PORTMANTEAU LIKELIHOOD RATIO TESTS FOR MODEL SELECTION *

BY NAOYA KATAYAMA †

Abstract. This paper provides an extension of Vuong's (1989, Econometrica, 57, pp.307–333) model selection test to the multivariate case. We use the Kullback–Leibler Information Criterion (KLIC) to measure the closeness of a model to the truth to provide a diagnosis of many competing models where the models are not correctly specified. After investigating the asymptotic joint distribution of the likelihood ratio (LR) statistics, we propose LR-based portmanteau test statistics for model selection among many competing models. These tests are easy to compute and are established under any model case, including nested, strictly nested, or overlapping; and are useful where several function forms and exogenous variables give the candidate models. The case of a nested structure yields a test of the best model and the confidence set of the best model from the standpoint of KLIC. We also discuss the use of information criteria instead of LR statistics.

KEYWORDS. Likelihood ratio tests; model selection; portmanteau tests; generalized likelihood ratio tests; goodness-of-fit; information criterion.

1 INTRODUCTION

In this paper, we propose some new tests for model selection by extending Vuong's (1989) results to the multivariate case. Vuong (1989) presented likelihood ratio (LR) based statistics for testing the null hypothesis that two competing models are equally close to the true data generating process (DGP) from the standpoint of the Kullback–Leibler (1951) Information Criterion (KLIC) against the alternative hypothesis that one model is closer.

The results in Vuong (1989) have been extended and applied in a number of ways, including Lien and Vuong's (1987) application to normal linear regression models, Vuong and Wang's (1993) application to Pearson chi-square type statistics, and Rivers and Vuong's (2002) extension to nonlinear dynamic models. Following findings in Nishii (1988) and Vuong (1989), Sin and White (1996) examine the sufficient conditions for consistency of various information criteria in a very wide class of models.

However, none of this work has considered the generalization of Vuong's (1989) results to the case of many competing models. As Vuong (1989, Section 8) pointed out, this is an important problem in model selection. Hence, the purpose of this paper is to extend Vuong's (1989) asymptotic results to the multivariate case and examine the testing problem: the null hypothesis that m models are equally close to the true DGP from the standpoint of the KLIC against the alternative hypothesis that at least one model is closer where m is a positive integer larger than two. This test is useful when some models estimate closer values for the information criteria, e.g., the Akaike (1973) Information Criterion (AIC), the Takeuchi (1976) Information Criterion (TIC), the Schwarz (1978) Information Criterion (SIC) and the Hannan–Quinn (1979) Information Criterion (HQIC).

On considering this problem, we have to face the brutal truth that there is a serious problem with computational cost, even though the asymptotic joint cumulative distribution functions (j.c.d.f.) of the

*Revision Date: April 2, 2008. First Version: April 2008

†Faculty of Economics, Kyushu University, 6-19-1, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan. E-Mail: kata@kyudai.jp

test statistics are established. This is because computation of the j.c.d.f. requires computation of numerical multiple integration. Therefore, the usual multiple tests are limited for practical use.

Some methods were presented for this problem: (i) Numerical integration rules by Shephard (1991a, 1991b). Shephard (1991a, 1991b) established a unified computation framework for the j.c.d.f. by inverting its joint characteristic function. This technique is based on the multivariate inversion formula and is useful when the joint characteristic function is inexpensive to evaluate and the dimension of integration is reasonably small (see Shephard (1991b, Section 7)). (ii) Simulation experimentation techniques that approximate the (asymptotic) j.c.d.f.. However, this technique generally requires a large computational cost. (iii) Use of various inequalities in multiple comparisons to reduce the dimensions of the j.c.d.f.. This technique is most popular in multiple tests, though the technique of multiple comparison uses inequality of probability. In summary, these three methods are not suitable for Vuong's (1989) test in the multivariate case with many competing models. Hence, we take a different approach to that suggested and propose (iv) portmanteau test statistics to test the composite hypothesis.

Portmanteau test statistics were first proposed in the 1970s and since then have been practically employed in various time series models. These are based on the sum of the first m residual autocorrelations to test that the noise of the fitted model behaves like an independent process and the first m autocorrelations are zero. Li (2004) reviews applications of these statistics in various time series models.

Among nested models, our proposed statistics also relate to Hosoya's (1984, 1986, 1989) generalized likelihood ratio (GLR) test. The GLR test conducts a simultaneous LR test with equal marginal error rate and constructs a nested confidence set based on the test in the nested models

Our portmanteau test statistics are based on the sum of (squared) $m-1$ LR test statistics, as LR test statistics measure the relative metric among models from the standpoint of KLIC. In particular, among nested models, these test statistics yield new goodness-of-fit test statistics when we suppose the models are not always correctly specified. Every portmanteau test needs at most one-dimensional numerical integration to compute the probability. It therefore overcomes the problem of computational cost.

The paper is organized as follows. In Section 2, we provide a basic framework. Section 3 derives the asymptotic joint distribution of the LR test statistics and the variance statistics presented in Vuong (1989). In Section 4, we derive the asymptotic theory of tests based on the portmanteau test statistics: namely, the sum of (squared) $m-1$ LR test statistics and the sum of variance test statistics where the competing models are nested, strictly nonnested, or overlapping. We then derive, for the case where all models are nested, the confidence set of the best model based on the GLR test. We also examine a mixed case; that is, partially nested, partially nonnested, or partially overlapping. Section 5 summarizes the results. The mathematical proofs are given in the Appendix.

For ease of understanding, we employ Vuong's (1989) notations, assumptions, definitions and so forth. Throughout this paper, let ∇ be the gradient operator such that, for appropriate vector x and y, $\nabla_x f(x) = \partial f(x)/\partial x$ and $\nabla^2_{xy'} f(x) = \partial^2 f(x,y)/\partial x \partial y'$. A matrix $\mathbf{0}$ denotes an appropriate dimensional matrix where all elements are zeros. All convergences are given by sample size n going to infinity. Therefore, we often omit these arguments for brevity.

2 BASIC FRAMEWORK

In this section, we briefly set the basic framework of the paper, where the DGP and the models are expressed in a similar framework to White (1982) and Vuong (1989). White (1994, Chapter 2 and Section 3.1) provides a rigorous discussion of the framework for dependent random vector observations.

Let X_t be a q-dimensional random vector on a complete probability space (X, σ_X) where σ_X is the Borel σ-field by the open sets of $X \equiv \mathbb{R}^q$. The vector X_t is partitioned as $X_t = (Y_t', Z_t')'$ where Y_t and Z_t are respectively q_y and q_z dimensional vectors with $q = q_y + q_z$. Let (Y, σ_Y) and (Z, σ_Z) be the measurable spaces associated with Y_t and Z_t. Let H_X^0 be the true joint distribution of X_t. We shall be interested in the true conditional distribution $H^0_{Y|Z}(\cdot\,|\,\cdot)$ of Y_t given by Z_t. Let H_Z^0 be the true marginal distribution of Z_t, and ν_Y be the a σ-finite measure on (Y, σ_Y).

The DGP $X_t, t = 1, 2, \ldots, n$ satisfies the following assumption:

Assumption A1 (Data Generation Process) *(i) The q-dimensional random vectors $X_t = (Y_t', Z_t')'$, $t = 1, 2, \ldots$, are independent and identically distributed (i.i.d.) with common true distribution H_X^0 on (X, σ).*

(ii) For H_Z^0-almost all z, $H_{Y|Z}^0(\cdot|z)$ has a Radon–Nikodym density $h^0(\cdot|z)$ relative to ν_Y, which is strictly positive for ν_Y-almost all y.

Throughout this paper, we consider m parametric families of conditional distributions defined on $\sigma_Y \times Z$ for Y_t given Z_t:

$$F_i \equiv \left\{ F_{Y|Z}^{(i)}(\cdot|\cdot;\theta_i); \theta_i \in \Theta_i \subset \mathbb{R}^{p_i} \right\}, \quad i = 1, 2, \ldots, m, \tag{2.1}$$

where p_is are positive integers such that $p_i \leq p_j$, $i < j$.

For ease of exposition, we collect the following assumptions used in Vuong (1989). To reduce complexity, we set the assumptions for a typical i, j of (2.1).

Assumption A2 *(i) (a) For a every $\theta_i \in \Theta_i$ and for H_Z^0-almost all z, the conditional distribution $F_{Y|Z}^{(i)}(\cdot|z,\theta_i)$ has a Radon–Nikodym density $f_i(\cdot|z;\theta_i)$ relative to ν_Y, which is strictly positive for ν_Y-almost all y. (b) Θ_i is a compact subset of $\mathbb{R}^{p_i}$, and the conditional density $f_i(y|z;\theta_i)$ is continuous in θ_i for H_X^0-almost all (y,z).*

(ii) (a) For H_X^0-almost all (y,z), $|\log f_i(y|z;\theta_i)|$ is dominated by an H_X^0-integrable function independent of $\theta_i \in \Theta_i$. (b) The function $z_f \equiv \int \log f_i(y|z;\theta_i) H_X^0(dx)$ has a unique maximum on $\theta_i^* \in \Theta_i$.

(iii) (a) For H_X^0-almost all (y,z), $\log f_i(y|z;\cdot)$ is twice continuously differentiable on Θ_i. (b) For H_X^0-almost all (y,z), $|\nabla_{\theta_i} \log f_i(y|z;\theta_i) \cdot \nabla_{\theta_j'} \log f_j(y|z;\theta_j)|$, $|\nabla^2_{\theta_i\theta_i'} \log f_i(y|z;\theta_i)|$ and $|\log f_i(y|z;\theta_i)|^2$ are dominated by H_X^0-integrable functions independent of $\theta_i \in \Theta_i$ and $\theta_j \in \Theta_j$.

(iv) (a) θ_i^* is an interior point of Θ_i. (b) θ_i^* is a regular point of $A_i(\theta_i)$, where

$$A_i(\theta_i) \equiv \mathrm{E}\left[\nabla^2_{\theta_i\theta_i'} \log f_i(Y_t|Z_t;\theta_i) \right] \tag{2.2}$$

and $\mathrm{E}[\cdot]$ denotes the expectation with respect to H_X^0.

From White (1994, Theorem 2.11), Assumption A2-(i) ensures existence of $\widehat{\theta}_{in}$ such that

$$L_n^{(i)}(\widehat{\theta}_{in}) = \sup_{\theta_i \in \Theta_i} L_n^{(i)}(\theta_i),$$

where $L_n^{(i)}(\theta)$ is defined as the sum of the quasi-log-likelihood functions:

$$L_n^{(i)}(\theta) \equiv \sum_{t=1}^{n} \log f_i(Y_t|Z_t;\theta).$$

From White (1994, Theorem 2.13), Assumptions A2-(i) and A2-(iii)-(a) ensure $\nabla_{\widehat{\theta}_{in}} L_n^{(i)}(\widehat{\theta}_{in}) = \mathbf{0}$. Assumption A2-(ii) ensures the existence of the global minimizer of KLIC in Θ_i:

$$\theta_i^* = \underset{\theta_i \in \Theta_i}{\operatorname{argmin}}\, \mathrm{E}\left[\log \frac{h^0(Y_t|Z_t)}{f_i(Y_t|Z_t;\theta_i)} \right].$$

Assumptions A2-(ii)-(a) and A2-(iii) ensure existence of all moments with respect to H_X^0 appear in this paper by using the Cauchy–Schwarz inequality. Assumptions A2-(iii) and A2-(iv)-(b) ensure the existence of (2.2) and the matrices:

$$B_{ij}(\theta_i, \theta_j) \equiv \mathrm{E}\left[\nabla_{\theta_i} \log f_i(Y_t|Z_t;\theta_i) \cdot \nabla_{\theta_j'} \log f_j(Y_t|Z_t;\theta_j) \right] = B_{ji}(\theta_j, \theta_i)' \tag{2.3}$$

$$C_{ij}(\theta_i^*, \theta_j^*) \equiv A_i(\theta_i^*)^{-1} B_{ij}(\theta_i^*, \theta_j^*) A_j(\theta_j^*)^{-1}. \tag{2.4}$$

For simplicity of notation, let $A_i^* \equiv A_i(\theta_i^*)$, $B_{ij}^* \equiv B_{ij}(\theta_i^*, \theta_j^*)$, $B_{ji}^* \equiv B_{ij}^*{}'$ and $C_{ij}^* \equiv A_i^{*-1} B_{ij}^* A_j^{*-1}$. As Vuong (1989, Lemma A and Section 2) notes, Assumptions A1 and A2 ensure $\widehat{\theta}_{in}$ is consistent for θ_i^*. From this, Assumption A2-(iii) and Jennrich (1969, Theorem 2), A_i^*, B_{ij}^* and C_{ij}^* are consistently estimated by:

$$
\begin{aligned}
A_{in}(\widehat{\theta}_{in}) &\equiv \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \log f(Y_t|Z_t;\widehat{\theta}_{in})}{\partial \theta_i \theta_i'}, \\
B_{ijn}(\widehat{\theta}_{in}, \widehat{\theta}_{jn}) &\equiv \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \log f(Y_t|Z_t;\widehat{\theta}_{in})}{\partial \theta_i} \frac{\partial \log f(Y_t|Z_t;\widehat{\theta}_{jn})}{\partial \theta_j'}, \\
C_{ijn}(\widehat{\theta}_{in}, \widehat{\theta}_{jn}) &\equiv A_{in}(\widehat{\theta}_i)^{-1} B_{ijn}(\widehat{\theta}_{in}, \widehat{\theta}_{jn}) A_{jn}(\widehat{\theta}_j)^{-1}.
\end{aligned} \tag{2.5}
$$

From White (1982, Theorem 3.1), Assumption A2-(iv) imposes negative definiteness of $A_i(\theta_i^*)$.

The remainder of this section summarizes the distribution of the quadratic form of normal random vectors used in this paper.

The following is from Vuong (1989, Definition 1):

Definition 1 (Weighted Sums of Chi-Square Distributions) Let $\mathscr{Z} = (\mathscr{Z}_1, \mathscr{Z}_2, \ldots, \mathscr{Z}_m)'$ be a vector of m independent standard normal variables, and let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$ be a vector of m real numbers. Then, the random variable $\sum_{i=1}^m \lambda_i \mathscr{Z}_i^2$ is distributed as a weighted sum of chi-squares with parameters (m, λ). The cumulative distribution function (c.d.f.) is denoted by $\mathscr{M}_m(\cdot;\lambda)$.

The following is from Vuong (1989, Lemma 3.2):

Lemma 2.1 (Vuong, 1989, Lemma 3.2) *Let U be a vector of m random variables distributed as $\mathscr{N}_m(\mathbf{0}, \Sigma)$ with* $\operatorname{rank}\Sigma \leq m$. *Let A be a $m \times m$ real symmetric matrix. Then:*

(i) $U'AU \sim \mathscr{M}_m(\cdot;\lambda)$, where λ is the vector of eigenvalues of $A\Sigma$. Moreover, the eigenvalues are all real and nonnegative if A is positive semidefinite.

(ii) In addition, if $A\Sigma$ is an idempotent matrix with $\operatorname{rank}(A\Sigma) = r \leq m$, *$U'AU \sim \chi^2(r)$ where $\chi^2(r)$ is a chi-squared random variable with r degrees of freedom.*

Part (i) of Lemma 2.1 is due to Vuong (1989, Lemma 3.2). Part (ii) of Lemma 2.1 is well known and obvious from Part (i) and Definition 1.

In this paper, multivariate versions of the weighted sums of chi-square distributions also appear. We define this distribution as follows:

Definition 2 (Multivariate Weighted Sums of Chi-Square Distributions) Let U be a vector of m random variables distributed as $\mathscr{N}_m(\mathbf{0}, \Sigma)$ with $\operatorname{rank}\Sigma \leq m$. Let A_i be a $m \times m$ real symmetric matrix, $i = 1, 2, \ldots, k$. Then the random vector $(U'A_1U, U'A_2U, \ldots, U'A_kU)'$ is distributed as a multivariate weighted sum of chi-squares with parameters $(m, A_1, \ldots, A_k, \Sigma)$. Its j.c.d.f. is denoted by $\mathscr{M}_{m,k}(\cdot; A_1, \ldots, A_k, \Sigma)$ and its joint characteristic function is given by

$$
\phi(t_1, \ldots, t_k) \equiv \prod_{j=1}^{m} \left\{ 1 - 2i\lambda_j \left(\sum_{l=1}^{k} t_l A_l \Sigma \right) \right\}^{-1/2} \tag{2.6}
$$

where $\lambda_j(\sum_{l=1}^k t_l A_l \Sigma)$ denotes jth eigenvalue of $\sum_{l=1}^k t_l A_l \Sigma$.

Remark 1 (Calculation of $\mathscr{M}_m(\cdot;\lambda)$ and $\mathscr{M}_{m,k}(\cdot; A_1, \ldots, A_k, \Sigma)$) Moments and independence conditions of $U'A_iU$s are obtained from Lemma A.1. Calculation of $\mathscr{M}_m(\cdot;\lambda)$ is given by Imhof's (1961) formula. Some approximations of $\mathscr{M}_m(\cdot;\lambda)$ are given by, e.g., Johnson et al. (1994, Section 18.8). Calculation of $\mathscr{M}_{m,k}(\cdot; A_1, \ldots, A_k, \Sigma)$ is conducted using Shephard's (1991a,1991b) multivariate version of Imhof's (1961) formula. However, in general, this method is expensive unless the value of k is small. See Shephard (1991b, Section 4).

3 ASYMPTOTIC JOINT DISTRIBUTION OF THE LIKELIHOOD RATIO AND VARIANCE STATISTICS

To evaluate the goodness-of-fit of the model, we employ LR statistics to measure the KLIC. In this section, we obtain the joint asymptotic distribution of the LR statistics and variance statistics discussed in Vuong (1989) under general conditions.

Let $LR_n(i,j)$ be a LR statistic F_i against F_j:

$$LR_n(i,j) \equiv L_n^{(j)}(\widehat{\theta}_{jn}) - L_n^{(i)}(\widehat{\theta}_{in}) = \sum_{t=1}^{n} \log \frac{f_j(Y_t|Z_t;\widehat{\theta}_{jn})}{f_i(Y_t|Z_t;\widehat{\theta}_{in})} \tag{3.1}$$

for $i < j$. Similarly, let $E^*(i,j)$ be a relative metric between F_i and F_j from the standpoint of KLIC:

$$E^*(i,j) \equiv \mathrm{E}\left[\log f_j(Y_t|Z_t;\theta_j^*)\right] - \mathrm{E}\left[\log f_i(Y_t|Z_t;\theta_i^*)\right]. \tag{3.2}$$

To proceed, we prepare the notation and lemmas given below and in the Appendix. Let $p \equiv \sum_{i=1}^{m} p_i$,

$$\begin{aligned}
\widehat{\theta}_n &\equiv (\widehat{\theta}'_{1n}, \widehat{\theta}'_{2n}, \ldots, \widehat{\theta}'_{mn})', \\
\theta^* &\equiv (\theta_1^{*\prime}, \theta_2^{*\prime}, \ldots, \theta_m^{*\prime})', \\
LR_{mn} &\equiv [LR_n(1,2), LR_n(1,3), \ldots, LR_n(1,m)]', \\
E^* &\equiv [E^*(1,2), E^*(1,3), \ldots, E^*(1,m)]', \\
\widehat{U}_n &\equiv \sqrt{n}\left[(\widehat{\theta}_n - \theta^*)', (LR_{mn}/n - E^*)'\right]'.
\end{aligned}$$

Given Assumptions A1 and A2, it follows from Vuong (1989), among others, that $\widehat{\theta}_n$ and LR_{mn}/n are consistent for θ^* and E^*, respectively. The following lemma is a generalization of Vuong (1989, Lemma A):

Lemma 3.1 (Asymptotic Normality of Estimates) *Under Assumptions A1 and A2, it holds that, as* $n \to \infty$,

$$\widehat{U}_n \xrightarrow{d} U, \tag{3.3}$$

where U *is a* $p+m-1$ *vector of normal random variables with mean zero and the following possibly singular variance covariance matrix:*

$$\begin{aligned}
\Sigma &\equiv \begin{bmatrix} \Sigma_{\theta\theta} & \Sigma_{\theta L} \\ \Sigma'_{\theta L} & \Sigma_{LL} \end{bmatrix}, \\
\Sigma_{\theta\theta} &\equiv \left[C_{i,j}(\theta_i^*, \theta_j^*)\right]_{i,j=1,\ldots,m}, \\
\Sigma_{\theta L} &\equiv \left[\mathrm{Cov}\left(-A_i^{*-1}\frac{\partial \log f_i(Y_t|Z_t;\theta_i^*)}{\partial \theta_i^*},\ \log\frac{f_{j+1}(Y_t|Z_t;\theta_{j+1}^*)}{f_1(Y_t|Z_t;\theta_1^*)}\right)\right]_{i=1,2,\ldots,m,\, j=1,2,\ldots,m-1}, \\
\textit{and } \Sigma_{LL} &\equiv \left[\mathrm{Cov}\left(\log\frac{f_{i+1}(Y_t|Z_t;\theta_{i+1}^*)}{f_1(Y_t|Z_t;\theta_1^*)},\ \log\frac{f_{j+1}(Y_t|Z_t;\theta_{j+1}^*)}{f_1(Y_t|Z_t;\theta_1^*)}\right)\right]_{i,j=1,2,\ldots,m-1},
\end{aligned}$$

where $\mathrm{Cov}[\cdot]$ *denotes the covariance with respect to* H_X^0.

Following from Vuong (1989, Theorem 3.3), when $f_1(\cdot|\cdot;\theta_1^*) = f_i(\cdot|\cdot;\theta_i^*)$ for some i, $LR_n(1,i) = O_p(1)$, $E^*(1,i) = 0$ and the corresponding element of Σ is zero. Hence, Σ_{LL} is possibly singular.

Remark 2 (Consistent Estimator of Σ) Following Jennrich (1969, Theorem 2) and Assumption A2-(iii), a strong consistent estimator of Σ is obtained from the sample analogs of the submatrices of Σ. $\Sigma_{\theta\theta}$ is estimated from (2.5). Typical (i,j)th submatrices of $\Sigma_{\theta L}$ and Σ_{LL} are respectively estimated from:

$$\begin{aligned}\sigma_{ijn}^{\theta L}(\widehat{\theta}_{in},\widehat{\theta}_{jn}) \equiv & -\frac{1}{n}A_{in}(\widehat{\theta}_{in})^{-1}\sum_{t=1}^{n}\frac{\partial \log f_i(Y_t|Z_t;\widehat{\theta}_{in})}{\partial\theta_i}\log\frac{f_{j+1}(Y_t|Z_t;\widehat{\theta}_{(j+1)n})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\\ & +\left\{A_{in}(\widehat{\theta}_{in})^{-1}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \log f_i(Y_t|Z_t;\widehat{\theta}_{in})}{\partial\theta_i}\right\}\left\{\frac{1}{n}\sum_{t=1}^{n}\log\frac{f_{j+1}(Y_t|Z_t;\widehat{\theta}_{(j+1)n})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\right\}\end{aligned} \tag{3.4}$$

$$\begin{aligned}\text{and } \sigma_{ijn}^{LL}(\widehat{\theta}_{in},\widehat{\theta}_{jn}) \equiv & \frac{1}{n}\sum_{t=1}^{n}\log\frac{f_{i+1}(Y_t|Z_t;\widehat{\theta}_{(i+1)n})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\log\frac{f_{j+1}(Y_t|Z_t;\widehat{\theta}_{(j+1)n})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\\ & -\left\{\frac{1}{n}\sum_{t=1}^{n}\log\frac{f_{i+1}(Y_t|Z_t;\widehat{\theta}_{(i+1)n})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\right\}\left\{\frac{1}{n}\sum_{t=1}^{n}\log\frac{f_{j+1}(Y_t|Z_t;\widehat{\theta}_{(j+1)n})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\right\}.\end{aligned} \tag{3.5}$$

Following from Lemma 3.1 and Lemma A.1-(i), we obtain:

Theorem 3.1 (Asymptotic Distribution of the LR Statistics) *Under Assumptions A1 and A2, it holds that, as $n \to \infty$,*

$$D_n(LR_{mn}/n - E^*) \xrightarrow{d} (\mathscr{U}_2,\mathscr{U}_3,\ldots,\mathscr{U}_m)' \tag{3.6}$$

where $D_n \equiv \mathrm{diag}(d_{2n},d_{3n},\ldots,d_{mn})$,

$$d_{in} \equiv \begin{cases} 2n & \text{if } f_1(\cdot|\cdot;\theta_1^*) = f_i(\cdot|\cdot;\theta_i^*)\\ \sqrt{n} & \text{if } f_1(\cdot|\cdot;\theta_1^*) \neq f_i(\cdot|\cdot;\theta_i^*),\end{cases}$$

$$\mathscr{U}_i \equiv \begin{cases} U'\mathscr{A}_iU & \text{if } f_1(\cdot|\cdot;\theta_1^*) = f_i(\cdot|\cdot;\theta_i^*)\\ b_i'U & \text{if } f_1(\cdot|\cdot;\theta_1^*) \neq f_i(\cdot|\cdot;\theta_i^*),\end{cases}$$

$i = 2,3,\ldots,m$, $\mathscr{A}_i$ is a $(p+m-1)\times(p+m-1)$ matrix defined by $\mathscr{A}_i = \mathscr{I}_i'Q_i\mathscr{I}_i$, Q_i and $\mathscr{I}_i$ are a $(p_1+p_i)\times(p_1+p_i)$ matrix and $(p_1+p_i)\times(p+m-1)$ matrix, respectively, which are given by:

$$Q_i \equiv \begin{bmatrix} A_1^* & \mathbf{0}\\ \mathbf{0} & -A_i^*\end{bmatrix},$$

$$\mathscr{I}_2 \equiv \begin{bmatrix} \mathbb{I}_{p_1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbb{I}_{p_2} & \mathbf{0}\end{bmatrix}$$

$$\text{and } \mathscr{I}_i \equiv \begin{bmatrix} \mathbb{I}_{p_1} & O_1 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & O_i & \mathbb{I}_{p_i} & \mathbf{0}\end{bmatrix},\ i \geq 3,$$

respectively, O_j, $j = 1,i$, is a $p_j \times (p_2+\cdots+p_{i-1})$ zero matrix, b_i is a $(p+m-1)$-vector where the $(p+i-1)$th component is one and zero otherwise, and U is given by Lemma 3.1.

From Vuong (1989, Lemma 4.1) and Lemma 3.1, $f_1(\cdot|\cdot;\theta_1^*) = f_i(\cdot|\cdot;\theta_i^*)$, $i = 2,3,\ldots,m$, if and only if the values of the diagonal elements of Σ_{LL} are all zero. It follows that the variance estimator $\sigma_{iin}^{LL}(\widehat{\theta}_{in},\widehat{\theta}_{in})$, $i = 1,2,\ldots,m-1$, given by (3.5) are natural statistics to test whether $f_1(\cdot|\cdot;\theta_1^*) = f_i(\cdot|\cdot;\theta_i^*)$ for all i.

Hence, we investigate the following statistics:

$$\begin{aligned}
\widehat{\omega}_n^2(1,i) &\equiv \sigma_{iin}^{LL}(\widehat{\theta}_{in},\widehat{\theta}_{in})\\
&= \frac{1}{n}\sum_{t=1}^{n}\left\{\log\frac{f_i(Y_t|Z_t;\widehat{\theta}_{in})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\right\}^2 - \left\{\frac{1}{n}\sum_{t=1}^{n}\log\frac{f_i(Y_t|Z_t;\widehat{\theta}_{in})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\right\}^2, \qquad (3.7)\\
\widetilde{\omega}_n^2(1,i) &\equiv \widehat{\omega}_n^2(1,i) - \{LR_n(1,i)/n\}^2\\
&= \frac{1}{n}\sum_{t=1}^{n}\left\{\log\frac{f_i(Y_t|Z_t;\widehat{\theta}_{in})}{f_1(Y_t|Z_t;\widehat{\theta}_{1n})}\right\}^2 \qquad (3.8)
\end{aligned}$$

for $i=2,3,\dots,m$ and

$$\omega_{mn}^2 \equiv \left[\widehat{\omega}_n^2(1,2),\widehat{\omega}_n^2(1,3),\dots,\widehat{\omega}_n^2(1,m)\right]'. \qquad (3.9)$$

Unless $E^*(1,i)=0$, and following from Vuong (1989, Lemma 4.2), $\widetilde{\omega}_n^2(1,i)$s are not strong consistent estimators of the diagonal elements of Σ_{LL}: Under Assumptions A1 and A2,

$$\widetilde{\omega}_n^2(1,i) \xrightarrow{a.s.} \mathrm{E}\left[\log\frac{f_i(Y_t|Z_t;\theta_i^*)}{f_1(Y_t|Z_t;\theta_1^*)}\right]^2 + \left[\mathrm{E}\left[\log\frac{f_i(Y_t|Z_t;\theta_i^*)}{f_1(Y_t|Z_t;\theta_1^*)}\right]\right]^2.$$

The following result considers the joint asymptotic distribution of LR_{mn} and ω_{mn}^2. This is a generalization of Vuong (1989, Theorem 4.3):

Corollary 3.1 (Asymptotic Distribution of the LR Statistics and the Variance Statistics)

(i) Under Assumptions A1 and A2, when $f_1(\cdot|\cdot;\theta_1^*)=f_i(\cdot|\cdot;\theta_i^*)$, $i=2,3,\dots,m$, it holds that as $n\to\infty$,

$$(2LR_{mn}', n\omega_{mn}^{2\;\prime})' \xrightarrow{d} \mathscr{M}_{p+m-1,2m-2}(\cdot;\mathscr{A}_2,\dots,\mathscr{A}_m,\mathscr{B}_2,\dots,\mathscr{B}_m,\Sigma) \qquad (3.10)$$

where $\mathscr{B}_i$ is a $(p+m-1)\times(p+m-1)$ matrix defined by $\mathscr{B}_i \equiv \mathscr{I}_i'V_i\mathscr{I}_i$, $\mathscr{A}_i$ and $\mathscr{I}_i$ are given by Theorem 3.1, V_i is a $(p_1+p_i)\times(p_1+p_i)$ matrix given by

$$V_i = \begin{bmatrix} B_{11}^* & -B_{1i}^* \\ -B_{i1}^* & B_{ii}^* \end{bmatrix},\ i=2,3,\dots,m.$$

(ii) Results in (i) still hold if $\widehat{\omega}_n^2(1,i)$s in ω_{mn}^2 are replaced by $\widetilde{\omega}_n^2(1,i)$s.

Remark 3 (Computation of the j.c.d.f. of the LR Statistics and the Variance Statistics) We now examine the asymptotic j.c.d.f. of $(2LR_n(1,i),\widehat{\omega}_n^2(1,i))'$. It is easy to check $V_i = Q_i\mathscr{I}_i\Sigma\mathscr{I}_i'Q_i$, $\mathscr{B}_i=\mathscr{A}_i\Sigma\mathscr{A}_i$, and $\mathscr{B}_i\Sigma=(\mathscr{A}_i\Sigma)^2$, since these are similar to the results in the proof of Vuong (1989, Theorem 4.3). Let $\underline{\mathscr{A}}_i$ be the first $p\times p$ submatrix of $\mathscr{A}_i$. Then we obtain:

$$\mathscr{A}_i\Sigma = \begin{bmatrix} \underline{\mathscr{A}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \Sigma_{\theta\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \underline{\mathscr{A}}_i\Sigma_{\theta\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathscr{B}_i\Sigma = \begin{bmatrix} (\underline{\mathscr{A}}_i\Sigma_{\theta\theta})^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad (3.11)$$

$$\underline{\mathscr{A}}_i\Sigma_{\theta\theta} = \begin{bmatrix} B_{11}^*A_1^{*-1} & \cdots & B_{1m}^*A_m^{*-1} \\ -B_{i1}^*A_1^{*-1} & \cdots & -B_{im}^*A_m^{*-1} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix},\ i=2; \qquad = \begin{bmatrix} B_{11}^*A_1^{*-1} & \cdots & B_{1m}^*A_m^{*-1} \\ O_1' & \cdots & O_m' \\ -B_{i1}^*A_1^{*-1} & \cdots & -B_{im}^*A_m^{*-1} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix},\ i\ge 3,$$

where O_j, $j=1,2,\dots,m$, is a $p_j\times(p_2+\cdots+p_{i-1})$ zero matrix. Therefore, the nonzero eigenvalue of $\mathscr{A}_i\Sigma$, λ_A, is an eigenvalue of $\underline{\mathscr{A}}_i\Sigma_{\theta\theta}$. It follows that the nonzero eigenvalues of $\mathscr{A}_i\Sigma$ are nonzero eigenvalues of:

$$W_i \equiv \begin{bmatrix} B_{11}^*A_1^{*-1} & B_{1i}^*A_i^{*-1} \\ -B_{i1}^*A_1^{*-1} & -B_{ii}^*A_i^{*-1} \end{bmatrix}, \qquad (3.12)$$

which is a similar expression to equation (3.6) in Vuong (1989). It follows from Lemma A.1-(ii) that $LR_n(1,i)$ and $\widetilde{\omega}_n^2(1,i)$ are not generally asymptotically independent. However, we note that an eigenvalue of $t_1(\mathscr{A}_i\Sigma)+t_2(\mathscr{A}_i\Sigma)^2$ is $t_1\lambda_A+t_2\lambda_A^2$ from the Frobenius Theorem where $t_1,t_2\in\mathbb{R}$. Therefore, we can easily compute the asymptotic j.c.d.f. of $(2LR_n(1,i),\widehat{\omega}_n^2(1,i))'$ from (2.6) and equation (1) in Shephard (1991b).

Remark 4 (Independence of $\{\mathscr{U}_i\}$) Necessary and sufficient conditions for the asymptotic independence of the two normalized LR statistics or the variance statistics are obtained from a generalized Craig–Sakamoto Theorem (see Lemma A.1). It follows that under the conditions given in Theorem 3.1, if $f_1(\cdot|\cdot;\theta_1^*)=f_i(\cdot|\cdot;\theta_i^*)=f_j(\cdot|\cdot;\theta_i^*)$, $LR_n(1,i)$ and $LR_n(i,j)$ are asymptotically independent if and only if $\Sigma_{\theta\theta}\underline{\mathscr{A}}_i\Sigma_{\theta\theta}(\underline{\mathscr{A}}_j-\underline{\mathscr{A}}_i)\Sigma_{\theta\theta}=\mathbf{0}$. Therefore, in general, normalized $LR_n(i,j)$ and $LR_n(j,k)$ for $i<j<k$ are not asymptotically independent. This is different from the standard theory of multiple tests under nested hypotheses when the hypothesized models contain a correct model. See, e.g., Gourieroux and Monfort (1989, Chapter 19). In addition, by Lemma A.1-(iv) and (v), if $f_1(\cdot|\cdot;\theta_1^*)=f_i(\cdot|\cdot;\theta_i^*)\neq f_j(\cdot|\cdot;\theta_j^*)$ for some $i\neq j$, $\mathscr{U}_i$ and $\mathscr{U}_j$ are uncorrelated, but are not independent in general.

4 PORTMANTEAU LIKELIHOOD RATIO TESTS IN MANY COMPETING MODELS

This section considers testing for equivalence by $E^*(1,i)$, $i=2,3,\ldots,m$ by using the LR test statistics and variance statistics discussed in the previous section. We also consider the application of an information criterion:

$$IC_n(i,j)\equiv LR_n(i,j)-\widehat{c}_{ijn},\ i<j, \tag{4.1}$$

where $\widehat{c}_{ijn}$ is nonnegative and depends on i,j,n. $\widehat{c}_{ijn}$ imposes a penalty to encourage the selection of a parsimonious model. Candidates of $\widehat{c}_{ijn}$ are estimates or nonstochastic numbers such as $(j-i)$ (AIC), $\mathrm{tr}\{B_{iin}(\widehat{\theta}_{in})A_{in}(\widehat{\theta}_{in})^{-1}\}-\mathrm{tr}\{B_{jjn}(\widehat{\theta}_{jn})A_{jn}(\widehat{\theta}_{jn})^{-1}\}$ (TIC), $(j-i)\log(n)/2$ (SIC), or $(j-i)c\{\log\log(n)\}$ with $c>1$ (HQIC), which suggest the choice of the model F_i if $IC_n(i,j)\leq 0$ and the choice of the model F_j otherwise.

4.1 Testing Hypotheses

We consider the following hypothesis and definitions.

Definition 3 (Hypothesis for Equivalence Test of KLIC) We say that: when $E^*(i,j)>0$, F_j is better than F_i, when $E^*(i,j)<0$, F_i is better than F_j and when $E^*(i,j)=0$, F_i and F_j are equivalent. Then:

$$H_0:\ E^*(1,2)=E^*(1,3)=\cdots=E^*(1,m)=0 \tag{4.2}$$

meaning that F_i, $i=1,2,\ldots,m$, are equivalent, against:

$$H_A:\ E^*(1,i)\neq 0 \text{ for some } i=2,3,\ldots,m, \tag{4.3}$$

meaning that F_1 and F_i are not equivalent, or equivalently, F_1 is better or worse than F_i for some $i=2,3,\ldots,m$.

This test is useful when we conjecture the competing m models are almost equally well fitted.

We also consider the following hypothesis and definitions.

Definition 4 (Hypothesis for Equivalence Test of $f_i(\cdot|\cdot;\theta_i^*)$s) Let $\omega_*^2(1,i)$ be $(i-1)$th diagonal element of Σ_{LL} for $i=2,3,\ldots,m$, where Σ_{LL} is given in Lemma 3.1. Then:

$$H_0^{\omega}:\ \omega_*^2(1,2)=\omega_*^2(1,3)=\cdots=\omega_*^2(1,m)=0 \tag{4.4}$$

meaning that $f_i(\cdot|\cdot;\theta_i^*)$, $i = 1,2,\dots,m$, are equivalent, against

$$H_A^{\omega}: \ \omega_*^2(1,i) \neq 0 \text{ for some } i = 2,3,\dots,m, \tag{4.5}$$

meaning that $f_1(\cdot|\cdot;\theta_1^*) \neq f_i(\cdot|\cdot;\theta_i^*)$ for some $i = 2,3,\dots,m$.

We note that, given Assumptions A1 and A2, H_0^{ω} is equivalent to $f_i(\cdot|\cdot;\theta_i^*) = f_j(\cdot|\cdot;\theta_j^*)$ for any i,j by Vuong (1989, Lemma 4.1). Also, testing H_0^{ω} is equivalent to testing $\Sigma_{LL} = \mathbf{0}$ or testing $\mathrm{tr}\Sigma_{LL} = 0$.

4.2 Nested Models

We now consider the case of a nested structure. We propose two tests for model selection based on the LR test and variance test statistics.

First, we provide a formal definition of the nested model given by Vuong (1989, Definition 4):

Definition 5 (Nested Models) m conditional models F_i, $i = 1,2,\dots,m$, are nested if and only if:

$$F_i \subset F_{i+1} \ \text{ for } i = 1,2,\dots,m-1. \tag{4.6}$$

Similarly to Vuong (1989, Assumption A8), we assume:

Assumption A3 *There exists C^2-function $\phi_{ij}(\cdot)$ from Θ_i to Θ_j such that for any $\theta_i \in \Theta_i$:*

$$f_i(\cdot|\cdot;\theta_i) = f_j(\cdot|\cdot;\phi_{ij}(\theta_i)) \tag{4.7}$$

for $(v_Y \times H_Z^0)$— almost any (y,z) and $i < j$.

The following lemma is due to Vuong (1989, Lemma 7.1):

Lemma 4.1 (Vuong (1989, Lemma 7.1)) *Given Assumptions A1-(ii), A2-(i), A2-(ii) and A3, the following statements are equivalent: for any $i < j$,*

(i) $\theta_j^* = \phi_{ij}(\theta_i^*)$,

(ii) $\theta_j^* \in \phi_{ij}(\Theta_i)$,

(iii) $\mathrm{E}[\log f_i(Y_t|Z_t;\theta_i^*)] = \mathrm{E}[\log f_j(Y_t|Z_t;\theta_j^*)]$,

(iv) $f_i(\cdot|\cdot;\theta_i^*) = f_j(\cdot|\cdot;\theta_j^*)$.

Lemma 4.1 shows that the hypothesis for the KLIC equivalence test in Definition 3 coincides with the hypothesis for the equivalence test of $f_i(\cdot|\cdot;\theta_i^*)$s in Definition 4. Also, following from Vuong (1989, Section 7), hypothesis H_0 against H_A are equivalent to the hypothesis $H_0^{(1)}$ against $H_A^{(1)}$, which are defined as follows:

Definition 6 (Hypothesis for Test of the Best Model among Nested Models) We say that for the nested models F_j, $j = i,i+1,\dots,m$, when $E^*(i,j) = 0$ for any $j = i+1,i+2,\dots,m$ and some $i = 1,2,\dots,m-1$, F_i is the best model among F_j, $j = i,i+1,\dots,m$.

Given Assumptions A1-(ii), A2-(i), A2-(ii) and A3, for the nested models F_j for any $j = i,i+1,\dots,m$ and some $i = 1,2,\dots,m-1$,

$$H_0^{(i)}: \theta_m^* \in \phi_{im}(\Theta_i), \tag{4.8}$$

meaning that F_i is the best model among F_j, $j = i,i+1,\dots,m$ from the stand point of KLIC and the principle of parsimony, against:

$$H_A^{(i)}: \theta_m^* \notin \phi_{im}(\Theta_i) \text{ and } \bigcup_{j=i+1}^{m} H_0^{(j)}, \tag{4.9}$$

meaning that F_i is not best among F_j, $j = i,i+1,\dots,m$, or equivalently, F_i is worse than some F_j, $j = i,i+1,\dots,m$, where $H_0^{(m)}$ denotes $\theta_m^* \in \Theta_m$.

As argued in Vuong (1989, Section 7), this test provides a diagnosis that the smaller model is equivalent or worse than the larger model because $E^*(i,j)$ are nonnegative for any $i<j$ under Assumption A3. If $H_0^{(i)}$ is true, $H_0^{(j)}$s, $j>i$ are also true and F_i is a suitable model from the principle of parsimony: thus, F_i is the best model among the larger models. In other words, this test is a new goodness-of-fit test among nested and possibly incorrect models.

For simplicity, we only discuss the case of $H_0^{(1)}$ against $H_A^{(1)}$. However, the following arguments readily apply to the case of $H_0^{(i)}$ against $H_A^{(i)}$, $1<i<m$.

From Corollary 3.1-(i) and Vuong (1989, Section 7), $2LR_{mn}$ converges to a multivariate weighted sum of chi-square distributions under H_0. However, calculation of the rejection region of all $LR_n(1,i)$ statistics, $i=2,3,\ldots,m$ entails computational cost when the value of m is large. However, both $LR_n(1,i)$ and $E^*(1,i)$ are nonnegative and $LR_n(1,i)$ takes large values when $E^*(1,i)>0$, as in Vuong (1989, Lemma 3.1). Therefore, we propose a sum of likelihood ratio (SLR) statistic:

$$SLR_{mn} \equiv 2\sum_{i=2}^{m} LR_n(1,i) \tag{4.10}$$

for testing H_0.

We have two interpretations on the use of this statistic.

First, we note that H_0 indicates testing for the goodness-of-fit test of F_1 in terms of the KLIC under the Definition 3 because H_0 means $H_0^{(1)}$ by Lemma 4.1. It follows that testing H_0 by SLR_{mn} is a new portmanteau test among nested and possibly misspecified models. In time series analysis, the portmanteau test statistic is known as a goodness-of-fit test statistic and is defined by the sum of squares of the first m residual autocorrelations, $\widehat{\rho}_i,\ i=1,2,\ldots,m$. Box & Pierce (1970) first presented this statistic, which is supposed to behave as the sum of squares of the first m sample autocorrelations of the white noise process: $\rho_i,\ i=1,2,\ldots,m$. Namely, the supposed model is assumed to be correctly specified and $\widehat{\rho}_i-\rho_i \xrightarrow{p} 0$ under the null hypothesis. It is also known that $(\widehat{\rho}_1,\widehat{\rho}_2,\ldots,\widehat{\rho}_m)'$ is a vector of Lagrange multipliers by a test for the null hypothesis that the disturbances are independent and the alternative hypothesis is that they are an mth order autoregressive (or moving-average) model. See, e.g., Godfrey (1991, Sections 3.5 and 4.4). While we consider the model selection problem where the models are not always correctly specified, we have $LR_n(1,i)/n \xrightarrow{a.s.} E^*(1,i)$ from Vuong (1989, Lemma 3.1). Now, we suppose $E^*(1,i)=0$ in place of $\rho_i=o_p(1)$ and use the sum of $LR_n(1,i)$s in place of the sum of squared $\widehat{\rho}_i$s.

The second interpretation is a generalization of Hosoya's GLR test. See Hosoya (1984, 1986, 1989). The framework of Hosoya's GLR test assumes that the nested models are correctly specified and reject the null model when $LR_n(1,i)>c_i$ for some $c_i>0$ and $i=2,3,\ldots,m$. Hosoya's analysis also reveals that the GLR test is more powerful than the standard LR test and presents a few numerical integration methods to save computational cost. Hosoya (1989, Section 4) provides some illustrations of the GLR tests. However, the joint probability of many $\{LR_n(1,i)>c_i\}$s requires substantial computational cost. Unfortunately, our procedure cannot use Hosoya's numerical integration methods because we assume that the models are not correctly specified and $LR_n(1,i)$ and $LR_n(i,j)$, $i<j$, are not asymptotically independent from Remark 4. However, the portmanteau statistic (4.10) overcomes this difficulty.

The asymptotic distribution of SLR_{mn} is given by the following result:

Theorem 4.1 (Asymptotic Distribution of the SLR Test Statistics) *Given Assumptions A1, A2 and A3:*

(i) Under $H_0^{(1)}$, it holds that, as $n\to\infty$,

$$SLR_{mn} \xrightarrow{d} \mathscr{M}_{p_m}(\,\cdot\,;\lambda_{SLR}), \tag{4.11}$$

where λ_{SLR} is the vector of p_m nonnegative eigenvalues of $\Sigma_{SLR}\equiv\sum_{i=2}^{m}\underline{W}_i$; $\underline{W}_i$, $i=2,3,\ldots,m$, is $p_m\times p_m$ matrix defined by

$$\underline{W}_i = B_{mm}^*\psi_{mi}^*(\psi_{i1}^*A_1^{*-1}\psi_{1i}^*-A_i^{*-1})\psi_{im}^*; \tag{4.12}$$

and $\psi_{ij}^* = \psi_{ij}(\theta_i^*) \equiv \nabla_{\theta_i^*}\phi_{ij}'(\theta_i^*)$, $\psi_{ji}^* \equiv \psi_{ij}^{*\prime}$ *for* $i < j$ *and* $\psi_{ii}^* \equiv \mathbb{I}_{p_i}$ *for* $i = 1, 2, \ldots, m$.

(ii) Under $H_A^{(1)}$, $SLR_{mn} \xrightarrow{d} \infty$, *as* $n \to \infty$.

This test is one-sided. It is carried out by computing the eigenvalue of the sample analog of Σ_{SLR}.

As noted in White (1982) and Vuong (1989), if the information matrix equivalence holds for the larger model, one has the following corollary as a generalization of Vuong (1989, Corollary 7.3).

Corollary 4.1 (Asymptotic Distribution of the SLR Test Statistics Given $A_m^* + B_{mm}^* = \mathbf{0}$) *Given Assumptions A1, A2 and A3 with* $p_i < p_j$ *for* $i < j$, *suppose that* $A_m^* + B_{mm}^* = \mathbf{0}$:

(i) Under $H_0^{(1)}$, *it holds that, as* $n \to \infty$, $LR_n(1,i)$ *and* $LR_n(i,j)$, $1 < i < j \le m$, *are asymptotically independent and*

$$SLR_{mn} \xrightarrow{d} \sum_{i=2}^{m} \left(\sum_{j=p_1+1}^{p_i} \mathscr{Z}_j^2 \right), \tag{4.13}$$

where $\{\mathscr{Z}_j\}$ *is i.i.d.* $\mathscr{N}(0,1)$.

(ii) Under $H_A^{(1)}$, $SLR_{mn} \xrightarrow{d} \infty$, *as* $n \to \infty$.

From White (1982, Theorem 3.3), the information matrix equivalence $A_m^* + B_{mm}^* = \mathbf{0}$ holds if all models are correctly specified. It follows that the results in Corollary 4.1 are also derived from those of Hosoya's (1984, 1986, 1989) GLR test.

As argued in Vuong (1989, Section 7), from Lemma 4.1, H_0^{ω} and H_0 are equivalent. Hence, the variance test statistics $\widehat{\omega}_n^2(1,i)$s or $\widetilde{\omega}_n^2(1,i)$s are also useful for testing H_0. We consider the sum of variance (SV) test statistic as a new test statistic:

$$SV_{mn} \equiv n \sum_{i=2}^{m} \widehat{\omega}_n^2(1,i). \tag{4.14}$$

Since $SV_{mn}/n \xrightarrow{a.s.} \mathrm{tr}\Sigma_{LL}$, one interpretation of the SV test statistic is testing for $\mathrm{tr}\Sigma_{LL} = 0$, which is an equivalent test of H_0^{ω}. The asymptotic distribution of SV_{mn} is given by the following:

Theorem 4.2 (Asymptotic Distribution of the SV Test Statistics) *Given Assumptions A1 and A2:*

(i) Under H_0^{ω}, *it holds that, as* $n \to \infty$,

$$SV_{mn} \xrightarrow{d} \mathscr{M}_p(\cdot\,; \lambda_{SV}), \tag{4.15}$$

where λ_{SV} *is the vector of* p *nonnegative eigenvalues of* $\Sigma_{SV}\Sigma_{\theta\theta} = \sum_{i=2}^{m}(\underline{\mathscr{A}}_i\Sigma_{\theta\theta})^2$ *and*

$$\Sigma_{SV} = \begin{bmatrix} (m-1)B_{11}^* & -B_{12}^* & -B_{13}^* & \cdots & -B_{1m}^* \\ -B_{21}^* & B_{22}^* & \mathbf{0} & \cdots & \mathbf{0} \\ -B_{31}^* & \mathbf{0} & B_{33}^* & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ -B_{m1}^* & \mathbf{0} & \cdots & \mathbf{0} & B_{mm}^* \end{bmatrix}. \tag{4.16}$$

(ii) Under H_A^{ω}, $SV_{mn} \xrightarrow{d} \infty$, *as* $n \to \infty$.

(iii) The results in (i) and (ii) still hold if $\widehat{\omega}_n^2(1,i)$*s in* SV_{mn} *are replaced by* $\widetilde{\omega}_n^2(1,i)$*s.*

This test is also one-sided, carried out by computing the eigenvalue of the sample analog of $\Sigma_{SV}\Sigma_{\theta\theta}$. Computation of λ_{SV} may require more computational cost than λ_{SLR} because the corresponding matrix is of a higher dimension.

We note that this theorem holds without definition and assumptions concerning the nested models because it is based on Corollary 3.1. Put differently, a remarkable feature of the SV test statistic is that its asymptotic properties are independent of the model structure.

If all models are correctly specified, then the limiting distribution reduces to the same distribution as the SLR statistic given in Corollary 4.1. The following corollary is a generalization of Vuong (1989, Corollary 7.5):

Corollary 4.2 (Asymptotic Distribution of the SV Test Statistics Given $A_m^* + B_{mm}^* = \mathbf{0}$) *Given Assumptions A1 and A2 with $p_i < p_j$ for $i < j$, suppose that $A_m^* + B_{mm}^* = \mathbf{0}$:*

(i) Under H_0^{ω}, it holds that, as $n \to \infty$,

$$SV_{mn} \xrightarrow{d} \sum_{i=2}^{m} \left(\sum_{j=p_1+1}^{p_i} \mathscr{Z}_j^2 \right), \tag{4.17}$$

where $\{\mathscr{Z}_j\}$ is i.i.d. $\mathscr{N}(0,1)$.

(ii) Under H_A^{ω}, $SV_{mn} \xrightarrow{d} \infty$, as $n \to \infty$.

(iii) The results in (i) and (ii) still hold if $\widehat{\omega}_n^2(1,i)$s in SV_{mn} are replaced by $\widetilde{\omega}_n^2(1,i)$s.

Corollaries 7.3 and 7.5 of Vuong (1989) examine the case of $m = 2$ and show that if the larger model is correctly specified, both the LR test statistic and the variance test statistic converge to a common chi-squared distribution under H_0. Corollaries 4.1 and 4.2 herein show that both the SLR test statistic and the SV test statistic converge to a common sum of the chi-squared distributions for $m > 2$.

We have a few remarks, as follows.

Remark 5 (Moments of Asymptotic Distribution of the SLR statistics and the SV statistics) We now examine the asymptotic mean and variance of SLR_{mn} and SV_{mn} under the null hypothesis. Let $\mathrm{AE}[\cdot]$ and $\mathrm{AVAR}[\cdot]$ denote the mean and variance of the asymptotic distribution, respectively, and let ι_k be a k-vector defined by $\iota_k \equiv (1,1,\ldots,1)'$. Then, from Lemma 2.1, Remark 3, Lemma A.1-(v), asymptotic mean and variance of SLR_{mn} and SV_{mn} are as follows:

(i) Under the conditions in Theorem 4.1, $\mathrm{AE}[SLR_{mn}] = \mathrm{tr}(\sum_{i=2}^{m} \underline{\mathscr{A}}_i \Sigma_{\theta\theta}) = \sum_{i=2}^{m} \mathrm{tr}(\underline{W}_i) = \iota_{p_m}' \lambda_{SLR}$ and $\mathrm{AVAR}[SLR_{mn}] = 2\,\mathrm{tr}\{(\sum_{i=2}^{m} \underline{\mathscr{A}}_i \Sigma_{\theta\theta})^2\} = 2\lambda_{SLR}' \lambda_{SLR}$.

(ii) Under the conditions in Theorem 4.2, $\mathrm{AE}[SV_{mn}] = \mathrm{tr}\{\sum_{i=2}^{m} (\underline{\mathscr{A}}_i \Sigma_{\theta\theta})^2\} = \iota_p' \lambda_{SV}$ and $\mathrm{AVAR}[SV_{mn}] = 2\,\mathrm{tr}[\{\sum_{i=2}^{m} (\underline{\mathscr{A}}_i \Sigma_{\theta\theta})^2\}^2] = 2\lambda_{SV}' \lambda_{SV}$.

(iii) Under the conditions in Corollary 4.1, $\mathrm{AE}[SLR_{mn}] = \mathrm{AE}[SV_{mn}] = \sum_{i=2}^{m} \mathrm{tr}(\underline{W}_i) = \sum_{i=2}^{m} (p_i - p_1)$ and $\mathrm{AVAR}[SLR_{mn}] = \mathrm{AVAR}[SV_{mn}] = 2\sum_{i=2}^{m} (m-i+1)^2 (p_{i+1} - p_i)$, where the last equality is follows from (A.15).

Therefore, the eigenvalues λ_{SLR} and λ_{SV} play an important role in the stability of the asymptotic distribution of SLR_{mn} and SV_{mn}.

Remark 6 (An Interpretation of the Information Criterion) We discuss an application of the LR test and the SLR test from the information criterion given in (4.1). Selection of model F_1 by the information criterion is given by $IC_n(1,i) \le 0$ for all $i > 1$. When $\widehat{c}_{1in} \xrightarrow{p} c_{1i} < \infty$, under the assumptions in Theorem 4.1-(i), the probability of the selection of model F_1 by the information criterion is given by:

$$\begin{aligned} \Pr(IC_n(1,i) \le 0,\ i = 2,3,\ldots,m) &= \Pr(2LR_n(1,i) \le 2\widehat{c}_{1in},\ i = 2,3,\ldots,m) \\ &\to \mathscr{M}_{p+m-1,m-1}(2c_{12},\ldots,2c_{1m};\mathscr{A}_2,\ldots,\mathscr{A}_m,\Sigma) \end{aligned} \tag{4.18}$$

from Corollary 3.1-(i). This probability is less than:

$$\begin{aligned}\Pr\left(\sum_{i=2}^{m} IC_n(1,i) \le 0\right) &= \Pr\left(SLR_{mn} \le 2\sum_{i=2}^{m}\widehat{c}_{1in}\right)\\ &\to \mathscr{M}_{p_m}\left(2\sum_{i=2}^{m} c_{1i}\,;\,\lambda_{SLR}\right)\end{aligned} \tag{4.19}$$

from Theorem 4.1-(i). While, by Lemma A.1-(v) and Lemma A.2,

$$\mathrm{AE}[2LR_{mn}] = \left[\mathrm{tr}(B_{11}^{*}A_{1}^{*-1}) - \mathrm{tr}(B_{22}^{*}A_{2}^{*-1}),\ldots,\mathrm{tr}(B_{11}^{*}A_{1}^{*-1}) - \mathrm{tr}(B_{mm}^{*}A_{m}^{*-1})\right]' \tag{4.20}$$

$$\text{and } \mathrm{AE}[SLR_{mn}] = \mathrm{tr}\,\Sigma_{SLR} = \sum_{i=2}^{m}\left\{\mathrm{tr}(B_{11}^{*}A_{1}^{*-1}) - \mathrm{tr}(B_{ii}^{*}A_{i}^{*-1})\right\}. \tag{4.21}$$

Therefore, when we choose the TIC penalty as $\widehat{c}_{1in}$, the asymptotic critical values given in (4.18) and (4.19) equal $2\,\mathrm{AE}[2LR_{mn}]$ and $2\,\mathrm{AE}[SLR_{mn}]$, respectively. In addition, when $A_i^{*}+B_{ii}^{*}=\mathbf{0}$, $\mathrm{AE}[2LR_{mn}] = \mathrm{AE}[n\omega_{mn}^{2}] = (p_2-p_1,\ldots,p_m-p_1)'$ and $\mathrm{AE}[SLR_{mn}] = \mathrm{AE}[SV_{mn}] = \sum_{i=2}^{m}(p_i-p_1)$. Therefore, when we choose the AIC penalty as $\widehat{c}_{1i}$, the asymptotic critical values given in (4.18) and (4.19) equal $2\,\mathrm{AE}[2LR_{mn}] = 2\,\mathrm{AE}[n\omega_{mn}^{2}]$ and $2\,\mathrm{AE}[SLR_{mn}] = 2\,\mathrm{AE}[SV_{mn}]$, respectively. This interpretation is also applicable to Vuong (1989, Theorem 7.2 and Corollaries 7.3 and 7.5).

Remark 7 (Confidence Sets) In the case where there are several competing hypotheses $H_0^{(i)}$, $i=1,2,\ldots,m$, following Hosoya (1984, Section 3; 1989, Section 4), we can establish the confidence set for the best model as follows: Let $\widehat{I}_m$ denote the set of indices where $H_0^{(i)}$, $i=1,2,\ldots,m$, are not rejected by the SLR test or the SV test at significance level $\alpha\in(0,1)$. Let $\widehat{i}_m$ denote the lowest integer in $\widehat{I}_m$. Namely, $\widehat{I}_m\subset\{1,2,\ldots,m\}$ and $\widehat{i}_m=\min\widehat{I}_m$. Suppose that $H_0^{(i_m^*)}$ is true. Since $\Pr(\widehat{i}_m > i_m^*\,|\,H_0^{(i_m^*)}) \le \Pr(i_m^* \text{ is rejected}\,|\,H_0^{(i_m^*)}) = \alpha$, we obtain:

$$\Pr(\widehat{i}_m \le i_m^* \le m\,|\,H_0^{(i_m^*)}) \ge 1-\alpha \quad \text{for } 1\le i_m^*\le m. \tag{4.22}$$

Therefore, the $100(1-\alpha)\%$ confidence set of the model index is $\{\,i\,|\,\widehat{i}_m\le i\le m, i\in\mathbb{N}\,\}$.

4.3 Strictly Nonnested Models

We consider the case where the models F_is are strictly nonnested. The following definition is from Vuong (1989, Section 5).

Definition 7 (Strictly Nonnested Models) m conditional models F_i, $i=1,2,\ldots,m$, are strictly nonnested if and only if:

$$F_i\cap F_j=\phi \text{ for any } i\ne j. \tag{4.23}$$

As Vuong (1989, Section 5) noted, when the competing models are strictly nonnested, all models must be misspecified under our null hypothesis H_0.

From Theorem 3.1 and Vuong (1989, Section 5), $n^{-1/2}LR_{mn}$ converges to a multivariate normal distribution under H_0. However, calculation of the rejection region of all normalized $LR_n(1,i)$ statistics, $i=2,3,\ldots,m$ entails computational cost when the value of m is large. Therefore, we propose a sum of squared likelihood ratio (SSLR) statistic:

$$SSLR_{mn} \equiv \frac{1}{n}\sum_{i=2}^{m} LR_n^2(1,i) \tag{4.24}$$

for testing H_0.

Theorem 4.3 (Asymptotic Distribution of the SSLR Statistics) *Given Assumptions A1 and A2, if F_i, $i = 1,2,\ldots,m$ are strictly nonnested, then we obtain as $n \to \infty$:*

(i) Under H_0,

$$n^{-1/2}LR_{mn} \xrightarrow{d} \mathscr{N}_{m-1}(\mathbf{0}, \Sigma_{LL}), \tag{4.25}$$

$$\text{and } SSLR_{mn} \xrightarrow{d} \mathscr{M}_{m-1}(\,\cdot\,;\lambda_{LL}), \tag{4.26}$$

where λ_{LL} is the vector of nonnegative eigenvalues of Σ_{LL}. In addition, if Σ_{LL} is nonsingular,

$$Q_{mn} \equiv LR'_{mn}\Sigma_{LLn}(\widehat{\theta}_n)^{-1}LR_{mn}/n \xrightarrow{d} \chi^2(m-1), \tag{4.27}$$

where $\Sigma_{LLn}(\widehat{\theta}_n) = [\sigma^{LL}_{ijn}(\widehat{\theta}_{in}, \widehat{\theta}_{jn})]$ and $\sigma^{LL}_{ijn}(\widehat{\theta}_{in}, \widehat{\theta}_{jn})$, $i, j = 1,2,\ldots,m-1$, is given by (3.5).

(ii) Under H_A, $SSLR_{mn} \xrightarrow{d} \infty$. In addition, if Σ_{LL} is nonsingular, $Q_{mn} \xrightarrow{d} \infty$.

(iii) If $\widehat{c}_{1in} = o_p(n^{1/2})$ where $\widehat{c}_{1in}$ is defined in (4.1)*, properties (i) and (ii) still hold if $LR_n(1,i)$ in LR_{mn} is replaced by $IC_n(1,i)$, $i = 2,3,\ldots,m$.*

The test with $SSLR_{mn}$ and Q_{mn} is one-sided. The test with $SSLR_{mn}$ is conducted by computing the eigenvalues of $\Sigma_{LLn}(\widehat{\theta}_n)$, which is a consistent estimator of the eigenvalues of Σ_{LL}.

4.4 Overlapping Models

We consider the case where the models F_i and F_j are all overlapping. The following definition is due to Vuong (1989, Section 6).

Definition 8 (Overlapping Models) m conditional models F_i, $i = 1,2,\ldots,m$, are overlapping if and only if

$$F_i \cap F_j \neq \phi, \tag{4.28}$$

$$F_i \not\subseteq F_j \text{ and } F_i \not\supseteq F_j \tag{4.29}$$

for any $i \neq j$.

Condition (4.28) says that F_i and F_j have some common conditional distributions for Y_t given Z_t for H_Z^0 for almost all z, while condition (4.29) states that neither model is nested in the other.

However, as discussed in Vuong (1989, Section 6), we cannot establish a direct method by LR_{mn} to test H_0 because we cannot discriminate whether $f_1(\,\cdot\,|\,\cdot\,;\theta_1^*) = f_i(\,\cdot\,|\,\cdot\,;\theta_i^*)$.

Nevertheless, we can use Theorem 4.2 to test H_0^{ω} for overlapping models and the following:

Corollary 4.3 (Asymptotic Distribution of the SSLR statistics) *Given Assumptions A1 and A2,*

(i) Under $H_0 - H_0^{\omega}$, $n^{-1/2}LR_{mn} \neq o_p(1)$, (4.25) *and* (4.26) *hold.*

(ii) Under H_A, $SSLR_{mn} \xrightarrow{d} \infty$.

(iii) If $\widehat{c}_{1in} = o_p(n^{1/2})$ where $\widehat{c}_{1in}$ is defined in (4.1)*, properties (i) and (ii) still hold if $LR_n(1,i)$ in LR_{mn} and* (4.24) *is replaced by $IC_n(1,i)$, $i = 2,3,\ldots,m$.*

Both tests are one-sided and conducted by the arguments discussed below in Theorems 4.2 and 4.3.

Part (i) of Corollary 4.3 utilizes Lemma 3.1. When at least one pair of the models satisfies $f_1(\,\cdot\,|\,\cdot\,;\theta_1^*) \neq f_i(\,\cdot\,|\,\cdot\,;\theta_i^*)$, $n^{-1/2}LR_{mn}$ converges to a nondegenerate normal distribution with mean zero and variance Σ_{LL}, which is possibly singular. Because of this, even though we do not possess concrete information about Σ_{LL} and a set of models that satisfies $f_1(\,\cdot\,|\,\cdot\,;\theta_1^*) = f_i(\,\cdot\,|\,\cdot\,;\theta_i^*)$, we can conduct the $H_0 - H_0^{\omega}$ test by using $SSLR_{mn}$, which reflects information about Σ_{LL}.

As a result, we adopt Vuong's (1989, Section 6) sequential procedure for testing H_0:

(i) Test H_0^ω against H_A^ω by using SV_{mn}. If H_0^ω is not rejected, we conclude that we accept H_0^ω and H_0.

(ii) If H_0^ω is rejected, test $H_0 - H_0^\omega$ against H_1 by using $SSLR_{mn}$. If $H_0 - H_0^\omega$ is not rejected, we accept H_0. If $H_0 - H_0^\omega$ is rejected, we conclude H_A.

This sequential procedure has a significance level asymptotically bounded above by the maximum of the asymptotic significance level α_1 and α_2 used for the SV test and SSLR test. This is because similarly to Vuong (1989, Section 6),

$$\begin{aligned}\Pr(H_0 \text{ rejected}\,|\,H_0) &\leq \max\{\Pr(SV_{mn} > c_1\,|\,H_0^\omega), \Pr(SSLR_{mn} > c_2\,|\,H_0 - H_0^\omega)\} \\ &\to \max(\alpha_1, \alpha_2) \end{aligned} \tag{4.30}$$

for some $c_1, c_2 > 0$. Therefore, if $\alpha_i = \alpha \in (0,1),\ i = 1,2$, the significance level of the procedure, as a test of H_0, is asymptotically no larger than α.

Among the nested or overlapping models, if H_0^ω is true, a common set of the models $G \equiv \bigcap_{i=1}^m F_i$ is a suitable model from the principle of parsimony. This is proven by the following corollary of Lemma 4.1:

Corollary 4.4 (Corollary of Lemma 4.1) *Suppose that m conditional models are nested or overlapping such that $F_i \cap F_j \neq \phi$, $i = 1,2,\ldots,m$. Let:*

$$G \equiv \bigcap_{i=1}^m F_i = \{G_{Y|Z}(\cdot\,|\,\cdot\,;\gamma); \gamma \in \Gamma \subset \mathbb{R}^q\},$$

where $q \leq p_1$ and the conditional distribution $G_{Y|Z}(\cdot\,|z;\gamma)$ has a Radon–Nikodym density $g(\cdot\,|z;\gamma)$ relative to ν_Y. Let $\gamma^* \in \Gamma$ be a pseudo-true value of γ for the conditional model G. Assume that there exists C^2-function $\pi_i(\cdot)$ from Γ to Θ_i such that for any $\gamma \in \Gamma$: $g(\cdot\,|\,\cdot\,;\gamma) = f_i(\cdot\,|\,\cdot\,;\pi_i(\gamma))$ for $(\nu_Y \times H_Z^0)$— almost any (y,z) and $i = 1,2,\ldots,m$. Given Assumptions A1-(ii), A2-(i), A2-(ii), and similar assumptions that are made on G, the following statements are equivalent:

(i) $\theta_i^* = \pi_i(\gamma^*)$,

(ii) $\theta_i^* \in \pi_i(\Gamma)$,

(iii) $\mathrm{E}[\log g(Y_t|Z_t;\gamma^*)] = \mathrm{E}[\log f_i(Y_t|Z_t;\theta_i^*)]$,

(iv) $g(\cdot\,|\,\cdot\,;\gamma^*) = f_i(\cdot\,|\,\cdot\,;\theta_i^*)$,

(v) H_0^ω

for any $i = 1,2,\ldots,m$.

Since $G \subset F_i$, $\mathrm{E}[\log g(Y_t|Z_t;\gamma^*)] \leq \mathrm{E}[\log f_i(Y_t|Z_t;\theta_i^*)]$ for any $i = 1,2,\ldots,m$. However H_0^ω is equivalent to Corollary 4.4-(iii). Therefore, and similarly to Section 4.2, hypothesis H_0^ω against H_A^ω is equivalent to hypothesis Corollary 4.4-(iii) against $\mathrm{E}[\log g(Y_t|Z_t;\gamma^*)] < \mathrm{E}[\log f_i(Y_t|Z_t;\theta_i^*)]$ for some i, which provides a diagnosis whether a common set of the models is the best model among larger models.

As Vuong (1989, p.322) pointed out, when at least one model is correctly specified, we have $H_0 = H_0^\omega$ and directly construct a model selection test based on the SV statistics from Theorem 4.2 or the SLR statistics from the following corollary:

Corollary 4.5 (Asymptotic Distribution of the SLR Statistics when a Model is Correctly Specified) *Given Assumptions A1 and A2, if F_i, $i = 1,2,\ldots,m$, are overlapping and at least one model is correctly specified, then:*

(i) Under H_0,

$$SLR_{mn} \xrightarrow{d} \mathscr{M}_p(\cdot\,;\lambda_{A\Sigma}) \tag{4.31}$$

where $\lambda_{A\Sigma}$ is the vector of p eigenvalue of $\sum_{i=2}^{m} \underline{\mathscr{A}}_i \Sigma_{\theta\theta}$.

(ii) Under H_A, if the model F_1 is correctly specified, $SLR_{mn} \xrightarrow{d} -\infty$.

(iii) Under H_A, if the model F_i is correctly specified for some $i = 2, 3, \ldots, m$, $SLR_{mn} \xrightarrow{d} \infty$.

This test is two-sided and conducted similarly to the arguments in Vuong (1989, p.322). A matrix $\sum_{i=2}^{m} \underline{\mathscr{A}}_i$ is given by (A.11). However, we note that when the models are misspecified, under $H_0 - H_0^{\omega}$, the SV statistics and the SLR statistics diverge from Theorem 4.2-(ii) and Corollary 4.3-(i). In this case, when the researcher wrongly decides that at least one model is correctly specified, both the SV statistics and the SLR statistics are frequently rejected: this is contrary to (4.15) and (4.31), respectively. Therefore, we should not use these direct tests without strong evidence about the DGP.

4.5 Mixed Structure Models and Summary of Methodology

We finally consider the case where the models are partially nested, strictly nonnested, or overlapping as based on the concepts developed in this section.

Sections 4.2, 4.3 and 4.4 consider that all models are nested, strictly nonnested, or overlapping. However, cases exist of mixed models such that some models are strictly nonnested and other models are nested. It is often complicated to analyze separately all types of models for the multiple tests with individual LR tests. In addition, the asymptotic j.c.d.f. of individual LR statistics calls for substantial computational cost. Therefore, it is practically impossible to evaluate the significance level of the multiple tests with individual LR tests.

Unexpectedly, our methods are easy to conduct by combining the results found previously. To proceed, we assume that Assumptions A1 and A2. We also assume Assumption A3 for nested models.

Then, the test H_0 cases in mixed structure models (whether at least one pair of models is strictly nonnested): First case. At least one pair of models is strictly nonnested: We can directly conduct test H_0 by using $SSLR_{mn}$ as in Section 4.3. This is because H_A^{ω} is true, $n^{-1/2}LR_{mn}$ and $SSLR_{mn}$ are both $O_p(1)$ from Lemma 3.1 and $\Sigma_{LL} \neq \mathbf{0}$. Second case. The other mixed cases are where some models are nested and some models are overlapping. Here, and based on Theorem 4.2 and Corollary 4.3, we conduct the sequential procedure discussed in Section 4.4. The significance level of this test is given by (4.30).

In summary, our unified procedure for testing H_0 against H_A is as follows:

(i) All models are nested: we conduct the test for H_0 using SLR_{mn} or SV_{mn}. We can interpret test H_0 as a test for the best model among the nested models. In addition, this test establishes the confidence set of the index of the model that possesses the best model from the standpoint of KLIC. See Section 4.2.

(ii) At least one pair of models is strictly nonnested: we conduct a test for H_0 using $SSLR_{mn}$. See Section 4.3.

(iii) For the remaining cases, such as where some models are nested and some are overlapping: we conduct a sequential test that H_0^{ω} using SV_{mn} before conducting a hypothesis test H_0 using $SSLR_{mn}$. We can interpret test H_0^{ω} as a diagnosis for a common set of the models among larger models. See Section 4.4.

In each case, our procedure only requires computable matrix operation and, at most, one-dimensional numerical integration to evaluate the probabilities, such as the significance levels and p-values. Remark 1 provides references for the accurate or approximate method to compute these probabilities.

5 CONCLUSION

In this paper, we propose a unified approach to test whether competing models are equally close to the true DGP from the standpoint of KLIC, where the models are (partially) nested, strictly nonnested, and/or overlapping. Our proposed test is based on the sum of (squared) LR test statistic: this comprises an extension of Vuong (1989) to the multivariate case. This test is workable under many competing models because it overcomes the problems of complicated model structures and computational cost associated with higher dimensional numerical integration. This test is useful where several functional forms and exogenous variables provide the candidate models.

Finally, we leave some remaining problems for future research: (i) Our framework only considers that the DGP is i.i.d. from Assumption A1-(i) similarly to Vuong (1989). As argued in Section 1, many studies have extended Vuong's (1989) results to various cases. For example, it is important to relax Assumption A1-(i) for the dependent data as in Domowitz and White (1982), Rivers and Vuong (2002), Sin and White (1996) and White (1994). (ii) Except for the case of a nested structure, we cannot establish a test of the best model with the information criterion. Let $F_{i_{IC}}$ be a model selected by the information criterion. Then in our notation, H_0^{IC}: $E^*(h,i_{IC}) \geq 0$ for any $h < i_{IC}$ and $E^*(i_{IC}, j) \leq 0$ for any $j > i_{IC}$, against H_A^{IC}: $E^*(h,i_{IC}) < 0$ for some $h < i_{IC}$ or $E^*(i_{IC}, j) > 0$ for some $j > i_{IC}$. Since $H_0^{\omega} \subset H_0 \subset H_0^{IC}$, when H_0^{ω} or H_0 is not rejected, we conclude that H_0^{IC} is not rejected; and when H_0 is rejected, we have to doubt H_0^{IC} and conduct another test such as $H_0^{IC} - H_0$ against H_A^{IC}. Relating to this problem, (iii) Except in the case of a nested structure, we cannot use the principle of parsimony. Therefore, we cannot propose an appropriate principle to proceed the model selection process when H_0 is accepted. For example, the model F_1 is not always the best model to interpret and analyze when it is a nonlinear function of θ_1. Further, even if p_2 is larger than p_1, if the model F_2 is a linear regression model it would be more tractable than F_1. Though it may depend on the researcher's arbitrariness, some will wish for a logical principle for mixed structure models.

ACKNOWLEDGMENTS

This paper is supported by Grant-in-Aid for Young Scientists (Start-up) of the Japan Society for the Promotion of Science.

The author wishes to thank the editors, the referees, Associate Professor Chor-Yiu Sin and Professor Masanobu Taniguchi for their valuable comments and suggestions. The author would also like to convey his sincere thanks to Professor Quang H. Vuong and Professor Halbert White for the inspiration provided in Vuong (1989) and White (1982) and related publications.

The author dedicates this paper to the memory of his recently deceased grandfather, Yoichi Katayama, and family dog, Tsubasa.

APPENDIX

Except when explicitly mentioned, let $f_{ti}(\theta_i^*) \equiv f_i(Y_t|Z_t;\theta_i^*)$ and $LR_n^*(i,j) \equiv L_n^{(j)}(\theta_j^*) - L_n^{(i)}(\theta_i^*) = \sum_{t=1}^n \log\{f_{tj}(\theta_j^*)/f_{ti}(\theta_i^*)\}$, $i,j = 1,2,\ldots,m$. $X_n \overset{\mathrm{a}}{=} Y_n$ denotes $X_n - Y_n = o_p(1)$, as $n \to \infty$, where X_n and Y_n are appropriate dimensional random matrices.

Proof of Lemma 3.1: From the proof of Vuong (1989, Lemma A and Theorem 3.3), we obtain:

$$\sqrt{n}(\widehat{\theta}_{in} - \theta_i^*) \overset{\mathrm{a}}{=} -A_i^{*-1} n^{-1/2} \nabla_{\theta_i^*} L_n^{(i)}(\theta_i^*), \tag{A.1}$$

$$\text{and } LR_n(1,i) = LR_n^*(1,i) + O_p(1), \quad i = 1,2,\ldots,m. \tag{A.2}$$

When $f_1(\cdot|\cdot;\theta_1^*) \neq f_i(\cdot|\cdot;\theta_i^*)$ for all $i \neq 1$, following from equation (A.9) of Vuong (1989), (A.1), (A.2) and the multivariate Central Limit Theorem (see, e.g., Rao (2002, Section 2c.5)), we obtain

$$\begin{aligned}\widehat{U}_n &\overset{\mathrm{a}}{=} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Bigg[-A_1^{*-1}\nabla_{\theta_1^*}\log f_{t1}(\theta_1^*), -A_2^{*-1}\nabla_{\theta_2^*}\log f_{t2}(\theta_2^*), \ldots, -A_m^{*-1}\nabla_{\theta_m^*}\log f_{tm}(\theta_m^*),\\ &\qquad \log\frac{f_{t2}(\theta_2^*)}{f_{t1}(\theta_1^*)} - E^*(1,2), \log\frac{f_{t3}(\theta_3^*)}{f_{t1}(\theta_1^*)} - E^*(1,3), \ldots, \log\frac{f_{tm}(\theta_m^*)}{f_{t1}(\theta_1^*)} - E^*(1,m)\Bigg]'\\ &\xrightarrow{d} U. \end{aligned} \tag{A.3}$$

If $f_1(\cdot|\cdot;\theta_1^*) = f_i(\cdot|\cdot;\theta_i^*)$ for some $i > 1$, then $LR_n^*(1,i) = \log\{f_{ti}(\theta_i^*)/f_{t1}(\theta_1^*)\} = E^*(1,i) = 0$, and $LR_n(1,i) = o_p(n^{1/2})$ from Vuong (1989, Lemma 3.1). Therefore, $(m+i)$th element of U and corresponding elements of Σ are zero, which proves the lemma. *Q.E.D.*

We prove the following lemma where (ii)–(iv) are due to a generalized Craig–Sakamoto Theorem (see, e.g., Provost (1996)):

Lemma A.1 (Linear and Quadratic Forms of Normal Variables) *Let U_n and U be p-dimensional random vectors such that $U_n \xrightarrow{d} U$ as $n \to \infty$, $U \sim \mathcal{N}_p(\mathbf{0},\Sigma)$, $\mathrm{rank}(\Sigma) \leq p$, and let A_i be a $p \times p$ real symmetric matrix, b_j be a p-dimensional real vector, $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,s$. Then the following holds:*

(i) *It holds that, as $n \to \infty$,*

$$\begin{aligned}&(U_n'A_1U_n, U_n'A_2U_n, \ldots, U_n'A_rU_n, b_1'U_n, b_2'U_n, \ldots, b_s'U_n)'\\ &\quad\xrightarrow{d} (U'A_1U, U'A_2U, \ldots, U'A_rU, b_1'U, b_2'U, \ldots, b_s'U)'.\end{aligned}$$

(ii) *$U'A_iU$ and $U'A_jU$, $i \neq j$, are independently distributed if and only if $\Sigma A_i \Sigma A_j \Sigma = \mathbf{0}$.*

(iii) *$b_i'U$ and $b_j'U$, $i \neq j$, are independently distributed if and only if $b_i'\Sigma b_j = 0$.*

(iv) *$U'A_iU$ and $b_j'U$ are independently distributed if and only if $\Sigma A_i \Sigma b_j = \mathbf{0}$.*

(v) $\mathrm{E}[U'A_iU] = \mathrm{tr}A_i\Sigma$, $\mathrm{Cov}[U'A_iU, U'A_jU] = 2\,\mathrm{tr}(A_i\Sigma A_j\Sigma)$, $\mathrm{Cov}[b_i'U, b_j'U] = b_i'\Sigma b_j$ *and* $\mathrm{Cov}[U'A_iU, b_j'U] = 0$.

Proof. First, we shall prove (i). Let $t = (t_1,t_2,\ldots,t_r)'$ and $u = (u_1,u_2,\ldots,u_s)'$ be real vectors, and let $A = \sum_{i=1}^r t_iA_i$ and $b = \sum_{i=1}^s u_ib_i$. Then, we obtain, as $n \to \infty$,

$$\begin{aligned}\sum_{i=1}^r t_iU_n'A_iU_n + \sum_{i=1}^s u_ib_i'U_n &= U_n'AU_n + b'U_n\\ &\xrightarrow{d} U'AU + b'U\\ &= \sum_{i=1}^r t_iU'A_iU + \sum_{i=1}^s u_ib_i'U,\end{aligned}$$

where the convergence follows from Serfling (1980, Section 1.7) since $\nabla_U(U'AU + b'U) = 2AU + b$. It follows Part (i) from Cramér–Wold device. The results in Part (v) are well known. See, e.g., Hocking (1996, Corollary 2.3). The remainder of the proof is obvious from, e.g., Provost (1996, Section 3). *Q.E.D.*

Proof of Theorem 3.1: From equation (A.7) of Vuong (1989), we obtain:

$$LR_n(1,i) = LR_n^*(1,i) + \widehat{\theta}_{in}^{*\prime} Q_i \widehat{\theta}_{in}^*/2 + o_p(1) \tag{A.4}$$

for $i = 2, 3, \ldots, m$, where

$$\widehat{\theta}_{in}^* \equiv \sqrt{n} \begin{bmatrix} \widehat{\theta}_{1n} - \theta_1^* \\ \widehat{\theta}_{in} - \theta_i^* \end{bmatrix} = \mathscr{I}_i \widehat{U}_n. \tag{A.5}$$

It follows that

$$\widehat{\theta}_{in}^{*\prime} Q_i \widehat{\theta}_{in}^* = \widehat{U}_n{}' \mathscr{A}_i \widehat{U}_n \tag{A.6}$$

and

$$\begin{aligned} d_{in}\{LR_n(1,i)/n - E^*(1,i)\} &\overset{\text{a}}{=} d_{in}\{LR_n^*(1,i)/n - E^*(1,i)\} + d_{in}\widehat{\theta}_{in}^{*\prime} Q_i \widehat{\theta}_{in}^*/2n \\ &\overset{\text{a}}{=} \begin{cases} \widehat{U}_n{}' \mathscr{A}_i \widehat{U}_n & \text{if } f_1(\cdot\,|\,\cdot\,;\theta_1^*) = f_i(\cdot\,|\,\cdot\,;\theta_i^*) \\ b_i' \widehat{U}_n & \text{if } f_1(\cdot\,|\,\cdot\,;\theta_1^*) \neq f_i(\cdot\,|\,\cdot\,;\theta_i^*). \end{cases} \end{aligned} \tag{A.7}$$

Therefore, the desired result follows from Lemma 3.1 and Lemma A.1-(i). *Q.E.D.*

Proof of Corollary 3.1: From the proof of Vuong (1989, Theorem 4.3) and (A.5), we obtain:

$$n\widehat{\omega}_n^2(1,i) \overset{\text{a}}{=} n\widetilde{\omega}_n^2(1,i) \overset{\text{a}}{=} \widehat{\theta}_{in}^{*\prime} V_i \widehat{\theta}_{in}^* \xrightarrow{d} U' \mathscr{B}_i U. \tag{A.8}$$

Hence, we obtain the result from (A.7) and Lemma A.1-(i). *Q.E.D.*

The following lemma is due to Vuong (1989, Lemma B):

Lemma A.2 (Vuong (1989, Lemma B)) *Given Assumptions A1-(ii), A2 and A3, we have under $H_0^{(1)}$, for any $i < j < k$,*

(i) $A_i^* = \psi_{ij}^* A_j^* \psi_{ji}^*$ *and* $B_{ii}^* = \psi_{ij}^* B_{jj}^* \psi_{ji}^*$.

(ii) $B_{ij}^* = \psi_{ij}^* B_{jj}^*$ *and* $B_{ji}^* = B_{jj}^* \psi_{ji}^*$.

(iii) $\operatorname{rank} \psi_{ij}^* = p_i$.

(iv) $\psi_{ik}^* = \psi_{ij}^* \psi_{jk}^*$ *and* $\psi_{ki}^* = \psi_{kj}^* \psi_{ji}^*$.

Proof. Parts (i)–(iii) are due to Vuong (1989, Lemma B). Part (iv) follows from the chain rule of $\psi_{ik}^* = \psi_{ij}^* \circ \psi_{jk}^*$. See, e.g., Magnus and Neudecker (1999). *Q.E.D.*

Proof of Theorem 4.1: We shall only prove Part (i) as Part (ii) is obvious from Vuong (1989, Theorem 7.2 (ii)). From Theorem 3.1, we obtain:

$$2LR_{mn} \xrightarrow{d} \mathscr{M}_{p+m-1,m-1}(\,\cdot\,;\mathscr{A}_2, \ldots, \mathscr{A}_m, \Sigma) \tag{A.9}$$

$$\text{and } SLR_{mn} \xrightarrow{d} U' \sum_{i=2}^{m} \mathscr{A}_i U. \tag{A.10}$$

Therefore, similarly to the proof of Theorem 7.2 (i) of Vuong (1989), we show that the nonzero eigenvalues of $\sum_{i=2}^{m}\mathscr{A}_i\Sigma$ are the nonzero eigenvalues of Σ_{SLR}. Since:

$$\sum_{i=2}^{m}\mathscr{A}_i\Sigma=\begin{bmatrix}\underline{\mathscr{A}}\Sigma_{\theta\theta} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},\ \text{where } \underline{\mathscr{A}}=\sum_{i=2}^{m}\underline{\mathscr{A}}_i=\begin{bmatrix}m'A_1^* & & & \\ & -A_2^* & & \mathbf{0}\\ \mathbf{0} & & \ddots & \\ & & & -A_m^*\end{bmatrix} \tag{A.11}$$

and $m'=m-1$, using Lemma A.2, the eigenvalues of $\sum_{i=2}^{m}\mathscr{A}_i\Sigma$ solve:

$$\begin{aligned}0&=\det(\underline{\mathscr{A}}\Sigma_{\theta\theta}-\lambda\mathbb{I}_p)\\&=\det\begin{bmatrix}m'\psi_{1m}^*B_{mm}^*\psi_{m1}^*A_1^{*-1}-\lambda\mathbb{I}_{p_1} & m'\psi_{12}^*B_{22}^*A_2^{*-1} & \cdots & m'\psi_{1m}^*B_{mm}^*A_m^{*-1}\\ -B_{22}^*\psi_{21}^*A_1^{*-1} & -B_{22}^*A_2^{*-1}-\lambda\mathbb{I}_{p_2} & \cdots & -\psi_{2m}^*B_{mm}^*A_m^{*-1}\\ -B_{33}^*\psi_{31}^*A_1^{*-1} & -B_{33}^*\psi_{32}^*A_2^{*-1} & \cdots & -\psi_{3m}^*B_{mm}^*A_m^{*-1}\\ \vdots & \vdots & \ddots & \vdots\\ -B_{mm}^*\psi_{m1}^*A_1^{*-1} & -B_{mm}^*\psi_{m2}^*A_2^{*-1} & \cdots & -B_{mm}^*A_m^{*-1}-\lambda\mathbb{I}_{p_m}\end{bmatrix}\end{aligned} \tag{A.12}$$

By adding to the first-row matrices the last-row matrices premultiplied by the $m'\psi_{1m}^*$, the first-row matrices becomes $(-\lambda\mathbb{I}_{p_1},\mathbf{0},\cdots,\mathbf{0},-\lambda m'\psi_{1m}^*)$ from Lemma A.2. Next, adding to the last-column matrices the first-column matrices postmultiplied by $-m'\psi_{1m}^*$, $0=\det(\underline{\mathscr{A}}\Sigma_{\theta\theta}-\lambda\mathbb{I}_p)$ becomes:

$$0=\det\begin{bmatrix}-B_{22}^*A_2^{*-1}-\lambda\mathbb{I}_{p_2} & -\psi_{23}^*B_{33}^*A_3^{*-1} & \cdots & -\psi_{2m}^*B_{mm}^*A_m^{*-1}+m'B_{22}^*\psi_{21}^*A_1^{*-1}\psi_{1m}^*\\ -B_{33}^*\psi_{32}^*A_2^{*-1} & -B_{33}^*A_3^{*-1}-\lambda\mathbb{I}_{p_3} & \cdots & -\psi_{3m}^*B_{mm}^*A_m^{*-1}+m'B_{33}^*\psi_{31}^*A_1^{*-1}\psi_{1m}^*\\ \vdots & \vdots & \ddots & \vdots\\ -B_{mm}^*\psi_{m2}^*A_2^{*-1} & -B_{mm}^*\psi_{m3}^*A_3^{*-1} & \cdots & -B_{mm}^*A_m^{*-1}-\lambda\mathbb{I}_{p_m}+m'B_{mm}^*\psi_{m1}^*A_1^{*-1}\psi_{1m}^*\end{bmatrix}.$$

Similarly, by adding to the first-row matrices the last-row matrices premultiplied by the $-\psi_{2m}^*$, the first-row matrices becomes $(-\lambda\mathbb{I}_{p_2},\mathbf{0},\cdots,\mathbf{0},\lambda\psi_{2m}^*)$ from Lemma A.2. Next, adding to the last-column matrices the first-column matrices postmultiplied by ψ_{2m}^*, $0=\det(\underline{\mathscr{A}}\Sigma_{\theta\theta}-\lambda\mathbb{I}_p)$ becomes:

$$0=\det\begin{bmatrix}-B_{33}^*A_3^{*-1}-\lambda\mathbb{I}_{p_3} & \cdots & -\psi_{3m}^*B_{mm}^*A_m^{*-1}+m'B_{33}^*\psi_{31}^*A_1^{*-1}\psi_{1m}^*-B_{33}^*\psi_{32}^*A_2^{*-1}\psi_{2m}^*\\ \vdots & \ddots & \vdots\\ -B_{mm}^*\psi_{m3}^*A_3^{*-1} & \cdots & -B_{mm}^*A_m^{*-1}-\lambda\mathbb{I}_{p_m}+m'B_{mm}^*\psi_{m1}^*A_1^{*-1}\psi_{1m}^*-B_{mm}^*\psi_{m2}^*A_2^{*-1}\psi_{2m}^*\end{bmatrix}.$$

Repeating these operations, $0=\det(\underline{\mathscr{A}}\Sigma_{\theta\theta}-\lambda\mathbb{I}_p)$ becomes:

$$0=\det\left\{B_{mm}^*\left(m'\psi_{m1}^*A_1^{*-1}\psi_{1m}^*-\sum_{i=2}^{m-1}\psi_{mi}^*A_i^{*-1}\psi_{im}^*-A_m^{*-1}\right)-\lambda\mathbb{I}_{p_m}\right\}, \tag{A.13}$$

which establishes (4.11) and (4.12) from Lemma A.2. *Q.E.D.*

Proof of Corollary 4.1: We shall first prove Part (i). To prove asymptotic independence of $LR_n(1,i)$ and $LR_n(i,j)$, from Remark 4, it is sufficient to show that

$$\underline{\mathscr{A}}_i\Sigma_{\theta\theta}(\underline{\mathscr{A}}_j-\underline{\mathscr{A}}_i)\Sigma_{\theta\theta}=\mathbf{0}. \tag{A.14}$$

For simplicity, we shall only prove the case of $(i,j)=(2,3)$, however the case of general (i,j)s can be treated similarly. Using (3.11), $A_m^* + B_{mm}^* = \mathbf{0}$ and Lemma A.2, we have

$$\underline{\mathscr{A}}_2\Sigma_{\theta\theta} = \begin{bmatrix} B_{11}^*A_1^{*-1} & \cdots & B_{1m}^*A_m^{*-1} \\ -B_{21}^*A_1^{*-1} & \cdots & -B_{2m}^*A_m^{*-1} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix},\ (\underline{\mathscr{A}}_3 - \underline{\mathscr{A}}_2)\Sigma_{\theta\theta} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} \\ B_{21}^*A_1^{*-1} & \cdots & B_{2m}^*A_m^{*-1} \\ -B_{31}^*A_1^{*-1} & \cdots & -B_{3m}^*A_m^{*-1} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix},$$

$$\begin{aligned}&\underline{\mathscr{A}}_2\Sigma_{\theta\theta}(\underline{\mathscr{A}}_3 - \underline{\mathscr{A}}_2)\Sigma_{\theta\theta} \\ &\quad = \begin{bmatrix} B_{12}^*A_2^{*-1}B_{21}^*A_1^{*-1} - B_{13}^*A_3^{*-1}B_{31}^*A_1^{*-1} & \cdots & B_{12}^*A_2^{*-1}B_{2m}^*A_m^{*-1} - B_{13}^*A_3^{*-1}B_{3m}^*A_m^{*-1} \\ -B_{22}^*A_2^{*-1}B_{21}^*A_1^{*-1} + B_{23}^*A_3^{*-1}B_{31}^*A_1^{*-1} & \cdots & -B_{22}^*A_2^{*-1}B_{2m}^*A_m^{*-1} + B_{23}^*A_3^{*-1}B_{3m}^*A_m^{*-1} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \\ &\quad = \mathbf{0}.\end{aligned}$$

By (A.14), $LR_n(i-1,i)$, $i=2,3,\dots,m$, are asymptotically independent and from Vuong (1989, Corollary 7.3), $LR_n(i-1,i) \xrightarrow{d} \chi^2(p_i - p_{i-1})$. It follows that

$$\begin{aligned} SLR_{mn} &= 2\sum_{i=2}^{m}\sum_{j=1}^{i-1} LR_n(j,j+1) \\ &= 2\sum_{i=2}^{m}(m-i+1)LR_n(i-1,i) \\ &\xrightarrow{d} \sum_{i=2}^{m}(m-i+1)\left(\sum_{j=p_{i-1}+1}^{p_i}\mathscr{Z}_j^2\right) \\ &= \sum_{i=2}^{m}\left(\sum_{j=p_1+1}^{p_i}\mathscr{Z}_j^2\right), \end{aligned} \tag{A.15}$$

which establishes (4.13). Part (ii) is identical to Theorem 4.1-(ii). *Q.E.D.*

Proof of Theorem 4.2: We shall first prove Part (i). From Corollary 3.1 and (A.8), we obtain:

$$SV_{mn} \overset{\mathrm{a}}{=} n\sum_{i=2}^{m}\widetilde{\omega}_n^2(1,i) \xrightarrow{d} U'\sum_{i=2}^{m}\mathscr{B}_i U. \tag{A.16}$$

Therefore, it is sufficient to evaluate the eigenvalues of $\sum_{i=2}^{m}\mathscr{B}_i\Sigma$ from Lemma 2.1. From Remark 3, we obtain:

$$\sum_{i=2}^{m}\mathscr{B}_i\Sigma = \begin{bmatrix}\Sigma_{SV} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\Sigma_{\theta\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} = \begin{bmatrix}\Sigma_{SV}\Sigma_{\theta\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} = \begin{bmatrix}\sum_{i=2}^{m}(\underline{\mathscr{A}}_i\Sigma_{\theta\theta})^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \tag{A.17}$$

which proves Part (i). Parts (ii) and (iii) follow from (A.8) and Lemma 4.2 of Vuong (1989). *Q.E.D.*

Proof of Corollary 4.2: We shall prove Part (i) as Parts (ii) and (iii) are identical to Parts (ii) and (iii) of Theorem 4.2, respectively. It is sufficient to show that

$$SLR_{mn} - SV_{mn} \xrightarrow{p} 0 \tag{A.18}$$

from Corollary 4.1-(i) and Slutsky's theorem. From Corollary 3.1, we have:

$$SLR_{mn} - SV_{mn} \xrightarrow{d} U'\sum_{i=2}^{m}(\mathscr{A}_i - \mathscr{B}_i)U.$$

Therefore, it is sufficient to show that all eigenvalues of $\sum_{i=2}^{m}(\mathscr{A}_i - \mathscr{B}_i)\Sigma$ are zero from Lemma 2.1-(i). We note that first-row and first-column matrices of $\Sigma_{SV}\Sigma_{\theta\theta}$ are zero matrices and typical (i, j)th submatrix of $\Sigma_{SV}\Sigma_{\theta\theta}$ is $B^*_{i1}A_1^{*-1}\psi^*_{1j} - B^*_{ij}A_j^{*-1}$ for $2 \le i, j \le m$ by Lemma A.2. Combining this and (A.12), the eigenvalues of $\sum_{i=2}^{m}(\mathscr{A}_i - \mathscr{B}_i)\Sigma$ solve:

$$0 = \det\begin{bmatrix} m'\mathbb{I}_{p_1} + \lambda\mathbb{I}_{p_1} & m'\psi^*_{12} & \cdots & m'\psi^*_{1m} \\ B^*_{21}A_1^{*-1} & B^*_{21}A_1^{*-1}\psi^*_{12} + \lambda\mathbb{I}_{p_2} & \cdots & B^*_{21}A_1^{*-1}\psi^*_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ B^*_{m1}A_1^{*-1} & B^*_{m1}A_1^{*-1}\psi^*_{12} & \cdots & B^*_{m1}A_1^{*-1}\psi^*_{1m} + \lambda\mathbb{I}_{p_m} \end{bmatrix},$$

where $m' = m-1$. Conducting the same matrix operations from (A.12) in the proof of Theorem 4.1, we obtain all eigenvalues of $\sum_{i=2}^{m}(\mathscr{A}_i - \mathscr{B}_i)\Sigma$ are zero, which proves (A.18). *Q.E.D.*

Proof of Theorem 4.3: We first prove Part (i). From Theorem 3.1, we obtain:

$$n^{-1/2}LR_{mn} \xrightarrow{d} BU, \tag{A.19}$$

where B is a $(m-1)\times(p+m-1)$ matrix given by

$$B = \begin{bmatrix} b'_2 \\ b'_3 \\ \vdots \\ b'_m \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbb{I}_{m-1} \end{bmatrix}.$$

Therefore, Part (i) follows from $BU \sim \mathscr{N}_{m-1}(\mathbf{0}, B\Sigma B')$ and Lemma 2.1. Part (ii) is obvious from equations (5.7) and (5.8) in Vuong (1989, Theorem 5.1). Part (iii) is straightforward from $n^{-1/2}IC_n(1,i) \stackrel{\text{a}}{=} n^{-1/2}LR_n(1,i)$. *Q.E.D.*

Proof of Corollary 4.3: We shall prove Part (i). From Lemma 3.1, (4.25) hold. However, $H_0 - H_0^{\omega}$ means $\Sigma_{LL} \neq \mathbf{0}$. Part (ii) is straightforward from Vuong (1989, Lemma 3.1). Part (iii) is straightforward from $n^{-1/2}IC_n(1,i) \stackrel{\text{a}}{=} n^{-1/2}LR_n(1,i)$. *Q.E.D.*

Proof of Corollary 4.4: We shall prove that (v)⇒(iii)⇒ (ii)⇒(i)⇒ (iv)⇒(v).

(v)⇒(iii): Since $G \subset F_i$, $\mathrm{E}[\log g(Y_t|Z_t;\gamma^*)] \le \mathrm{E}[\log f_i(Y_t|Z_t;\theta_i^*)]$ for any $i = 1,2,\ldots,m$. Since $F^{(i)}_{Y|Z}(\cdot|\cdot;\theta_i^*) = F^{(j)}_{Y|Z}(\cdot|\cdot;\theta_j^*) \in F_j$ for any $i,j = 1,2,\ldots,m$, $F^{(i)}_{Y|Z}(\cdot|\cdot;\theta_i^*) \in G$ for any $i = 1,2,\ldots,m$, which implies that $\mathrm{E}[\log g(Y_t|Z_t;\gamma^*)] \ge \mathrm{E}[\log f_i(Y_t|Z_t;\theta_i^*)]$, and hence (iii) holds.

(iii)⇒(ii)⇒(i)⇒(iv): Obvious from Vuong (1989, Lemma 7.1) and $G \subset F_i$.

(iv)⇒(v): Obvious from H_0^{ω} meaning that $f_i(\cdot|\cdot;\theta_i^*)$, $i = 1,2,\ldots,m$, are equivalent. *Q.E.D.*

Proof of Corollary 4.5: Part (i) is straightforward from $H_0 = H_0^{\omega}$ and Corollary 3.1-(i). Parts (ii) and (iii) are straightforward from Vuong (1989, Theorem 6.3 (ii), (iii)). *Q.E.D.*

REFERENCES

AKAIKE, H. (1973): “Information Theory and an Extention of the Maximum Likelihood Principle,” in *Second International Symposium on Information Theory*, ed. by B. N. Petroc and F. Csake. Budapest: Akademiai Kiado, 267–281.

BOX, G. E. P. AND D.A. PIERCE (1970): "Distribution of Residual Autocorrelations in Autoregressive-integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–1526.

DOMOWITZ, I., AND H. WHITE (1982): "Misspecified Models with Dependent Observations," *Journal of Econometrics*, 20, 35–58.

GODFREY, L. G. (1991): *Misspecification Tests in Econometrics*. Cambridge, U.K.: Cambridge University Press.

GOURIEROUX, C. AND MONFORT, A. (1989): *Statistics and Econometric Models*, Vol. 2. Cambridge, U.K.: Cambridge University Press.

HANNAN, E.J. AND B.G. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of Royal Statistical Society*, Series B41, 190–195.

HOCKING, R. R. (1996): *Methods and Applications of Linear Models*. New York: John Wiley and Sons.

HOSOYA, Y. (1984): "Information Criteria and Tests for Time-Series Models," *Time Series Analysis: Theory and Practice 5*, ed. by Anderson, O. D. . Amsterdam: North-Holland, 39–52.

——— (1986): "A Simultaneous Test in the Presence of Nested Alternative Hypotheses," *Journal of applied probability*, 23A, 187–200.

——— (1989): "Hierarchical Statistical Models and a Generalized Likelihood Ratio Test," *Journal of the Royal Statistical Society*, B51, 435–447.

IMHOF, J. P. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

JENNRICH, R. I. (1969): "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

JOHNSON, N. L. , KOTZ, S. , AND BALAKRISHNAN, N. (1994): *Continuous Univariate Distributions*, Vol. 1. New York: John Wiley and Sons.

KULLBACK, S., AND R.A. LEIBLER (1951): "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79–86.

LI, W.K. (2004): *Diagnostic Checks in Time Series*. New York: Chapman & Hall.

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

MAGNUS, J.R. AND NEUDECKER, H. (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition. New York: John Wiley and Sons.

NISHII, R (1988): "Maximum Likelihood Principle and Model Selection When the True Model is Unspecified," *Journal of Multivariate Analysis*, 27, 392–403.

PROVOST, S.B. (1996): "On Craig's Theorem and its Generalizations," *Journal of Statistical Planning and Inference*, 53, 311–321.

Rao, C.R. (2002): *Linear Statistical Inference and Its Applications*, Second Edition. New York: John Wiley and Sons.

RIVERS, D. AND Q.H. VUONG (2002): "Model Selection for Nonlinear Dynamic Models," *Econometrics Journal*, 5, 1–39.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

SERFLING, R.J. (1980): *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.

SHEPHARD, N.G. (1991a): "From Characteristic Function to Distribution Function: A Simple Framework for the Theory," *Econometric Theory*, 7, 519–529.

——— (1991b): "Numerical Integration Rules for Multivariate Inversions," *Journal of Statistical Computation and Simulation*, 39, 37–46.

SIN, C.Y., AND H. WHITE (1996): "Information Criteria for Selecting Possibly Misspecified Parametric Models," *Journal of Econometrics*, 71, 207–225.

TAKEUCHI, K. (1976): “Distribution of Informational Statistics and a Criterion of Model Fitting,” *Suri-Kagaku (Mathematic Sciences)*, 153, 12–18 (In Japanese).

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

VUONG, Q.H., AND W. WANG (1993): “Minimum Chi-Square Estimation and Tests for Model Selection,” *Journal of Econometrics*, 56, 141–168.

WHITE, H. (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 52, 143–161.

——— (1994): *Estimation, Inference and Specification Analysis*. Cambridge, U.K.: Cambridge University Press.