gamma})(-\tilde{\gamma}_1' V_{11} + f_t')\right] \\ &= -\tilde{z}\tilde{\gamma}_1' V_{11} + E\left[e^{r_t + \tilde{\gamma}_0 - (\tilde{\gamma}_1' V_{11} \tilde{\gamma}_1)/2 + \tilde{\gamma}_1' f_t} f_t'\right] \\ &= -\tilde{z}\tilde{\gamma}_1' V_{11} + (V_{21} + 1_n \tilde{\gamma}_1' V_{11}) \circ (\tilde{z} 1_{k-1}') \\ &= V_{21} \circ (\tilde{z} 1_{k-1}'). \end{aligned} \tag{A.98}$$

Therefore, we can write D as

$$D = [\tilde{z},\ V_{21} \circ (\tilde{z} 1_{k-1}')]. \tag{A.99}$$

Having expressed D, U and e as functions of $\tilde{z}$, $\tilde{z}$ can be obtained numerically by minimizing the distance between $\hat{\mu}_2$ and μ_2 subject to the k constraints in (A.88). Denote the solution for $\tilde{z}$ in this minimization problem by $\tilde{z}^*$. In the simulations (Tables 1.C and 1.D, Tables 3.C and 3.D (size experiments), Table 4.C (size experiments)), we use $\tilde{z}^*$ and set the mean of the returns for misspecified models as $\mu_2 = -\frac{v_r}{2} - X\tilde{\gamma}^* + \tilde{z}^*$, where $\tilde{\gamma}^*$ is obtained from (A.86).

References

Abel, A., 1990, Asset prices under habit formation and catching up with the Joneses. American Economic Review 80, 38–42.

Ahn, S. C. and C. Gadarowski, 2004, Small sample properties of the GMM specification test based on the Hansen-Jagannathan distance. Journal of Empirical Finance 11, 109–132.

Almeida, C. and R. Garcia, 2012, Assessing misspecified asset pricing models with empirical likelihood estimators. Journal of Econometrics 170, 519–537.

Andrews, D. W. K., 1993, An introduction to econometric applications of empirical process theory for dependent random variables. Econometric Reviews 12, 183–216.

Andrews, D. W. K., 1994, Empirical process methods in econometrics, in: R. F. Engle and D. L. McFadden, (Eds.), Handbook of Econometrics. North-Holland, Amsterdam, pp. 2247–2294.

Andrews, D. W. K. and G. Soares, 2010, Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–158.

Brown, D. P. and M. Gibbons, 1985, A simple econometric approach for utility-based asset pricing models. Journal of Finance 40, 359–381.

Chen, X., H. Hong and M. Shum, 2007, Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. Journal of Econometrics 141, 109–140.

Chen, X. and S. C. Ludvigson, 2009, Land of addicts? An empirical investigation of habit-based asset pricing models. Journal of Applied Econometrics 24, 1057–1093.

Childs, D. R., 1967, Reduction of the multivariate normal integral to characteristic form. Biometrika 54, 293–300.

Cochrane, J. H., 2005, Asset pricing. Princeton University Press, Princeton.

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

Gallant, R. A., 1987, Nonlinear statistical models. Wiley, New York.

Golden, R. M., 2003, Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. Psychometrika 68, 229–249.

Gospodinov, N., R. Kan and C. Robotti, 2011, On the Hansen-Jagannathan distance with a no-arbitrage constraint. Discussion Paper. Federal Reserve Bank of Atlanta.

Gospodinov, N., R. Kan and C. Robotti, 2012, Further results on the limiting distribution of GMM sample moment conditions. Journal of Business and Economic Statistics 30, 494–504.

Hall, A. R. and D. Pelletier, 2011, Nonnested testing in models estimated via generalized method of moments. Econometric Theory 27, 443–456.

Hansen, L. P., J. C. Heaton and E. G. J. Luttmer, 1995, Econometric evaluation of asset pricing models. Review of Financial Studies 8, 237–274.

Hansen, L. P. and R. Jagannathan, 1997, Assessing specification errors in stochastic discount factor models. Journal of Finance 52, 557–590.

Hansen, P. R., 2005, A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365–380.

Imbens, G. W., R. H. Spady and P. Johnson, 1998, Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

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

Kan, R. and C. Robotti, 2008, Specification tests of asset pricing models using excess returns. Journal of Empirical Finance 15, 816–838.

Kan, R. and C. Robotti, 2009, Model comparison using the Hansen-Jagannathan distance. Review of Financial Studies 22, 3449–3490.

Kan, R., C. Robotti and J. Shanken, 2012, Pricing model performance and the two-pass cross-sectional regression methodology. Journal of Finance, forthcoming.

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

Kitamura, Y., 2000, Comparing misspecified dynamic econometric models using nonparametric likelihood. Working Paper. University of Pennsylvania.

Kudo, A., 1963, A multivariate analogue of the one-sided test. Biometrika 50, 403–418.

Lettau, M. and S. C. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.

Li, H., Y. Xu and X. Zhang, 2010, Evaluating asset pricing models using the second Hansen-Jagannathan distance. Journal of Financial Economics 97, 279–301.

Lien, D. and Q. H. Vuong, 1987, Selecting the best linear regression model: A classical approach. Journal of Econometrics 35, 3–23.

Marcellino, M. and B. Rossi, 2008, Model selection for nested and overlapping nonlinear, dynamic and possibly mis-specified models. Oxford Bulletin of Economics and Statistics 70, 867–893.

Nagel, S. and K. J. Singleton, 2011, Estimation and evaluation of conditional asset pricing models. Journal of Finance 66, 873–909.

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

Newey, W. K. and K. D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 55, 703–708.

Parker, J. A. and C. Julliard, 2005, Consumption risk and the cross section of expected returns. Journal of Political Economy 113, 185–222.

Pollard, D., 1982, A central limit theorem for k-means clustering. Annals of Probability 10, 919–926.

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

Smith, R. J., 1997, Alternative semi-parametric likelihood approaches to generalized method of moments estimation. Economic Journal 107, 503–519.

Sowell, F. B., 1996, Optimal tests for parameter instability in the generalized method of moments framework. Econometrica 64, 1085–1107.

Stock, J. H. and J. H. Wright, 2000, GMM with weak identification. Econometrica 68, 1055–1096.

Sun, H., 1988a, A general reduction method for n-variate normal orthant probability. Communications in Statistics – Theory and Methods 11, 3913–3921.

Sun, H., 1988b, A Fortran subroutine for computing normal orthant probabilities of dimensions up to nine. Communications in Statistics – Simulation and Computation 17, 1097–1111.

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

White, H., 2000, A reality check for data snooping. Econometrica 68, 1097–1126.

Wolak, F. A., 1987, An exact test for multiple inequality and equality constraints in the linear regression model. Journal of the American Statistical Association 82, 782–793.

Wolak, F. A., 1989, Testing inequality constraints in linear econometric models. Journal of Econometrics 31, 205–235.

Yogo, M., 2006, A consumption-based explanation of expected stock returns. Journal of Finance 61, 539–580.

Table 1. t-tests under model misspecification

Panel A: t-tests under potentially misspecified models (linear YOGO model)

	Size (r_{mkt})			Size ($\triangle c_{ndur}$)			Size ($\triangle c_{dur}$)		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.123	0.064	0.013	0.084	0.037	0.006	0.085	0.038	0.006
240	0.118	0.061	0.013	0.084	0.040	0.006	0.084	0.038	0.006
360	0.115	0.059	0.013	0.088	0.040	0.007	0.087	0.040	0.006
480	0.110	0.057	0.013	0.089	0.042	0.007	0.089	0.042	0.007
600	0.108	0.057	0.012	0.090	0.043	0.008	0.091	0.043	0.008

Panel B: t-tests under correctly specified models (linear YOGO model)

	Size (r_{mkt})			Size ($\triangle c_{ndur}$)			Size ($\triangle c_{dur}$)		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.217	0.137	0.043	0.284	0.192	0.068	0.293	0.200	0.074
240	0.245	0.162	0.063	0.317	0.226	0.098	0.330	0.237	0.106
360	0.256	0.175	0.073	0.337	0.248	0.118	0.356	0.265	0.130
480	0.265	0.183	0.078	0.350	0.261	0.131	0.372	0.283	0.148
600	0.269	0.187	0.081	0.358	0.269	0.140	0.385	0.295	0.160

Panel C: t-tests under potentially misspecified models (nonlinear YOGO model)

	Size ($\ln(R_{mkt})$)			Size ($\triangle c_{ndur}$)			Size ($\triangle c_{dur}$)		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.093	0.041	0.004	0.094	0.041	0.004	0.101	0.044	0.005
240	0.106	0.054	0.010	0.099	0.051	0.010	0.111	0.059	0.012
360	0.105	0.054	0.012	0.095	0.049	0.011	0.107	0.059	0.014
480	0.104	0.054	0.012	0.088	0.044	0.010	0.101	0.054	0.014
600	0.102	0.053	0.011	0.083	0.040	0.009	0.093	0.049	0.013

Panel D: t-tests under correctly specified models (nonlinear YOGO model)

	Size ($\ln(R_{mkt})$)			Size ($\triangle c_{ndur}$)			Size ($\triangle c_{dur}$)		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.219	0.136	0.039	0.281	0.182	0.049	0.293	0.190	0.052
240	0.256	0.172	0.068	0.332	0.242	0.111	0.355	0.265	0.126
360	0.274	0.191	0.082	0.354	0.268	0.141	0.383	0.299	0.166
480	0.287	0.204	0.091	0.365	0.280	0.155	0.397	0.313	0.188
600	0.299	0.214	0.097	0.368	0.285	0.161	0.408	0.325	0.197

The table presents the empirical size of t-tests of $H_0 : \gamma_i = 0$. We report results for different levels of significance (10%, 5% and 1% levels) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the factors and the gross returns (continuously compounded gross returns in the nonlinear case) are generated from a multivariate normal distribution. The various t-ratios are compared to the critical values from a standard normal distribution. Panels A and C report the sizes of t-tests under potentially misspecified models based on the asymptotic covariance matrix in (23) and (24), while Panels B and D report the sizes of t-tests under correctly specified models based on the asymptotic covariance matrix in (26) and (27).

Table 2. Specification tests

Panel A: Linear YOGO model

	HJ-distance test						LM test					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.269	0.177	0.065	0.991	0.982	0.947	0.122	0.058	0.009	0.982	0.959	0.847
240	0.164	0.094	0.026	1.000	1.000	0.999	0.103	0.049	0.009	1.000	1.000	0.998
360	0.139	0.076	0.019	1.000	1.000	1.000	0.099	0.049	0.009	1.000	1.000	1.000
480	0.127	0.068	0.016	1.000	1.000	1.000	0.098	0.048	0.009	1.000	1.000	1.000
600	0.120	0.064	0.015	1.000	1.000	1.000	0.097	0.048	0.009	1.000	1.000	1.000

Panel B: Nonlinear YOGO model

	HJ-distance test						LM test					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.267	0.173	0.062	0.986	0.970	0.910	0.136	0.069	0.012	0.995	0.985	0.921
240	0.164	0.095	0.026	0.997	0.995	0.983	0.106	0.053	0.009	1.000	1.000	1.000
360	0.140	0.077	0.019	0.999	0.998	0.994	0.101	0.050	0.009	1.000	1.000	1.000
480	0.128	0.069	0.016	1.000	0.999	0.997	0.100	0.050	0.009	1.000	1.000	1.000
600	0.123	0.065	0.015	1.000	0.999	0.998	0.101	0.049	0.009	1.000	1.000	1.000

The table presents the empirical size and power of the conventional HJ-distance test and the LM test in Theorem 1. We report results for different levels of significance (10%, 5% and 1% levels) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the factors and the gross returns (continuously compounded gross returns in the nonlinear case) are generated from a multivariate normal distribution.

Table 3. Model selection tests for nested SDFs

Panel A: Pairwise model comparison tests: Linear models

	Weighted χ^2 test						Wald test					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.089	0.038	0.005	0.337	0.198	0.046	0.107	0.051	0.009	0.468	0.334	0.129
240	0.090	0.040	0.006	0.508	0.346	0.114	0.109	0.054	0.010	0.650	0.526	0.285
360	0.093	0.043	0.006	0.636	0.478	0.195	0.107	0.054	0.010	0.760	0.657	0.423
480	0.094	0.045	0.007	0.734	0.588	0.286	0.107	0.054	0.011	0.836	0.750	0.537
600	0.093	0.045	0.007	0.807	0.682	0.377	0.105	0.053	0.011	0.887	0.819	0.633

Panel B: Multiple model comparison test: Linear models

	Size			Power		
T	10%	5%	1%	10%	5%	1%
120	0.085	0.038	0.006	0.370	0.243	0.082
240	0.087	0.041	0.006	0.572	0.441	0.212
360	0.089	0.041	0.008	0.715	0.598	0.357
480	0.091	0.044	0.008	0.811	0.715	0.492
600	0.092	0.045	0.008	0.880	0.806	0.610

Panel C: Pairwise model comparison tests: Nonlinear models

	Weighted χ^2 test						Wald test					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.059	0.020	0.001	0.280	0.146	0.024	0.092	0.036	0.003	0.397	0.240	0.047
240	0.072	0.027	0.003	0.460	0.304	0.093	0.125	0.063	0.011	0.635	0.508	0.254
360	0.074	0.030	0.003	0.574	0.418	0.165	0.126	0.069	0.015	0.742	0.638	0.405
480	0.073	0.030	0.003	0.661	0.512	0.236	0.121	0.066	0.016	0.812	0.727	0.519
600	0.071	0.029	0.003	0.733	0.597	0.313	0.115	0.063	0.016	0.861	0.792	0.611

Panel D: Multiple model comparison test: Nonlinear models

	Size			Power		
T	10%	5%	1%	10%	5%	1%
120	0.078	0.031	0.003	0.363	0.215	0.047
240	0.106	0.052	0.009	0.645	0.515	0.263
360	0.107	0.054	0.011	0.776	0.675	0.444
480	0.104	0.054	0.012	0.859	0.781	0.582
600	0.101	0.052	0.012	0.910	0.853	0.691

The table presents the empirical size and power of pairwise and multiple model comparison tests for nested linear (Panels A and B) and nonlinear (Panels C and D) models. In Panels A and C, we report results for the weighted chi-squared test in part (a) of Theorem 2 and the Wald test in part (b) of Theorem 2. Panels B and D are for the Wald test for multiple nested model comparison in (54). We report results for different levels of significance (10%, 5% and 1% levels) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the factors and the gross returns (continuously compounded gross returns in the nonlinear case) are generated from a multivariate normal distribution.

Table 4. Model selection tests for non-nested SDFs

Panel A: Pairwise tests of equality: Linear models

	Weighted χ^2 test						Wald test					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.079	0.035	0.005	0.648	0.561	0.357	0.081	0.037	0.006	0.776	0.651	0.374
240	0.083	0.038	0.006	0.887	0.855	0.758	0.082	0.037	0.006	0.979	0.957	0.849
360	0.087	0.040	0.007	0.958	0.947	0.912	0.085	0.040	0.007	0.999	0.997	0.981
480	0.088	0.042	0.007	0.984	0.980	0.968	0.088	0.042	0.007	1.000	1.000	0.998
600	0.091	0.042	0.007	0.993	0.992	0.987	0.089	0.042	0.007	1.000	1.000	1.000

Panel B: Pairwise ($p = 1$) and multiple ($p = 2$) model comparison tests: Linear models

	Normal test ($p = 1$)						LR test ($p = 2$)					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.134	0.069	0.011	0.433	0.312	0.128	0.134	0.066	0.010	0.359	0.244	0.086
240	0.109	0.052	0.007	0.622	0.489	0.250	0.104	0.047	0.005	0.526	0.394	0.174
360	0.104	0.049	0.007	0.752	0.636	0.387	0.100	0.046	0.006	0.668	0.540	0.296
480	0.102	0.049	0.008	0.841	0.745	0.512	0.097	0.046	0.007	0.772	0.662	0.418
600	0.102	0.048	0.008	0.900	0.825	0.623	0.097	0.046	0.007	0.848	0.758	0.532

Panel C: Pairwise tests of equality: Nonlinear models

	Weighted χ^2 test						Wald test					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.050	0.017	0.001	0.159	0.074	0.011	0.077	0.031	0.003	0.363	0.216	0.046
240	0.062	0.023	0.002	0.268	0.157	0.039	0.105	0.051	0.009	0.646	0.518	0.265
360	0.065	0.026	0.003	0.328	0.210	0.064	0.107	0.055	0.012	0.778	0.678	0.445
480	0.066	0.026	0.003	0.369	0.245	0.084	0.105	0.054	0.013	0.859	0.784	0.586
600	0.067	0.027	0.003	0.409	0.282	0.106	0.102	0.052	0.012	0.910	0.854	0.692

Panel D: Pairwise ($p = 1$) and multiple ($p = 2$) model comparison tests: Nonlinear models

	Normal test ($p = 1$)						LR test ($p = 2$)					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.119	0.052	0.005	0.773	0.649	0.354	0.106	0.043	0.003	0.686	0.542	0.253
240	0.105	0.046	0.004	0.914	0.858	0.682	0.095	0.037	0.003	0.880	0.808	0.587
360	0.103	0.047	0.005	0.956	0.922	0.809	0.095	0.040	0.004	0.938	0.892	0.752
480	0.102	0.048	0.006	0.976	0.955	0.879	0.095	0.042	0.005	0.965	0.937	0.840
600	0.100	0.048	0.007	0.987	0.974	0.923	0.096	0.043	0.005	0.981	0.963	0.896

The table presents the empirical size and power of pairwise and multiple model comparison tests for non-nested linear (Panels A and B) and nonlinear (Panels C and D) models. In Panels A and C, we report results for the weighted chi-squared test in part (a) of Theorem 3 and the Wald test in part (b) of Theorem 3. Panels B and D present results for the pairwise ($p = 1$) and multiple ($p = 2$) model comparison tests in (35) and (52), respectively. We report results for different levels of significance (10%, 5% and 1% levels) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the factors and the gross returns (continuously compounded gross returns in the nonlinear case) are generated from a multivariate normal distribution.