FEDERAL RESERVE BANK *of* ATLANTA | WORKING PAPER SERIES

# Chi-Squared Tests for Evaluation and Comparison of Asset Pricing Models

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2011-8
March 2011

**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 new pivotal specification and model comparison tests that are asymptotically chi-squared distributed. In addition, we develop modified versions of the existing model selection tests with improved finite-sample properties. Finally, we fill an important gap in the literature by providing formal tests of multiple model comparison.

JEL classification: C12, C13, G12

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

---

Nikolay Gospodinov gratefully acknowledges financial support from Fonds de recherche sur la société et la culture (FQRSC), Institut de Finance Mathématique de Montréal (IFM2), and the Social Sciences and Humanities Research Council of Canada. Raymond Kan gratefully acknowledges financial support from the National Bank Financial of Canada, the Social Sciences and Humanities Research Council of Canada, and the Center for Financial Innovation and Stability at the Federal Reserve Bank of Atlanta. The authors thank Esther Eiling, Wayne Ferson, Jonathan Fletcher, Eric Jondeau, B. Ravikumar, Sergei Sarkissian, Jonathan Wright, Chu Zhang, Guofu Zhou, seminar participants at 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, and the 2010 NBER-NSF Time Series 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. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Nikolay Gospodinov, Concordia University and CIREQ, Department of Economics, 1455 de Maisonneuve Boulevard West, Montreal, Quebec, Canada H3G 1M8, 514-848-2424 (ex. 3935), 514-848-4536 (fax), nikolay.gospodinov@concordia.ca; Raymond Kan, University of Toronto, Joseph L. Rotman School of Management, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, 416-978-4291, 416-978-5433, kan@chass.utoronto.ca; or Cesare Robotti (corresponding author), Federal Reserve Bank of Atlanta and EDHEC Risk Institute, Research Department, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470, 404-498-8543, 404-498-8810 (fax), cesare.robotti@atl.frb.org.

## 1. INTRODUCTION

It is common for financial economists to view all asset pricing models only as approximations of 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 missing (Chen and Ludvigson, 2009, p. 1080). As a result, researchers are still ranking alternative models by comparing their corresponding sample HJ-distances (see, for example, Parker and Julliard, 2005 and Chen and Ludvigson, 2009, among others) without any use of a formal statistical criterion that takes into account the sampling and model misspecification uncertainty. In this paper, we provide a fully-fledged statistical framework for estimation, evaluation and comparison of linear and nonlinear (potentially misspecified) asset pricing models based on the unconstrained HJ-distances. Given some unappealing theoretical properties of the constrained HJ-distance (Gospodinov, Kan and Robotti, 2010a), 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 (see Gospodinov, Kan and Robotti, 2010a, for details).

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 and the sample HJ-distance that are valid for misspecified models. 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 new Lagrange multiplier tests for individual and joint testing of correct specification of one or more asset pricing models. These new specification tests are asymptotically chi-squared distributed and enjoy 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 size and power improvements in setups 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 (chi-squared) 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 cases where they are used as pre-tests in sequential testing procedures for strictly non-nested and overlapping models. Fifth, we develop a test for multiple model comparison as well as a fast numerical algorithm for computing its asymptotic $p$-value.[2] Finally, we investigate the finite-sample performance of the proposed inference procedures using Monte Carlo simulations.

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 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. Some concluding

[1] This problem is further investigated rigorously in Gospodinov, Kan and Robotti (2010b).

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

remarks are provided in Section 5. Proofs are collected in the Appendix.

The paper adopts the following notation. Let $\overset{A}{\sim}$ stand for "asymptotically distributed as," $\chi^2_p$ 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 $||A|| = \sqrt{\mathrm{tr}(A'A)}$ be the Euclidean or Frobenius norm of a matrix $A$, where $\mathrm{tr}(\cdot)$ 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 independent chi-squared random variables with one degree of freedom.

## 2. HANSEN-JAGANNATHAN DISTANCE

### 2.1. *Population Hansen-Jagannathan Distance*

Let $x_t$ denote a vector of payoffs of $n$ 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 $\Gamma$ is the parameter space of $\gamma$.[5] 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 $\gamma \in \Gamma$. When the asset pricing model is misspecified, we are interested in measuring the degree of model misspecification. HJ suggest using

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

[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).

[5]In this paper, we present results for the case in which the candidate SDF depends on some unknown parameters, but it is straightforward to adapt our analysis to the case in which the SDF does not depend on parameters.

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

Instead of solving the above primal problem to obtain $\delta$, 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 $\lambda$ 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

$$\phi_t(\theta) \equiv y_t(\gamma)^2 - m_t(\theta)^2 - 2\lambda' q_{t-1} \tag{5}$$

and

$$m_t(\theta) \equiv y_t(\gamma) - \lambda' x_t. \tag{6}$$

Note that $y_t(\gamma^*)$ prices the $n$ test assets correctly if the vector of pricing errors is zero, i.e.,

$$e(\gamma^*) = E[x_t y_t(\gamma^*) - q_{t-1}] = 0_n. \tag{7}$$

In this case, $y_t(\gamma^*) \in \mathcal{M}$, $\lambda^* = 0_n$ and we refer to $\gamma^*$ as the true value.[6]

By rearranging the dual problem in (3), it is easy to show that

$$\lambda = U^{-1} e(\gamma^*) \tag{8}$$

and

$$\delta^2 = e(\gamma^*)' U^{-1} e(\gamma^*). \tag{9}$$

While the quadratic form in the pricing errors in (9) 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

[6]The optimization problem in (4) bears some 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 (2009) for further discussion of the class of GEL estimators in the context of asset pricing models.

largely ignored. In this paper, we explicitly exploit this information to develop Lagrange multiplier specification tests for individual and multiple models.

## 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 $\theta^*$ in (4) is obtained as the solution to the sample dual problem

$$\hat{\theta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\lambda} \end{bmatrix} = \arg\min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} \frac{1}{T} \sum_{t=1}^{T} \phi_t(\theta). \tag{10}$$

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

$$\hat{\gamma} = \arg\min_{\gamma \in \Gamma} e_T(\gamma)' \hat{U}^{-1} e_T(\gamma), \tag{11}$$

and

$$\hat{\lambda} = \hat{U}^{-1} e_T(\hat{\gamma}). \tag{12}$$

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.

We first introduce regularity conditions to ensure the stochastic equicontinuity of the sample HJ-distance and the consistency of $\hat{\theta}$.

Assumption A. *Assume that*

(i) $\phi_t(\theta)$ *is m-dependent,*

(ii) *the parameter space* $\Theta$ *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* $\theta^*$ *which is in the interior of* $\Theta$.

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

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\phi_t(\theta) - E[\phi_t(\theta)]\right| \xrightarrow{p} 0. \tag{13}$$

The $m$-dependence can be relaxed although results for empirical processes with more general dependence structure are still limited (see, for instance, Andrews, 1993 and Andrews and Pollard, 1994). Assumption A(vi) is an identification condition that ensures the uniqueness of the pseudo-true value $\theta^*$. The uniform convergence in (13) and Assumption A(vi) are sufficient for establishing the consistency of $\hat{\theta}$:

$$\hat{\theta} \xrightarrow{p} \theta^*. \tag{14}$$

Let

$$H \equiv \begin{bmatrix} H_{\gamma\gamma} & H'_{\lambda\gamma} \\ H_{\lambda\gamma} & H_{\lambda\lambda} \end{bmatrix} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 E[\phi_t(\theta^*)]}{\partial\theta\partial\theta'} \tag{15}$$

and

$$M \equiv \begin{bmatrix} M_{\gamma\gamma} & M'_{\lambda\gamma} \\ M_{\lambda\gamma} & M_{\lambda\lambda} \end{bmatrix} = \lim_{T\to\infty}\mathrm{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\phi_t(\theta^*)}{\partial\theta}\right]. \tag{16}$$

The next assumption provides conditions for the existence and uniform convergence of the limiting matrices in (15) and (16).

Assumption B. *Let* $\mathcal{N}(\theta^*)$ *be a neighborhood of* $\theta^*$. *Assume that*

(i) $E[\phi_t(\theta)]$ *is twice continuously differentiable in* $\theta$ *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

$$\bar{v}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}v_t(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T}(h_t(\theta) - E[h_t(\theta)]). \tag{17}$$

The next assumption 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 (15) and (16). Using the fact that

$$\frac{\partial \phi_t(\theta^*)}{\partial \gamma} = 2[y_t(\gamma^*) - m_t(\theta^*)]\frac{\partial y_t(\gamma^*)}{\partial \gamma}, \tag{18}$$

$$\frac{\partial \phi_t(\theta^*)}{\partial \lambda} = 2[x_t m_t(\theta^*) - q_{t-1}], \tag{19}$$

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

$$H_{\gamma\gamma} = 2E\left[(y_t(\gamma^*) - m_t(\theta^*))\frac{\partial^2 y_t(\gamma^*)}{\partial \gamma \partial \gamma'}\right], \tag{20}$$

$$H_{\lambda\gamma} = 2E\left[x_t \frac{\partial y_t(\gamma^*)}{\partial \gamma'}\right], \tag{21}$$

$$H_{\lambda\lambda} = -2E\left[x_t x_t'\right] \equiv -2U, \tag{22}$$

and

$$M_{\lambda\lambda} = 4\sum_{j=-\infty}^{\infty} E\left[(x_t m_t(\theta^*) - q_{t-1})(x_{t+j}m_{t+j}(\theta^*) - q_{t+j-1})'\right]. \tag{23}$$

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 $M_{\lambda\lambda} = \sum_{j=-\infty}^{\infty} E[(x_t y_t(\gamma^*) - q_{t-1})(x_{t+j}y_{t+j}(\gamma^*) - q_{t+j-1})']$. 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

$$C = E\left[u_t \frac{\partial^2 y_t(\gamma^*)}{\partial\gamma\partial\gamma'}\right], \tag{24}$$

$$D = E\left[x_t \frac{\partial y_t(\gamma^*)}{\partial\gamma'}\right], \tag{25}$$

$$S = \sum_{j=-\infty}^{\infty} E\left[e_t(\gamma^*)e_{t+j}(\gamma^*)'\right], \tag{26}$$

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

The following lemma presents the asymptotic distributions of the sample squared HJ-distance under correctly specified and misspecified models.

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

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

$$T\hat{\delta}^2 \overset{A}{\sim} F_{n-k}(\xi), \tag{27}$$

*where the* $\xi_i$*'s are the eigenvalues of*

$$A = P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P, \tag{28}$$

*with* $P$ *being an* $n \times (n-k)$ *orthonormal matrix whose columns are orthogonal to* $U^{-\frac{1}{2}}D$.

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

$$\sqrt{T}(\hat{\delta}^2 - \delta^2) \overset{A}{\sim} N(0, \sigma_b^2), \tag{29}$$

*where* $\sigma_b^2 = \sum_{j=-\infty}^{\infty} E[b_t(\gamma^*)b_{t+j}(\gamma^*)]$ *and* $b_t(\gamma^*) = 2u_t(\gamma^*)y_t(\gamma^*) - u_t^2(\gamma^*) + \delta^2$.

The asymptotic distribution and matrix $A$ in part (a) of Lemma 1 coincide with the ones derived by Hansen, Heaton and Luttmer (1995) and Parker and Julliard (2005). To conduct inference, the covariance matrices in Lemma 1 should be replaced with consistent estimators. In particular, in part (a), we can replace $A$ with its sample analog

$$\hat{A} = \hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}, \tag{30}$$

where $\hat{S}$ is obtained using a nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator (see, for example, Newey and West, 1987 and Andrews, 1991), $\hat{P}$ is an orthonormal matrix whose columns are orthogonal to $\hat{U}^{-\frac{1}{2}}\hat{D}$ and $\hat{D} = \frac{1}{T}\sum_{t=1}^{T}\left[x_t\frac{\partial y_t(\hat{\gamma})}{\partial\gamma'}\right]$. Similarly, in part (b) we can use a HAC estimator to estimate the variance $\sigma_b^2$.

It has been documented (see Ahn and Gadarowski, 2004) that if we use $\hat{A}$ to estimate the eigenvalues $\xi_i$'s, the specification test in part (a) of Lemma 1 tends to overreject substantially when the number of test assets $n$ is large relative to the time series observations $T$. One way to reduce the overrejection problem is to use a different estimator of $S$. The consistent estimator of $S_A = M_{\lambda\lambda}/4$, denoted by $\hat{S}_A$, is a good alternative. While $\hat{S}_A$ converges to $S$ under the correctly specified model, $\hat{S}_A$ tends to be larger than $\hat{S}$ in finite samples, thus rendering the overrejection problem less severe.

Lemma 2 below establishes the asymptotic normality of the estimates of the SDF parameters and of the Lagrange multipliers, $\hat{\theta}$, based on the HJ-distance.

**Lemma 2.** *Under Assumptions A, B and C,*

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

$$\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{n+k}, \Sigma), \tag{31}$$

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

$$\begin{aligned} l_{1t} &= (C + D'U^{-1}D)^{-1}\left[D'U^{-1}e_t(\gamma^*) + \left\{\frac{\partial y_t(\gamma^*)}{\partial\gamma} - D'U^{-1}x_t\right\}u_t\right], \quad (32)\\ l_{2t} &= U^{-1}[Dl_{1t} - e_t(\gamma^*) + x_t u_t]. \quad (33)\end{aligned}$$

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

$$\sqrt{T}\Pi(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_n, \tilde{\Sigma}), \tag{34}$$

*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*

$$\begin{aligned} \tilde{l}_{1t} &= (D'U^{-1}D)^{-1}D'U^{-1}e_t(\gamma^*), \quad (35)\\ \tilde{l}_{2t} &= -P'U^{-\frac{1}{2}}e_t(\gamma^*), \quad (36)\end{aligned}$$

*and*

$$\Pi = \begin{bmatrix} I_k & 0_{k\times n} \\ 0_{(n-k)\times k} & P'U^{\frac{1}{2}} \end{bmatrix}. \tag{37}$$

The covariance matrices $\Sigma$ and $\tilde{\Sigma}$ in Lemma 2 can be consistently estimated using the sample analogs of (32)–(33) and (35)–(36), respectively. Tests of parameter restrictions based on the Wald or distance metric (likelihood ratio-type) statistics can be easily developed from the results in Lemma 2.

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 (33) simplifies to

$$l_{2t} = [U^{-1}D(D'U^{-1}D)^{-1}D' - I_n]U^{-1}e_t(\gamma^*). \tag{38}$$

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 $\alpha$ in the span of the column space of $D$, $\sqrt{T}\alpha'\hat{\lambda}$ is not asymptotically normal because $\alpha' l_{2t} = 0$.[7]

More generally, Gospodinov, Kan and Robotti (2010b) show that when $\alpha$ is in the span of the column space of $D$, then

$$T\alpha'\hat{\lambda} \xrightarrow{d} -\tilde{v}_1' v_2, \tag{39}$$

where $\tilde{v}_1$ and $v_2$ 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 2 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 $\gamma$ 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 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 2 to develop a Lagrange multiplier (LM) test for model specification.

**Theorem 1.** *Define the LM statistic as*

$$LM_{\hat{\lambda}} \equiv T\hat{\lambda}'\hat{U}^{\frac{1}{2}}\hat{P}\left(\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\right)^{-1}\hat{P}'\hat{U}^{\frac{1}{2}}\hat{\lambda}. \tag{40}$$

[7] It should be emphasized that when the SDF does not have parameters (as in the case of Proposition 4.1 of Hansen, Heaton and Luttmer, 1995), then $\sqrt{T}\hat{\lambda}$ has an asymptotic normal distribution even when $\delta = 0$.

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

$$LM_{\hat{\lambda}} \overset{A}{\sim} \chi^2_{n-k}. \tag{41}$$

Since $\delta = 0$ if and only if $\lambda = 0_n$, the LM test in Theorem 1 provides an alternative model specification test that measures the distance of the Lagrange multipliers from zero.[8] Similar arguments can be used for developing an asymptotically equivalent specification test on the model's pricing errors.

## 3. MODEL SELECTION TESTS

In this section, we refine the asymptotic theory for model comparison tests for strictly non-nested, nested and overlapping models and provide some new results including chi-squared versions of the model selection tests and multiple model comparison. 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 in the literature. First, since for nested models the 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 size distortions in finite samples. 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 strictly non-nested, nested and overlapping models that are easier to implement than the weighted chi-squared tests. Finally, we provide a multiple model comparison test that allows 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

$$\mathcal{F} = \{y_t^{\mathcal{F}}(\gamma_{\mathcal{F}})\ ; \gamma_{\mathcal{F}} \in \Gamma_{\mathcal{F}}\} \tag{42}$$

[8] A similar test, that uses the whole vector of Lagrange multipliers and a generalized inverse of their $n \times n$ asymptotic covariance matrix, is used by Smith (1997) and Imbens, Spady and Johnson (1998) in the context of GEL estimation of moment condition models.

and

$$\mathcal{G} = \{y_t^{\mathcal{G}}(\gamma_{\mathcal{G}})\,;\gamma_{\mathcal{G}} \in \Gamma_{\mathcal{G}}\}, \tag{43}$$

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

$$\delta_{\mathcal{F}}^2 = \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{F}}} E[\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}})] \tag{44}$$

$$\delta_{\mathcal{G}}^2 = \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} E[\phi_t^{\mathcal{G}}(\theta_{\mathcal{G}})], \tag{45}$$

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. If $\mathcal{F} \cap \mathcal{G} = \emptyset$, we have the case of strictly non-nested models. For nested models, we have $\mathcal{F} \subset \mathcal{G}$ or $\mathcal{G} \subset \mathcal{F}$. Finally, if $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \not\subset \mathcal{G}$, and $\mathcal{G} \not\subset \mathcal{F}$, we refer to $\mathcal{F}$ and $\mathcal{G}$ as overlapping models.

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:

$$\sqrt{T}(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \stackrel{A}{\sim} N(0, \sigma_d^2), \tag{46}$$

where

$$\sigma_d^2 = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}] \tag{47}$$

and $d_t = \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) - \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$.

Define

$$LR = \frac{T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)^2}{\hat{\sigma}_d^2}, \tag{48}$$

where $\hat{\sigma}_d^2$ is a consistent estimator of $\sigma_d^2$. Then, from (46) it follows that

$$LR \stackrel{A}{\sim} \chi_1^2. \tag{49}$$

It is important to emphasize that the results in (46) and (49) hold only if $\sigma_d^2 \neq 0$. To determine whether the use of the chi-squared test in (49) 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. 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 strictly non-nested models, we cannot have $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. As a result, we only have to test $H_0: \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ before using the test in (49).[9] For nested models, the test in (49) should not be performed because under $H_0: \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$, we must have $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{F}}^*)$. 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. Finally, 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 (49). We discuss the strictly non-nested, nested and overlapping cases in the following subsections.

### 3.1.1 *Strictly Non-Nested Models*

To test $H_0: \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ for strictly non-nested models, we can use the test statistic $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ based on the difference of the sample HJ-distances of models $\mathcal{F}$ and $\mathcal{G}$. Alternatively, using our results in Section 2, we can also develop an LM test that measures the distance of the Lagrange multipliers of the two models from zero. This will provide a joint test of correct model specification for models $\mathcal{F}$ and $\mathcal{G}$.

To set up the notation, define $e_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = x_t y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) - q_{t-1}$, $e_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = x_t y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) - q_{t-1}$, and

$$\mathcal{S} \equiv \begin{bmatrix} S_{\mathcal{F}} & S_{\mathcal{F}\mathcal{G}} \\ S_{\mathcal{G}\mathcal{F}} & S_{\mathcal{G}} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E\left[\tilde{e}_t \tilde{e}_{t+j}'\right], \tag{50}$$

where $\tilde{e}_t = [e_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*)', \ e_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)']'$. Let $P_{\mathcal{F}}$ and $P_{\mathcal{G}}$ denote orthonormal matrices with dimensions $n \times (n - k_1)$ and $n \times (n - k_2)$ whose columns are orthogonal to $U^{-\frac{1}{2}} D_{\mathcal{F}}$ and $U^{-\frac{1}{2}} D_{\mathcal{G}}$, respectively, where $D_{\mathcal{F}}$ ($D_{\mathcal{G}}$) is the $D$ matrix for model $\mathcal{F}$ ($\mathcal{G}$) defined in Section 2.3. Also, denote by $\hat{P}_{\mathcal{F}}$, $\hat{P}_{\mathcal{G}}$, $\hat{S}_{\mathcal{F}}$, $\hat{S}_{\mathcal{G}}$, $\hat{S}_{\mathcal{F}\mathcal{G}}$, $\hat{S}_{\mathcal{G}\mathcal{F}}$, $\hat{\lambda}_{\mathcal{F}}$, and $\hat{\lambda}_{\mathcal{G}}$ the sample counterparts of $P_{\mathcal{F}}$, $P_{\mathcal{G}}$, $S_{\mathcal{F}}$, $S_{\mathcal{G}}$, $S_{\mathcal{F}\mathcal{G}}$, $S_{\mathcal{G}\mathcal{F}}$, $\lambda_{\mathcal{F}}$ and $\lambda_{\mathcal{G}}$,

[9] 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. We offer such an example in the Appendix. See Kan and Robotti (2009) and Hall and Pelletier (2010) for further discussion of this point.

respectively. Finally, let

$$\tilde{\lambda}_{\mathcal{FG}} = \begin{bmatrix} \hat{P}'_{\mathcal{F}} \hat{U}^{\frac{1}{2}} \hat{\lambda}_{\mathcal{F}} \\ \hat{P}'_{\mathcal{G}} \hat{U}^{\frac{1}{2}} \hat{\lambda}_{\mathcal{G}} \end{bmatrix}. \tag{51}$$

The following theorem provides the appropriate asymptotic distributions of the difference in the sample squared HJ-distances when both models are correctly specified and an LM test of $H_0 : \lambda_{\mathcal{F}} = \lambda_{\mathcal{G}} = 0_n$ (which is equivalent to testing $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$).

**Theorem 2.** *Suppose that Assumptions A, B and C hold for each model and* $y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) \neq y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})$. *Then, under* $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$,

(a)

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

*where the* $\xi_i$*'s are the eigenvalues of the matrix*

$$\begin{bmatrix} P'_{\mathcal{F}} U^{-\frac{1}{2}} S_{\mathcal{F}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P'_{\mathcal{F}} U^{-\frac{1}{2}} S_{\mathcal{FG}} U^{-\frac{1}{2}} P_{\mathcal{G}} \\ P'_{\mathcal{G}} U^{-\frac{1}{2}} S_{\mathcal{GF}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P'_{\mathcal{G}} U^{-\frac{1}{2}} S_{\mathcal{G}} U^{-\frac{1}{2}} P_{\mathcal{G}} \end{bmatrix}, \tag{53}$$

(b)

$$LM_{\tilde{\lambda}_{\mathcal{FG}}} = T \tilde{\lambda}'_{\mathcal{FG}} \hat{\Sigma}^{-1}_{\tilde{\lambda}_{\mathcal{FG}}} \tilde{\lambda}_{\mathcal{FG}} \stackrel{A}{\sim} \chi^2_{2n-k_1-k_2}, \tag{54}$$

*where*

$$\hat{\Sigma}_{\tilde{\lambda}_{\mathcal{FG}}} = \begin{bmatrix} \hat{P}'_{\mathcal{F}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{F}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{F}} & \hat{P}'_{\mathcal{F}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{FG}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{G}} \\ \hat{P}'_{\mathcal{G}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{GF}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{F}} & \hat{P}'_{\mathcal{G}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{G}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{G}} \end{bmatrix}. \tag{55}$$

Since the eigenvalues $\xi_i$'s in part (a) of Theorem 2 can take on both positive and negative values, the test of the hypothesis $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ should be two-sided. The LM test in part (b) of Theorem 2 provides an alternative way of testing $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ (using the equivalence between $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ and $H_0 : \lambda_{\mathcal{F}} = \lambda_{\mathcal{G}} = 0_n$) but it is easier to implement and is expected to deliver power gains compared to the test in part (a). The reason is that the test in part (a) may have low power in finite samples when $\hat{\delta}^2_{\mathcal{F}} \approx \hat{\delta}^2_{\mathcal{G}} \neq 0$ although it is still consistent since under the alternative $\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}} = O_p(T^{-1/2})$ and $|T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}})| \to \infty$.

In summary, our proposed test of equality of the squared HJ-distances of two strictly non-nested models involves first testing whether the two models are both correctly specified using one of the tests in Theorem 2. If we reject, then we can perform the test in (49). Suppose that $\alpha_1$ and $\alpha_2$ are the asymptotic significance levels used in the pre-test $\delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ and in the chi-squared test in (49), respectively. Then, our sequential test has a significance level that is asymptotically bounded above by $\max[\alpha_1, \alpha_2]$. Thus, if $\alpha_1 = \alpha_2 = 0.05$, the significance level of this procedure, as a test of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}}$, is asymptotically no larger than 5%.

### 3.1.2 *Nested Models*

For nested models, $\sigma^2_d$ is zero by construction under the null of equal HJ-distances. Therefore, the chi-squared test in (49) cannot be used. In addition, for nested models, $\delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}}$ if and only if $y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})$, so we can simply test $H_0 : y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})$.

Let $H_{\mathcal{F}} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \frac{\partial^2 E[\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}})]}{\partial\theta_{\mathcal{F}}\partial\theta'_{\mathcal{F}}}$ and $M_{\mathcal{F}} = \lim_{T\to\infty} \mathrm{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial\phi_t(\theta^*_{\mathcal{F}})}{\partial\theta}\right]$ with $H_{\mathcal{G}}$ and $M_{\mathcal{G}}$ defined similarly. Marcellino and Rossi (2008) among others show that under $H_0 : \phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}}) = \phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}})$,[10]

$$T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}) \stackrel{A}{\sim} F_{2n+k_1+k_2}(\xi), \tag{57}$$

where the $\xi_i$'s are the eigenvalues of the matrix

$$\frac{1}{2}\begin{bmatrix} -H^{-1}_{\mathcal{F}} M_{\mathcal{F}} & -H^{-1}_{\mathcal{F}} M_{\mathcal{F}\mathcal{G}} \\ H^{-1}_{\mathcal{G}} M_{\mathcal{G}\mathcal{F}} & H^{-1}_{\mathcal{G}} M_{\mathcal{G}} \end{bmatrix}. \tag{58}$$

Several remarks regarding this inference procedure are in order. First, estimating the $\xi_i$'s from the sample counterpart of the matrix in (58) produces more nonzero estimated $\xi_i$'s than the theory suggests. In addition, the estimated $\xi_i$'s do not have the same sign. This is problematic because for nested models, the larger model has a smaller sample HJ-distance by construction. By not imposing the constraints that the $\xi_i$'s should have the same sign, the nonnegative test statistic $T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}})$ is compared with a distribution that can take on both positive and negative values. This could result in serious finite-sample distortions of the test. In the ensuing analysis, we will show that under $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}}$, some of the $\xi_i$'s are equal to zero and the nonzero $\xi_i$'s have the same sign.

Without loss of generality, we assume $\mathcal{F} \subset \mathcal{G}$. Suppose that the null hypothesis $H_0 : y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})$ 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 $\mathcal{G}$

[10] Alternatively, we can directly test $H_0 : \sigma^2_d = 0$. In this case,

$$T\hat{\sigma}^2_d \stackrel{A}{\sim} F_{2n+k_1+k_2}(\xi), \tag{56}$$

where the $\xi_i$'s are four times the squared eigenvalues of the matrix in (58) (see Golden, 2003).

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

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma_{\mathcal{G}}'} \tag{59}$$

as a $(k_2 - k_1) \times k_2$ derivative matrix of the parametric restrictions $\psi_{\mathcal{G}}$. For many models of interest, $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}})$ 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, $\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_1)\times k_1},\ I_{k_2-k_1}]$, which is a selector matrix that selects 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 $k \times k$ block of $\Sigma$ in part (a) of Lemma 2, $\Psi_*^{\mathcal{G}} \equiv \Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, 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 2.3. Finally, define the Wald test statistic

$$Wald_{\hat{\psi}_{\mathcal{G}}} = T\hat{\psi}_{\mathcal{G}}'(\hat{\Psi}^{\mathcal{G}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{G}}}\hat{\Psi}^{\mathcal{G}\prime})^{-1}\hat{\psi}_{\mathcal{G}}, \tag{60}$$

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 3 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 3.** *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)

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_2-k_1}(\xi), \tag{61}$$

*where the $\xi_i$'s are the eigenvalues of the matrix*

$$(\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1}\Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}, \tag{62}$$

(b)

$$Wald_{\hat{\psi}_{\mathcal{G}}} \overset{A}{\sim} \chi^2_{k_2-k_1}. \tag{63}$$

Part (a) of Theorem 3 shows that, under $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, only $k_2 - k_1$ of the eigenvalues of (58) are nonzero and they all have the same sign.[11] In practice, we need to estimate the $\xi_i$'s to

[11]It can also be shown that under $H_0 : \sigma_d^2 = 0$,

$$T\hat{\sigma}_d^2 \overset{A}{\sim} F_{k_2-k_1}(\xi), \tag{64}$$

construct the test. Using the sample version of the matrix in part (a) of Theorem 3 instead of the sample version of the matrix in (58) to estimate the $\xi_i$'s results in a substantial reduction of the number of estimated eigenvalues. In addition, the resulting estimated eigenvalues are guaranteed to be positive. The Wald test in part (b) of Theorem 3 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).

3.1.3 *Overlapping Models*

For overlapping models, the variance $\sigma_d^2$ can be zero when (i) $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ or (ii) both models are correctly specified.[12] Since Theorem 2 is applicable to the second scenario, here we only need to derive the test of $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$.

It is well known that for linear models, the equality of the SDFs implies zero 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

$$\Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}) = \frac{\partial \psi_{\mathcal{F}}(\gamma_{\mathcal{F}})}{\partial \gamma_{\mathcal{F}}'} \tag{65}$$

and

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma_{\mathcal{G}}'} \tag{66}$$

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 cases, $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ implies that a subset of the parameters of model $\mathcal{F}$ is equal to zero, and $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{G}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ implies that a subset of the parameters of model $\mathcal{G}$ is 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 $\Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}}$ be the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{F}\mathcal{G}} =$

where the $\xi_i$'s are four times the squared eigenvalues of the matrix in (62). Note the reduction in the number of eigenvalues compared to the test in (56) that does not impose parametric restrictions. The proof of this result is available from the authors upon request.

[12] Similar to the case of strictly non-nested models, it is possible for two overlapping SDFs to be both correctly specified. Examples are available upon request.

$[\hat{\gamma}_{\mathcal{F}}{}', \hat{\gamma}_{\mathcal{G}}{}']'$, $\tilde{H}_{\mathcal{F}} = (C^{\mathcal{F}} + D^{\mathcal{F}\prime} U^{-1} D^{\mathcal{F}})^{-1}$, $\tilde{H}_{\mathcal{G}} = (C^{\mathcal{G}} + D^{\mathcal{G}\prime} U^{-1} D^{\mathcal{G}})^{-1}$, $\Psi_*^{\mathcal{F}} = \Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}^*)$, $\Psi_*^{\mathcal{G}} = \Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ and

$$\Psi_*^{\mathcal{FG}} \equiv \begin{bmatrix} \Psi_*^{\mathcal{F}} & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi_*^{\mathcal{G}} \end{bmatrix}. \tag{67}$$

Define the Wald test statistic

$$Wald_{\hat{\psi}_{\mathcal{FG}}} = T\hat{\psi}_{\mathcal{FG}}'(\hat{\Psi}^{\mathcal{FG}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}\hat{\Psi}^{\mathcal{FG}\prime})^{-1}\hat{\psi}_{\mathcal{FG}}, \tag{68}$$

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

$$\hat{\Psi}^{\mathcal{FG}} = \begin{bmatrix} \Psi^{\mathcal{F}}(\hat{\gamma}_{\mathcal{F}}) & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi^{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}}) \end{bmatrix}, \tag{69}$$

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 $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ and $Wald_{\hat{\psi}_{\mathcal{FG}}}$ 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 4.** *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)

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_1+k_2-2k_3}(\xi), \tag{70}$$

*where the $\xi_i$'s are the eigenvalues of the matrix*

$$\begin{bmatrix} -(\Psi_*^{\mathcal{F}}\tilde{H}_{\mathcal{F}}\Psi_*^{\mathcal{F}\prime})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1} \end{bmatrix} \Psi_*^{\mathcal{FG}}\Sigma_{\hat{\gamma}_{\mathcal{FG}}}\Psi_*^{\mathcal{FG}\prime}, \tag{71}$$

(b)

$$Wald_{\hat{\psi}_{\mathcal{FG}}} \overset{A}{\sim} \chi^2_{k_1+k_2-2k_3}. \tag{72}$$

Unlike the case of nested models, the eigenvalues in part (a) of Theorem 4 are not always positive because $\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2$ can take on both positive and negative values. As a result, we need to perform a two-sided test of $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. 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 4.

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 two models are both correctly specified using either the test in part (a) of Theorem 2 or the chi-squared test in (54). It should be noted that the tests in part (a) of Theorem 2 and part (a) of Theorem 4 are both $O_p(T^{-1})$ and will have low power against each other. Furthermore, the test in part (a) of Theorem 4 will not be consistent against the alternative $H_1 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) \neq y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ when both models are correctly specified. As a result, our recommendation is to use the LM test in part (b) of Theorem 2 as a pre-test of whether the two models are both correctly specified. If the null is rejected, we can proceed with testing if the SDFs of the two models are equal using the tests in Theorem 4. Finally, if we still reject, we can then perform the chi-squared test in (49). 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 $\alpha_1$, $\alpha_2$, and $\alpha_3$ are the asymptotic significance levels used in these three tests.

The results in Theorems 3 and 4 offer substantial advantages over the inference procedure (57)–(58) in Section 3.1.2. 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 3 and 4 are expected to lead to improved finite-sample (size and power) 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. The non-nested model comparison test is a multivariate inequality test based on results in the statistics literature due to Wolak (1987, 1989).[13] Suppose we have $p + 1$ models. We are interested in testing the null hypothesis that the benchmark model, model 1 (we could think

[13] Kan, Robotti and Shanken (2010) 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.

of model 1 as model $\mathcal{F}$ in the pairwise model comparison subsection), performs at least as well as the other $p$ models. Let $\delta_i^2$ 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 states that some model has a smaller population squared HJ-distance than model 1.

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 that

$$\sqrt{T}(\hat{\rho} - \rho) \overset{A}{\sim} N(0_p, \Omega_{\hat{\rho}}). \tag{73}$$

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.[14] Let $\tilde{\rho}$ be the optimal solution in the following quadratic programming problem:

$$\min_{\rho} (\hat{\rho} - \rho)' \hat{\Omega}_{\hat{\rho}}^{-1} (\hat{\rho} - \rho) \quad \text{s.t.} \quad \rho \leq 0_r, \tag{74}$$

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

$$LR = T(\hat{\rho} - \tilde{\rho})' \hat{\Omega}_{\hat{\rho}}^{-1} (\hat{\rho} - \tilde{\rho}). \tag{75}$$

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 $\rho$, which is $\rho = 0_p$. Under this value, $LR$ follows a "chi-bar-squared distribution,"

$$LR \overset{A}{\sim} \sum_{i=0}^{p} w_i(\Omega_{\hat{\rho}}^{-1}) X_i. \tag{76}$$

where the $X_i$'s are independent $\chi^2$ random variables with $i$ degrees of freedom and $\chi_0^2$ is simply defined as the constant zero.[15] We use this procedure to obtain asymptotically valid $p$-values.

Before using the multivariate inequality test to compare a benchmark model with a set of alternative models, we remove those alternative models $i$ 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 $\delta^2$ of the nesting model will be at least as small. Finally, we also remove from consideration any alternative models that nest the benchmark, since the normality assumption on $\hat{\rho}_i$ does not hold under the null hypothesis that $\rho_i = 0$.

[14] As in Section 3.1, pre-tests of correct specification and equality of SDFs can be easily developed also for multiple models by generalizing (54) and (72) to the $p > 1$ case.

[15] An explicit formula for the weights $w_i(\Omega_{\hat{\rho}}^{-1})$ is given in Kudo (1963) and the computational details are available from the authors upon request.

Since the multivariate inequality test described above is no longer applicable when the benchmark is nested by some alternative models, a different multiple model comparison test is needed in this case. When the alternative models nesting the benchmark are nested within each other, we remove the "nested" models since the $\delta^2$ 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 3.1.2. The situation, however, becomes more complicated when the alternative models exhibit an overlapping structure.

Suppose that the benchmark (with $k_1$ parameter vector $\gamma_1$) is nested by model $i$ (with $k_i$ parameter vector $\gamma_i$, $i = 2, \ldots, p+1$). Similar to the setup of Section 3.1.2, 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 2, we can use the delta method to obtain the asymptotic distribution of $\hat{\psi}_i = \psi_i(\hat{\gamma}_i)$. Specifically, let

$$\psi = \begin{pmatrix} \psi_2(\gamma_2^*) \\ \vdots \\ \psi_{p+1}(\gamma_{p+1}^*) \end{pmatrix} \tag{77}$$

and denote by $\hat{\psi}$ a consistent estimator of $\psi$. 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

$$Wald_{\hat{\psi}} = T\hat{\psi}'\hat{\Sigma}_{\hat{\psi}}^{+}\hat{\psi} \overset{A}{\sim} \chi_l^2, \tag{78}$$

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.

To conclude, if the benchmark model is nested by some competing models, one should separate the set of competing models into two subsets. The first subset will include competing models that

[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 4.3 below.

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 multivariate inequality test in (76). 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 $\alpha$.

## 4. MONTE CARLO SIMULATIONS

In this section, we undertake a Monte Carlo experiment to explore the small-sample properties of all the test statistics discussed in the theoretical part of the paper. We focus on linear asset pricing models given their popularity in the literature. To make our simulations more realistic, we calibrate the parameters by using almost 50 years, 1952:2–2000:4, of quarterly gross returns on the three-month T-bill and the well-known 25 Fama-French size and book-to-market portfolios. The time-series sample size is taken to be $T = 120, 240, 360, 480$ and 600. These choices of $T$ reflect sample sizes that are typically encountered in empirical work. The factors and the returns on the test assets are drawn from a multivariate normal distribution (a more detailed description of the various simulation designs can be found in a separate appendix available on the authors' websites). We compare actual rejection rates over 100,000 iterations to the nominal 10%, 5% and 1% levels of our tests. The parameters in our simulation study are calibrated using actual data from the following models:[17]

**CAPM:** the capital asset pricing model

$$y_t^{CAPM} = \gamma_0 + \gamma_1 r_{mkt,t},$$

where $r_{mkt}$ is the excess return on the market portfolio;

**CCAPM:** the consumption CAPM

$$y_t^{CCAPM} = \gamma_0 + \gamma_1 c_{ndur,t},$$

where $c_{ndur}$ is the log growth rate of non-durable consumption;

[17]See Gospodinov, Kan and Robotti (2010a) for a detailed description of the various risk factors.

**YOGO:** the durable consumption CAPM of Yogo (2006)

$$y_t^{YOGO} = \gamma_0 + \gamma_1 r_{mkt,t} + \gamma_2 c_{ndur,t} + \gamma_3 c_{dur,t},$$

where $c_{dur}$ denotes the log consumption growth rate of durable goods;

**FF3:** the three-factor model of Fama and French (1993)

$$y_t^{FF3} = \gamma_0 + \gamma_1 r_{mkt,t} + \gamma_2 r_{smb,t} + \gamma_3 r_{hml,t},$$

where $r_{smb}$ is the return difference between portfolios of small and large stocks and $r_{hml}$ is the return difference between portfolios of high and low book-to-market ratios;

**LL:** the conditional consumption CAPM of Lettau and Ludvigson (2001)

$$y_t^{LL} = \gamma_0 + \gamma_1 c_{ndur,t} + \gamma_2 cay_{t-1} + \gamma_3 c_{ndur,t} cay_{t-1},$$

where *cay* is the consumption-wealth ratio.

### 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 two SDFs: CAPM and YOGO. CAPM is an example of models with traded factors since $r_{mkt}$ is a portfolio return. YOGO is an example of models with traded and non-traded factors since $c_{ndur}$ and $c_{dur}$ are macroeconomic factors. As we will see later on, our inference procedures and simulation results are sensitive to whether the included risk factors are traded or not.

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. The $t$-tests are constructed using the asymptotic covariance matrices in Lemma 2 and are compared against the critical values from a standard normal distribution. We refer to the $t$-tests based on (31) and (32) 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 (34) and (35). 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.

Table 1 presents the empirical size of both $t$-tests of the null hypothesis $H_0 : \gamma_i = 0$ for realistic values of the HJ-distance measure: $\delta = 0.6524$ for CAPM and $\delta = 0.6514$ for YOGO.[18] Panel A is for the $t$-tests under potentially misspecified models, while Panel B is for the $t$-tests under correctly specified models.

Table 1 about here

For CAPM, the empirical rejection rates of both $t$-tests are very close to the nominal size. In contrast, the finite-sample performances of these two tests are very different for YOGO. The $t$-test under potentially misspecified models is well-behaved in this scenario and its empirical size is close to the nominal level with only a slight underrejection.[19] On the contrary, the $t$-test under correctly specified models tends to overreject substantially. For example, the $t$-test on the SDF parameter of durable consumption rejects the null hypothesis 30% of the time at the 5% nominal level for $T = 600$. Interestingly, the presence of non-traded factors in YOGO also leads to significant size distortions of the $t$-test on the traded factor. Finally, in the YOGO case, the performances of the $t$-tests under correctly specified models deteriorate as $T$ increases.[20]

The difference in behavior between the two $t$-tests when the model contains non-traded factors 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

[18] To preserve space, we do not report simulation results for the $t$-ratios associated with the SDF intercept terms.

[19] 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).

[20] The size distortions of the $t$-tests under correctly specified models documented in Table 1 are somewhat conservative. In unreported simulation experiments, we further analyzed the size properties of the $t$-tests of $H_0 : \gamma_i = 0$. While the $t$-test under potentially misspecified models maintained its excellent size properties, there were several instances in which the $t$-test under correctly specified models exhibited even stronger overrejections. These were typically situations in which the factors and the returns were generated under stronger model misspecification.

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 in 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 and 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 $c_{ndur}$ and $c_{dur}$ 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 developed in Section 2.3. In particular, we assess the finite-sample performance of three specification tests: (i) the HJ-distance test in part (a) of Lemma 1 based on the matrix $S$, (ii) the HJ-distance test based on the matrix $S_A$, and (iii) the LM test in Theorem 1. To examine size, we generate returns such that the model holds exactly, i.e., we set $E[x_t] = (1_n - \mathrm{Cov}[x_t, y_t])\,/E[y_t]$, where $1_n$ is an $n$-vector of ones. The covariance matrix of the factors and returns and the factor means are chosen based on the covariance matrix and the factor means estimated from the data. To examine power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for CAPM and YOGO are 0.6524 and 0.6514, respectively. The empirical size and power of the three tests are presented in Table 2.

$\boxed{\text{Table 2 about here}}$

The overrejections of the HJ-distance test based on $S$ have already been documented in the literature (see, for example, Ahn and Gadarowski, 2004). This typically happens 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 different model specifications (24% for CAPM and 17.4% for YOGO at the 5% significance level for $T = 120$) and show that the empirical size approaches the nominal level of the test as $T$ increases (7.3% for CAPM and 6.4% for YOGO at the 5% significance level for $T = 600$).

To the best of our knowledge, the HJ-distance test based on $S_A$ is new to the literature. As we argued in Section 2.3, using $\hat{S}_A$ instead of $\hat{S}$ appears to be particularly important when there are fewer observations per moment condition. This HJ-distance test enjoys much better size properties although it tends to be somewhat conservative. As $T$ increases, the rejection rates approach the nominal size of the test.

Finally, our new LM test has excellent size properties across different models and for all sample sizes. It should be emphasized that the improved sizes of the HJ-distance test that uses $S_A$ and of the LM test are accompanied by impressive power performance, very similar to the one of the HJ-distance test that uses $S$. All tables in this section report actual power since computing size-adjusted power seems infeasible for several of our tests. In summary, the LM test in Theorem 1 appears to dominate the other two tests given its simplicity and superior size properties.

### 4.2. *Model Selection Tests for Strictly Non-Nested Models*

In Table 3.A, we evaluate the size and power properties of the tests of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$. We consider the weighted chi-squared test and the LM test in parts (a) and (b) of Theorem 2, respectively. In light of the discussion in Section 3, one should test whether models $\mathcal{F}$ and $\mathcal{G}$ are correctly specified before applying the LR test in (49).[21] To examine size, we consider two correctly specified one-factor models (with no intercept) whose factors are generated as $r_{mkt}$ plus two different normal noise terms. To analyze power, we set the return means equal to the means estimated from the sample and compare one of the two one-factor models described above with a model (also with no intercept) that contains a market factor contaminated with a normal noise term, $c_{ndur}$ and $c_{dur}$. In the size and power experiments, we set the mean of the market factor such that it prices the risk-free

[21] The finite-sample performance of this LR test will be evaluated in Section 4.4 in the context of overlapping models.

asset correctly. This guarantees that the HJ-distances of the two models in the power experiment are close to each other and to values typically encountered in empirical work. The noise terms here and in Section 4.4 have mean zero, standard deviation which is 10% of the standard deviation of $r_{mkt}$ and are independent of the returns and the market factor.

Table 3 about here

The size properties of the two tests are quite good when $T$ is bigger than 120. The weighted chi-squared test slightly overrejects for small sample sizes while the LM test slightly underrejects which, again, appears to be due to the relatively large number of test assets considered in the simulation experiment. The empirical size of the tests quickly approaches the nominal level as $T$ increases. Consistent with the discussion below Theorem 2, the LM test delivers nontrivial power gains compared to the weighted chi-squared test and rejects the null $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ with probability one for $T \geq 240$.

### 4.3. *Model Selection Tests for Nested Models*

In Panels B and C of Table 3, we report simulation results for pairwise and multiple nested model comparison tests. 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. The tests under investigation are the restricted weighted chi-squared test and Wald test in Theorem 3, the unrestricted weighted chi-squared test in (57)–(58) and the chi-squared multiple model comparison test described in Section 3.2.

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 nesting model's slope coefficients associated with the factors that do not belong to the benchmark are zero and both the benchmark and the nesting model are misspecified. To analyze power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for CCAPM and YOGO are 0.6768 and 0.6514, respectively. Turning to multiple model comparison, the size of the chi-squared test is evaluated by choosing the return means such that the nesting models' slope coefficients associated with the factors that do not belong to the benchmark are zero and the benchmark as well as the two nesting models are misspecified. To evaluate power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for CCAPM, YOGO and LL are 0.6768, 0.6514 and 0.6561, respectively.

Table 3.B shows that the test in (57)–(58), that does not impose the restrictions arising from the nested structure of the models, exhibits overrejections that are nontrivial even for $T = 600$ (14.7% and 8.0% at the 10% and 5% nominal levels, respectively). This is due to fact that estimating eigenvalues from the sample counterpart of the matrix in (58) produces more nonzero estimated eigenvalues than the theory suggests. The weighted chi-squared and Wald tests in Theorem 3 have very good size properties and high power (despite the small differences in HJ-distances between models), with the Wald test performing better overall. It should be stressed again that Table 3 reports actual (not size-adjusted) power and the seemingly similar power of the restricted and unrestricted weighted chi-squared tests is likely due to the overrejections of the unrestricted weighted chi-squared test in (57)–(58) under the null.

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

### 4.4. *Model Selection Tests for Overlapping Models*

The case of 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 4 and the unrestricted weighted chi-squared test in (57)–(58). The simulated data are generated using FF3 and YOGO. To evaluate size, we choose the return means such that the slope coefficients associated with the non-overlapping factors in FF3 and YOGO are zero and the two models are misspecified. 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. Table 3.D shows that all three tests have good size. In terms of power, however, the restricted weighted chi-squared test proposed in part (a) of Theorem 4 performs better than the unrestricted test in (57)–(58) while the chi-squared test in part (b) of Theorem 4 provides further power gains and dominates both weighted chi-squared tests. The high power of the Wald test appears to be particularly important given the fact that this test (along with the test that the models are jointly correctly specified) 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 LR test in (49). In the size computations, the data are simulated from two misspecified three-factor models with intercept, $r_{smb}$ and $r_{hml}$ as common factors and a non-overlapping part that is obtained from contaminating the market factor with two independent normal noise terms defined as in Section 4.2. In the power comparison, the two overlapping models are FF3 and YOGO. Table 3.E ($p = 1$ case) shows that the size properties of the LR 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 three three-factor models and employ the LR test in (76) for multiple model comparison. In the size comparison, we add another model with a constant, $r_{smb}$ and $r_{hml}$ as common factors and a non-overlapping part given by the market factor contaminated with independent normal noise. For power evaluation, we consider LL in addition to YOGO and the benchmark FF3. Table 3.E ($p = 2$ case) reveals the very good finite-sample properties of the LR test for comparing multiple asset pricing models.

Overall, these simulation results suggest that the tests developed in this paper should be fairly reliable for the sample sizes typically encountered in empirical work.

## 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 propose new pivotal specification and model comparison tests that are asymptotically chi-squared distributed. We also derive new versions of the weighted chi-squared specification and model comparison tests that are computationally efficient and possess improved finite-sample properties. Finally, we develop computationally attractive tests for multiple model comparison. The excellent size and power properties of the proposed tests are demonstrated with simulated data from popular asset pricing models. The simulation results clearly suggest that the standard tests for model specification and selection as well 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 extra restrictions implied by the structure of the models lead to substantially improved inference. Looking to the future, although our simulation results are encouraging, the small-sample properties of the test statistics proposed in this paper should be explored further.

APPENDIX

A.1. *Preliminary Lemma*

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

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

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

***Proof***. We start by expanding $E[\phi_t(\theta^*)] = \delta^2$ about $\hat{\theta}$. Since $\frac{1}{T}\sum_{t=1}^{T}\partial\phi_t(\hat{\theta})/\partial\theta = 0_{n+k}$, we obtain

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

where $\tilde{\theta}$ is an intermediate point between $\hat{\theta}$ and $\theta^*$. Let

$$\bar{h}^*_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}E[h_t(\theta)]. \tag{A.3}$$

A mean value expansion of $\bar{h}^*_T(\theta^*)$ about $\hat{\theta}$ yields

$$0_{n+k} = \sqrt{T}\bar{h}^*_T(\theta^*) = \sqrt{T}\bar{h}^*_T(\hat{\theta}) - \frac{\partial\bar{h}^*_T(\check{\theta})}{\partial\theta}\sqrt{T}(\hat{\theta}-\theta^*), \tag{A.4}$$

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

$$\sqrt{T}(\hat{\theta}-\theta^*) = H^{-1}\sqrt{T}\bar{h}^*_T(\hat{\theta}) + o_p(1). \tag{A.5}$$

Using the definition of $\bar{v}_T(\theta)$ in (17) and the first order condition of $\frac{1}{T}\sum_{t=1}^{T}h_t(\hat{\theta}) = 0_{n+k}$, it follows that

$$\sqrt{T}\bar{v}_T(\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(h_t(\hat{\theta}) - E[h_t(\hat{\theta})]\right) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}E[h_t(\hat{\theta})] = -\sqrt{T}\bar{h}^*_T(\hat{\theta}). \tag{A.6}$$

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

$$\sqrt{T}\bar{h}^*_T(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\hat{\theta}) = \sqrt{T}[\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})] - \sqrt{T}\bar{v}_T(\theta^*). \tag{A.7}$$

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

$$\sqrt{T}|\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})| \leq \sup_{\theta \in \Theta : |\theta - \theta^*| \leq \omega} \sqrt{T} \left| \bar{v}_T(\theta^*) - \bar{v}_T(\theta) \right|. \tag{A.8}$$

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

$$\sup_{\theta \in \Theta : |\theta - \theta^*| \leq \omega} \sqrt{T} \left| \bar{v}_T(\theta^*) - \bar{v}_T(\theta) \right| \xrightarrow{p} 0. \tag{A.9}$$

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

$$\sqrt{T}\bar{h}_T^*(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\theta^*) + o_p(1). \tag{A.10}$$

Finally, substituting (A.10) into (A.5) yields

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

Thus, from (A.11), the consistency of $\hat{\theta}$, and Assumption B(ii), we obtain

$$\begin{aligned} \hat{\delta}^2 - \delta^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\right) - \frac{1}{2}(\hat{\theta} - \theta^*)' \frac{\partial^2 E[\phi_t(\tilde{\theta})]}{\partial\theta\partial\theta'}(\hat{\theta} - \theta^*) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\theta^*) - E[\phi_t(\theta^*)]\right) - \frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right). \end{aligned} \tag{A.12}$$

This completes the proof. ■

A.2. *Proofs*

***Proof of Lemma 1.*** (a) From the definition of $H$ in (15), we can use the partitioned matrix inverse formula to obtain

$$H^{-1} = \begin{bmatrix} 2C & 2D' \\ 2D & -2U \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} \tilde{H} & \tilde{H}D'U^{-1} \\ U^{-1}D\tilde{H} & -U^{-1} + U^{-1}D\tilde{H}D'U^{-1} \end{bmatrix}, \tag{A.13}$$

where $\tilde{H} = (C + D'U^{-1}D)^{-1}$. Under the null hypothesis $H_0 : \delta = 0$, (A.1) in Lemma A.1 becomes

$$\hat{\delta}^2 = -\frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right) \tag{A.14}$$

since $\lambda^* = 0_n$ and $\phi_t(\gamma^*, 0_n) = E[\phi_t(\gamma^*, 0_n)] = 0$. Let $\bar{v}_T(\theta^*) = [\bar{v}_{1,T}(\theta^*)', \bar{v}_{2,T}(\theta^*)']'$, where $\bar{v}_{1,T}(\theta^*)$ denotes the first $k$ elements of $\bar{v}_T(\theta^*)$. Under the null, $\bar{v}_{1,T}(\theta^*) = 0_k$ and $C = 0_{k\times k}$.

Then, it follows that

$$\begin{aligned} T\hat{\delta}^2 &= -\frac{1}{2}\sqrt{T}\bar{v}_T(\theta^*)'H^{-1}\sqrt{T}\bar{v}_T(\theta^*) + o_p(1) \\ &= \frac{1}{4}\sqrt{T}\bar{v}_{2T}(\theta^*)'[U^{-1} - U^{-1}D(D'U^{-1}D)^{-1}D'U^{-1}]\sqrt{T}\bar{v}_{2,T}(\theta^*) + o_p(1) \\ &= \frac{1}{4}\sqrt{T}\bar{v}_{2T}(\theta^*)'U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}\sqrt{T}\bar{v}_{2,T}(\theta^*) + o_p(1) \end{aligned} \qquad \text{(A.15)}$$

by using the fact that $I_n - U^{-\frac{1}{2}}D(D'U^{-1}D)^{-1}D'U^{-\frac{1}{2}} = PP'$. Also, Assumptions A, B and C ensure that the empirical process $\sqrt{T}\bar{v}_{2,T}(\theta^*)$ obeys the central limit theorem and

$$\sqrt{T}\bar{v}_{2,T}(\theta^*) \overset{A}{\sim} N(0_n, M_{\lambda\lambda}). \qquad \text{(A.16)}$$

Thus, using the fact that $M_{\lambda\lambda} = 4S$ under the null, we obtain

$$T\hat{\delta}^2 \overset{A}{\sim} z'S^{\frac{1}{2}}U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}S^{\frac{1}{2}}z, \qquad \text{(A.17)}$$

where $z \sim N(0_n, I_n)$. Since $S^{\frac{1}{2}}U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}S^{\frac{1}{2}}$ has the same nonzero eigenvalues as $P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$, we have

$$T\hat{\delta}^2 \overset{A}{\sim} F_{n-k}(\xi), \qquad \text{(A.18)}$$

where the $\xi_i$'s are the eigenvalues of $P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$. This completes the proof of part (a).

(b) Now consider the case $\delta > 0$. In this situation, the asymptotic behavior of $\sqrt{T}(\hat{\delta}^2 - \delta^2)$ is determined by $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)])$, which converges weakly to a Gaussian process. Under Assumptions A, B and C, and since $E[\phi_t(\theta^*)] = \delta^2$, we have

$$\sqrt{T}(\hat{\delta}^2 - \delta^2) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)]) + o_p(1) \overset{A}{\sim} N(0, \sigma_b^2). \qquad \text{(A.19)}$$

This completes the proof of part (b). ■

***Proof of Lemma 2.*** (a) For $\delta > 0$ and under Assumptions A, B and C,

$$\sqrt{T}\bar{v}_T(\theta^*) \overset{A}{\sim} N(0_{n+k}, M). \qquad \text{(A.20)}$$

Then, combining (A.11) and (A.20), we obtain

$$\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{n+k}, H^{-1}MH^{-1}). \qquad \text{(A.21)}$$

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

$$H^{-1}MH^{-1} = \sum_{j=-\infty}^{\infty} E[l_t l'_{t+j}], \tag{A.22}$$

where

$$l_t \equiv \begin{bmatrix} l_{1t} \\ l_{2t} \end{bmatrix} = H^{-1}\frac{\partial \phi_t(\theta^*)}{\partial \theta}. \tag{A.23}$$

Using (A.13), (18) and (19), we can express $l_{1t}$ and $l_{2t}$ as

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

Using the definition of $m_t(\theta^*)$ and rearranging 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

$$\begin{aligned} l_{1t} &= \tilde{l}_{1t} = (D'U^{-1}D)^{-1}D'U^{-1}e_t(\gamma^*), & \text{(A.26)} \\ l_{2t} &= U^{-1}[Dl_{1t} - e_t(\gamma^*)]. & \text{(A.27)} \end{aligned}$$

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 2, we have

$$\sqrt{T}P'U^{\frac{1}{2}}\hat{\lambda} \overset{A}{\sim} N(0_{n-k}, P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P) \tag{A.28}$$

when $\delta = 0$, or equivalently

$$\sqrt{T}(P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P)^{-\frac{1}{2}}P'U^{\frac{1}{2}}\hat{\lambda} \overset{A}{\sim} N(0_{n-k}, I_{n-k}). \tag{A.29}$$

Then, under Assumptions A, B and C,

$$\begin{aligned} LM_{\hat{\lambda}} &= T\hat{\lambda}'\hat{U}^{\frac{1}{2}}\hat{P}\left(\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\right)^{-1}\hat{P}'\hat{U}^{\frac{1}{2}}\hat{\lambda} \\ &= T\hat{\lambda}'U^{\frac{1}{2}}P\left(P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P\right)^{-1}P'U^{\frac{1}{2}}\hat{\lambda} + o_p(1) \\ &\overset{A}{\sim} \chi^2_{n-k}. & \text{(A.30)} \end{aligned}$$

This completes the proof. ■

*Example of two strictly non-nested models that are both correctly specified.* Let $R$ be the gross returns on $N$ risky assets and $R_0$ be the gross return on the risk-free asset. Suppose that $R_p$ is the gross return on the tangency portfolio of the $N$ risky assets. Simple mean-variance mathematics gives

$$E[R] = R_0 1_N + \mathrm{Cov}[R, R_p]\gamma^*, \tag{A.31}$$

where $\gamma^* = (\mu_p - R_0)/\sigma_p^2$, with $\mu_p = E[R_p]$ and $\sigma_p^2 = \mathrm{Var}[R_p]$. In addition, assume that $R$ is multivariate normally distributed. Consider the following SDF

$$y^{\mathcal{F}}(\gamma_{\mathcal{F}}) = \frac{1}{R_0} \exp\left(a_0 - \gamma_{\mathcal{F}} R_p\right), \tag{A.32}$$

where

$$a_0 = -\frac{\gamma^{*2}\sigma_p^2}{2} + \gamma^* \mu_p. \tag{A.33}$$

Let $1_N$ denote an $N$-vector of ones. Using Stein's lemma, we can easily establish that

$$E[R_0 y^{\mathcal{F}}(\gamma^*)] = 1, \qquad E[R y^{\mathcal{F}}(\gamma^*)] = 1_N, \tag{A.34}$$

so that $y^{\mathcal{F}}(\gamma^*) \in \mathcal{M}$.

Now consider a factor $f = R_p + \epsilon$, where $\epsilon$ is a normal mean-zero measurement error independent of the returns. It follows that $\mu_f = E[f] = \mu_p$ and $\sigma_f^2 = \mathrm{Var}[f] > \sigma_p^2$. Consider an alternative SDF

$$y^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{1}{R_0} \exp\left(\tilde{a}_0 - \gamma_{\mathcal{G}} f\right), \tag{A.35}$$

where

$$\tilde{a}_0 = -\frac{\gamma^{*2}\sigma_f^2}{2} + \gamma^* \mu_p. \tag{A.36}$$

Using Stein's lemma again, we obtain

$$E[R_0 y^{\mathcal{G}}(\gamma^*)] = 1, \qquad E[R y^{\mathcal{G}}(\gamma^*)] = 1_N, \tag{A.37}$$

and $y^{\mathcal{G}}(\gamma^*)$ is also a correctly specified model. Note that $y^{\mathcal{F}}(\gamma_{\mathcal{F}})$ and $y^{\mathcal{G}}(\gamma_{\mathcal{G}})$ are two strictly non-nested models because there are no choices of $\gamma_{\mathcal{F}}$ and $\gamma_{\mathcal{G}}$ that can make these two SDFs identical. This example shows that we can have two strictly non-nested SDFs that are both correctly specified. ■

***Proof of Theorem 2.*** (a) From Lemma A.1 and under the null $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$, we can use (A.15) to obtain

$$
\begin{aligned}
&T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}) \\
&= \frac{1}{4}\begin{bmatrix} \sqrt{T}\bar{v}^{\mathcal{F}}_{2,T}(\theta^*_{\mathcal{F}}) \\ \sqrt{T}\bar{v}^{\mathcal{G}}_{2,T}(\theta^*_{\mathcal{G}}) \end{bmatrix}' \begin{bmatrix} U^{-\frac{1}{2}}P_{\mathcal{F}}P'_{\mathcal{F}}U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}}P_{\mathcal{G}}P'_{\mathcal{G}}U^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{T}\bar{v}^{\mathcal{F}}_{2,T}(\theta^*_{\mathcal{F}}) \\ \sqrt{T}\bar{v}^{\mathcal{G}}_{2,T}(\theta^*_{\mathcal{G}}) \end{bmatrix} + o_p(1). \quad \text{(A.38)}
\end{aligned}
$$

From Assumptions A, B and C, we have

$$
\begin{bmatrix} \sqrt{T}\bar{v}^{\mathcal{F}}_{2,T}(\theta^*_{\mathcal{F}}) \\ \sqrt{T}\bar{v}^{\mathcal{G}}_{2,T}(\theta^*_{\mathcal{G}}) \end{bmatrix} \overset{A}{\sim} N\left(0_{2n}, 4\mathcal{S}\right). \tag{A.39}
$$

Hence,

$$
T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}) \overset{A}{\sim} z'\mathcal{S}^{\frac{1}{2}} \begin{bmatrix} U^{-\frac{1}{2}}P_{\mathcal{F}}P'_{\mathcal{F}}U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}}P_{\mathcal{G}}P'_{\mathcal{G}}U^{-\frac{1}{2}} \end{bmatrix} \mathcal{S}^{\frac{1}{2}} z, \tag{A.40}
$$

where $z \sim N(0_{2n}, I_{2n})$. Furthermore, the nonzero eigenvalues of

$$
\mathcal{S}^{\frac{1}{2}} \begin{bmatrix} U^{-\frac{1}{2}}P_{\mathcal{F}}P'_{\mathcal{F}}U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}}P_{\mathcal{G}}P'_{\mathcal{G}}U^{-\frac{1}{2}} \end{bmatrix} \mathcal{S}^{\frac{1}{2}} \tag{A.41}
$$

are the same as the eigenvalues of the matrix

$$
\begin{bmatrix} P'_{\mathcal{F}}U^{-\frac{1}{2}} & 0_{(n-k_1)\times n} \\ 0_{(n-k_2)\times n} & P'_{\mathcal{G}}U^{-\frac{1}{2}} \end{bmatrix} \mathcal{S} \begin{bmatrix} U^{-\frac{1}{2}}P_{\mathcal{F}} & 0_{n\times(n-k_2)} \\ 0_{n\times(n-k_1)} & -U^{-\frac{1}{2}}P_{\mathcal{G}} \end{bmatrix}. \tag{A.42}
$$

Then, it follows that

$$
T(\hat{\delta}^2_{\mathcal{F}} - \hat{\delta}^2_{\mathcal{G}}) \overset{A}{\sim} F_{2n-k_1-k_2}(\xi), \tag{A.43}
$$

where the $\xi_i$'s are the eigenvalues of the matrix

$$
\begin{bmatrix} P'_{\mathcal{F}}U^{-\frac{1}{2}}S_{\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & -P'_{\mathcal{F}}U^{-\frac{1}{2}}S_{\mathcal{F}\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \\ P'_{\mathcal{G}}U^{-\frac{1}{2}}S_{\mathcal{G}\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & -P'_{\mathcal{G}}U^{-\frac{1}{2}}S_{\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \end{bmatrix}. \tag{A.44}
$$

This completes the proof of part (a).

(b) Using the result in part (b) of Lemma 2, it can be shown that when $\lambda_{\mathcal{F}} = \lambda_{\mathcal{G}} = 0_n$,

$$
\sqrt{T}\tilde{\lambda}_{\mathcal{F}\mathcal{G}} \overset{A}{\sim} N(0_{2n-k_1-k_2}, \Sigma_{\tilde{\lambda}_{\mathcal{F}\mathcal{G}}}), \tag{A.45}
$$

where

$$
\Sigma_{\tilde{\lambda}_{\mathcal{F}\mathcal{G}}} = \begin{bmatrix} P'_{\mathcal{F}}U^{-\frac{1}{2}}S_{\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & P'_{\mathcal{F}}U^{-\frac{1}{2}}S_{\mathcal{F}\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \\ P'_{\mathcal{G}}U^{-\frac{1}{2}}S_{\mathcal{G}\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & P'_{\mathcal{G}}U^{-\frac{1}{2}}S_{\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \end{bmatrix}. \tag{A.46}
$$

Using the fact that $\hat{\Sigma}_{\tilde{\lambda}_{\mathcal{FG}}}$ is a consistent estimator of $\Sigma_{\tilde{\lambda}_{\mathcal{FG}}}$, we have

$$LM_{\tilde{\lambda}_{\mathcal{FG}}} = T\tilde{\lambda}'_{\mathcal{FG}}\hat{\Sigma}^{-1}_{\tilde{\lambda}_{\mathcal{FG}}}\tilde{\lambda}_{\mathcal{FG}} \overset{A}{\sim} \chi^2_{2n-k_1-k_2}. \tag{A.47}$$

This completes the proof of part (b). ■

***Proof of Theorem 3.*** (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

$$\frac{\partial \phi_t^{\mathcal{F}}(\theta^*_{\mathcal{F}})}{\partial \lambda_{\mathcal{F}}} = 2[x_t m_t^{\mathcal{F}}(\theta^*_{\mathcal{F}}) - q_{t-1}] = 2[x_t m_t^{\mathcal{G}}(\theta^*_{\mathcal{G}}) - q_{t-1}] = \frac{\partial \phi_t^{\mathcal{G}}(\theta^*_{\mathcal{G}})}{\partial \lambda_{\mathcal{G}}} \tag{A.48}$$

and

$$\bar{v}^{\mathcal{F}}_{2,T}(\theta^*_{\mathcal{F}}) = \bar{v}^{\mathcal{G}}_{2,T}(\theta^*_{\mathcal{G}}). \tag{A.49}$$

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

$$H_0 : \gamma^*_{\mathcal{G}} = g(\gamma^*_{\mathcal{F}}), \tag{A.50}$$

where $g(\cdot)$ is a twice continuously differentiable function from $\Gamma_{\mathcal{F}}$ to $\Gamma_{\mathcal{G}}$ (see Gallant, 1987 and Vuong, 1989).[22] Denote by

$$G(\gamma_{\mathcal{F}}) = \frac{\partial g(\gamma_{\mathcal{F}})}{\partial \gamma'_{\mathcal{F}}} \tag{A.51}$$

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

$$\Psi^{\mathcal{G}}(\gamma^*_{\mathcal{G}})G(\gamma^*_{\mathcal{F}}) = \Psi^{\mathcal{G}}(g(\gamma^*_{\mathcal{F}}))G(\gamma^*_{\mathcal{F}}) = 0_{(k_2-k_1)\times k_1}. \tag{A.52}$$

Define the matrices

$$\mathbb{S} = [\Psi^{\mathcal{G}}_*,\ 0_{(k_2-k_1)\times n}], \qquad \mathbb{Q} = \begin{bmatrix} G(\gamma^*_{\mathcal{F}}) & 0_{k_2\times n} \\ 0_{n\times k_1} & I_n \end{bmatrix} \tag{A.53}$$

and note that $\mathbb{SQ} = 0_{(k_2-k_1)\times(n+k_1)}$. Then, using (A.49) and (A.50), it follows that (see Lemma B in Vuong, 1989)

$$\bar{v}^{\mathcal{F}}_T(\theta^*_{\mathcal{F}}) = \mathbb{Q}'\bar{v}^{\mathcal{G}}_T(\theta^*_{\mathcal{G}}) \tag{A.54}$$

and

$$H_{\mathcal{F}} = \mathbb{Q}'H_{\mathcal{G}}\mathbb{Q}. \tag{A.55}$$

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

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

$$\begin{aligned}
& T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \\
= & -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)' H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
= & -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'\mathbb{Q}(\mathbb{Q}' H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
= & \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-\frac{1}{2}}[I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\ \mathbb{Q}(\mathbb{Q}' H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}' H_{\mathcal{G}}^{\frac{1}{2}}] H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \qquad (A.56)
\end{aligned}$$

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)

$$I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\mathbb{Q}(\mathbb{Q}' H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}' H_{\mathcal{G}}^{\frac{1}{2}} = H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}. \qquad (A.57)$$

Substituting (A.57) into (A.56) yields

$$\begin{aligned}
T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) &= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-\frac{1}{2}}[H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}] H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
&= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \qquad (A.58)
\end{aligned}$$

Furthermore, invoking

$$\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \overset{A}{\sim} N\left(0_{n+k_2}, M_{\mathcal{G}}\right), \qquad (A.59)$$

we have

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} \frac{1}{2}z'\left[M_{\mathcal{G}}^{\frac{1}{2}} H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}^{\frac{1}{2}}\right] z, \qquad (A.60)$$

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 2. 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

$$\frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}' = \frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}\Sigma_{\hat{\theta}_{\mathcal{G}}}\mathbb{S}' = (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1}\Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}, \qquad (A.61)$$

we conclude that

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_2-k_1}(\xi), \qquad (A.62)$$

where the $\xi_i$'s are the eigenvalues of the matrix in (A.61). 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,

$$\sqrt{T}\hat{\psi}_{\mathcal{G}} \overset{A}{\sim} N(0_{k_2-k_1}, \Psi_*^{\mathcal{G}} \Sigma_{\hat{\gamma}_{\mathcal{G}}} \Psi_*^{\mathcal{G}\prime}). \tag{A.63}$$

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 4.*** (a) Define the following matrices

$$\mathbb{S}_{\mathcal{F}} = \left[\Psi_*^{\mathcal{F}},\ 0_{(k_1-k_3)\times n}\right], \qquad \mathbb{S}_{\mathcal{G}} = \left[\Psi_*^{\mathcal{G}},\ 0_{(k_2-k_3)\times n}\right]. \tag{A.64}$$

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 3 to obtain

$$T(\hat{\delta}_{\mathcal{H}}^2 - \hat{\delta}_{\mathcal{F}}^2) = \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)' H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}'(\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}')^{-1}\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) + o_p(1) \tag{A.65}$$

and

$$T(\hat{\delta}_{\mathcal{H}}^2 - \hat{\delta}_{\mathcal{G}}^2) = \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}'(\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}')^{-1}\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \tag{A.66}$$

Taking the difference yields

$$\begin{aligned} T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) &= -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)' H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}'(\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}')^{-1}\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ &\quad +\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}'(\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}')^{-1}\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \end{aligned} \tag{A.67}$$

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

$$\begin{bmatrix} \sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ \sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \end{bmatrix} \overset{A}{\sim} N\left(0_{2n+k_1+k_2}, \mathbb{M}\right), \tag{A.68}$$

where

$$\mathbb{M} = \begin{bmatrix} M_{\mathcal{F}} & M_{\mathcal{F}\mathcal{G}} \\ M_{\mathcal{G}\mathcal{F}} & M_{\mathcal{G}} \end{bmatrix} = \lim_{T\to\infty} \mathrm{Var}\begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)}{\partial \theta_{\mathcal{F}}} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)}{\partial \theta_{\mathcal{G}}} \end{bmatrix}. \tag{A.69}$$

Hence,

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} z'\left[\frac{1}{2}\mathbb{M}^{\frac{1}{2}}\begin{pmatrix} -H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}'(\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}')^{-1}\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}'(\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}')^{-1}\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \end{pmatrix}\mathbb{M}^{\frac{1}{2}}\right] z, \tag{A.70}$$

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.70) has the same nonzero eigenvalues as the matrix

$$\begin{aligned}&\frac{1}{2}\begin{bmatrix} -(\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\mathbb{S}'_{\mathcal{F}})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}'_{\mathcal{G}})^{-1}\end{bmatrix} \\ &\times\begin{bmatrix} \mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1} & 0_{(k_1-k_3)\times(n+k_2)} \\ 0_{(k_2-k_3)\times(n+k_1)} & \mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\end{bmatrix}\mathbb{M}\begin{bmatrix} H_{\mathcal{F}}^{-1}\mathbb{S}'_{\mathcal{F}} & 0_{(n+k_1)\times(k_2-k_3)} \\ 0_{(n+k_2)\times(k_1-k_3)} & H_{\mathcal{G}}^{-1}\mathbb{S}'_{\mathcal{G}}\end{bmatrix}.\end{aligned} \tag{A.71}$$

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

$$\Sigma_{\hat{\theta}_{\mathcal{FG}}} = \begin{bmatrix} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1}\end{bmatrix}\mathbb{M}\begin{bmatrix} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1}\end{bmatrix} \tag{A.72}$$

is the asymptotic covariance matrix of $[\hat{\theta}'_{\mathcal{F}},\ \hat{\theta}'_{\mathcal{G}}]'$, the matrix in (A.71) can be written as

$$\begin{bmatrix} -(\Psi_*^{\mathcal{F}}\tilde{H}_{\mathcal{F}}\Psi_*^{\mathcal{F}\prime})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1}\end{bmatrix}\Psi_*^{\mathcal{FG}}\Sigma_{\hat{\gamma}_{\mathcal{FG}}}\Psi_*^{\mathcal{FG}\prime}. \tag{A.73}$$

Therefore,

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_1+k_2-2k_3}(\xi), \tag{A.74}$$

where the $\xi_i$'s are the eigenvalues of the matrix in (A.73). This completes the proof of part (a).

(b) By the delta method,

$$\sqrt{T}\hat{\psi}_{\mathcal{FG}} \overset{A}{\sim} N(0_{k_1+k_2-2k_3}, \Psi_*^{\mathcal{FG}}\Sigma_{\hat{\gamma}_{\mathcal{FG}}}\Psi_*^{\mathcal{FG}\prime}). \tag{A.75}$$

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

$$Wald_{\hat{\psi}_{\mathcal{FG}}} = T\hat{\psi}'_{\mathcal{FG}}(\hat{\Psi}^{\mathcal{FG}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}\hat{\Psi}^{\mathcal{FG}\prime})^{-1}\hat{\psi}_{\mathcal{FG}}, \tag{A.76}$$

we obtain immediately

$$Wald_{\hat{\psi}_{\mathcal{FG}}} \overset{A}{\sim} \chi^2_{k_1+k_2-2k_3}. \tag{A.77}$$

This completes the proof of part (b). ■

REFERENCES


AHN, S. C. and GADAROWSKI, C. (2004), "Small Sample Properties of the Model Specification Test based on the Hansen-Jagannathan Distance," *Journal of Empirical Finance*, **11**, 109–132.

ALMEIDA, C. and GARCIA, R. (2009), "Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators,"(Discussion Paper, EDHEC Business School).

ANDREWS, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, **59**, 817–858.

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* (Amsterdam: North-Holland) 2247–2294.

ANDREWS, D. W. K. and POLLARD, D. (1994), "An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes," *International Statistical Review*, **62**, 119–132.

CHEN, X. and LUDVIGSON, S. C. (2009), "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics*, **24**, 1057–1093.

COCHRANE, J. H. (2005), *Asset Pricing* (Princeton: Princeton University Press).

FAMA, E. F. and FRENCH, K. R. (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* (New York: Wiley).

GOLDEN, R. M. (2003), "Discrepancy Risk Model Selection Test Theory for Comparing Possibly Misspecified or Nonnested Models," *Psychometrika*, **68**, 229–249.

GOSPODINOV, N., KAN, R. and ROBOTTI, C. (2010a), "On the Hansen-Jagannathan Distance with a No-Arbitrage Constraint,"(Discussion Paper, Federal Reserve Bank of Atlanta).

GOSPODINOV, N., KAN, R. and ROBOTTI, C. (2010b), "Further Results on the Limiting Distribution of GMM Sample Moment Conditions,"(Discussion Paper, Federal Reserve Bank of Atlanta).

HALL, A. R. and PELLETIER, D. (2010), "Nonnested Testing in Models Estimated via Generalized Method of Moments," *Econometric Theory*, forthcoming.

HANSEN, L. P., HEATON, J. C. and LUTTMER, E. G. J. (1995), "Econometric Evaluation of Asset Pricing Models," *Review of Financial Studies*, **8**, 237–274.

HANSEN, L. P. and JAGANNATHAN, R. (1997), "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, **52**, 557–590.

IMBENS, G. W., SPADY, R. H. and JOHNSON, P. (1998), "Information Theoretic Approached to Inference in Moment Condition Models," *Econometrica*, **66**, 333–357.

JAGANNATHAN, R. and WANG, Z. (1996), "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, **51**, 3–53.

KAN, R. and ROBOTTI, C. (2008), "Specification Tests of Asset Pricing Models Using Excess Returns," *Journal of Empirical Finance*, **15**, 816–838.

KAN, R. and ROBOTTI, C. (2009), "Model Comparison Using the Hansen-Jagannathan Distance," *Review of Financial Studies*, **22**, 3449–3490.

KAN, R., ROBOTTI, C. and SHANKEN, J. (2010), "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology," (Discussion Paper, Federal Reserve Bank of Atlanta).

KAN, R. and ZHANG, C. (1999), "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, **54**, 103–127.

KUDO, A. (1963), "A Multivariate Analogue of the One-Sided Test," *Biometrika*, **50**, 403–418.

LETTAU, M., and LUDVIGSON, S. C. (2001), "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, **109**, 1238–1287.

LI, H., XU, Y. and ZHANG, X. (2010), "Evaluating Asset Pricing Models Using the Second Hansen-Jagannathan Distance," *Journal of Financial Economics*, **97**, 279–301.

LIEN, D. and VUONG, Q. H. (1987), "Selecting the Best Linear Regression Model: A Classical Approach," *Journal of Econometrics*, **35**, 3–23.

MARCELLINO, M. and ROSSI, B. (2008), "Model Selection for Nested and Overlapping Nonlinear, Dynamic and Possibly Mis-Specified Models," *Oxford Bulletin of Economics and Statistics*, **70**, 867–893.

NEWEY, W. K. and WEST, K. D. (1987), "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, **55**, 703–708.

PARKER, J. A. and JULLIARD, C. (2005), "Consumption Risk and the Cross Section of Expected Returns," *Journal of Political Economy*, **113**, 185–222.

RIVERS, D. and VUONG, Q. H. (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 WRIGHT, J. H. (2000), "GMM with Weak Identification," *Econometrica*, **68**, 1055–1096.

VUONG, Q. H. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, **57**, 307–333.

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*

| | CAPM | | | YOGO | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size ($mkt$) | | | Size ($mkt$) | | | Size ($ndur$) | | | Size ($dur$) | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.097 | 0.047 | 0.008 | 0.076 | 0.032 | 0.004 | 0.083 | 0.037 | 0.006 | 0.084 | 0.039 | 0.006 |
| 240 | 0.098 | 0.048 | 0.009 | 0.078 | 0.033 | 0.004 | 0.085 | 0.039 | 0.006 | 0.083 | 0.038 | 0.006 |
| 360 | 0.097 | 0.048 | 0.009 | 0.082 | 0.037 | 0.005 | 0.089 | 0.042 | 0.006 | 0.087 | 0.040 | 0.006 |
| 480 | 0.098 | 0.049 | 0.010 | 0.085 | 0.039 | 0.006 | 0.090 | 0.044 | 0.007 | 0.088 | 0.041 | 0.007 |
| 600 | 0.098 | 0.049 | 0.010 | 0.089 | 0.041 | 0.006 | 0.092 | 0.044 | 0.008 | 0.091 | 0.042 | 0.007 |

Panel B: *t-tests under correctly specified models*

| | CAPM | | | YOGO | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size ($mkt$) | | | Size ($mkt$) | | | Size ($ndur$) | | | Size ($dur$) | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.098 | 0.048 | 0.008 | 0.157 | 0.086 | 0.018 | 0.282 | 0.189 | 0.066 | 0.291 | 0.199 | 0.073 |
| 240 | 0.098 | 0.048 | 0.009 | 0.194 | 0.117 | 0.033 | 0.317 | 0.225 | 0.097 | 0.331 | 0.239 | 0.106 |
| 360 | 0.098 | 0.049 | 0.009 | 0.221 | 0.140 | 0.046 | 0.337 | 0.248 | 0.120 | 0.357 | 0.265 | 0.130 |
| 480 | 0.099 | 0.049 | 0.010 | 0.236 | 0.153 | 0.057 | 0.352 | 0.263 | 0.132 | 0.374 | 0.283 | 0.147 |
| 600 | 0.099 | 0.049 | 0.010 | 0.248 | 0.166 | 0.064 | 0.360 | 0.272 | 0.142 | 0.385 | 0.296 | 0.158 |

*Notes:* The table presents the empirical size of the $t$-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 returns are generated from a multivariate normal distribution. The various $t$-ratios are compared to the critical values from a standard normal distribution. Panel A reports results for $t$-tests under potentially misspecified models based on the asymptotic covariance matrix in (31) and (32), while Panel B reports results for $t$-tests under correctly specified models based on the asymptotic covariance matrix in (34) and (35).

## TABLE 2

*Model specification tests*

Panel A: *HJ-distance test using S*

| | CAPM | | | | | | YOGO | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | | | Power | | | Size | | | Power | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.343 | 0.240 | 0.101 | 1.000 | 0.999 | 0.995 | 0.267 | 0.174 | 0.064 | 0.991 | 0.982 | 0.946 |
| 240 | 0.202 | 0.123 | 0.037 | 1.000 | 1.000 | 1.000 | 0.166 | 0.096 | 0.027 | 1.000 | 1.000 | 0.999 |
| 360 | 0.162 | 0.092 | 0.025 | 1.000 | 1.000 | 1.000 | 0.139 | 0.077 | 0.019 | 1.000 | 1.000 | 1.000 |
| 480 | 0.145 | 0.079 | 0.020 | 1.000 | 1.000 | 1.000 | 0.127 | 0.068 | 0.016 | 1.000 | 1.000 | 1.000 |
| 600 | 0.134 | 0.073 | 0.017 | 1.000 | 1.000 | 1.000 | 0.122 | 0.064 | 0.014 | 1.000 | 1.000 | 1.000 |

Panel B: *HJ-distance test using* $S_A$

| | CAPM | | | | | | YOGO | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | | | Power | | | Size | | | Power | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.061 | 0.019 | 0.001 | 0.988 | 0.958 | 0.716 | 0.052 | 0.017 | 0.001 | 0.941 | 0.869 | 0.563 |
| 240 | 0.083 | 0.036 | 0.005 | 1.000 | 1.000 | 1.000 | 0.072 | 0.030 | 0.004 | 1.000 | 0.999 | 0.995 |
| 360 | 0.089 | 0.041 | 0.006 | 1.000 | 1.000 | 1.000 | 0.080 | 0.036 | 0.005 | 1.000 | 1.000 | 1.000 |
| 480 | 0.091 | 0.042 | 0.007 | 1.000 | 1.000 | 1.000 | 0.084 | 0.039 | 0.006 | 1.000 | 1.000 | 1.000 |
| 600 | 0.093 | 0.044 | 0.008 | 1.000 | 1.000 | 1.000 | 0.086 | 0.040 | 0.007 | 1.000 | 1.000 | 1.000 |

Panel C: *LM test*

| | CAPM | | | | | | YOGO | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | | | Power | | | Size | | | Power | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.117 | 0.054 | 0.008 | 0.995 | 0.984 | 0.905 | 0.121 | 0.058 | 0.009 | 0.982 | 0.959 | 0.846 |
| 240 | 0.111 | 0.054 | 0.010 | 1.000 | 1.000 | 1.000 | 0.104 | 0.050 | 0.009 | 1.000 | 1.000 | 0.999 |
| 360 | 0.106 | 0.053 | 0.010 | 1.000 | 1.000 | 1.000 | 0.099 | 0.049 | 0.009 | 1.000 | 1.000 | 1.000 |
| 480 | 0.104 | 0.051 | 0.010 | 1.000 | 1.000 | 1.000 | 0.099 | 0.049 | 0.009 | 1.000 | 1.000 | 1.000 |
| 600 | 0.103 | 0.052 | 0.010 | 1.000 | 1.000 | 1.000 | 0.099 | 0.047 | 0.009 | 1.000 | 1.000 | 1.000 |

*Notes:* The table presents the empirical size and power of three tests of $H_0 : \delta^2 = 0$. Panel A is for the test in part (a) of Lemma 1 that uses the matrix $S$. Panel B is for the HJ-distance test based on $S_A$. Finally, Panel C is for 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 returns are generated from a multivariate normal distribution.

TABLE 3

*Model selection tests*

Panel A: *Joint tests of correct specification for two strictly non-nested SDFs*

| | Weighted $\chi^2$ test | | | | | | LM test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | | | Power | | | Size | | | Power | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.187 | 0.113 | 0.035 | 0.585 | 0.499 | 0.336 | 0.049 | 0.013 | 0.000 | 0.831 | 0.663 | 0.247 |
| 240 | 0.138 | 0.077 | 0.020 | 0.806 | 0.747 | 0.611 | 0.078 | 0.031 | 0.004 | 1.000 | 1.000 | 1.000 |
| 360 | 0.127 | 0.069 | 0.016 | 0.913 | 0.881 | 0.791 | 0.088 | 0.038 | 0.005 | 1.000 | 1.000 | 1.000 |
| 480 | 0.119 | 0.063 | 0.014 | 0.961 | 0.944 | 0.892 | 0.092 | 0.042 | 0.006 | 1.000 | 1.000 | 1.000 |
| 600 | 0.115 | 0.060 | 0.013 | 0.984 | 0.975 | 0.946 | 0.093 | 0.043 | 0.007 | 1.000 | 1.000 | 1.000 |

Panel B: *Pairwise model comparison tests for nested SDFs*

| | Restricted weighted $\chi^2$ test | | | Wald test | | | Unrestricted weighted $\chi^2$ test | | |
|---|---|---|---|---|---|---|---|---|---|
| | Size | | | Size | | | Size | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.123 | 0.065 | 0.014 | 0.108 | 0.052 | 0.010 | 0.189 | 0.112 | 0.032 |
| 240 | 0.106 | 0.053 | 0.010 | 0.101 | 0.047 | 0.007 | 0.180 | 0.106 | 0.030 |
| 360 | 0.102 | 0.050 | 0.010 | 0.099 | 0.046 | 0.007 | 0.166 | 0.095 | 0.025 |
| 480 | 0.099 | 0.049 | 0.009 | 0.099 | 0.047 | 0.007 | 0.154 | 0.086 | 0.022 |
| 600 | 0.098 | 0.048 | 0.009 | 0.100 | 0.048 | 0.008 | 0.147 | 0.080 | 0.021 |

| | Power | | | Power | | | Power | | |
|---|---|---|---|---|---|---|---|---|---|
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.337 | 0.198 | 0.046 | 0.469 | 0.334 | 0.130 | 0.311 | 0.188 | 0.050 |
| 240 | 0.505 | 0.343 | 0.112 | 0.647 | 0.524 | 0.284 | 0.451 | 0.311 | 0.115 |
| 360 | 0.634 | 0.474 | 0.195 | 0.760 | 0.656 | 0.421 | 0.563 | 0.419 | 0.183 |
| 480 | 0.732 | 0.585 | 0.285 | 0.836 | 0.751 | 0.535 | 0.657 | 0.518 | 0.256 |
| 600 | 0.807 | 0.681 | 0.377 | 0.888 | 0.821 | 0.634 | 0.737 | 0.610 | 0.333 |

Panel C: *Multiple model comparison test for nested SDFs*

| | Wald test | | | | | |
|---|---|---|---|---|---|---|
| | Size | | | Power | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.155 | 0.084 | 0.020 | 0.372 | 0.246 | 0.083 |
| 240 | 0.118 | 0.059 | 0.012 | 0.573 | 0.441 | 0.215 |
| 360 | 0.110 | 0.055 | 0.011 | 0.713 | 0.597 | 0.359 |
| 480 | 0.109 | 0.054 | 0.010 | 0.809 | 0.713 | 0.491 |
| 600 | 0.108 | 0.053 | 0.010 | 0.875 | 0.802 | 0.606 |

TABLE 3 (continued)

*Model selection tests*

Panel D: *Pairwise tests of equality for overlapping SDFs*

| | Restricted weighted $\chi^2$ test | | | Wald test | | | Unrestricted weighted $\chi^2$ test | | |
|---|---|---|---|---|---|---|---|---|---|
| | Size | | | Size | | | Size | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.091 | 0.042 | 0.007 | 0.109 | 0.053 | 0.010 | 0.101 | 0.051 | 0.010 |
| 240 | 0.091 | 0.043 | 0.007 | 0.099 | 0.048 | 0.008 | 0.108 | 0.057 | 0.013 |
| 360 | 0.093 | 0.044 | 0.007 | 0.098 | 0.047 | 0.009 | 0.108 | 0.058 | 0.013 |
| 480 | 0.094 | 0.045 | 0.008 | 0.099 | 0.048 | 0.009 | 0.106 | 0.056 | 0.013 |
| 600 | 0.095 | 0.045 | 0.008 | 0.100 | 0.049 | 0.009 | 0.105 | 0.054 | 0.012 |

| | Power | | | Power | | | Power | | |
|---|---|---|---|---|---|---|---|---|---|
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.649 | 0.560 | 0.359 | 0.776 | 0.654 | 0.374 | 0.418 | 0.327 | 0.171 |
| 240 | 0.888 | 0.857 | 0.758 | 0.980 | 0.957 | 0.851 | 0.722 | 0.663 | 0.528 |
| 360 | 0.959 | 0.948 | 0.914 | 0.999 | 0.997 | 0.982 | 0.870 | 0.838 | 0.760 |
| 480 | 0.984 | 0.980 | 0.968 | 1.000 | 1.000 | 0.998 | 0.941 | 0.925 | 0.883 |
| 600 | 0.993 | 0.992 | 0.987 | 1.000 | 1.000 | 1.000 | 0.973 | 0.964 | 0.944 |

Panel E: *Pairwise and multiple model comparison tests for overlapping distinct SDFs*

| | $p = 1$ | | | | | | $p = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | | | Power | | | Size | | | Power | | |
| $T$ | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 120 | 0.134 | 0.069 | 0.011 | 0.386 | 0.259 | 0.085 | 0.134 | 0.066 | 0.009 | 0.466 | 0.331 | 0.121 |
| 240 | 0.109 | 0.053 | 0.007 | 0.600 | 0.461 | 0.199 | 0.106 | 0.048 | 0.005 | 0.699 | 0.562 | 0.273 |
| 360 | 0.104 | 0.050 | 0.008 | 0.759 | 0.640 | 0.371 | 0.100 | 0.046 | 0.006 | 0.851 | 0.752 | 0.482 |
| 480 | 0.102 | 0.050 | 0.008 | 0.857 | 0.768 | 0.538 | 0.100 | 0.046 | 0.006 | 0.928 | 0.869 | 0.669 |
| 600 | 0.102 | 0.048 | 0.008 | 0.916 | 0.856 | 0.673 | 0.098 | 0.046 | 0.007 | 0.967 | 0.935 | 0.803 |

*Notes:* The table presents the empirical size and power of pairwise and multiple model comparison tests for strictly non-nested (Panel A), nested (Panels B and C) and overlapping (Panels D and E) models. In Panel A, we report simulation results for the weighted chi-squared and LM tests in parts (a) and (b) of Theorem 2, respectively. In Panel B, we report, in the order, results for the restricted weighted chi-squared test in part (a) of Theorem 3, the Wald test in part (b) of Theorem 3 and the unrestricted weighted chi-squared test in (57)–(58). Panel C is for the Wald test for multiple nested model comparison analyzed in Section 3.2. Panel D reports results for the restricted weighted chi-squared test in part (a) of Theorem 4, the Wald test in part (b) of Theorem 4 and the unrestricted weighted chi-squared test in (57)–(58). Finally, Panel E presents results for the pairwise ($p = 1$) and multiple ($p = 2$) model comparison tests in (49) and (76), 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 returns are generated from a multivariate normal distribution.

# Chi-Squared Tests for Evaluation and Comparison of Asset Pricing Models

NIKOLAY GOSPODINOV, RAYMOND KAN, and CESARE ROBOTTI

APPENDIX

In this appendix, we provide a detailed description of some simulation designs that, for brevity, were not included in the paper. These simulation designs are used to study the empirical size properties of $t$-tests of $H_0 : \gamma_i = 0$ based on Lemma 2 and of the pairwise model comparison tests for nested and overlapping models based on Theorem 3 and 4. The remaining simulation designs are already described in the paper.

*SDF Parameter Estimates*

Denote by $x_t$ an $n$-vector of gross returns on the test assets and by $f_t$ a $K$-vector of risk factors. Let $Y_t = [f_t', \ x_t']'$. Without loss of generality, assume that $\mu_1 = E[f_t] = 0_K$. In addition, let $\mu_2 = E[x_t]$,

$$V = \mathrm{Var}[Y_t] = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{1}$$

and $U = V_{22} + \mu_2\mu_2'$. We consider linear SDFs of the form

$$y_t = \gamma_0 + \gamma_1' f_t, \tag{2}$$

where $\gamma \equiv [\gamma_0, \ \gamma_1']'$ is a $(K + 1)$-vector of SDF parameters. The $\gamma$ vector that minimizes the population HJ-distance is given by

$$\gamma = (D'U^{-1}D)^{-1}D'U^{-1}1_n = (D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}1_n, \tag{3}$$

where $1_n$ is an $n$-vector of ones and $D = [\mu_2, \ V_{21}]$.

For a given $V$ (chosen based on the covariance matrix estimated from the data, i.e., $V = \hat{V}$), we are interested in how to set $\mu_2$ such that the SDF parameter associated with a given risk factor is equal to zero. Without loss of generality, in the following analysis we show how to set $\mu_2$ such that the SDF parameter associated with the first risk factor is equal to zero, i.e., $e_2'\gamma = 0$, where $e_2 = [0, \ 1, \ 0_{K-1}']'$.

For the case of correctly specified models, we start with $\hat{D} = [\hat{\mu}_2, \ V_{21}]$, where $\hat{\mu}_2$ is the vector of return means estimated from the data. Then, we obtain

$$\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{bmatrix} = (\hat{D}'V_{22}^{-1}\hat{D})^{-1}\hat{D}'V_{22}^{-1}1_n \tag{4}$$

and the pricing errors are given by

$$\hat{e} = \hat{D}\hat{\gamma} - 1_n. \tag{5}$$

Let $\tilde{\gamma} = \begin{bmatrix} \tilde{\gamma}_0 \\ \tilde{\gamma}_1 \end{bmatrix}$ be the same as $\hat{\gamma}$ except that its second element is set equal to zero. By setting

$$\mu_2 = \frac{1_n - V_{21}\tilde{\gamma}_1}{\tilde{\gamma}_0}, \tag{6}$$

we have

$$D\tilde{\gamma} = 1_n, \tag{7}$$

and the model is correctly specified.

We now turn to the more relevant case of misspecified models and show how to set $\mu_2$ such that $e_2'\gamma = 0$. For an $n$-vector $z$, we assume that

$$\mu_2 = \frac{1_n - V_{21}\tilde{\gamma}_1 + z}{\tilde{\gamma}_0}, \tag{8}$$

which implies that

$$D\tilde{\gamma} - z = 1_n. \tag{9}$$

The issue is how $z$ should be chosen. In order for $\tilde{\gamma}$ to be the pseudo parameters, we need

$$\tilde{\gamma} = (D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}1_n = \tilde{\gamma} - (D'V_{22}^{-1}D)^{-1}D'V_{22}^{-1}z. \tag{10}$$

This suggests that $z$ has to satisfy

$$D'V_{22}^{-1}z = 0_{K+1}. \tag{11}$$

The last $K$ equations are given by $V_{12}V_{22}^{-1}z = 0_K$, which means that $z$ has to be in the span of the null space of $V_{12}V_{22}^{-1}$. The first equation implies

$$\mu_2'V_{22}^{-1}z = 0. \tag{12}$$

Together with the restrictions $V_{12}V_{22}^{-1}z = 0_K$, it implies that $z$ has to satisfy

$$z'V_{22}^{-1}z = z'V_{22}^{-1}1_n. \tag{13}$$

Every vector in the null space of $V_{12}V_{22}^{-1}$ with a proper normalization will satisfy these $K + 1$ constraints. However, randomly picking one of these vectors may lead to unrealistic $\mu_2$ values. To obtain a vector of return means that is reasonable, we choose $z$ based on the following minimization problem:

$$\min_{z}(\mu_2 - \hat{\mu}_2)' V_{22}^{-1}(\mu_2 - \hat{\mu}_2) \tag{14}$$

$$\text{s.t.} \quad V_{12}V_{22}^{-1}z = 0_K, \tag{15}$$

$$z'V_{22}^{-1}z = z'V_{22}^{-1}1_n. \tag{16}$$

The optimal solution for the above minimization problem is $z = \hat{e}$, where $\hat{e}$ is defined in (5). To show this, note that from $\hat{e} = \hat{D}\hat{\gamma} - 1_n$, we have

$$\hat{\mu}_2 = \frac{1_n - V_{21}\hat{\gamma}_1 + \hat{e}}{\hat{\gamma}_0}. \tag{17}$$

Using that $\tilde{\gamma}_0 = \hat{\gamma}_0$, we have

$$\mu_2 - \hat{\mu}_2 = \frac{V_{21}(\tilde{\gamma}_1 - \hat{\gamma}_1) + z - \hat{e}}{\hat{\gamma}_0} \tag{18}$$

and we can write the objective function as

$$\min_{z} \frac{(\tilde{\gamma}_1 - \hat{\gamma}_1)' V_{12}V_{22}^{-1}V_{21}(\tilde{\gamma}_1 - \hat{\gamma}_1) + 2(z - \hat{e})'V_{22}^{-1}V_{21}(\tilde{\gamma}_1 - \hat{\gamma}_1) + (z - \hat{e})'V_{22}^{-1}(z - \hat{e})}{\tilde{\gamma}_0^2}. \tag{19}$$

From the first order condition, we know that $V_{12}V_{22}^{-1}\hat{e} = 0_K$. Also, from constraint (15), we have $V_{12}V_{22}^{-1}z = 0_K$. Therefore, the second term in the objective function vanishes. Since only the third term depends on $z$, the objective function is minimized by setting $z = \hat{e}$ because $\hat{e}$ satisfies the constraints (15) and (16).

*Nested and Overlapping Models*

We start with the case of nested models. Partition the $\gamma$ vector in (3) as $\gamma = [\gamma_0,\ \gamma_1',\ \gamma_2']'$, where $\gamma_0$ is a scalar, $\gamma_1$ is a $K_1$-vector and $\gamma_2$ is a $K_2$-vector with $K_2 = K - K_1$. We propose a method for setting $\mu_2$ such that $\gamma_2 = 0_{K_2}$.

Without loss of generality, we assume $\mu_1 = 0_K$ and as a result, $D = [\mu_2,\ V_{21}]$. Let $V_{21} = [V_{21}^a,\ V_{21}^b]$, where $V_{21}^a$ and $V_{21}^b$ are $n \times K_1$ and $n \times K_2$ matrices, respectively. Premultiplying both sides of (3) by $D'V_{22}^{-1}D$, we have

$$(D'V_{22}^{-1}D)\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ 0_{K_2} \end{bmatrix} = D'V_{22}^{-1}1_n. \tag{20}$$

This leads to two equations:

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

$$V_{12} V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0_K. \tag{22}$$

The second equation suggests that

$$\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n \tag{23}$$

is orthogonal to $V_{22}^{-1}V_{21}$. This implies that

$$\mu_2\gamma_0 = 1_n - V_{21}^a\gamma_1 + cz, \tag{24}$$

where $z$ is an $n$-vector such that $V_{12}V_{22}^{-1}z = 0_K$ and $c$ is a scalar.

When we choose $z = 0_n$, we have

$$\mu_2 = \frac{1}{\gamma_0}\left(1_n - V_{21}^a\gamma_1\right) \tag{25}$$

and both the nested and nesting models are correctly specified.

When $z \neq 0_n$, we can normalize $z$ such that $z'z = 1$. Premultiplying (21) by $\gamma_0$, we have

$$\begin{aligned} & \gamma_0\mu_2' V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0 \\ \Rightarrow\quad & (1_n - V_{21}^a\gamma_1 + cz)' V_{22}^{-1}(cz) = 0 \\ \Rightarrow\quad & c = -\frac{z'V_{22}^{-1}1_n}{z'V_{22}^{-1}z}. \end{aligned} \tag{26}$$

Therefore, for a given vector $z$ with $z'z = 1$, orthogonal to $V_{22}^{-1}V_{21}$, we can set

$$\mu_2 = \frac{1}{\gamma_0}\left(1_n - V_{21}^a\gamma_1 - \frac{z'V_{22}^{-1}1_n}{z'V_{22}^{-1}z}z\right) \tag{27}$$

so that $\gamma_2 = 0_{K_2}$ and both models have the same HJ-distance.

Turning to the case of overlapping models, let $f = [f_1',\ f_2',\ f_3']'$, where $f_i$ is $K_i \times 1$ and $K = K_1 + K_2 + K_3$. We assume that the first model consists of $f_1$ and $f_2$ and the second model consists of $f_1$ and $f_3$. For the first model, we have $D_1 = [\mu_2,\ V_{21}^a,\ V_{21}^b]$, where $V_{21}^a = E[x_t f_{1t}']$ and $V_{21}^b = E[x_t f_{2t}']$. For the second model, we have $D_2 = [\mu_2,\ V_{21}^a,\ V_{21}^c]$, where $V_{21}^c = E[x_t f_{3t}']$. As in the nested models case, for the SDF parameters associated with $f_2$ to be zero, we need

$$\mu_2' V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0, \tag{28}$$

$$V_{21}^{a\,\prime} V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0_{K_1}, \tag{29}$$

$$V_{21}^{b\,\prime} V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0_{K_2}. \tag{30}$$

Similarly, for the SDF parameters associated with $f_3$ to be zero, we need

$$\mu_2' V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0, \tag{31}$$

$$V_{21}^{a\,\prime} V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0_{K_1}, \tag{32}$$

$$V_{21}^{c\,\prime} V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0_{K_3}. \tag{33}$$

Combining these two sets of conditions and letting $V_{21} = [V_{21}^a,\ V_{21}^b,\ V_{21}^c]$, we have the same conditions as in the nested models case:

$$\mu_2' V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0, \tag{34}$$

$$V_{21}' V_{22}^{-1}(\mu_2\gamma_0 + V_{21}^a\gamma_1 - 1_n) = 0_K. \tag{35}$$

This implies that $\mu_2$ must satisfy

$$\mu_2\gamma_0 = 1_n - V_{21}^a\gamma_1 + cz, \tag{36}$$

where $z$ is an $n$-vector such that $V_{12}V_{22}^{-1}z = 0_K$ and $c$ is a scalar.

When we set $z = 0_n$, we have the case of correctly specified models

$$\mu_2 = \frac{1}{\gamma_0}\left(1_n - V_{21}^a\gamma_1\right). \tag{37}$$

When $z \neq 0_n$, we can normalize $z$ such that $z'z = 1$. Following the same derivation as before, we can set

$$\mu_2 = \frac{1}{\gamma_0}\left(1_n - V_{21}^a\gamma_1 - \frac{z'V_{22}^{-1}1_n}{z'V_{22}^{-1}z}z\right) \tag{38}$$

and both models will be misspecified but yet have the same HJ-distance.