Econometrics Journal (2002), volume **5**, pp. 1–39.

Model selection tests for nonlinear dynamic models

DOUGLAS RIVERS[†] AND QUANG VUONG[‡*]

[†]*Department of Political Science, Stanford University, Stanford, CA 94305-6044*

[‡]*Department of Economics, University of Southern California, Los Angeles, CA 90089-0253, USA*

Received: November 1999

Summary This paper generalizes Vuong (1989) asymptotically normal tests for model selection in several important directions. First, it allows for incompletely parametrized models such as econometric models defined by moment conditions. Second, it allows for a broad class of estimation methods that includes most estimators currently used in practice. Third, it considers model selection criteria other than the models' likelihoods such as the mean squared errors of prediction. Fourth, the proposed tests are applicable to possibly misspecified nonlinear dynamic models with weakly dependent heterogeneous data. Cases where the estimation methods optimize the model selection criteria are distinguished from cases where they do not. We also consider the estimation of the asymptotic variance of the difference between the competing models' selection criteria, which is necessary to our tests. Finally, we discuss conditions under which our tests are valid. It is seen that the competing models must be essentially nonnested.

Keywords: *Model selection tests, Nonnested hypotheses, Nonlinear dynamic models, Goodness-of-fit, Mean square prediction error.*

1. INTRODUCTION

Most procedures for choosing between two competing econometric models take the following form: each econometric model is estimated by a method that solves some optimization problem; the models are then compared by defining an appropriate goodness-of-fit or selection criterion for each model; and the better-fitting model according to this criterion is selected. In some cases the method of estimation for each model maximizes the goodness-of-fit criterion used for model selection. For instance, when the competing models are fully parametrized and estimated by maximum likelihood, some popular procedures for model selection are based on Akaike (1973, 1974) information criterion (AIC), Schwarz (1978) information criterion (SIC), or Hannan and Quinn (1979) criterion. In other cases, a different goodness-of-fit criterion is used for model selection. This arises when the competing models are estimated using the same sample and compared on their out-of-sample mean squared errors of prediction (MSEP). See Linhart and Zucchini (1986) for various other model selection criteria and procedures.

These model selection procedures are not entirely satisfactory. Since model selection criteria depend on sample information, their actual values are subject to statistical variations. As a consequence a model with a higher model selection criterion value may not outperform *significantly*

*Corresponding author. E-mail: qvuong@usc.edu

its competitor. When the competing models are fully parametrized, nonnested and estimated by maximum likelihood (ML) and when the observations are independent and identically distributed (i.i.d.), Linhart (1988) and Vuong (1989) independently proposed a general testing procedure that takes into account statistical variations and which relies on some convenient asymptotically standard normal tests for model selection based on the familiar likelihood ratio (LR) test statistic.[1] Such tests are testing the null hypothesis that the competing models are as close to the data generating process (DGP) against the alternative hypotheses that one model is closer to the DGP where closeness of a model is measured according to the Kullback–Leibler (1951) information criterion (KLIC). Thus, as in classical nested hypothesis testing, outcomes of the tests provide information on the strength of the statistical evidence for the choice of a model based on its goodness-of-fit.[2]

Though quite general, the applicability of Vuong's model selection tests is currently limited for various reasons. First, because they are based on the likelihood function, these tests require that the competing models be completely parametrized. For instance, this implies that error terms in competing nonlinear regression models or simultaneous equations models must be specified to belong to some parametric family of distributions. As a consequence, Vuong's tests cannot be used to discriminate between two econometric models defined by moment conditions, or more generally, between two competing models that are incompletely specified.

The second limitation arises from the method of estimation. While ML estimation is quite natural when the model selection criterion is KLIC (see e.g. White (1982)), there are various reasons that may lead a researcher to use an estimation method other than ML. For instance, for computational simplicity, robustness reasons, or by necessity because the competing models are incompletely specified, one may use an instrumental variable (IV) estimator or more generally a generalized method of moment (GMM) estimator (see Hansen (1982)), a robust estimator (see Huber (1981), Hampel *et al.* (1986)) or other extremum estimators (see Amemiya (1985), Gallant and White (1988)), a semiparametric estimator (see Andrews (1994), Newey and McFadden (1994), Powell (1994)), etc. Thus it is useful to provide a model selection testing framework that allows for a wide variety of estimation techniques.

Third, the maximum value of the likelihood (possibly adjusted) is not the only model selection criterion used in practice. For instance, when dealing with qualitative dependent variables models, alternative model selection criteria are Pearson-type goodness-of-fit statistics (see e.g. Moore (1978), Heckman (1981)). In linear regression models, criteria based on the in-sample MSEP are widely used (see e.g. Mallows (1973), Amemiya (1980)). When comparing the relative performance of macroeconometric models, a frequent criterion is the out-of-sample MSEP (see e.g. Meese and Rogoff (1983), Fair and Shiller (1990)). Another approach to the use of out-of-sample forecast performance for time series models is illustrated in Findley *et al.* (1998). Along these lines, recent contributions on model selection based on out-of-sample predictability are Diebold and Mariano (1995), West (1994, 1996), Granger and Pesaran (2000), and White (2000).

[1] As noted in Vuong (1989), the LR statistic can also be adjusted by some correction factors such as those proposed by Akaike (1973, 1974), Schwarz (1978), and Hannan and Quinn (1979) to reflect the parsimony of each competing model. For a recent contribution on penalizing the LR statistic, see Sin and White (1996).

[2] Applications of Vuong's test, as it is called in the econometric literature, have appeared in empirical work. For instance, it has been used to test for the presence of collusion in Gasmi *et al.* (1992), for the presence of asymmetric information in Wolak (1994), for distributional assumptions in Paarsch (1997), and for discriminating a structural nonlinear model from linear counterparts in Caballero and Engel (1999).

Moreover, when the models are incompletely specified, the use of criteria other than the ML values becomes necessary. For instance, Sargan (1958) and Pesaran and Smith (1994) have proposed to use the value of the IV criterion function when the competing models are simultaneous equations models estimated by IV methods. Recently, for robustness reasons, Martin (1980), Ronchetti (1985) and Machado (1993) have proposed robust versions of the AIC and SIC criteria by replacing the likelihood part by the extremal value of the sample objective function defining the robust estimator used. See also Konishi and Kitagawa (1996) who propose a generalized information criterion that can be used with robust parameter estimates. Though the preceding criteria have a goodness-of-fit appeal, a model selection criterion need not have such a property. For instance, the precision or MSE of the parameter estimates of interest in the competing models can be a criterion for model selection (see e.g. Torro-Vizcarrondo and Wallace (1968)). This list of criteria is clearly not exhaustive. It suggests, however, than the choice of a model selection criterion depends on the researcher and the purpose of the econometric modelling.

Fourth, Vuong's tests are derived for competing models that are completely static and for observations that are i.i.d. Clearly, the i.i.d. assumption is restrictive when considering time series data. Moreover, dynamic models are frequently considered in empirical work. For instance, a classical question is to determine the order of an ARMA process (see e.g. Hannan and Quinn (1979)). Some generalizations of Vuong's tests to time series models have been undertaken recently and independently from the present work. Findley (1990, 1991) considers essentially the case of competing Gaussian ARMA models when the true DGP is a strictly stationary process. Findley and Wei (1993) provide a generalization to some dynamic regression models.

The goal of the present paper is thus clear. It is to generalize Vuong (1989) model selection tests in several important directions. First, the present paper allows for incompletely specified models such as econometric models defined by moment conditions. Second, it allows for a broad class of estimation methods that includes most estimators used in practice such as the ML estimator, minimum chi-square estimators, GMM estimators, as well as other extremum estimators and some semiparametric estimators. In particular, we shall require that the estimators be $\sqrt{n}$-consistent. See Gallant and White (1988) and Newey and McFadden (1994) for the class of parametric and semiparametric estimators considered here. Third, the present paper allows for model selection criteria other than the models' likelihoods. An important example is the out-of-sample MSEP. Lastly, our tests are obtained for weakly dependent heterogeneous data. This permits the application of our tests to the selection of nonlinear dynamic models in times series situations.

At the outset, it is useful to stress that a distinctive feature of our approach is that both competing models may be misspecified. In particular, our approach does not require that either competing model be correctly specified under the null hypothesis under test. Such a feature reflects the observation that one can seldom specify a statistical model that can describe accurately the data in empirical work, especially in Social Sciences. See e.g. Nakamura *et al.* (1990) for a similar point of view. As we shall see, however, this does not prevent model comparisons.

Not requiring correct specification of the competing models also contrasts with the tests of Cox (1961, 1962), which led to the development of a vast econometric literature on testing nonnested hypotheses (see Pesaran (1974) and the surveys by MacKinnon (1983) and McAleer (1987)).[3] Indeed, under the null hypothesis under test in Cox's approach, one of the competing models is correctly specified. Moreover Cox's tests are more difficult to compute than ours

[3] It is worth noting that extensions of Cox's tests followed the lines described previously, namely extensions to time series models and incompletely specified models estimated by methods other than ML. See Walker (1967), Davidson and MacKinnnon (1981), Ericsson (1983), Godfrey (1983), Gourieroux *et al.* (1983) and Mizon and Richard (1986), among others.

because they require a consistent estimate of the asymptotic mean of the test statistic under the null hypothesis. This also raises some theoretical difficulties when the competing models are incompletely specified. See e.g. Ghysels and Hall (1990) and Smith (1992).

The paper is divided into six more sections. Section 2 describes in some detail the general model selection framework for nonlinear dynamic models. A series of hypotheses about the asymptotic fit of the models are put forth. A basic result on the asymptotic properties of our test statistics is established under general conditions on the model selection criteria and the estimators used. The next two sections seek more primitive conditions ensuring that the conditions of this basic result hold. As the preceding examples of model selection procedures suggest, we distinguish two cases. Section 3 considers the case where model selection criteria can be viewed as optimands of the Gallant and White (1988) type although estimators that are employed are not those that maximize these criteria. In contrast, Section 4 specializes to the case where models are estimated by maximizing some criteria that are used subsequently for model selection. This section also covers the case where models are estimated by means of two-stage procedures in which the second stage involves optimizing goodness-of-fit conditional upon preliminary estimates of the nuisance parameters. Section 5 considers estimation of the asymptotic variance of the difference in goodness-of-fit despite possible misspecification of the competing models. This step is necessary for the construction of our test statistics. Section 6 discusses a critical condition for our test statistics to be asymptotically normal. Essentially, this condition requires that the estimated models be nonnested. Throughout, our theoretical results are illustrated with the comparison between two nonnested autoregressive models based on their in-sample and out-of-sample MSEP. Section 7 concludes. An appendix collects the proofs of our results.

2. TESTS STATISTICS AND GENERAL RESULTS

Two econometric models, $\mathcal{M}_1$ and $\mathcal{M}_2$, are estimated using data generated by an unknown stochastic process. The DGP satisfies

Assumption 1. *$\{X_t\}_{t=-\infty}^{\infty}$ is a p-dimensional stochastic process on a complete probability space $(\Omega, \mathcal{F}, P)$.*

To simplify the notation, we use ω to indicate the whole sequence $\{X_t\}_{t=-\infty}^{+\infty}$. Hereafter, there will be various functions indexed by n where ω appears as an argument. In most cases, these functions depend on ω through the vector $(X_1, \ldots, X_n)$ corresponding to the period for which the competing models are compared. This need not be always the case, for instance when estimation of the competing models uses a different sample (see below).

For $j = 1, 2$, let γ^j denote the features of interest for model $\mathcal{M}_j$ and let Γ_j denote the set of possible values for γ^j. As usual, we require

Assumption 2. *For $j = 1, 2$, Γ_j is a compact subset of $\mathbb{R}^{k_j}$.*

Thus γ^j and Γ_j can be viewed as the parameter vector of interest and the parameter space associated with model $\mathcal{M}_j$, respectively.

Let $\hat{\gamma}_n^j$ denote an estimator of γ^j. At this point, the method of estimation plays no role. The estimator $\hat{\gamma}_n^j$ can be obtained from the same sample $(X_1, \ldots, X_n)$ used for model comparison, i.e. the competing models are compared within sample. Alternatively, one may use a prior sample for estimation and the sample $(X_1, \ldots, X_n)$ to assess the performance of the estimated competing models as in West (1994). One can also sequentially update $\hat{\gamma}_n^j$ with increasing estimation samples and compare (say) the one-period ahead predictability over the sample $(X_1, \ldots, X_n)$ of the sequentially updated models as in West (1996). We require, however, an assumption about the asymptotic behavior of the estimators, which is satisfied in general.

Assumption 3. *For $j = 1, 2$, $\{\hat{\gamma}_n^j\}_{n=1}^{\infty}$ is a sequence of random vectors on $(\Omega, \mathcal{F}, P)$ such that $\hat{\gamma}_n^j \in \Gamma_j$ for each n and there exists a nonstochastic sequence $\{\bar{\gamma}_n^j\}_{n=1}^{\infty}$ uniformly interior to Γ_j for which $\hat{\gamma}_n^j - \bar{\gamma}_n^j \xrightarrow{as} 0$ as $n \to \infty$.*

Note that $\bar{\gamma}_n^j$ need not have a limit as $n \to \infty$. In the literature, the $\bar{\gamma}_n^j$ are referred to as pseudo-true values. See Domowitz and White (1982), Bates and White (1985) and Gallant and White (1988) among others. The i.i.d. case considered in White (1982) and Vuong (1989) is considerably simpler because the pseudo-true values do not vary with the sample size. For dynamic model selection problems, it can be useful to allow for parameter drift. This can occur because the data generating process is nonstationary. But it may also occur with stationary data because of the sequence of specified models or estimating methods used.

Next, most if not all known model selection procedures are defined by model selection criteria (see e.g. Linhart and Zucchini (1986)). Typically, such criteria are goodness-of-fit criteria. Hereafter, we view them as lack-of-fit criteria, i.e. the opposite of goodness-of-fit criteria. As seen in the introduction, however, selection criteria may emphasize other aspects such as the precision of parameter estimates of interest (see e.g. Torro-Vizcarrondo and Wallace (1968)). To be general, we consider selection criteria of the form $Q_n^j(\omega, \gamma^j)$ for each competing model. Note that the criterion typically depends on the features of interest γ_j for model $\mathcal{M}_j$. It may include some Akaike (1973) type correction factor for parsimony reasons as in Sin and White (1996). It is generally random because of its dependence on ω. This is the case when selection criteria use sample information, i.e. are statistics.

A first assumption on selection criteria covered by our theory is

Assumption 4. *For $j = 1, 2$, let $\{Q_n^j : \Omega \times \Gamma_j \to \mathbb{R}\}_{n=1}^{\infty}$ be a sequence of functions such that $Q_n^j(\cdot, \gamma^j)$ is measurable $\mathcal{F}/\mathcal{B}$ for every $\gamma^j \in \Gamma_j$. There exists an equicontinuous sequence of nonstochastic functions $\{\bar{Q}_n^j : \Gamma_j \to \mathbb{R}\}_{n=1}^{\infty}$ such that*

$$\sup_{\gamma^j \in \Gamma_j} \mid Q_n^j(\omega, \gamma^j) - \bar{Q}_n^j(\gamma^j) \mid \xrightarrow{as} 0$$

as $n \to \infty$.

Typically, $\bar{Q}_n^j(\gamma^j)$ is the expectation of $Q_n^j(\omega, \gamma^j)$ with respect to ω. Thus Assumption 4 generally follows from a uniform strong law of large numbers (see e.g. Jennrich (1969), Andrews (1987)). As a matter of fact, many model selection criteria satisfy Assumption 4 under suitable regularity conditions (see Section 3). The most well known is minus the model log-likelihood $-(1/n) \log f_n^j(X_1, \ldots, X_n, \gamma^j)$ where f_n^j is the joint density of the first n observations associated with the parameter value γ^j of model $\mathcal{M}_j$. Another common criterion is the MSEP based on n predictions $(1/n) \sum_{t=1}^n \|Y_t - f_t^j(X_1, \ldots, X_{t-1}, W_t, \gamma^j)\|^2$ where $X_t = (Y_t', W_t')'$ and $\{f_t^j\}$ is a sequence of predictor functions known up to γ^j as in West (1994, 1996).

Given lack-of-fit criteria $Q_n^1(\omega, \gamma^1)$ and $Q_n^2(\omega, \gamma^2)$, and estimates $\hat{\gamma}_n^1$ and $\hat{\gamma}_n^2$, which may come from the comparison sample $(X_1, \ldots, X_n)$ or from another sample such as a prior one, a frequent procedure is to select the model with the smallest lack-of-fit measure $Q_n^j(\omega, \hat{\gamma}_n^j)$. Such a model selection procedure can be given an asymptotic interpretation. Given Assumptions 1–4, it follows from Domowitz and White (1982, Theorem 2.3) that $Q_n^j(\omega, \hat{\gamma}_n^j) - \bar{Q}_n^j(\bar{\gamma}_n^j) \stackrel{as}{\to} 0$, for $j = 1, 2$. Hence

$$\Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) - \Delta \bar{Q}_n(\bar{\gamma}_n^1, \bar{\gamma}_n^2) \stackrel{as}{\to} 0, \tag{1}$$

where $\Delta Q_n(\omega, \gamma^1, \gamma^2) = Q_n^1(\omega, \gamma^1) - Q_n^2(\omega, \gamma^2)$ and $\Delta \bar{Q}_n(\gamma^1, \gamma^2) = \bar{Q}_n^1(\gamma^1) - \bar{Q}_n^2(\gamma^2)$. The quantity $\Delta \bar{Q}_n(\bar{\gamma}_n^1, \bar{\gamma}_n^2)$ can be interpreted as the difference between the asymptotic lack-of-fit of the competing models.

In view of the preceding remark, we shall consider various hypotheses comparing the (asymptotic) lack-of-fit of the competing models. Following Vuong (1989), the null hypothesis H_0 is that $\mathcal{M}_1$ and $\mathcal{M}_2$ are *asymptotically equivalent* when

$$H_0 : \lim_{n\to\infty} \{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\} = 0.$$

The first alternative hypothesis is that $\mathcal{M}_1$ is *asymptotically better* than $\mathcal{M}_2$ when

$$H_1 : \limsup_{n\to\infty} \{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\} < 0.$$

Similarly, the second alternative hypothesis is that $\mathcal{M}_2$ is asymptotically better than $\mathcal{M}_1$:

$$H_2 : \liminf_{n\to\infty} \{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\} > 0.$$

Note that the limit of $\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)$ as $n \to \infty$ may not exist under the alternative hypotheses H_1 and H_2. Consequently, in the case of dynamic models with time series data, there is a third alternative, which is that $\mathcal{M}_1$ and $\mathcal{M}_2$ are *asymptotically incomparable*:

$$H_3 : \liminf_{n\to\infty} \{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\} \le 0 \le \limsup_{n\to\infty} \{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\}$$

with at least one inequality being strict. In this case, there is some asymptotically nonnegligible region of the data for which $\mathcal{M}_1$ fits better than $\mathcal{M}_2$ or vice versa without the models being asymptotically equivalent.[4] Such a situation contrasts with the i.i.d. case considered by Vuong (1989) where the hypotheses H_0, H_1 and H_2 are exhaustive. This is because $\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)$ does not depend on n so that its limit necessarily exists and H_3 is void. More generally, H_3 is void under strict stationarity of the DGP as $\bar{Q}_n^j(\cdot)$ and $\bar{\gamma}_n^j$ will typically not depend on n.

As argued in the introduction, a simple numerical comparison of the sample values of the respective lack-of-fit criteria is not entirely satisfactory for it does not take into account sample variability. Significance of the difference in lack-of-fit needs to be assessed. To do so, we propose some tests with good asymptotic properties. We require additional regularity conditions. The first condition bears on lack-of-fit criteria and is similar to Assumption 4.

Assumption 5. *For $j = 1, 2$, $Q_n^j(\omega, \cdot)$ and $\bar{Q}_n^j(\cdot)$ are continuously differentiable, and*

$$\sup_{\gamma^j \in \Gamma_j} \left| \frac{\partial Q_n^j(\omega, \gamma^j)}{\partial \gamma^j} - \frac{\partial \bar{Q}_n^j(\gamma^j)}{\partial \gamma^j} \right| \stackrel{as}{\to} 0$$

as $n \to \infty$. Moreover, $\partial \bar{Q}_n^j(\cdot)/\partial \gamma^j$ is equicontinuous and $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \gamma^j$ is bounded.

[4] Findley (1990) proposes an interesting graphical procedure that addresses this issue when the competing models are Gaussian ARMA or ARIMA models.

The next one is a joint assumption on estimation methods and lack-of-fit criteria. Typically, it follows from a multivariate Central Limit Theorem.

Assumption 6. *For some integer $s > 0$, there exists a bounded sequence of nonstochastic $s \times (k+2)$ matrices $\{C_n\}$ and a sequence of $s \times 1$ random vectors $\{Z_n\}$ on $(\Omega, \mathcal{F}, P)$ with $\sqrt{n}Z_n \Rightarrow N(0, I_s)$ such that*

$$\sqrt{n}U_n = C_n'\sqrt{n}Z_n + o_P(1), \tag{2}$$

where $k = k_1 + k_2$ and

$$U_n = \begin{pmatrix} Q_n^1(\omega, \bar{\gamma}_n^1) - \bar{Q}_n^1(\bar{\gamma}_n^1) \\ \hat{\gamma}_n^1 - \bar{\gamma}_n^1 \\ Q_n^2(\omega, \bar{\gamma}_n^2) - \bar{Q}_n^2(\bar{\gamma}_n^2) \\ \hat{\gamma}_n^2 - \bar{\gamma}_n^2 \end{pmatrix}. \tag{3}$$

In particular, Assumption 6 requires that estimators be $\sqrt{n}$-asymptotically normal. This is satisfied by a large class of common econometric estimators such as extremum estimators (see e.g. Amemiya (1985), Gallant and White (1988) and Section 4 below) as well as some semiparametric estimators (see Newey and McFadden (1994, Section 8)). Assumption 6 also requires that lack-of-fit criteria evaluated at pseudo-true values be $\sqrt{n}$-asymptotically normal. The latter condition will be verified in Section 3 for a large class of model selection criteria.

Let $\sigma_n^2 = L_n' C_n' C_n L_n$ where

$$L_n = \left(1, \frac{\partial \bar{Q}_n^1(\bar{\gamma}_n^1)}{\partial \gamma^{1'}}, -1, -\frac{\partial \bar{Q}_n^2(\bar{\gamma}_n^2)}{\partial \gamma^{2'}}\right)'. \tag{4}$$

Assumption 7. $\liminf_n \sigma_n^2 > 0$.

It turns out that σ_n^2 is the asymptotic variance of the difference in lack-of-fit. The condition $\liminf_n \sigma_n^2 > 0$ is crucial and is discussed in Section 6.

Lastly, we need a consistent estimator of this variance.

Assumption 8. *There exists a sequence of random variables $\{\hat{\sigma}_n^2\}_{n=1}^{\infty}$ on $(\Omega, \mathcal{F}, P)$ such that $\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0$ as $n \to \infty$.*

In Section 5, we shall propose some consistent estimators of σ_n^2.

We are now in a position to define our test statistic as a suitably normalized difference of the sample lack-of-fit criteria:

$$T_n = \frac{\sqrt{n}}{\hat{\sigma}_n}\{Q_n^1(\omega, \hat{\gamma}_n^1) - Q_n^2(\omega, \hat{\gamma}_n^2)\}. \tag{5}$$

Our model selection test involves comparing values of T_n with critical values of a standard normal distribution. Let α denote the desired (asymptotic) size of the test and $z_{\alpha/2}$ the value of the inverse standard normal distribution function evaluated at $1 - \alpha/2$. If $T_n < -z_{\alpha/2}$, we reject H_0 in favor of H_1; If $T_n > z_{\alpha/2}$, we reject H_0 in favor of H_2; Otherwise, we accept H_0.

An asymptotic justification of the proposed test is given in the next theorem. Define the hypotheses

$$H_0^* : \lim_{n\to\infty} \sqrt{n}\{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\} = 0,$$

$$H_1^* : \lim_{n\to\infty} \sqrt{n}\{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\} = -\infty,$$

$$H_2^* : \lim_{n\to\infty} \sqrt{n}\{\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)\} = +\infty.$$

Under H_0^*, the competing models are $\sqrt{n}$*-asymptotically equivalent*. This constitutes a strengthening of H_0. Note also that H_1^* (say) contains the important case where $\bar{Q}_n^1(\bar{\gamma}_n^1) - \bar{Q}_n^2(\bar{\gamma}_n^2)$ has a finite and strictly negative limit as $n \to \infty$. We have

Theorem 1. *Given Assumptions* 1–8, *then* σ_n^2 *is bounded and*

(i) under H_0^*, $T_n \Rightarrow N(0, 1)$,
(ii) under H_1^*, $T_n \overset{as}{\to} -\infty$,
(iii) under H_2^*, $T_n \overset{as}{\to} +\infty$.

Since $H_0^* \subset H_0$, Theorem 1 shows that our test has correct asymptotic size on a subset of the null hypothesis of asymptotic equivalence H_0. The set H_0^*, however, contains important situations such as $\bar{Q}_n^1(\bar{\gamma}_n^1) = \bar{Q}_n^2(\bar{\gamma}_n^2)$ for all n sufficiently large. Theorem 1 also shows that the test is consistent against H_1^* and H_2^*. Since H_1 implies H_1^* and H_2 implies H_2^*, the test is consistent against a larger set of alternatives than H_1 and H_2. Moreover, if $\bar{Q}_n^j(\cdot)$ and $\bar{\gamma}_n^j$ do not depend on n, as is typically the case under strict stationarity of the DGP, then $H_j^* = H_j$ for $j = 0, 1, 2$, and our proposed test has the desired asymptotic size under the null hypothesis of interest H_0 and is consistent against the alternatives H_1 and H_2.

3. MODEL SELECTION TESTS WITHOUT LACK-OF-FIT MINIMIZATION

Theorem 1 is derived under general assumptions. In this section and the next we seek more primitive assumptions on estimators and lack-of-fit criteria that will imply Assumptions 3–6. We postpone the discussion of Assumptions 7 and 8 to Sections 6 and 5, respectively. Here we focus on the assumptions on the model selection criteria, namely Assumptions 4–6.

The framework of the preceding section is too general for our purpose. First, we need to be more precise on how the data are generated, i.e. on the DGP. We complement Assumption 1 by

Assumption 9. *For some* $r > 2$, $\{X_t\}$ *is a* ϕ*- or* α*-mixing sequence such that* ϕ_m *is of size* $-r/(r-1)$ *or* α_m *is of size* $-2r/(r-2)$, *respectively.*

Definitions of ϕ-mixing, α-mixing and size can be found in Gallant and White (1988, Chapter 3) among others. Assumption 9 allows quite general time dependence and heterogeneity of the data generating process, though extensions of our results to even more general processes such as mixingales can be entertained as in Gallant and White (1988).

Second, since we focus here on the assumptions concerning the model selection criteria, we retain the assumptions on the estimators, namely Assumption 3 and the appropriate part of Assumption 6. Specifically, the part of Assumption 6 concerning the estimators $\hat{\gamma}_n^j$ is replaced by

Assumption 10. *For* $j = 1, 2$, *there exist* $k_j \times 1$ *random vectors* $\{Y_{nt}^j; t = 1, \dots, n, n = 1, 2, \dots\}$ *on* $(\Omega, \mathcal{F}, P)$ *with mean zero, bounded* r*th absolute moments, and near-epoch dependent upon* $\{X_t\}$ *of size* -1 *such that*

$$\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) = -\frac{1}{\sqrt{n}} A_{jn}^+ \sum_{t=1}^n Y_{nt}^j + o_P(1) \tag{6}$$

where $\{A_{jn}^+; n = 1, 2, \dots\}$ *are bounded nonstochastic symmetric* $k_j \times k_j$ *matrices.*

A definition and properties of near-epoch dependence can be found in Gallant and White (1988, Definition 3.13 and Chapter 4). Many econometric estimators defined by optimizing a stochastic function satisfy Assumption 10 (see Gallant and White (1988)). Some $\sqrt{n}$-asymptotically normal semiparametric estimators also possess such an asymptotically linear representation as shown in Newey and McFadden (1994, Section 8).

Now, we turn to model selection criteria. Many common criteria actually take the form of an optimand of the type considered by Gallant and White (1988). For instance, this is the case for the log-likelihood and the MSEP seen earlier as well as for other criteria cited in the introduction and discussed below. Hence, it is natural to restrict the class of criteria by imposing the following primitive assumption.

Assumption 11. *For* $j = 1, 2$,

$$Q_n^j(\omega, \gamma^j) = d_j\{M_n^j(\omega, \gamma^j), \gamma^j\} \tag{7}$$

where $M_n^j(\omega, \gamma^j) = (1/n)\sum_{t=1}^n m_t^j(\omega, \gamma^j)$ *and*

(i) $m_t^j : \Omega \times \Gamma_j \to \mathbb{R}^{q_j}$ *is measurable* $\mathcal{F}/\mathcal{B}^{q_j}$ *for each* γ^j *and continuously differentiable on* Γ_j. *Also,* $d_j : \mathbb{R}^{q_j} \times \Gamma_j \to \mathbb{R}$ *is continuously differentiable on* $\mathbb{R}^{q_j} \times \Gamma_j$,

(ii) $\{m_t^j(\omega, \gamma^j)\}$ *and* $\{\partial m_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ *are almost surely Lipschitz-*L_1 *on* Γ_j,

(iii) $\{m_t^j(\omega, \gamma^j)\}$ *and* $\{\partial m_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ *are* r*-dominated on* Γ_j *uniformly in* t,

(iv) $\{m_t^j(\omega, \gamma^j)\}$ *and* $\{\partial m_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ *are near-epoch dependent upon* $\{X_t\}$ *of size* -1 *and* $-1/2$, *respectively, where the first dependence is uniform on* Γ_j.

The form (7) of the model selection criteria and the requirements (i)–(iv) are reminiscent of the optimands and the regularity conditions placed on them that are considered by Gallant and White (1988), though to simplify, we assume that $d_j(\cdot)$ is not indexed by n. Definitions of a random function that is almost surely Lipschitz-L_1 and r-dominated can be found in Gallant and White (1988, Definitions 3.5 and 3.16).

As in the previous section, however, estimation plays no role. Specifically, we have in mind situations where a researcher has estimated two competing models *via* some estimation methods satisfying the regularity Assumption 10 and wishes to compare the estimated models according to their criterion values $Q_n^j(\omega, \hat{\gamma}_n^j)$. In contrast to the next section, the estimation methods used in this section need not optimize the selection criteria. Such situations are actually frequent. A time-series example using the out-of-sample MSEP as a model selection criterion is fully worked out after Theorem 2. Many other examples can be given. For instance, an econometric example is provided by the comparison of two competing nonlinear simultaneous equations models based on their out-of-sample MSEP, where each model is defined by a set of orthogonality conditions and, as in Andrews and Fair (1988) and Gallant and White (1988), estimated by GMM. Other examples also arise when estimating competing models by ML or other methods on ungrouped data and evaluating their lack-of-fit by some Pearson type chi-square statistics on grouped data (see e.g. Heckman (1981)). Such situations are considered in Vuong and Wang (1993b).

Let $q = q_1 + q_2$ and $k = k_1 + k_2$. Define the $(q + k) \times (q + k)$ matrix

$$V_n = \frac{1}{n}\mathrm{Var}\sum_{t=1}^n \begin{pmatrix} m_t^1(\omega, \bar{\gamma}_n^1) - Em_t^1(\omega, \bar{\gamma}_n^1) \\ Y_{nt}^1 \\ m_t^2(\omega, \bar{\gamma}_n^2) - Em_t^2(\omega, \bar{\gamma}_n^2) \\ Y_{nt}^2 \end{pmatrix}. \tag{8}$$

Let $R_n = (R_n^{1\prime}, -R_n^{2\prime})'$ where R_n^j is the $(q_j + k_j)$-dimensional vector

$$R_n^j = \begin{pmatrix} \partial \bar{d}_{jn}/\partial m^j \\ -A_{jn}^{+}\{E\partial M_n^j(\omega, \bar{\gamma}_n^j)'/\partial \gamma^j \cdot \partial \bar{d}_{jn}/\partial m^j \ + \ \partial \bar{d}_{jn}/\partial \gamma^j\} \end{pmatrix}, \tag{9}$$

and $\partial \bar{d}_{jn}/\partial m^j$ and $\partial \bar{d}_{jn}/\partial \gamma^j$ are the partial derivatives of $d_j(m^j, \gamma^j)$ with respect to m^j and γ^j evaluated at $(EM_n^j(\omega, \bar{\gamma}_n^j), \bar{\gamma}_n^j)$.

The next result specializes Theorem 1 to model selection criteria of the form (7). It replaces the general Assumptions 4–6 by the more primitive Assumptions 9–11, and gives an expression for the asymptotic variance σ_n^2.

Theorem 2. *Given Assumptions* 1–3 *and* 7–11, *suppose that* V_n *is uniformly of rank* s *for some* $s > 0$. *The conclusions of Theorem* 1 *then hold with* $\sigma_n^2 = R_n' V_n R_n$ *where* $V_n = O(1)$ *and* $R_n = O(1)$.

From (8), the asymptotic variance σ_n^2 depends generally on how $\bar{\gamma}_n^j$ is estimated through the asymptotic variance of $\hat{\gamma}_n^j$. However, when $d_j(m_j, \gamma_j)$ does not depend on γ_j and $E\{\partial M_n^j (\omega, \bar{\gamma}_n^j)/\partial \gamma^{j\prime}\} = 0$, we have $R_n^j = (\partial \bar{d}_{jn}/\partial m^{j\prime}, 0)'$ (see also Section 4). The asymptotic variance σ_n^2 then depends only on the asymptotic variance of the (m^1, m^2) components in (8). That is, $\overline{\gamma}_n^j$ can be treated as if it is *known* since sampling uncertainty due to estimation of $\overline{\gamma}_n^j$ becomes asymptotically irrelevant for testing the null hypothesis H_0^* of $\sqrt{n}$-asymptotic equivalence of the competing models. In particular, under weak stationarity of the DGP, the condition $E\{\partial M_n^j(\omega, \bar{\gamma}_n^j)/\partial \gamma^{j\prime}\} = 0$ appears in West (1994, 1996) and is shown to be satisfied by the out-of-sample MSEP in (non)linear regressions estimated by (non)linear least squares.

The following example illustrates the preceding result as well as the verification of its assumptions.[5]

Example. Consider the problem of choosing between the following autoregressive (AR) models for the univariate process $\{Y_t\}_{-\infty}^{+\infty}$:

$$\begin{aligned} \text{AR}(1) &: Y_t = \gamma^1 Y_{t-1} + \varepsilon_{1t} \\ \text{AR}(2) &: Y_t = \gamma^2 Y_{t-2} + \varepsilon_{2t} \end{aligned}$$

where it is specified that $\gamma^j \in \Gamma_j = [-1, 1]$ and that ε_{jt} are uncorrelated with mean zero and variance $\tau_j^2 \in \Upsilon_j \subset [0, +\infty)$, for $j = 1, 2$.

A model selection procedure frequently used in macroeconomic modelling is to compare the out-of-sample MSEP of the estimated competing models. Specifically, suppose one has a sample of $n = 2n^*$ observations $(Y_1, \ldots, Y_n)$, and the first half is used for estimation while the second half is reserved for model comparison.[6] For $j = 1, 2$ the out-of-sample MSEP for the AR(j) model is $Q_n^j(\omega, \hat{\gamma}_n^j)$ where $Q_n^j(\omega, \gamma^j) = (1/n^*) \sum_{t=n^*+1}^{n} (Y_t - \gamma^j Y_{t-j})^2$ and $\hat{\gamma}_n^j$ is some estimator of γ^j based on the first half of the sample. Here, we take $\hat{\gamma}_n^j$ to be the

[5] We are grateful to a referee for suggesting this example. Other model selection problems can be worked out similarly such as choosing between an AR(1) model and a MA(1) model. In particular, the latter problem has been treated differently using some Cox-type tests for nonnested hypotheses (see e.g. Walker (1967), King and McAleer (1987)).

[6] To simplify, we assume that the sample size used for estimation is equal to the out-of-sample size used for model selection. Appropriate changes can accommodate an out-of-sample size p that increases at the same rate as n. See also West (1994, 1996) for other situations such as $\lim_{n\to\infty} p/n = 0$ or ∞.

Yule–Walker estimator of the jth-order autocorrelation coefficient ρ_j^o, namely $\hat{\gamma}_n^j = \hat{\rho}_j \equiv \left(\sum_{t=1+j}^{n^*} Y_t Y_{t-j}\right)/\left(\sum_{t=1}^{n^*} Y_t^2\right)$ (see e.g. Fuller (1976, p. 327)). In particular, $\hat{\gamma}_n^j$ belongs to Γ_j though it does not minimize the out-of-sample MSEP criterion $Q_n^j(\omega, \gamma^j)$.

We now verify that Assumptions 1–3 and 9–11 of Theorem 2 are satisfied. Assumptions 7–8 will be discussed in Sections 6 and 5, respectively. Hereafter, we assume that $\{Y_t\}$ is generated by a finite ARMA with roots outside the unit circle and i.i.d. Gaussian innovations.[7] In particular, $\{Y_t\}$ is strictly stationary (see e.g. Hayashi (2000, Propositions 6.1 and 6.5)) and α-mixing of arbitrary size from Ibragimov and Linnik (1971).[8] For reasons seen below, define $X_t = (Y_t, Y_t^*)'$ where $Y_t^* = 0$ if $|t|$ is odd, and $Y_t^* = Y_{t/2}$ if $|t|$ is even or zero. Given the definition of $\{X_t\}$ and Γ_j, Assumptions 1, 2 and 9 are thus trivially satisfied.

Turning to Assumption 11, which bears on the out-of-sample MSEP, let $M_n^j(\omega, \gamma^j) = (M_{1n}^j(\omega, \gamma^j), M_{2n}^j(\omega, \gamma^j))'$, where

$$M_{1n}^j(\omega, \gamma^j) = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \gamma^j Y_{t-j})^2 \quad \text{and} \quad M_{2n}^j(\omega, \gamma^j) = \frac{1}{n}\sum_{t=1}^{n}(Y_t^* - \gamma^j Y_{t-2j}^*)^2.$$

Note that $M_{2n}^j(\omega, \gamma^j) = (1/n)\sum_{t=1}^{n^*}(Y_t - \gamma^j Y_{t-j})^2$ from the definition of Y_t^*. Let $d_j(m_1, m_2) = 2(m_1 - m_2)$. It is easy to see that the out-of-sample MSEP criterion $Q_n^j(\omega, \gamma^j)$ is of the form (7) with $m_t^j(\omega, \gamma^j) = (m_{1t}^j(\omega, \gamma^j), m_{2t}^j(\omega, \gamma^j))' = ((Y_t - \gamma^j Y_{t-j})^2, (Y_t^* - \gamma^j Y_{t-2j}^*)^2)'$. Moreover, $m_t^j(\cdot)$ and $d_j(\cdot)$ clearly satisfy Assumption 11(i). Regarding Assumption 11(ii)–(iv), we verify them for $\{m_{1t}(\omega, \gamma^j)\}$. Similar arguments apply to $\{m_{2t}(\omega, \gamma^j)\}$.

Now, $\{m_{1t}^j(\omega, \gamma^j)\}$ is almost surely Lipschitz-L_1 on Γ_j from Gallant and White (1988, pp. 21–22). We have $\partial m_{1t}^j(\omega, \gamma^j)/\partial\gamma^j = -2(Y_t - \gamma^j Y_{t-j})Y_{t-j}$. Hence, $|\partial m_{1t}^j(\omega, \gamma^j)/\partial\gamma^j - \partial m_{1t}^j(\omega, \gamma_o^j)/\partial\gamma^j| = 2Y_{t-j}^2|\gamma^j - \gamma_o^j|$, showing that $\{\partial m_{1t}^j(\omega, \gamma^j)/\partial\gamma^j\}$ is almost surely Lipschitz-L_1 on Γ_j since $E(Y_{t-j}^2)$ is constant and finite. Hence Assumption 11(ii) is satisfied. Next, using $|\gamma^j| \leq 1$ together with Minskowski and Cauchy–Schwarz inequalities, we have

$$\|m_{1t}^j(\omega, \gamma^j)\|_r \leq (\|Y_t\|_{2r} + \|Y_{t-j}\|_{2r})^2$$
$$\|\partial m_{1t}^j(\omega, \gamma^j)/\partial\gamma^j\|_r \leq 2(\|Y_t\|_{2r}\|Y_{t-j}\|_{2r} + \|Y_{t-j}\|_{2r}^2)$$

whenever $r > 1$, where $\|\cdot\|_r$ is the L_r norm. This establishes Assumption 11(iii) since $E(Y_t^{2r})$ is constant and finite (see Gallant and White (1988, p. 33)).[9] Lastly, because they involve at most j lags of X_t, $\{m_{1t}(\omega, \gamma^j)\}$ and $\{\partial m_{1t}(\omega, \gamma^j)/\partial\gamma^j\}$ are near-epoch dependent on $\{X_t\}$ of any size. Such near-epoch dependences are clearly uniform on Γ_j, establishing Assumption 11(iv).

It remains to verify Assumptions 3 and 10, which bear on the estimators $\hat{\gamma}_n^j$, $j = 1, 2$. Following the argument in Gallant and White (1988, p. 49), $\hat{\gamma}_n^j = \hat{\rho}_j$ is a strongly consistent estimator of $\left\{\sum_{t=1+j}^{n^*} E(Y_t Y_{t-j})\right\}/\left\{\sum_{t=1}^{n^*} E(Y_t^2)\right\}$, for $j = 1, 2$. Let $\bar{\gamma}_n^j$ be $n^*/(n^* - j)$ times

[7] This assumption is stronger than necessary, but greatly facilitates the verification of the assumptions. Whenever possible, we indicate when it can be weakened.

[8] Gaussianity can be relaxed as non-Gaussian ARMA(p, q) processes are also α-mixing of arbitrary size under appropriate conditions. See Pham and Tran (1980).

[9] The preceding argument shows that stationarity and Gaussianity can be weakened for Assumption 11(ii), (iii) to hold as it suffices that EY_t^{2r} be uniformly bounded for some $r > 1$.

the latter quantity. Because $n^*/(n^* - j) \to 1$ as $n \to \infty$, it follows that $\hat{\gamma}_n^j - \bar{\gamma}_n^j \xrightarrow{as} 0$. Moreover, because the process $\{Y_t\}$ is weakly stationary, it is easy to see that $\bar{\gamma}_n^j = \rho_j^o$, the jth-order autocorrelation coefficient of $\{Y_t\}$, which is constant and in the interior of Γ_j. Hence Assumption 3 is satisfied. Moreover, we have

$$\begin{aligned}\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) &= -\frac{2n^*}{\sum_{t=1}^{n^*} Y_t^2} \frac{1}{\sqrt{2n^*}} \left\{ \sum_{t=1}^{n^*} (\rho_j^o Y_t^2 - Y_t Y_{t-j}) + \sum_{t=1}^{j} Y_t Y_{t-j} \right\} \\ &= -\frac{2}{E(Y_t^2)} \frac{1}{\sqrt{2n^*}} \sum_{t=1}^{n^*} (\rho_j^o Y_t^2 - Y_t Y_{t-j}) + o_P(1)\end{aligned}$$

since $(1/n^*) \sum_{t=1}^{n^*} Y_t^2 \xrightarrow{as} E(Y_t^2)$ and $(1/\sqrt{n^*}) \sum_{t=1}^{n^*} (\rho_j^o Y_t^2 - Y_t Y_{t-j}) = O_P(1)$ by a variety of Laws of Large Numbers and Central Limit Theorems (see e.g. Theorems 3.15 and 5.3 in Gallant and White (1988) for near-epoch dependent functions of mixing processes), and $(1/\sqrt{n}) \sum_{t=1}^{j} Y_t Y_{t-j} \xrightarrow{P} 0$. Hence, the estimator $\hat{\gamma}_n^j$ satisfies the asymptotic linear representation (6) with $A_{jn}^+ = 2/E(Y_t^2)$ and $Y_{nt}^j = (\rho_j^o Y_t^2 - Y_t Y_{t-j}) \mathbb{I}(t \le n^*)$, where $\mathbb{I}(\cdot)$ is the indicator of the event in parentheses. Thus, because $\{A_{jn}^+\}$ is constant and $\{Y_{nt}^j\}$ is near-epoch dependent on $\{Y_t\}$ of any size and hence near-epoch dependent on $\{X_t\}$ of size -1, Assumption 10 is satisfied.

Provided Assumptions 7 and 8 are satisfied and V_n is of uniform rank (see Sections 6 and 5), Theorem 2 applies to the out-of-sample MSEP for comparing the above AR(1) and AR(2) models. That is, the quantity

$$T_n = \frac{\sqrt{2}}{\hat{\sigma}_n \sqrt{n^*}} \sum_{t=n^*+1}^{n} \{(Y_t - \hat{\rho}_1 Y_{t-1})^2 - (Y_t - \hat{\rho}_2 Y_{t-2})^2\} \tag{10}$$

can be used as a model selection statistic for testing the null hypothesis H_0^* of $\sqrt{n}$-asymptotic equivalence. In particular, from $\bar{Q}_n^j(\gamma^j) = d_j\{EM_n^j(\omega, \gamma^j), \gamma_j\}$ (see the proof of Theorem 2) and weak stationarity of $\{Y_t\}$, we have $\bar{Q}_n^j(\gamma^j) = E(Y_t - \gamma^j Y_{t-j})^2 = \gamma_Y(0)(1 + \gamma^{j2}) - 2\gamma_Y(j)\gamma^j$, where $\gamma_Y(\cdot)$ is the autocovariance function of $\{Y_t\}$. Hence, $\bar{Q}_n^j(\bar{\gamma}_n^j) = \gamma_Y(0)(1 - \rho_j^{o2})$, which is independent of n. Therefore, the hypotheses H_0, H_1 and H_2 are identical to H_0^*, H_1^* and H_2^*, respectively, which reduce to

$$H_0 : |\rho_1^o| = |\rho_2^o|, \qquad H_1 : |\rho_1^o| > |\rho_2^o|, \qquad H_2 : |\rho_1^o| < |\rho_2^o|. \tag{11}$$

Moreover, because $E\{\partial M_n^j(\omega, \bar{\gamma}_n^j)/\partial \gamma^{j\prime}\} = 0$ and $d_j(\cdot)$ does not depend on γ^j, the asymptotic variance σ_n^2 does not depend on the sampling variability of $\hat{\rho}_j$ so that ρ_j^o can be treated as known in the computation of σ_n^2, as noted after Theorem 2. See also Section 5.

4. MODEL SELECTION TESTS WITH LACK-OF-FIT MINIMIZATION

Up to now, the methods used to estimate the competing models need not optimize the criteria for model selection. It frequently happens, however, that estimators minimize the chosen lack-of-fit measures. The most common situation arises when the competing models are fully parametric and estimated by ML methods. A frequent criterion then is the model log-likelihood possibly adjusted (see e.g. Akaike (1973, 1974)). When the observations are i.i.d. this situation is analyzed in Vuong (1989) for general parametric models and in Lien and Vuong (1987) for normal linear regressions. Other examples arise when estimating fully parametric models by minimum chi-square methods and using Pearson type statistics as a criterion for model selection (see Vuong and Wang (1991, 1993a)).

When the competing models are not fully parametrized, an important econometric example is given by nonlinear simultaneous equation models, where each competing model is defined by a set of implicit simultaneous equations and a set of orthogonality conditions. Each model is then estimated by nonlinear IV or GMM (see Amemiya (1985), Hansen (1982) and Gallant and White (1988) among others). Following Sargan (1958), Newey and West (1987a) and recently Pesaran and Smith (1994), when both competing models are overidentified, the value of the GMM optimand evaluated at the GMM estimator can be the basis for hypothesis testing and more generally model selection in nested and nonnested situations.

The simultaneous equation example is interesting because the lack-of-fit criterion depends on some nuisance parameters associated with the weighting matrix used in GMM estimation. To include such situations in our analysis, we partition the parameter vector γ^j into the parameter vector of interest θ^j and the vector of nuisance parameters τ^j. Similarly to Assumption 11, model selection criteria are now assumed to satisfy

Assumption 12. *For $j = 1, 2$, $\Gamma_j = \Theta_j \times \Upsilon_j$ where Θ_j and Υ_j are compact subsets of $\mathbb{R}^{k_j-h_j}$ and $\mathbb{R}^{h_j}$. Moreover,*

$$Q_n^j(\omega, \gamma^j) = d_j\{M_n^j(\omega, \theta^j), \theta^j, \tau^j\} \tag{12}$$

where $\gamma^j = (\theta^{j\prime}, \tau^{j\prime})'$, $M_n^j(\omega, \theta^j) = (1/n)\sum_{t=1}^n m_t^j(\omega, \theta^j)$ and

- *(i) $m_t^j : \Omega \times \Theta_j \to \mathbb{R}^{q_j}$ is measurable $\mathcal{F}/\mathcal{B}^{q_j}$ for each θ^j and twice continuously differentiable on Θ_j. Also, $d_j : \mathbb{R}^{q_j} \times \Gamma_j \to \mathbb{R}$ is twice continuously differentiable on $\mathbb{R}^{q_j} \times \Gamma_j$,*
- *(ii) $\{m_t^j(\omega, \theta^j)\}$, $\{\partial m_t^j(\omega, \theta^j)/\partial\theta^{j\prime}\}$ and $\{\partial^2 m_t^j(\omega, \theta^j)/\partial\theta^j\partial\theta^{j\prime}\}$ are almost surely Lipschitz-L_1 on Θ_j,*
- *(iii) $\{m_t^j(\omega, \theta^j)\}$, $\{\partial m_t^j(\omega, \theta^j)/\partial\theta^{j\prime}\}$ and $\{\partial^2 m_t^j(\omega, \theta^j)/\partial\theta^j\partial\theta^{j\prime}\}$ are r-dominated on Θ_j uniformly in t,*
- *(iv) $\{m_t^j(\omega, \theta^j)\}$, $\{\partial m_t^j(\omega, \theta^j)/\partial\theta^{j\prime}\}$ and $\{\partial^2 m_t^j(\omega, \theta^j)/\partial\theta^j\partial\theta^{j\prime}\}$ are near-epoch dependent upon $\{X_t\}$ of size -1, -1 and $-1/2$, respectively, where the first two dependencies are uniform on Θ_j.*

Conditions (i)–(iv) are similar to those used by Gallant and White (1988) and strengthen conditions (i)–(iv) of Assumption 11. These authors consider the case where optimands are of the form $Q_n^j(\omega, \theta^j) = d_j\{M_n^j(\omega, \theta^j)\}$. Bates and White (1985) consider the case where $Q_n^j(\omega, \gamma^j) = d_j\{M_n^j(\omega, \theta^j), \tau^j\}$. These are special cases of (12). On the other hand, Andrews and Fair (1988) consider the case where $Q_n^j(\omega, \gamma^j) = d_j\{M_n^j(\omega, \theta^j, \tau^j), \tau^j\}$. The present formulation was preferred because it includes minimum Pearson chi-square estimation (see Vuong and Wang (1991, 1993a)).[10]

[10]The general case where $Q_n^j(\omega, \gamma^j) = d_j\{M_n^j(\omega, \theta^j, \tau^j), \theta^j, \tau^j\}$ was not treated to economize on proofs and notations, but follows similarly. Moreover, to simplify, $d_j(\cdot)$ is again assumed independent of n.

In this section, estimators $\hat{\theta}_n^j$ of the parameters of interest are obtained by minimizing model selection criteria conditional upon some preliminary estimates $\hat{\tau}_n^j$. That is,

$$\hat{\theta}_n^j = \arg\min_{\theta^j \in \Theta_j} d_j\{M_n^j(\omega, \theta^j), \theta^j, \hat{\tau}_n^j\}, \tag{13}$$

for $j = 1, 2$. The lack-of-fit associated with model $\mathcal{M}_j$ is then $Q_n^j(\omega, \hat{\gamma}_n^j) = d_j\{M_n^j(\omega, \hat{\theta}_n^j), \hat{\theta}_n^j, \hat{\tau}_n^j\}$. Conditions on the asymptotic behavior of the nuisance parameter estimators are required. We assume

Assumption 13. *For $j = 1, 2$, let $\hat{\tau}_n^j : \Omega \to \Upsilon_j$ be such that there exists a nonstochastic sequence $\{\bar{\tau}_n^j\}_{n=1}^{\infty}$ uniformly interior to Υ_j for which $\hat{\tau}_n^j - \bar{\tau}_n^j \xrightarrow{as} 0$ as $n \to \infty$. Moreover, there exist $h_j \times 1$ random vectors $\{Y_{2nt}^j; t = 1, \ldots, n, n = 1, 2, \ldots\}$ on $(\Omega, \mathcal{F}, P)$ with mean zero, bounded rth absolute moments, and near-epoch dependent upon $\{X_t\}$ of size -1 such that*

$$\sqrt{n}(\hat{\tau}_n^j - \bar{\tau}_n^j) = -\frac{1}{\sqrt{n}} A_{2jn}^{+} \sum_{t=1}^{n} Y_{2nt}^j + o_P(1) \tag{14}$$

where $\{A_{2jn}^{+}; n = 1, 2, \ldots\}$ are bounded nonstochastic symmetric $h_j \times h_j$ matrices.

Assumption 13 is standard and is satisfied by many optimization estimators (see Gallant and White (1988)). Unlike Bates and White (1985) and Andrews and Fair (1988), however, we do not impose conditions on cross partial derivatives with respect to θ^j and τ^j of the lack-of-fit criterion or optimand (12) so that estimation of the nuisance parameters may affect the asymptotic distribution of the estimators $\hat{\theta}_n^j$. This extension is useful for minimum chi-square estimation and model selection tests based on Pearson type chi-square statistics (see Vuong and Wang (1991, 1993a)).

We need an identification condition similar to those used in the literature.

Assumption 14. *Let $\bar{Q}_n^j(\theta^j, \tau^j) = d_j\{EM_n^j(\omega, \theta^j), \theta^j, \tau^j\}$ and let $A_{1jn} = \partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^j \partial\theta^{j\prime}$.*

(i) The sequence $\{\bar{Q}_n^j(\cdot, \bar{\tau}_n^j)\}$ has identifiably unique minimizers $\{\bar{\theta}_n^j\}$ uniformly interior to Θ_j.
(ii) The sequence of matrices A_{1jn} is uniformly positive definite.

For a definition of identifiably uniqueness, see Domowitz and White (1982) or Gallant and White (1988).

Let $h = h_1 + h_2$. Define the $(q + h) \times (q + h)$ matrix

$$V_n = \frac{1}{n} \mathrm{Var} \sum_{t=1}^{n} \begin{pmatrix} m_t^1(\omega, \bar{\theta}_n^1) - Em_t^1(\omega, \bar{\theta}_n^1) \\ Y_{2nt}^1 \\ m_t^2(\omega, \bar{\theta}_n^2) - Em_t^2(\omega, \bar{\theta}_n^2) \\ Y_{2nt}^2 \end{pmatrix}. \tag{15}$$

Let $R_n = (R_n^{1\prime}, -R_n^{2\prime})'$ where R_n^j is the $(q_j + h_j)$-dimensional vector

$$R_n^j = \begin{pmatrix} \partial\bar{d}_{jn}/\partial m^j \\ -A_{2jn}^{+}\partial\bar{d}_{jn}/\partial\tau^j \end{pmatrix}, \tag{16}$$

and $\partial \bar{d}_{jn}/\partial m^j$ and $\partial \bar{d}_{jn}/\partial \tau^j$ are the partial derivatives of $d_j(m^j, \theta^j, \tau^j)$ with respect to m^j and τ^j evaluated at $(EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j)$.

The next theorem gives the basic result for model selection when the estimators used minimize (possibly in two steps) the lack-of-fit criteria. Relative to Theorem 1, it replaces the general Assumptions 2–6 by the more primitive Assumptions 9 and 12–14. The theorem also gives the corresponding expression for the asymptotic variance σ_n^2.

Theorem 3. *Given Assumptions* 1, 7–9 *and* 12–14, *suppose that* V_n *is uniformly of rank* s *for some* $s > 0$. *The conclusions of Theorem* 1 *then hold with* $\sigma_n^2 = R_n' V_n R_n$ *where* R_n *and* V_n *are now defined by* (16) *and* (15).

Note that the asymptotic variance σ_n^2 does not depend on the first partial derivative of $d_j(m^j, \theta^j, \tau^j)$ with respect to θ^j. More importantly, it is easy to see that σ_n^2 is the same as if $\bar{\theta}_n^j$, $j = 1, 2$ were *known*. That is, when estimators of θ^j minimize the chosen model selection criteria, sampling uncertainty due to estimation of $\overline{\theta}_n^j$ becomes asymptotically irrelevant for testing the null hypothesis H_0^* of $\sqrt{n}$-asymptotic equivalence of the competing models. This is the case when the competing models are estimated by ML and compared on the basis of their likelihood values possibly adjusted as in Sin and White (1996).[11] This is also the case for competing (non)linear regressions estimated by (non)linear least squares and compared *via* their in-sample MSEP, and hence their out-of-sample MSEP under covariance stationarity of the DGP, as in West (1994, 1996). See the example below.

In contrast, when $\partial \bar{d}_{jn}/\partial \tau^j \neq 0$, then σ_n^2 and the asymptotic distribution of the difference in lack-of-fit $\Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2)$ depend on how the nuisance parameters τ^j are estimated. This is so whether or not $\partial^2 \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \theta^j \partial \tau^{j\prime} = 0$, i.e. whether or not estimation of the nuisance parameters τ^j affects the asymptotic distributions of the estimators $\hat{\theta}_n^j$ of the parameters of interest. Such a result is surprising, but in fact agrees with Theorem 2 for the special case where $m_t^j(\cdot)$ does not depend on θ^j and $h_j = k_j$ so that $\tau^j = \gamma^j$.

Example (Continued). Consider again the problem of choosing between the AR(1) and AR(2) models of Section 3. Instead of the out-of-sample MSEP, we use here the in-sample MSEP $\tilde{Q}_n^j(\omega, \tilde{\gamma}_n^j)$, where $\tilde{\gamma}_n^j$ minimizes $\tilde{Q}_n^j(\omega, \gamma^j) = (1/(n-j)) \sum_{t=1+j}^n (Y_t - \gamma^j Y_{t-j})^2$ over Γ_j for $j = 1, 2$. Note that $\tilde{\gamma}_n^j$ is the least-squares estimator of γ^j constrained to the compact Γ_j, and is in general not equal to the jth autocorrelation estimator $\hat{\rho}_j$. As before, to verify easily the assumptions of Theorem 3, we assume that $\{Y_t\}$ is generated by a finite ARMA with roots outside the unit circle and i.i.d. Gaussian innovations. Thus Assumptions 1 and 9 are satisfied with $X_t = Y_t$.

Let $Q_n^j(\omega, \gamma^j) = \{(n-j)/n\}\tilde{Q}_n^j(\omega, \gamma^j)$. Hence, $\tilde{\gamma}_n^j$ also minimizes $Q_n^j(\omega, \gamma^j)$ over Γ_j for $j = 1, 2$. Because there are no nuisance parameters, Assumption 13 holds trivially. Moreover, $Q_n^j(\omega, \gamma^j)$ is of the form (12) with $\gamma^j = \theta^j$, $d_j(m) = m$, and $m_t^j(\omega, \theta^j) = (Y_t - \gamma^j Y_{t-j})^2$ if $t \geq j+1$ and equal to zero if $t \leq j$. Thus, as in Section 3, $d_j(\cdot)$, $\{m_t^j(\omega, \theta^j)\}$ and $\{\partial m_t^j(\omega, \theta^j)/\partial \theta^j\}$ satisfy Assumptions 12(i)–(iv). Since $\partial^2 m_t^j(\omega, \theta^j)/\partial \theta^j \partial \theta^{j\prime} = 2Y_{t-j}^2$, it follows that Assumption 12 is satisfied. Next, consider Assumption 14. We have $\bar{Q}_n^j(\gamma^j) = \{(n-j)/n\}E(Y_t -$

[11] Sin and White (1996) provide conditions on the penalty functions ensuring weak or strong consistency of the adjusted likelihood criterion. Thus, combining their results with ours delivers a likelihood-based procedure that is consistent both as a model selection criterion and a model selection test of H_0^*.

$\gamma^j Y_{t-j})^2$ using the weak stationarity of $\{Y_t\}$. Thus $\bar{Q}_n^j(\gamma^j)$ is minimized uniquely at $\bar{\gamma}_n^j = \rho_j^o$, which is independent of n and belongs to $(-1, +1)$. Also, A_{1jn} is uniformly positive whenever $n > j$ as $A_{1jn} = 2\{(n-j)/n\}\gamma_Y(0)$. Hence, Assumption 14 is satisfied.

Provided Assumptions 7 and 8 are satisfied and V_n is of uniform rank (see Sections 5 and 6), Theorem 3 applies to the criterion $Q_n^j(\omega, \tilde{\gamma}_n^j)$ for comparing the AR(1) and AR(2) models. Because $\tilde{Q}_n^j(\omega, \tilde{\gamma}_n^j) = Q_n^j(\omega, \tilde{\gamma}_n^j)\{1 + O(1/n)\}$, Theorem 3 also applies to the in-sample MSEP, i.e. the quantity

$$\tilde{T}_n = \frac{\sqrt{n}}{\tilde{\sigma}_n}\left\{\frac{1}{n-1}\sum_{t=2}^{n}(Y_t - \tilde{\gamma}_n^1 Y_{t-1})^2 - \frac{1}{n-2}\sum_{t=3}^{n}(Y_t - \tilde{\gamma}_n^2 Y_{t-2})^2\right\} \tag{17}$$

can be used as a model selection statistic for testing the null hypothesis H_0^* of $\sqrt{n}$-asymptotic equivalence. Moreover, because $\bar{Q}_n^j(\bar{\gamma}_n^j)$ is the same as for the out-of-sample MSEP evaluated at the autocorrelation estimator (see Section 3), the same remarks apply.[12] Namely, the hypotheses H_0, H_1 and H_2 are identical to H_0^*, H_1^* and H_2^*, respectively, and reduce to (11). Moreover, as noted after Theorem 3, because $\tilde{\gamma}_n^j$ minimizes $Q_n^j(\omega, \gamma^j)$, the asymptotic variance σ_n^2 does not depend on the sampling variability of the (constrained) least-squares estimators $\tilde{\gamma}_n^j$, and hence can be computed as if ρ_j^o is known. See also Section 5.

5. CONSISTENT VARIANCE ESTIMATION

For our proposed tests to be operational, it is necessary to have a consistent estimator of the asymptotic variance σ_n^2 (see Assumption 8). The next results are derived for situations where estimators do not necessarily optimize the selection criteria (see Section 3). Situations where estimators do optimize the selection criteria (see Section 4) are studied similarly. From Theorem 2 we know that $\sigma_n^2 = R_n' V_n R_n$ where V_n and R_n are given by (8) and (9), respectively. Thus it suffices to construct some consistent estimators of V_n and R_n. A consistent estimator of the $(q+k)$-dimensional vector R_n is obtained as usual by its sample analog evaluated at $(\hat{\gamma}_n^1, \hat{\gamma}_n^2)$ (see (19) below).

The difficulty is to obtain a consistent estimator of the $(q+k) \times (q+k)$ variance covariance matrix V_n. When observations are i.i.d. and the competing models are nondynamic, constructing a consistent estimator of the asymptotic variance σ_n^2 is straightforward (see Vuong (1989), Vuong and Wang (1991, 1993a,b)). This is because $\bar{\gamma}_n^j$ and Y_{nt}^j do not depend on n and because the $(q+k)$-dimensional vectors appearing in (8) are i.i.d. Thus the matrix V_n is just the population variance covariance matrix of this $(q+k)$-dimensional vector. Hence a simple consistent estimator of V_n is its sample analog evaluated at the estimates $(\hat{\gamma}_n^1, \hat{\gamma}_n^2)$.

When observations are dependent and heterogeneous, consistent estimation of asymptotic variances is more complex but has been solved under general conditions (see Newey and West (1987b)), Gallant and White (1988, Chapter 6), and Andrews (1991) among others). In particular, an important condition is that the estimated model be correctly specified or that the DGP

[12] Findley (1990) notes that comparing the (in-sample) log-likelihood values is also equivalent to comparing the one-step MSEP when the competing models are Gaussian ARMA or ARIMA models. Diebold and Mariano (1995) allow for more general losses than the MSEP, though their results require either that the parameters of the competing models be known or $\lim_{n\to\infty} p/n = 0$, as noted by West (1996).

be stationary. However, even under our null hypothesis H_0^*, both competing models can be misspecified, while stationarity has not been assumed. The contribution of this section is to show that consistent estimation of the asymptotic variance of our test statistic is still possible in some important situations for weakly dependent and heterogeneous DGPs.

First, we strengthen Assumption 10 on the estimators $\hat{\gamma}_n^j$.

Assumption 15. *For $j = 1, 2$, there exists a sequence of k_j-dimensional functions $\{\delta_t^j : \Omega \times \Gamma_j \to \mathbb{R}^{k_j}\}_{t=1}^{+\infty}$ satisfying*

(i) $\delta_t^j(\cdot,\cdot)$ *is measurable $\mathcal{F}/\mathcal{B}^{k_j}$ for each γ^j and continuously differentiable on Γ_j,*
(ii) $\{\delta_t^j(\omega, \gamma^j)\}$ *and* $\{\partial\delta_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ *are almost surely Lipschitz-L_1 on Γ_j,*
(iii) $\{\delta_t^j(\omega, \gamma^j)\}$ *and* $\{\partial\delta_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ *are $2r$-dominated on Γ_j uniformly in t,*
(iv) $\{\delta_t^j(\omega, \gamma^j)\}$ *is near-epoch dependent upon $\{X_t\}$ of size $-2(r-1)/(r-2)$ uniformly on Γ_j,*

such that

$$\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) = -\frac{1}{\sqrt{n}} A_{jn}^+ \sum_{t=1}^{n} \{\delta_t^j(\omega, \bar{\gamma}_n^j) - E\delta_t^j(\omega, \bar{\gamma}_n^j)\} + o_P(1), \tag{18}$$

where $\{A_{jn}^+; n = 1, 2, \ldots\}$ are bounded nonstochastic symmetric $k_j \times k_j$ matrices. Moreover, there exists a sequence of random matrices $\{\hat{A}_{jn}^+; n = 1, 2, \ldots\}$ such that $\hat{A}_{jn}^+ - A_{jn}^+ \xrightarrow{P} 0$.

A comparison of (6) and (18) gives $Y_{nt}^j = \delta_t^j(\omega, \bar{\gamma}_n^j) - E\delta_t^j(\omega, \bar{\gamma}_n^j)$. Note that we allow $E\delta_t^j(\omega, \bar{\gamma}_n^j) \neq 0$. Extremum estimators that are $\sqrt{n}$-asymptotically normal typically satisfy Assumption 15 (see Gallant and White (1988)).

Second, we strengthen Assumption 11 on model selection criteria.

Assumption 16. *Assumption 11 holds with (iii) and (iv) strengthened to*

(iii) $\{m_t^j(\omega, \gamma^j)\}$ *and* $\{\partial m_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ *are $2r$-dominated on Γ_j uniformly in t,*
(iv) $\{m_t^j(\omega, \gamma^j)\}$ *and* $\{\partial m_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ *are near-epoch dependent upon $\{X_t\}$ of size $-2(r-1)/(r-2)$ and $-1/2$, respectively, where the first dependence is uniform on Γ_j.*

Next, we follow Newey and West (1987b) and Gallant and White (1988) and we introduce a truncation lag and some weights.

Assumption 17. *$\{m_n\}$ is a sequence of integers such that $m_n \to +\infty$ as $n \to +\infty$ and $m_n = o(n^{1/4})$.*

Assumption 18. *Given a sequence $\{m_n\}$, define $w_{n\tau} = \sum_{t=\tau+1}^{m_n+1} a_{nt}a_{n,t-\tau}$ where $\{a_{nt}; t = 1, \ldots, m_n+1, n = 1, 2, \ldots\}$ is a triangular array such that $|w_{nt}| \leq \Delta$ for some $\Delta < \infty$ and all $n = 1, 2, \ldots$ and $\tau = 0, 1, \ldots, m_n$. Moreover, for each τ, $w_{n\tau} \to 1$ as $n \to \infty$.*

Assumptions 17 and 18 are identical to assumptions TL and WT of Gallant and White (1988). See also Andrews (1991) for weaker assumptions.

We are now in a position to define the class of variance estimators that are considered. Using (9), a consistent estimator of R_n is, as usual, $\hat{R}_n = (\hat{R}_n^{1\prime}, -\hat{R}_n^{2\prime})'$ where $\hat{R}_n^j$ is the $(q_j + k_j)$-dimensional vector

$$\hat{R}_n^j = \begin{pmatrix} \partial \hat{d}_{jn}/\partial m^j \\ -\hat{A}_{jn}^{+}(\partial M_n^j(\omega, \hat{\gamma}_n^j)'/\partial \gamma^j \cdot \partial \hat{d}_{jn}/\partial m^j + \partial \hat{d}_{jn}/\partial \gamma^j) \end{pmatrix}, \tag{19}$$

and $\partial \hat{d}_{jn}/\partial m^j$ and $\partial \hat{d}_{jn}/\partial \gamma^j$ are the partial derivatives of $d_j(m^j, \gamma^j)$ with respect to m^j and γ^j evaluated at $(M_n^j(\omega, \hat{\gamma}_n^j), \hat{\gamma}_n^j)$. Define the $(q + k) \times (q + k)$ matrix

$$\hat{V}_n = \frac{w_{n0}}{n} \sum_{t=1}^{n} \hat{U}_{nt}\hat{U}'_{nt} + \frac{1}{n} \sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+1}^{n} (\hat{U}_{nt}\hat{U}'_{n,t-\tau} + \hat{U}_{n,t-\tau}\hat{U}'_{nt}) \tag{20}$$

where $\hat{U}_{nt}$ is the $(q + k)$-dimensional vector

$$\hat{U}_{nt} = (m_t^1(\omega, \hat{\gamma}_n^1)', \delta_t^1(\omega, \hat{\gamma}_n^1)', m_t^2(\omega, \hat{\gamma}_n^2)', \delta_t^2(\omega, \hat{\gamma}_n^2)')'. \tag{21}$$

We then define $\hat{\sigma}_n^2 = \hat{R}'_n \hat{V}_n \hat{R}_n$.

The next theorem is the main result of this section. Define the $(q + k) \times (q + k)$ matrix

$$\Lambda_n = \frac{w_{n0}}{n} \sum_{t=1}^{n} \mu_{nt}\mu'_{nt} + \frac{1}{n} \sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+1}^{n} (\mu_{nt}\mu'_{n,t-\tau} + \mu_{n,t-\tau}\mu'_{nt}) \tag{22}$$

where μ_{nt} is the $(q + k)$-dimensional vector

$$\mu_{nt} = (Em_t^1(\omega, \bar{\gamma}_n^1)', E\delta_t^1(\omega, \bar{\gamma}_n^1)', Em_t^2(\omega, \bar{\gamma}_n^2)', E\delta_t^2(\omega, \bar{\gamma}_n^2)')'. \tag{23}$$

Lastly, let $\bar{\mu}_n = (1/n) \sum_{t=1}^{n} \mu_{nt} = (\bar{\mu}_n^{1\prime}, \bar{\mu}_n^{2\prime})'$. Note that $\bar{\mu}_n \neq 0$ in general. Indeed, $EM_n^j(\omega, \bar{\gamma}_n^j)$, which composes the first q_j elements of $\bar{\mu}_n^j$, is typically nonzero. In contrast, the last k_j elements of $\bar{\mu}_n^j$ are typically equal to zero as $\bar{\gamma}_n^j$ satisfies $E(1/n) \sum_{t=1}^{n} \delta_t^j(\omega, \bar{\gamma}_n^1) = 0$ in general.

Theorem 4. *Given Assumptions 1–3, 9 and 15–18, suppose that, under H_0^*, (i) $R'_n\bar{\mu}_n = O(n^{-1/4})$ and (ii) there exists a sequence $\{d_n\}$ such that $|R'_n(\mu_{nt} - \bar{\mu}_n)| \leq d_n$ for every $t = 1, \ldots, n$ with $d_n = O(n^{-1/8})$. Then $\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0$ under H_0^*.*[13]

As noted by Gallant and White (1988, Chapter 6), in the presence of heterogeneous observations and misspecifications, $\hat{V}_n$ *overestimates* in general V_n since $\hat{V}_n - (V_n + \Lambda_n) \xrightarrow{P} 0$, where Λ_n is positive semidefinite (see step 2 in the proof). Moreover, as noted by these authors, Λ_n is not guaranteed to be bounded. The contribution of Theorem 4 is thus to provide some sufficient conditions that ensure the consistency of $\hat{\sigma}_n^2$ to σ_n^2 so that asymptotically valid inferences based on T_n can be performed. Because $R_n = O(1)$, condition (i) requires that a linear combination of

[13] The rates $n^{-1/4}$ and $n^{-1/8}$ arise from the rate of m_n in Assumption 17. As its proof shows, Theorem 4 actually holds for any rate of m_n that guarantees the consistency of $\hat{V}_n$ for V_n, provided $R'_n\bar{\mu}_n = o(1/m_n)$ in (i) and $d_n = o(1/\sqrt{m_n})$ in (ii). In particular, Andrews (1991) shows that the optimal rate of m_n for the Bartlett weights $w_{n\tau}$ used by Newey and West (1987b) is $O(n^{1/3})$, and hence does not satisfy Assumption 17. See also Andrews (1991) for optimal weights and data-dependent automatic determination of m_n.

$\bar{\mu}_n$ vanishes, while condition (ii) controls the fluctuations of the individual means μ_{nt} around the overall mean $\bar{\mu}_n$.[14]

Condition (i) is satisfied in important situations such as selecting models estimated by ML based on their possibly adjusted likelihood values, or selecting models by GMM based on their GMM criteria. Specifically in ML estimation, we have $d_j(m^j, \gamma^j) = m^j$ so that $\bar{Q}_n^j(\bar{\gamma}_n^j) = EM_n^j(\omega, \bar{\gamma}_n^j)$. Moreover, $\hat{\gamma}_n^j = \arg\min Q_n^j(\omega, \gamma^j)$ and $\bar{\gamma}_n^j$ satisfies $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \gamma^j = 0$. Therefore R_n^j is zero except for its first component which is equal to one. Hence $R_n^{j\prime}\bar{\mu}_n^j = EM_n^j(\omega, \bar{\gamma}_n^j) = \bar{Q}_n^j(\bar{\gamma}_n^j)$. It follows that condition (i) is satisfied under H_0^*. In GMM estimation we have $d_j\ (m^j, \gamma^j) = m^{j\prime}P_j m^j$ where P_j is a $q_j \times q_j$ matrix so that $\bar{Q}_n^j(\bar{\gamma}_n^j) = EM_n^j(\omega, \bar{\gamma}_n^j)' P_j EM_n^j(\omega, \bar{\gamma}_n^j)$. Moreover, $\bar{\gamma}_n^j$ satisfies $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \gamma^j = 0$. Thus $R_n^{j\prime} = (2EM_n^j(\omega, \bar{\gamma}_n^j)' P_j, 0)$. Hence $R_n^{j\prime}\bar{\mu}_n^j = 2EM_n^j(\omega, \bar{\gamma}_n^j)' P_j EM_n^j(\omega, \bar{\gamma}_n^j) = 2\bar{Q}_n^j(\bar{\gamma}_n^j)$ so that condition (i) again holds under H_0^*. The latter case includes selecting (non)linear regressions estimated by (non)linear least squares based on their in-sample or out-of-sample MSEP under covariance stationarity as in West (1994, 1996). See the example below. More generally, when the last k_j elements of $\bar{\mu}_n^j$ are zero, as noted before Theorem 4, then condition (i) reduces to $n^{1/4}\{\partial \bar{d}_{1n}/\partial m^{1\prime} EM_n^1(\omega, \bar{\gamma}_n^1) - \partial \bar{d}_{2n}/\partial m^{2\prime} EM_n^2(\omega, \bar{\gamma}_n^2)\} = O(1)$, which must hold under H_0^*.

Regarding condition (ii), note that Assumptions 15(iii) and 16(iii) already imply $|\mu_{nt} - \bar{\mu}_n| \leq K < \infty$ for every n, t so that $|R_n'(\mu_{nt} - \bar{\mu}_n)|$ is bounded uniformly in n, t because $R_n = O(1)$. Thus condition (ii) strengthens this requirement. For instance, in the above ML case, this condition requires that the deviation of $Em_t^1(\omega, \bar{\gamma}_n^1) - Em_t^2(\omega, \bar{\gamma}_n^2)$ from its mean $EM_n^1(\omega, \bar{\gamma}_n^1) - EM_n^2(\omega, \bar{\gamma}_n^2)$, for $t = 1, \ldots, n$ not only remains bounded but decreases (at the rate $n^{-1/8}$) to zero as $n \to \infty$. Second, condition (ii) is clearly satisfied if $Em_t^j(\omega, \gamma^j) = Em_s^j(\omega, \gamma^j)$ and $E\delta_t^j(\omega, \gamma^j) = E\delta_s^j(\omega, \gamma^j)$, for every $t, s = 1, 2, \ldots, \gamma^j \in \Gamma^j$, and for $j = 1, 2$. This holds when $\{m_t^j(\omega, \gamma^j), t = 1, 2, \ldots\}$ and $\{\delta_t^j(\omega, \gamma^j), t = 1, 2, \ldots\}$ are first-order stationary processes.

Example (Continued). We consider the case where the out-of-sample MSEP is used as a model selection criterion. Because Assumptions 1–3 and 9 have been already verified in Section 3, it remains to verify Assumptions 15–18 and conditions (i)–(ii) in order to apply Theorem 4. Assumptions 17 and 18 are satisfied by letting m_n grow at a rate slower than $n^{1/4}$ and by choosing appropriate weights w_{nt} such as the Bartlett weights $w_{nt} = 1 - \{t/(m_n + 1)\}$ (see also footnote 13). Assumption 16 is verified by using an argument similar to that used for verifying Assumption 11 in Section 3.

To verify Assumption 15, we use the asymptotic linear representation of $\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j)$ obtained in Section 3 and the definition of Y_t^*. These give

$$\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) = -\frac{2}{E(Y_t^2)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\rho_j^o Y_t^{*2} - Y_t^* Y_{t-2j}^*) + o_P(1).$$

[14]Note that $\sigma_n^2 = \mathrm{Var}\ (1/\sqrt{n}) \sum_{t=1}^n R_n' U_{nt}$, where U_{nt} is defined as in (21) but with $\bar{\gamma}_n^j$ replacing $\hat{\gamma}_n^j$. Hence, from $E(U_{nt}) = \mu_{nt}$ it can be easily shown that $\sigma_n^2 = (1/n) \sum_{t=1}^n E(R_n' U_{nt})^2 + (2/n) \sum_{\tau=1}^n \sum_{t=1}^{n-1} E(R_n' U_{nt} R_n' U_{n,t-\tau}) + o(1)$, if $\sqrt{n} R_n' \bar{\mu}_n = o(1)$. The latter condition, however, is not sufficient to ensure the consistency of $\hat{\sigma}_n^2$ to σ_n^2 because the near-epoch dependence of $R_n' U_{nt}$ on X_t does not guarantee that the *raw* moment $E(R_n' U_{nt} R_n' U_{n,t-\tau})$ vanishes as τ increases, when $E(R_n' U_{nt}) \neq 0$. On the other hand, $E(R_n' U_{nt}) = 0$ for all n, t trivially implies conditions (i)–(ii).

Let $\delta_t^j(\omega, \gamma^j) = \gamma^j Y_t^{*2} - Y_t^* Y_{t-2j}^* = (\gamma^j Y_{t/2}^2 - Y_{t/2} Y_{t/2-j})\mathbb{I}(|t| \text{ even})$. Note that $E\{\delta_t^j(\omega, \bar{\gamma}_n^j)\} = 0$ because $\bar{\gamma}_n^j = \rho_j^o$ and $\{Y_t\}$ is stationary (see Section 3). Using an argument similar to that used for verifying Assumption 10 in Section 3, it is easy to see that Assumption 15(i)–(iv) hold, where $X_t = (Y_t, Y_t^*)'$. Moreover, let $A_{jn}^+ = 2/E(Y_t^2)$ and $\hat{A}_{jn}^+ = 2/\hat{E}(Y_t^2)$, where $\hat{E}(Y_t^2)$ is any consistent estimator of $E(Y_t^2)$, such as $\hat{E}(Y_t^2) = (1/n)\sum_{t=1}^n Y_t^2$ by any Law of Large Numbers for stationary ARMA or stationary and ergodic processes (see e.g. Hayashi (2000, p. 101)). It follows that Assumption 15 holds.

We now turn to conditions (i)–(ii). From Section 3 recall that $m_t^j(\omega, \gamma^j) = ((Y_t - \gamma^j Y_{t-j})^2, (Y_t^* - \gamma^j Y_{t-2j}^*)^2)'$. Combining this with the above definition of $\delta_t^j(\omega, \gamma^j)$, the definition of Y_t^*, and the weak stationarity of $\{Y_t\}$ gives $\mu_{nt}^{j\prime} = \gamma_Y(0)(1 - \rho_j^{o2})(1, \mathbb{I}(|t| \text{ even}), 0)$. Moreover, from the definitions of $M_{1n}^j(\omega, \gamma^j)$ and $M_{2n}^j(\omega, \gamma^j)$ given in Section 3, it is easy to see that $E\partial M_{kn}^j(\omega, \rho_j^o)/\partial\gamma^j = 0$ for $k = 1, 2$. Hence, $R_n^{j\prime} = (2, -2, 0)$ since $d_j(m, \gamma^j) = 2(m_1 - m_2)$. Therefore, $R_n'\mu_{nt} = R_n^{1\prime}\mu_{nt}^1 - R_n^{2\prime}\mu_{nt}^2 = 2\gamma_Y(0)(\rho_2^{o2} - \rho_1^{o2})(1 - \mathbb{I}(|t|\text{even}) = 0$ under $H_0^* = H_0$ (see (11)). It follows that conditions (i) and (ii) are satisfied under H_0^*.

Theorem 4 thus applies and delivers a consistent estimator $\hat{\sigma}_n^2 = \hat{R}_n'\hat{V}_n\hat{R}_n$ of σ_n^2. Specifically, simple algebra shows that $\hat{\sigma}_n^2$ is given by (20), where $\hat{U}_{nt}$ is replaced by the scalar $\hat{R}_n'\hat{U}_{nt} = \hat{R}_n^{1\prime}\hat{U}_{nt}^1 - \hat{R}_n^{2\prime}\hat{U}_{nt}^2$ with

$$\begin{aligned}\hat{R}_n^{j\prime}\hat{U}_{nt}^j &= 2\{(Y_t - \hat{\rho}_j Y_{t-j})^2 - (Y_t^* - \hat{\rho}_j Y_{t-2j}^*)^2\} \\ &\quad + \frac{8\sum_{t=n^*+1}^n (Y_t - \hat{\rho}_j Y_{t-j})Y_{t-j}}{\sum_{t=1}^n Y_t^2}(\hat{\rho}_j Y_t^* - Y_{t-2j}^*)Y_t^*\end{aligned}$$

and $Y_t^* = Y_{t/2}\mathbb{I}(|t| \text{ even})$. As a matter of fact, we can propose a simpler consistent estimator of σ_n^2 by exploiting the fact that $R_n'\mu_{nt} = 0$ under H_0. Namely, from the alternative expression for σ_n^2 given in footnote 14, we have $R_n'U_{nt} = R_n^{1\prime}U_{nt}^1 - R_n^{2\prime}U_{nt}^2$, where $R_n^{j\prime}U_{nt}^j = 2\{(Y_t - \rho_j^o Y_{t-j})^2 - (Y_t^* - \rho_j^o Y_{t-2j}^*)^2\}$. Hence, $\sum_{t=1}^n R_n^{j\prime}U_{nt}^j = 2\sum_{t=n^*+1}^n (Y_t - \rho_j^o Y_{t-j})^2$ using the definition of Y_t^*. Thus

$$\sigma_n^2 = \operatorname{Var}\frac{\sqrt{2}}{\sqrt{n^*}}\sum_{t=n^*+1}^n \{(Y_t - \rho_1^o Y_{t-1})^2 - (Y_t - \rho_2^o Y_{t-2})^2\}. \tag{24}$$

As expected from the remark after Theorem 2, the asymptotic variance σ_n^2 can be computed by neglecting the estimation uncertainty arising from $\hat{\rho}_j$, i.e. as if ρ_j^o is known (see (10)). Moreover, because $R_n'\mu_{nt} = 0$ under H_0, the expectation of the term in braces is zero under H_0. Hence from Newey and West (1987b) and Gallant and White (1988), a simpler consistent estimator $\tilde{\sigma}_n^2$ of σ_n^2 is given by four times the expression (20), where $\hat{U}_{nt}$ is replaced by the difference in squared prediction errors $\{(Y_t - \hat{\rho}_1 Y_{t-1})^2 - (Y_t - \hat{\rho}_2 Y_{t-2})^2\}$ with the first and last sums starting from $t = n^* + 1$ and $t = \tau + n^* + 1$.[15]

[15] Similarly, for the in-sample MSEP studied in Section 4, the estimator $\tilde{\sigma}_n^2$ appearing in (17) can be taken to be given by (20), where $\hat{U}_{nt}$ is replaced by the difference in squared prediction errors $\{(Y_t - \tilde{\gamma}_n^1 Y_{t-1})^2 - (Y_t - \tilde{\gamma}_n^2 Y_{t-2})^2\}$.

6. ON THE POSITIVE ASYMPTOTIC VARIANCE

It remains to discuss Assumption 7, namely $\liminf_n \sigma_n^2 > 0$. Similar assumptions appear in Vuong (1989) for likelihood-based criteria in the static case and West (1996) for out-of-sample MSEP-based criteria. The purpose of this section is to characterize situations for which this assumption is violated. More precisely, we consider cases when $\lim_n \sigma_n^2 = 0$. By considering subsequences of σ_n^2, our results can be modified to obtain necessary and sufficient conditions for $\liminf_n \sigma_n^2 > 0$. Hereafter, we maintain H_0^* since we are interested in the asymptotic distribution of our test statistic T_n under the null hypothesis. Moreover, we adopt the general framework of Section 2.

Our first result shows the importance of Assumption 7.

Lemma 1. *Given Assumptions* 1–6*, suppose that H_0^* holds.*

(i) Then $\sigma_n^2 = o(1)$ if and only if $\sqrt{n}\{Q_n^1(\omega, \hat{\gamma}_n^1) - Q_n^2(\omega, \hat{\gamma}_n^2)\} = o_P(1)$.
(ii) In addition, assume that $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial\gamma^{j\prime} = o(1)$ for $j = 1, 2$. Then $\sigma_n^2 = o(1)$ if and only if $\sqrt{n}\{Q_n^1(\omega, \bar{\gamma}_n^1) - Q_n^2(\omega, \bar{\gamma}_n^2)\} = o_P(1)$.

Lemma 1 extends Vuong (1989, Lemma 4.1) to dynamic situations and general model selection criteria. Part (i) shows that the $\sqrt{n}$-asymptotic normality of our test statistic hold only if $\sigma_n^2 \neq o(1)$. Part (ii) specializes to the case where estimation methods for both competing models (whether nested or nonnested) optimize their respective model selection criteria. As seen in Section 4, examples are ML estimation with log-likelihood-type criteria and GMM estimation with GMM criterion functions. The necessary and sufficient condition (ii) can then be interpreted as requiring that the estimated models are $\sqrt{n}$-asymptotically *identical*. This condition is, of course, satisfied when $Q_n^1(\omega, \bar{\gamma}_n^1) = Q_n^2(\omega, \bar{\gamma}_n^2)$ almost surely for n sufficiently large.

The last remark suggests that Assumption 7 is violated when the competing models are nested and estimated by optimizing the same selection criterion. This is confirmed by the next result. We define

Definition. *Model* 1 *is nested in model* 2 *according to the chosen selection criteria if there exists a sequence of continuously differentiable functions $\{h_n(\cdot); n = 1, 2, \ldots\}$ from Γ_1 to Γ_2 such that, for n sufficiently large, $Q_n^1(\omega, \gamma^1) = Q_n^2\{\omega, h_n(\gamma^1)\}$, for all $(\omega, \gamma^1) \in \Omega \times \Gamma_1$.*

In particular, this definition applies when the competing models are nested in the usual sense and the same criterion is used to compare these models. Note, however, that it is not sufficient to consider only nested models.

Theorem 5. *Given Assumptions* 1–6*, suppose that model* 1 *is nested in model* 2 *according to the selection criteria. Moreover, suppose that* (i) $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial\gamma^{j\prime} = o(1)$ *for* $j = 1, 2$, (ii) $\sqrt{n}\partial Q_n^2(\omega, \ddot{\gamma}_n^2)/\partial\gamma^{2\prime} = O_p(1)$ *for any nonstochastic sequence $\{\ddot{\gamma}_n^2\}$ such that $\ddot{\gamma}_n^2 - \bar{\gamma}_n^2 \to 0$, and* (iii) *$\bar{\gamma}_n^2$ is the identifiably unique minimizer of $\bar{Q}_n^2(\cdot)$ on Γ_2. Then $\sigma_n^2 = o(1)$ under H_0^*.*

As in Lemma 1(ii), we consider the case where estimators $\hat{\gamma}_n^j$ optimize the corresponding selection criterion so their limits $\bar{\gamma}_n^j$ satisfy condition (i). Conditions (ii) and (iii) are typically satisfied by extremum estimators. In particular, $\sqrt{n}\partial Q_n^2(\omega, \ddot{\gamma}_n^2)/\partial\gamma^{2\prime}$ is in general asymptotically normal under sequences of local alternatives $\ddot{\gamma}_n^2$ converging to $\bar{\gamma}_n^2$.

Theorem 5 shows that our model selection testing procedure based on T_n is not valid when the competing models are nested according to the chosen selection criteria. For instance, this is the case for nested regression models estimated by nonlinear least squares and compared using their in-sample or out-of-sample MSEPs. This result is in agreement with previous work in nested situations (see e.g. Hansen (1982), Andrews and Fair (1988), Gallant and White (1988)), which indicates that $n\{Q_n^1(\omega, \hat{\gamma}_n^1) - Q_n^2(\omega, \hat{\gamma}_n^2)\}$ is asymptotically chi-square distributed under the null hypothesis that the smaller model is correctly specified. More generally, when $\sigma_n^2 = o(1)$ Marcellino (2000) has shown that $n\{Q_n^1(\omega, \hat{\gamma}_n^1) - Q_n^2(\omega, \hat{\gamma}_n^2)\}$ follows a weighted chi-square distribution (see Vuong (1989, Definition 1)).

In view of the importance of Assumption 7, a natural question is whether one can test $\sigma_n^2 = o(1)$ using an extension of the variance test proposed in Vuong (1989) for the static likelihood case. When the selection criteria are of the likelihood type, i.e. $Q_n^j(\omega, \gamma^j) = (1/n)\sum_{t=1}^n m_t^j(\omega, \gamma^j)$, and $\hat{\gamma}^j$ minimizes $Q_n^j(\omega, \gamma^j)$, recent work by Golden (2000) for the strict stationary case and Marcellino (2000) for the case where $m_t^1(\omega, \bar{\gamma}_n^1) - m_t^2(\omega, \bar{\gamma}_n^2)$ is a martingale difference sequence has shown that $n\hat{\sigma}_n^2$ follows asymptotically a weighted chi-square distribution under the null hypothesis $\sigma_n^2 = o(1)$. This result will continue to hold in our general framework.

Example (Continued). When the out-of-sample MSEP is used as a model selection criterion, we have $Q_n^j(\omega, \gamma^j) = (1/n^*)\sum_{t=n^*+1}^n (Y_t - \gamma^j Y_{t-j})^2$. Because Assumptions 1–6 holds by the verification of Assumptions 1–3 and 9–11 in Section 3, Lemma 1 applies. Hence, from $\hat{\gamma}^j = \hat{\rho}_j$ and Lemma 1(i), $\sigma_n^2 = o(1)$ if and only if $(\sqrt{2}/\sqrt{n^*})\sum_{t=n^*+1}^n \{(Y_t - \hat{\rho}_1 Y_{t-1})^2 - (Y_t - \hat{\rho}_2 Y_{t-2})^2\} = o_P(1)$ under $H_0^* = H_0$. Moreover, because $\hat{\rho}_j$ is consistent for $\bar{\gamma}_n^j = \rho_j^o$ and $\bar{Q}_n^j(\gamma^j) = \gamma_Y(0)(1 + \gamma^{j2}) - 2\gamma_Y(j)\gamma^j$, as shown is Section 3, then $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \gamma^{j\prime} = 0$, i.e. $\hat{\rho}_j$ minimizes asymptotically the out-of-sample MSEP. Hence, by Lemma 1(ii), $\sigma_n^2 = o(1)$ if and only if $(\sqrt{2}/\sqrt{n^*})\sum_{t=n^*+1}^n \{(Y_t - \rho_1^o Y_{t-1})^2 - (Y_t - \rho_2^o Y_{t-2})^2\} = o_P(1)$ under H_0. These results agree with the direct computation of the asymptotic variance σ_n^2 given in (24).

As indicated above, one might want to test $\sigma_n^2 = o(1)$, i.e. that the so-called *long-run* variance of the stationary process $\{d_t\} \equiv \{(Y_t - \rho_1^o Y_{t-1})^2 - (Y_t - \rho_2^o Y_{t-2})^2\}$ is zero. This is obviously the case when the first and second autocorrelations ρ_1^o and ρ_2^o of the stationary process $\{Y_t\}$ are zero, though $\sigma_n^2 = o(1)$ may hold for other DGPs under $H_0^* = H_0$ given in (11). From Section 5, $\tilde{\sigma}_n^2$ can be used to test $\sigma_n^2 = o(1)$ as the former is a consistent estimator of the latter, which is equal to $2\sum_{\tau=-\infty}^{+\infty} \gamma_d(\tau)$ when the autocovariances $\{\gamma_d(\tau)\}$ of the process $\{d_t\}$ are summable (see e.g. Hayashi (2000, p. 401)). In particular, recent results obtained by Golden (2000) and Marcellino (2000) under some additional assumptions indicate that $n\tilde{\sigma}_n^2$ follows asymptotically a weighted chi-square distribution under $\sigma_n^2 = o(1)$ with weights that can be estimated consistently.[16]

7. CONCLUSION

This paper offers a general testing framework for assessing the statistical significance of the difference in model selection criterion values for two competing models under weak assumptions on the data generating process. Such a testing framework encompasses the static likelihood-based situations studied in Vuong (1989) as well as the out-of-sample prediction-based criteria considered in West (1994, 1996). The competing models must be essentially nonnested but can be dynamic and incompletely specified. Our results allow for a wide class of $\sqrt{n}$-asymptotically

[16] Similar results hold when using the in-sample MSEP for choosing between the two competing AR models.

normal estimators and model selection criteria. Moreover, the methods used to estimate the competing models need not optimize the selection criteria used for model selection. Thus different samples can be used for model estimation and comparison. Situations where sampling uncertainty due to parameter estimation is asymptotically irrelevant for testing model equivalence are stressed. In particular, this is the case when the employed estimators optimize (possibly asymptotically) the model selection criteria used for model comparison.

To conclude, we make three remarks. First, our testing framework allows for the comparison of two competing models only. This is restrictive in practice. Extension to more than two models raises a problem of multiple comparison. Recently, Shimodaira (1998) has extended Vuong (1989) setting to multiple competing models through the use of confidence intervals, and White (2000) has extended West's (1994, 1996) prediction framework while establishing the validity of bootstrapping in such a multiple testing situation. Second, our results do not apply to the comparison of nonparametric models. Extension of our testing framework to nonparametric situations is possible as shown by Lavergne and Vuong (1996, 2000) for nonparametric regressions. For a survey of selection of regressors in parametric and nonparametric regressions, see Lavergne (1998).

Third, we have not attempted to address the choice of model selection criteria. Indeed no single index may be universally superior as each reflects the particular features of interest to a researcher. As Amemiya (1980) wrote '... all of the criteria considered are based on a somewhat arbitrary assumption which cannot be fully justified, and that by slightly varying the loss function and the decision strategy one can indefinitely go on inventing new criteria'. Recently, however, Granger and Pesaran (2000) have argued for a closer link between forecast evaluation and decision theory. Moreover, though our framework allows for a large class of criteria, one should be cautious in the choice of such criteria as systematic applications of our results may lead to nonsensical outcomes. This can be the case if different criteria are used across competing models.

ACKNOWLEDGEMENTS

We thank E. Guerre, P. Lavergne, A. Monfort, two referees and the Editor as well as seminar participants at the Université de Montréal, Stanford University, Université de Toulouse, Université d'Aix–Marseille, the World Congress of the Econometric Society, Barcelona, August 1990 and Malinvaud econometric seminar, Paris, November 1990. The second author is grateful to J. M. Dufour and E. Ghysels for a visit at the Université de Montréal in April 1990, which led to an early version (Rivers and Vuong, 1991) containing the basic results of the current paper. Financial support from the National Science Foundation under Grant SBR-9631212 is gratefully acknowledged.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the likelihood ratio principle. In B. N. Petrov and F. Csaki (eds), *Proceedings of the Second International Symposium of Information Theory*. pp. 257–81. Budapest: Akademiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions and Automatic Control AC-19*, 716–23.

Amemiya, T. (1980). Selection of regressors. *International Economic Review 21*, 331–54.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica 55*, 1465–71.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent matrix estimation. *Econometrica 59*, 817–58.

Andrews, D. W. K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica 62*, 43–72.

Andrews, D. W. K. and R. C. Fair (1988). Inferences in econometric models with structural change. *Review of Economic Studies 55*, 615–40.

Bates, C. and H. White (1985). A unified theory of consistent estimation for parametric models. *Econometric Theory 1*, 151–78.

Caballero, R. J. and E. M. R. A. Engel (1999). Explaining investment dynamics in U.S. manufacturing: a generalized (S, s) approach. *Econometrica 67*, 783–826.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 105–23.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B 24*, 406–24.

Davidson, R. and J. G. MacKinnnon (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica 49*, 781–93.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–63.

Domowitz, I. and H. White (1982). Misspecified models with dependent observations. *Journal of Econometrics 20*, 35–58.

Ericsson, N. R. (1983). Asymptotic properties of instrumental variables statistics for testing non-nested hypotheses. *Review of Economic Studies 50*, 287–304.

Fair, R. C. and R. J. Shiller (1990). Comparing information in forecasts from econometric models. *American Economic Review 80*, 375–89.

Findley, D. F. (1990). Making Difficult Model Comparisons, mimeo, U.S. Bureau of the Census.

Findley, D. F. (1991). Convergence of finite multistep predictors from incorrect models and its role in model selection. *Note di Matematica XI*, 145–55.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B. C. Chen (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics 16*, 127–77.

Findley, D. F. and C. Z. Wei (1993). Moment bound for deriving time series CLT's and model selection procedures. *Statistica Sinica 3*, 453–80.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: Wiley.

Gallant, A. R. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. New York: Basil Blackwell.

Gasmi, F., J. J. Laffont, and Q. Vuong (1992). Econometric analysis of collusive behavior in a soft drink industry. *Journal of Economics and Management Strategy 1*, 277–311.

Ghysels, E. and A. Hall (1990). Testing nonnested Euler conditions with quadrature-based methods of approximation. *Journal of Econometrics 46*, 273–308.

Godfrey, L. G. (1983). Testing non-nested models after estimation by instrumental variables or least squares. *Econometrica 51*, 355–65.

Golden, M. (2000). Discrepancy Risk Model Selection Test Theory for Comparing Possibly Misspecified or Nonnested Models, mimeo, University of Texas, Dallas.

Gourieroux, C., A. Monfort and A. Trognon (1983). Testing nested or non-nested hypotheses. *Journal of Econometrics 21*, 83–115.

Granger, C. W. J. and M. H. Pesaran (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting 19*, 537–60.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.

Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B 41*, 190–5.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–54.

Hayashi, F. (2000). *Econometrics*. Princeton: Princeton University Press.

Heckman, J. (1981). Heterogeneity and state dependence. In S. Rosen (ed.), *Studies in Labor Markets*. pp. 91–139. Chicago: University of Chicago Press.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Ibragimov, I. and Y. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics 40*, 633–43.

King, M. L. and M. McAleer (1987). Further results on testing AR(1) against MA(1) disturbances in the linear regression model. *Review of Economic Studies 54*, 649–63.

Konishi, S. and G. Kitagawa (1996). Generalized information criterion in model selection. *Biometrika 83*, 875–90.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics 22*, 79–86.

Lavergne, P. (1998). Selection of regressors in econometrics: parametric and nonparametric methods. *Econometric Reviews 17*, 227–73.

Lavergne, P. and Q. Vuong (1996). Nonparametric selection of regressors: the nonnested case. *Econometrica 64*, 207–19.

Lavergne, P. and Q. Vuong (2000). Nonparametric significance testing. *Econometric Theory 16*, 576–601.

Lien, D. and Q. Vuong (1987). Selecting the best linear regression model: a classical approach. *Journal of Econometrics, Annals 35*, 3–23.

Linhart, H. (1988). A test whether two AIC's differ significantly. *South African Statistical Journal 22*, 153–61.

Linhart, H. and W. Zucchini (1986). *Model Selection*. New York: Wiley.

Machado, J. A. F. (1993). Robust model selection and M-estimation. *Econometric Theory 9*, 478–93.

MacKinnon, J. G. (1983). Model specification tests against non-nested alternatives. *Econometric Reviews 2*, 85–110.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics 15*, 661–76.

Marcellino, M. (2000). Model Selection for Non-linear Dynamic Models, mimeo, Universita Bocconi.

Martin, R. D. (1980). Robust estimation of autoregressive models. In D. R. Brillinger and G. C. Tiao (eds), *Directions in Time Series*. pp. 228–62. Hayward: Institute of Mathematical Statistics.

McAleer, M. (1987). Specification tests for separate models: a survey. In M. L. King and D. E. A. Giles (eds), *Specification Analysis in the Linear Model*. pp. 146–95. London: Routledge and Kegan Paul.

Meese, R. A. and K. Rogoff (1983). Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics 14*, 3–24.

Mizon, G. E. and J. F. Richard (1986). The encompassing principle and its applications to testing non-nested hypotheses. *Econometrica 54*, 657–78.

Moore, D. (1978). Chi-square tests. In R. V. Hogg (ed.), *Studies in Statistics*, vol. 19, The Mathematical Association of America.
Nakamura, A. O., M. Nakamura, and H. Orcutt–Duleep (1990). Alternative approaches to model choice. *Journal of Economic Behavior and Organization 14*, 97–125.
Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (eds), *Handbook of Econometrics*, vol. 4, pp. 2111–245. Amsterdam: North Holland.
Newey, W. K. and K. D. West (1987a). Hypothesis testing with efficient method of moments estimators. *International Economic Review 28*, 777–87.
Newey, W. K. and K. D. West (1987b). A simple positive semi-definite heteroskedasticty and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–8.
Paarsch, H. J. (1997). Deriving an estimate of the optimal reserve price: an application to British Columbian timber sales. *Journal of Econometrics 78*, 333–57.
Pesaran, M. H. (1974). On the general problem of model selection. *Review of Economic Studies 41*, 153–71.
Pesaran, M. H. and R. J. Smith (1994). A generalized R^2 criterion for regression models estimated by the instrumental variables method. *Econometrica 62*, 705–10.
Pham, T. and L. Tran (1980). The Strong Mixing Properties of the Autoregressive Moving Average Time Series Models, Seminaire de Statistique, Grenoble.
Powell, J. L. (1994). Estimation of semiparametric models. In R. F. Engle and D. McFadden (eds), *Handbook of Econometrics*, vol. 4, pp. 2443–521. Amsterdam: North-Holland.
Rivers, D. and Q. Vuong (1991). Model selection tests for nonlinear dynamic models, Working Paper 9108, INRA-ESR, Toulouse.
Ronchetti, E. (1985). Robust model selection in regression. *Statistics and Probability Letters 3*, 21–3.
Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*, 393–415.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461–4.
Shimodaira, H. (1998). An application of multiple comparison techniques to model selection. *Annals of the Institute of Statistical Mathematics 50*, 1–15.
Sin, C. Y. and H. White (1996). Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics 71*, 207–25.
Smith, R. J. (1992). Non-nested tests for competing models estimated by generalized method of moments. *Econometrica 60*, 973–80.
Torro-Vizcarrondo, C. and T. D. Wallace (1968). A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association 63*, 558–72.
Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 307–33.
Vuong, Q. and W. Wang (1991). Tests for model selection using power divergency statistics, Working Paper 9106, INRA–ESR, Toulouse.
Vuong, Q. and W. Wang (1993a). Minimum chi-square estimation and tests for model selection. *Journal of Econometrics 56*, 141–68.
Vuong, Q. and W. Wang (1993b). Selecting estimates using chi-square statistics. *Annales d'Economie et de Statistique 30*, 143–64.
Walker, A. M. (1967). Some tests of separate families of hypotheses in time series analysis. *Biometrika 54*, 39–68.
West, K. D. (1994). Asymptotic Inference about predictive ability, mimeo, University of Wisconsin.
West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*, 1067–84.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.
White, H. (1984). *Asymptotic Theory for Econometricians*. New York: Academic.
White, H. (2000). A reality check for data snooping. *Econometrica 68*, 1097–126.
Wolak, F. (1994). An econometric analysis of the asymmetric information regulator utility interaction. *Annales d'Economie et de Statistique 34*, 13–69.

APPENDIX

Proof of Theorem 1. A Taylor expansion of $Q_n^j(\omega, \hat{\gamma}_n^j)$ around $\bar{\gamma}_n^j$ gives

$$Q_n^j(\omega, \hat{\gamma}_n^j) = Q_n^j(\omega, \bar{\gamma}_n^j) + \frac{\partial Q_n^j(\omega, \ddot{\gamma}_n^j)}{\partial \gamma^{j\prime}}(\hat{\gamma}_n^j - \bar{\gamma}_n^j),$$

where $\ddot{\gamma}_n^j$ is in the line segment $[\hat{\gamma}_n^j, \bar{\gamma}_n^j]$ for $j = 1, 2$. Multiplying by $\sqrt{n}$ and adding and subtracting a term, we obtain

$$\begin{aligned}\sqrt{n} Q_n^j(\omega, \hat{\gamma}_n^j) = \sqrt{n} Q_n^j(\omega, \bar{\gamma}_n^j) &+ \frac{\partial \bar{Q}_n^j(\bar{\gamma}_n^j)}{\partial \gamma^{j\prime}}\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) \\ &+ \left\{\frac{\partial Q_n^j(\omega, \ddot{\gamma}_n^j)}{\partial \gamma^{j\prime}} - \frac{\partial \bar{Q}_n^j(\bar{\gamma}_n^j)}{\partial \gamma^{j\prime}}\right\}\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j).\end{aligned}$$

Now, $\ddot{\gamma}_n^j - \bar{\gamma}_n^j \overset{as}{\to} 0$ because $\hat{\gamma}_n^j - \bar{\gamma}_n^j \overset{as}{\to} 0$ by Assumption 3. Thus, given Assumptions 1, 2 and 5, it follows from Domowitz and White (1982, Theorem 2.3) that $\partial Q_n^j(\omega, \ddot{\gamma}_n^j)/\partial \gamma^{j\prime} - \partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \gamma^{j\prime} \overset{as}{\to} 0$. Moreover, by Assumption 6, $\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) = O_P(1)$. Hence we obtain

$$\sqrt{n} Q_n^j(\omega, \hat{\gamma}_n^j) = \sqrt{n} Q_n^j(\omega, \bar{\gamma}_n^j) + \frac{\partial \bar{Q}_n^j(\bar{\gamma}_n^j)}{\partial \gamma^{j\prime}}\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) + o_P(1). \tag{A.1}$$

Subtracting $\sqrt{n}\bar{Q}_n^j(\bar{\gamma}_n^j)$ from both sides, and then subtracting the resulting equations for $j = 1, 2$ from each other, we obtain in matrix notation

$$\begin{aligned}\sqrt{n}\{\Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) - \Delta \bar{Q}_n(\bar{\gamma}_n^1, \bar{\gamma}_n^2)\} &= L_n' \sqrt{n} U_n + o_P(1) \\ &= L_n' C_n' \sqrt{n} Z_n + o_P(1),\end{aligned} \tag{A.2}$$

where the first equality follows from (3) and (4), and the second equality uses Assumption 6 and the boundedness of L_n, which is implied by Assumption 5.

Now we apply White (1984, Corollary 4.24) with $b_n = \sqrt{n} Z_n$, $V_n = I_s$ and $A_n = C_n L_n$. Since $L_n = O_P(1)$ and $C_n = O_P(1)$ by Assumptions 5–6, then $A_n = O_P(1)$. Moreover, A_n is of (full column) rank one for all n sufficiently large by Assumption 7. Thus, if $\sigma_n^2 = L_n' C_n' C_n L_n$, then $\sigma_n^2 = O(1)$ and $\sigma_n^{-1} L_n' C_n' \sqrt{n} Z_n \Rightarrow N(0, 1)$. Since $\sigma_n^{-1} = O(1)$ by Assumption 7, multiplying (A.2) by σ_n^{-1} gives

$$\frac{\sqrt{n}}{\sigma_n}\{\Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) - \Delta \bar{Q}_n(\bar{\gamma}_n^1, \bar{\gamma}_n^2)\} \Rightarrow N(0, 1). \tag{A.3}$$

Finally, we have

$$\begin{aligned}\hat{\sigma}_n^{-1}\sqrt{n}\Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) = \frac{\hat{\sigma}_n^{-1}}{\sigma_n^{-1}}[&\sigma_n^{-1}\sqrt{n}\{\Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) - \Delta \bar{Q}_n(\bar{\gamma}_n^1, \bar{\gamma}_n^2)\} \\ &+ \sigma_n^{-1}\sqrt{n}\Delta \bar{Q}_n(\bar{\gamma}_n^1, \bar{\gamma}_n^2)].\end{aligned}$$

Since $\sigma_n = O(1)$ and $\sigma_n^{-1} = O(1)$, it follows from Assumption 8 that $\hat{\sigma}_n^{-1}/\sigma_n^{-1} \overset{P}{\to} 1$. Moreover, because $1/M < \sigma_n^{-1} < M$ for n sufficiently large and some $M > 0$, statements (i)–(iii) follow immediately from (5) and (A.3).

Proof of Theorem 2. We shall show that the conditions of Theorem 2 imply those of Theorem 1. It suffices to show that Assumptions 4–6 are implied by the conditions of Theorem 2. We then shall show that $\sigma_n^2 = R_n' V_n R_n$ as required.

Step 1: Verification of Assumptions 4 *and* 5. Consider the quantities $M_n^j(\omega, \gamma^j)$, $\partial M_n^j(\omega, \gamma^j)/\partial \gamma^{j\prime}$ and their expectations. From Assumptions 1, 2, 9 and 11, it follows from Gallant and White (1988, Theorem 3.18) that

$$M_n^j(\omega, \gamma^j) - E M_n^j(\omega, \gamma^j) \xrightarrow{as} 0 \quad \text{uniformly on } \Gamma_j, \tag{A.4}$$

$$\frac{\partial M_n^j(\omega, \gamma^j)}{\partial \gamma^{j\prime}} - E \frac{\partial M_n^j(\omega, \gamma^j)}{\partial \gamma^{j\prime}} \xrightarrow{as} 0 \quad \text{uniformly on } \Gamma_j, \tag{A.5}$$

where $E M_n^j(\omega, \gamma^j)$ and $E \partial M_n^j(\omega, \gamma^j)/\partial \gamma^{j\prime}$ are continuous on Γ_j uniformly in n, i.e. equicontinuous. Moreover, by Assumption 11(iii), the latter two functions are uniformly bounded, i.e. there exists an M finite such that $E \mid M_n^j(\omega, \gamma^j) \mid < M$ and $E \mid \partial M_n^j(\omega, \gamma^j)/\partial \gamma^{j\prime} \mid < M$ for all γ^j and all n. Now define the nonstochastic function $\bar{Q}_n^j(\gamma^j) = d_j\{E M_n^j(\omega, \gamma^j), \gamma^j\}$ from $\Gamma_j \to \mathbb{R}$ for $j = 1, 2$ where $E M_n^j(\omega, \gamma^j) = (1/n) \sum_{t=1}^n E m_t^j(\omega, \gamma^j)$. Because $d_j(\cdot)$ is continuous and $E M_n^j(\omega, \gamma^j)$ is uniformly bounded and equicontinuous, then $\bar{Q}_n^j(\cdot)$ is equicontinuous. Moreover, because Γ_j is compact, (A.4) implies that $Q_n^j(\omega, \gamma^j) - \bar{Q}_n^j(\omega, \gamma^j) \xrightarrow{as} 0$ uniformly on Γ_j. This completes the verification of Assumption 4.

Turning to Assumption 5, from the LDC theorem we have $\partial E m_t^j(\omega, \gamma^j)/\partial \gamma^{j\prime} = E \partial m_t^j(\omega, \gamma^j)/\partial \gamma^{j\prime}$, for $t = 1, 2, \ldots$. Thus $\bar{Q}_n^j(\cdot)$ is continuously differentiable on Γ_j, and we have by the chain rule

$$\frac{\partial \bar{Q}_n^j(\gamma^j)}{\partial \gamma^{j\prime}} = \frac{\partial d_j\{E M_n^j(\omega, \gamma^j), \gamma^j\}}{\partial m^{j\prime}} \cdot E \frac{\partial M_n^j(\omega, \gamma^j)}{\partial \gamma^{j\prime}} + \frac{\partial d_j\{E M_n^j(\omega, \gamma^j), \gamma_j\}}{\partial \gamma^{j\prime}}. \tag{A.6}$$

In fact, $\partial \bar{Q}_n^j(\cdot)/\partial \gamma^{j\prime}$ is equicontinuous because each term of (A.6) is equicontinuous. The latter statement follows from the continuous differentiability of $d_j(\cdot)$ and the uniform boundedness and equicontinuity of $E M_n^j(\omega, \gamma^j)$ and $E \partial M_n^j(\omega, \gamma^j)/\partial \gamma^{j\prime}$. In particular, $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \gamma^{j\prime}$ is bounded as required. Lastly, to prove uniform convergence, note that

$$\frac{\partial Q_n^j(\omega, \gamma^j)}{\partial \gamma^{j\prime}} = \frac{\partial d_j\{M_n^j(\omega, \gamma^j), \gamma^j\}}{\partial m^{j\prime}} \cdot \frac{\partial M_n^j(\omega, \gamma^j)}{\partial \gamma^{j\prime}} + \frac{\partial d_j\{M_n^j(\omega, \gamma^j), \gamma^j\}}{\partial \gamma^{j\prime}}. \tag{A.7}$$

Since $d_j(\cdot)$ is continuously differentiable and Γ_j is compact, (A.4) and (A.5) imply

$$\frac{\partial d_j\{M_n^j(\omega, \gamma^j), \gamma^j\}}{\partial m^{j\prime}} - \frac{d_j\{E M_n^j(\omega, \gamma^j), \gamma^j\}}{\partial m^{j\prime}} \xrightarrow{as} 0 \quad \text{uniformly on } \Gamma_j,$$

$$\frac{\partial d_j\{M_n^j(\omega, \gamma^j), \gamma^j\}}{\partial \gamma^{j\prime}} - \frac{d_j\{E M_n^j(\omega, \gamma^j), \gamma^j\}}{\partial \gamma^{j\prime}} \xrightarrow{as} 0 \quad \text{uniformly on } \Gamma_j.$$

Thus, because $d_j\{E M_n^j(\omega, \gamma^j), \gamma^j\}/\partial m^{j\prime}$ and $E \partial M_n^j(\omega, \gamma^j)/\partial \gamma^{j\prime}$ are both $O(1)$, we obtain from (A.4)–(A.7) that $\partial Q_n^j(\omega, \gamma^j)/\partial \gamma^{j\prime} - \partial \bar{Q}_n^j(\gamma^j)/\partial \gamma^{j\prime} \xrightarrow{as} 0$ uniformly on Γ_j. This completes the verification of Assumption 5.

Step 2: Verification of Assumption 6. A Taylor expansion gives

$$\begin{aligned} Q_n^j(\omega, \bar{\gamma}_n^j) - \bar{Q}_n^j(\bar{\gamma}_n^j) &= d_j\{M_n^j(\omega, \bar{\gamma}_n^j), \bar{\gamma}_n^j\} - d_j\{EM_n^j(\omega, \bar{\gamma}_n^j), \bar{\gamma}_n^j\} \\ &= \frac{\partial d_j(\ddot{m}_n^j, \bar{\gamma}_n^j)}{\partial m^{j\prime}}\{M_n^j(\omega, \bar{\gamma}_n^j) - EM_n^j(\omega, \bar{\gamma}_n^j)\} \end{aligned} \tag{A.8}$$

where $\ddot{m}_n^j$ is in the line segment $[M_n^j(\omega, \bar{\gamma}_n^j), EM_n^j(\omega, \bar{\gamma}_n^j)]$. Now $M_n^j(\omega, \bar{\gamma}_n^j) - EM_n^j(\omega, \bar{\gamma}_n^j) \stackrel{as}{\to} 0$ because of (A.4), Assumption 3 and Domowitz and White (1982, Theorem 2.3). Hence $\ddot{m}_n^j - EM_n^j(\omega, \bar{\gamma}_n^j) \stackrel{as}{\to} 0$. Since $EM_n^j(\omega, \bar{\gamma}_n^j) = O(1)$ and Γ_j is compact, then $\partial d_j(\ddot{m}_n^j, \bar{\gamma}_n^j)/\partial m^{j\prime} - \partial d_j\{EM_n^j(\omega, \bar{\gamma}_n^j), \bar{\gamma}_n^j\}/\partial m^{j\prime} \stackrel{as}{\to} 0$. Thus, *provided* $\sqrt{n}\{M_n^j(\omega, \bar{\gamma}_n^j) - EM_n^j(\omega, \bar{\gamma}_n^j)\} = O_P(1)$, which is proved later, (A.8) becomes

$$\sqrt{n}\{Q_n^j(\omega, \bar{\gamma}_n^j) - \bar{Q}_n^j(\bar{\gamma}_n^j)\} = \frac{\partial \bar{d}_{jn}}{\partial m^{j\prime}}\sqrt{n}\{M_n^j(\omega, \bar{\gamma}_n^j) - EM_n^j(\omega, \bar{\gamma}_n^j)\} + o_P(1).$$

Hence, stacking up and using Assumption 10, we obtain

$$\sqrt{n}U_n = \begin{pmatrix} \partial\bar{d}_{1n}/\partial m^{1\prime} & 0 & 0 & 0 \\ 0 & -A_{1n}^+ & 0 & 0 \\ 0 & 0 & \partial\bar{d}_{2n}/\partial m^{2\prime} & 0 \\ 0 & 0 & 0 & -A_{2n}^+ \end{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^n \begin{pmatrix} \tilde{m}_{nt}^1 \\ Y_{nt}^1 \\ \tilde{m}_{nt}^2 \\ Y_{nt}^2 \end{pmatrix} + o_P(1), \tag{A.9}$$

where U_n is defined in Assumption 6, and $\tilde{m}_{nt}^j = m_t^j(\omega, \bar{\gamma}_n^j) - Em_t^j(\omega, \bar{\gamma}_n^j)$.

We shall show that $\sqrt{n}\, U_n$ satisfies Assumption 6. To do so, define

$$W_{jn} = \operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \tilde{m}_{nt}^j\right) \qquad \text{and} \qquad B_{jn} = \operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n Y_{nt}^j\right).$$

First, by Assumptions 10 and 11 with $r > 2$, we have $W_{jn} < \infty$ and $B_{jn} < \infty$. Moreover, $W_{jn} = O(1)$ and $B_{jn} = O(1)$. The proof of the latter follows the proof that $B_n^o = O(1)$ in Gallant and White (1988, pp. 86–87). Specifically, define $Z_{nt} = \lambda'\tilde{m}_{nt}^j$ (or $Z_{nt} = \lambda' Y_{nt}^j$) where $\lambda \in \mathbb{R}^{q_j}$ (or $\lambda \in \mathbb{R}^{k_j}$), $\lambda'\lambda = 1$ so that $\operatorname{Var}(1/\sqrt{n})\sum_{t=1}^n Z_{nt} = \lambda' W_{jn}\lambda$ (or $\lambda' B_{jn}\lambda$). Now, since $EZ_{nt} = 0$,

$$\operatorname{Var}\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{nt} = \frac{1}{n}E\left\{\left(\sum_{t=1}^n Z_{nt}\right)^2\right\} \le \frac{1}{n}E\left\{\max_{1\le j\le n}\left(\sum_{t=1}^j Z_{nt}\right)^2\right\}.$$

But, given Assumptions 1, 9, 11, (or 10), it follows from Gallant and White (1988, Lemma 3.14) that Z_{nt} is a mixingale of size -1 and *a fortiori* of size $-1/2$ with $c_{nt} \le \Delta \le \infty$ for all n, t. Therefore, applying McLeish's inequality (Gallant and White (1988, Theorem 3.11)), we have $\operatorname{Var}(1/\sqrt{n})\sum_{t=1}^n Z_{nt} \le (1/n)(K\Delta^2 n) = K\Delta^2 < \infty$. Thus, $\lambda' W_{jn}\lambda$ (or $\lambda' B_{jn}\lambda$) is $O(1)$ for arbitrary $\lambda \in \mathbb{R}^{q_j}$ or ($\lambda \in \mathbb{R}^{k_j}$), $\lambda'\lambda = 1$, implying that $W_{jn} = O(1)$ (or $B_{jn} = O(1)$).

Next, define $V_{nt} = (V_{nt}^{1\prime}, V_{nt}^{2\prime})'$ where $V_{nt}^j = (\tilde{m}_{nt}^{j\prime}, Y_{nt}^{j\prime})'$ so that V_n is the covariance matrix of $(1/\sqrt{n})\sum_{t=1}^n V_{nt}$ by (8). Thus, by the Cauchy–Schwartz inequality, it follows from the preceding properties of W_{jn} and B_{jn} that $V_n < \infty$ and $V_n = O(1)$. Moreover, by assumption, V_n is uniformly of rank $s > 0$. Hence, for every $n = 1, 2, \ldots$, there exists a $(k+q) \times s$ matrix P_n that is uniformly of full column rank such that $V_n = P_n P_n'$. Since $V_n = O(1)$, then $P_n = O(1)$ also. Let $Z_n = (P_n' \ P_n)^{-1} P_n'(1/n) \sum_{t=1}^n V_{nt}$ so that, almost surely,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_{nt} = P_n \sqrt{n} Z_n. \tag{A.10}$$

We shall show that $\sqrt{n} Z_n \Rightarrow N(0, I_s)$. The proof is similar to the proof that $(1/\sqrt{n}) \sum_{t=1}^n B_n^{o-1/2} M_{nt}^o \Rightarrow N(0, I)$ in Gallant and White (1988, p. 87).

Specifically, define $Z_{nt} = \lambda'(P_n' \ P_n)^{-1} P_n' V_{nt}$. where $\lambda \in \mathbb{R}^s$, $\lambda' \lambda = 1$. Thus, $E Z_{nt} = 0$. By Assumptions 10 and 11, we have $\|V_{nt}^j\|_r \leq \Delta < \infty, \ \ r > 2$. Moreover, because P_n is $O(1)$ and P_n is uniformly of full-column rank, then $(P_n' \ P_n)^{-1} = O(1)$ and $(P_n' P_n)^{-1} P_n' = O(1)$. Thus $\|Z_{nt}\|_r \leq \Delta'$ for $r > 2$. In addition, by Assumptions 10 and 11, $\{Z_{nt}\}$ is near-epoch dependent on $\{X_t\}$ of size -1 where $\{X_t\}$ is mixing with ϕ_m of size $-r/(r-1)$ or α_m of size $-2r/(r-2)$ by Assumption 9. Define

$$\begin{aligned} v_n^2 &= \operatorname{Var}\left(\sum_{t=1}^n Z_{nt}\right) \\ &= \lambda'(P_n' \ P_n)^{-1} P_n' \left(\operatorname{Var} \sum_{t=1}^n V_{nt} \right) P_n (P_n' \ P_n)^{-1} \lambda \\ &= n\lambda'(P_n' \ P_n)^{-1} P_n' \ V_n \ P_n (P_n' P_n)^{-1} \lambda \\ &= n\lambda'\lambda = n \end{aligned}$$

so that $v_n^{-2} = O(n^{-1})$. Hence, by Gallant and White (1988, Theorem 5.3),

$$v_n^{-1} \sum_{t=1}^n Z_{nt} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \lambda'(P_n' P_n)^{-1} P_n' V_{nt} \Rightarrow N(0, 1).$$

Thus, $\sqrt{n} Z_n = (1/\sqrt{n})(P_n' P_n)^{-1} P_n' \sum_{t=1}^n V_{nt} \Rightarrow N(0, I_s)$ by the Cramer–Wold device. In particular, because $P_n = O(1)$, (A.10) implies that $(1/\sqrt{n}) \sum_{t=1}^n \tilde{m}_{nt} = O_P(1)$, a condition which was needed to obtain (A.9).

Combining (A.9) and (A.10), we obtain $\sqrt{n} U_n = C_n' \sqrt{n} Z_n + o_P(1)$, where

$$C_n' = \begin{pmatrix} \partial \bar{d}_{1n}/\partial m^{1\prime} & 0 & 0 & 0 \\ 0 & -A_{1n}^+ & 0 & 0 \\ 0 & 0 & \partial \bar{d}_{2n}/\partial m^{2\prime} & 0 \\ 0 & 0 & 0 & -A_{2n}^+ \end{pmatrix} P_n.$$

Note that $C_n = O(1)$ because of Assumptions 10 and 11, $E M_n^j(\omega, \bar{\gamma}_n^j) = O(1)$, $\bar{\gamma}_n^j \in \Gamma_j$ compact, and $P_n = O(1)$. Thus Assumption 6 is verified.

Step 3: Computation of σ_n^2. We have

$$L_n' \ C_n' = \left(\frac{\partial \bar{d}_{1n}}{\partial m^{1\prime}}, -\frac{\partial \bar{Q}_n^1(\bar{\gamma}_n^1)}{\partial \gamma^{1\prime}} A_{1n}^+, -\frac{\partial \bar{d}_{2n}}{\partial m^{2\prime}}, \frac{\partial \bar{Q}_n^2(\bar{\gamma}_n^2)}{\partial \gamma^{2\prime}} A_{2n}^+ \right) P_n.$$

Using $P_n P_n' = V_n$ and formula (A.6) for $\partial \bar{Q}_n^j(\cdot)/\partial \gamma^{j\prime}$, we obtain the desired result from $\sigma_n^2 = L_n' C_n' C_n L_n$.

Proof of Theorem 3. We shall show that the conditions of Theorem 3 imply those of Theorem 2. Assumptions 2 and 11 are trivially implied by Assumption 12. Thus, we need to prove that Assumptions 3 and 10 are implied by the conditions of Theorem 3. The desired result will follow by application of Theorem 2.

Step 1: Verification of Assumption 3. Let $\hat{\gamma}_n^j = (\hat{\theta}_n^{j\prime}, \hat{\tau}_n^{j\prime})'$. The existence and measurability of $\hat{\theta}_n^j$ follows from Gallant and White (1988, Lemma 2.1) or Jennrich (1969) because $Q_n^j(\omega, \theta^j, \hat{\tau}_n^j(\omega))$ is measurable-$\mathcal{F}/\mathcal{B}$ for every $\theta^j \in \Theta_j$ and is continuous in θ^j for almost all ω by Assumption 12(i).

Define $\bar{\gamma}_n^j = (\bar{\theta}_n^{j\prime}, \bar{\tau}_n^{j\prime})'$. The fact that $\bar{\gamma}_n^j$ belongs to the interior of Γ_j uniformly in n follows from Assumptions 13 and 14(i). It remains to verify that $\hat{\gamma}_n^j - \bar{\gamma}_n^j \overset{\text{as}}{\to} 0$. In view of Assumption 13, it suffices to prove that $\hat{\theta}_n^j - \bar{\theta}_n^j \overset{\text{as}}{\to} 0$. By Assumption 12(iii) and $r > 2$, we have $\|EM_n^j(\omega, \theta^j)\| \le (1/n)\sum_{t=1}^n \|Em_t^j(\omega, \theta^j)\| \le \Delta$, for some $\Delta < \infty$, and all $\theta^j \in \Theta_j$ and $n = 1, 2, \ldots$. Also, by Assumptions 12(ii)–(iv), it follows from Gallant and White (1988, Theorem 3.18) that $EM_n^j(\omega, \theta^j)$ is continuous on Θ_j uniformly in n, and $M_n^j(\omega, \theta^j) - EM_n^j(\omega, \theta^j) \overset{as}{\to} 0$ uniformly on Θ_j. Hence, using Assumption 13, we have

$$\begin{pmatrix} M_n^j(\omega, \theta^j) \\ \theta^j \\ \hat{\tau}_n^j(\omega) \end{pmatrix} - \begin{pmatrix} EM_n^j(\omega, \theta^j) \\ \theta^j \\ \bar{\tau}_n^j \end{pmatrix} \overset{as}{\to} 0 \text{ uniformly on } \Theta_j.$$

The first vector is measurable on Ω and, for each $\omega \in \Omega$, is continuous in θ^j. The second vector is also continuous in θ^j. Since $EM_n^j(\omega, \theta^j)$ is bounded on Θ_j uniformly in n, $\theta^j \in \Theta_j$ compact, and $\bar{\tau}_n^j \in \Upsilon_j$ compact, it follows from Gallant and White (1988, Lemma 3.4) or Bates and White (1985, Lemma 2.4) that

$$Q_n^j\{\omega, \theta^j, \hat{\tau}_n^j(\omega)\} - \bar{Q}_n^j(\theta^j, \bar{\tau}_n^j) \overset{as}{\to} 0 \text{ uniformly on } \Theta_j,$$

where $\bar{Q}_n^j(\theta^j, \bar{\tau}_n^j) = d_j\{EM_n^j(\omega, \theta^j), \theta^j, \bar{\tau}_n^j\}$. Therefore, by Gallant and White (1988, Theorem 3.3) or Domowitz and White (1982, Theorem 2.2), it follows from Assumption 14(i) that $\hat{\theta}_n^j - \bar{\theta}_n^j \overset{as}{\to} 0$.

Step 2: Verification of Assumption 10. First, we consider the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n^j - \bar{\theta}_n^j)$. Because $\bar{\theta}_n^j$ belongs to the interior of Θ_j uniformly in n by Assumption 14, and $\hat{\theta}_n^j - \bar{\theta}_n^j \overset{as}{\to} 0$, we have $0 = \partial Q_n^j(\omega, \hat{\theta}_n^j, \hat{\tau}_n^j)/\partial\theta^j$ as. From a Taylor expansion around $(\bar{\theta}_n^j, \bar{\tau}_n^j)$ we obtain

$$0 = \frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^j} + \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial\theta^j\,\partial\theta^{j\prime}}(\hat{\theta}_n^j - \bar{\theta}_n^j) + \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial\theta^j\,\partial\tau^{j\prime}}(\hat{\tau}_n^j - \bar{\tau}_n^j), \tag{A.11}$$

where $\bar{\theta}_n^j \le \ddot{\theta}_n^j \le \hat{\theta}_n^j$ and $\bar{\tau}_n^j \le \ddot{\tau}_n^j \le \hat{\tau}_n^j$. Now

$$\begin{aligned}\frac{\partial Q_n^j(\omega, \theta^j, \tau^j)}{\partial\theta^{j\prime}} &= \frac{\partial d_j\{M_n^j(\omega, \theta^j), \theta^j, \tau^j\}}{\partial m^{j\prime}} \cdot \frac{\partial M_n^j(\omega, \theta^j)}{\partial\theta^{j\prime}} \\ &\quad + \frac{\partial d_j\{M_n^j(\omega, \theta^j), \theta^j, \tau^j\}}{\partial\theta^{j\prime}},\end{aligned} \tag{A.12}$$

$$
\begin{aligned}
\frac{\partial^2 Q_n^j(\omega,\theta^j,\tau^j)}{\partial\theta^j\,\partial\theta^{j\prime}} = {} & \frac{\partial M_n^j(\omega,\theta^j)'}{\partial\theta^j}\cdot\frac{\partial^2 d_j\{M_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial m^j\,\partial m^{j\prime}}\cdot\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^{j\prime}} \\
& + \frac{\partial^2 d_j\{M_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\theta^j\,\partial m^{j\prime}}\cdot\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^{j\prime}} \\
& + \left[\frac{\partial d_j\{M_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial m^{j\prime}}\otimes I_{p_j}\right]\cdot\frac{\partial^2 M_n^j(\omega,\theta^j)}{\partial\theta^j\partial\theta^{j\prime}} \\
& + \frac{\partial M_n^j(\omega,\theta^j)'}{\partial\theta^j}\cdot\frac{\partial^2 d_j\{M_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial m^j\,\partial\theta^{j\prime}} \\
& + \frac{\partial^2 d_j\{M_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\theta^j\,\partial\theta^{j\prime}},
\end{aligned}
$$

where $\partial^2 M_n^j(\omega,\theta^j)/\partial\theta^j\partial\theta^{j\prime} \equiv (\partial^2 M_{n1}^j(\omega,\theta^j)/\partial\theta^j\partial\theta^{j\prime},\ldots,\partial^2 M_{nq_j}^j(\omega,\theta^j)/\partial\theta^j\partial\theta^{j\prime})'$ using the same notation as in Gallant and White (1988, Lemma 5.2). Also

$$
\frac{\partial^2 Q_n^j(\omega,\theta^j,\tau^j)}{\partial\tau^j\,\partial\theta^{j\prime}} = \frac{\partial^2 d_j\{M_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\tau^j\,\partial m^{j\prime}}\cdot\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^{j\prime}} + \frac{\partial^2 d_j\{M_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\tau^j\,\partial\theta^{j\prime}}.
$$

On the other hand, from Assumption 12(iii) and the LDC theorem,

$$
\frac{\partial E M_n^j(\omega,\theta^j)}{\partial\theta^j} = E\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^j}, \qquad \frac{\partial^2 E M_n^j(\omega,\theta^j)}{\partial\theta^j\,\partial\theta^{j\prime}} = E\frac{\partial^2 M_n^j(\omega,\theta^j)}{\partial\theta^j\,\partial\theta^{j\prime}}.
$$

Thus, from $\bar{Q}_n^j(\theta^j,\tau^j) = d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}$, we obtain

$$
\begin{aligned}
\frac{\partial\bar{Q}_n^j(\theta^j,\tau^j)}{\partial\theta^{j\prime}} &= \frac{\partial d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial m^{j\prime}} E\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^{j\prime}} + \frac{\partial d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\theta^{j\prime}}, \\
\frac{\partial^2\bar{Q}_n^j(\theta^j,\tau^j)}{\partial\theta^j\,\partial\theta^{j\prime}} &= E\frac{\partial M_n^j(\omega,\theta^j)'}{\partial\theta^j}\cdot\frac{\partial^2 d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial m^j\,\partial m^{j\prime}}\cdot E\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^{j\prime}} \\
&\quad + \frac{\partial^2 d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\theta^j\,\partial m^{j\prime}}\cdot E\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^{j\prime}} \\
&\quad + \left[\frac{\partial d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial m^{j\prime}}\otimes I_{p_j}\right]\cdot E\frac{\partial^2 M_n^j(\omega,\theta^j)}{\partial\theta^j\partial\theta^{j\prime}} \\
&\quad + E\frac{\partial M_n^j(\omega,\theta^j)'}{\partial\theta^j}\cdot\frac{\partial^2 d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial m^j\,\partial\theta^{j\prime}} \\
&\quad + \frac{\partial^2 d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\theta^j\,\partial\theta^{j\prime}}, \\
\frac{\partial^2\bar{Q}_n^j(\theta^j,\tau^j)}{\partial\tau^j\,\partial\theta^{j\prime}} &= \frac{\partial^2 d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\tau^j\,\partial m^{j\prime}}\cdot E\frac{\partial M_n^j(\omega,\theta^j)}{\partial\theta^{j\prime}} + \frac{\partial^2 d_j\{EM_n^j(\omega,\theta^j),\theta^j,\tau^j\}}{\partial\tau^j\,\partial\theta^{j\prime}}.
\end{aligned}
$$

Note that every function in the right-hand side of any of the above three equations is bounded on Θ_j or $\Theta_j\times\Upsilon_j$ uniformly in n because of Assumption 12(iii). Hence $\partial\bar{Q}_n^j(\theta^j,\tau^j)/\partial\theta^{j\prime}$ and its derivatives are bounded on $\Theta_j\times\Upsilon_j$ uniformly in n.

Now, given Assumptions 1, 9 and 12, it follows from Gallant and White (1988, Theorem 3.18) that

$$
\begin{aligned}
M_n^j(\omega, \theta^j) - EM_n^j(\omega, \theta^j) &\overset{as}{\to} 0 \text{ uniformly on } \Theta_j, \\
\frac{\partial M_n^j(\omega, \theta^j)}{\partial \theta^j} - E\frac{\partial M_n^j(\omega, \theta^j)}{\partial \theta^j} &\overset{as}{\to} 0 \text{ uniformly on } \Theta_j, \\
\frac{\partial^2 M_n^j(\omega, \theta^j)}{\partial \theta^j\, \partial \theta^{j\prime}} - E\frac{\partial^2 M_n^j(\omega, \theta^j)}{\partial \theta^j\, \partial \theta^{j\prime}} &\overset{as}{\to} 0 \text{ uniformly on } \Theta_j,
\end{aligned}
$$

where $EM_n^j(\omega, \theta^j)$, $E\partial M_n^j(\omega, \theta^j)/\partial\theta^j$, and $E\partial^2 M_n^j(\omega, \theta^j)/\partial\theta^j\partial\theta^{j\prime}$ are continuous on Θ_j uniformly in n. Since these three functions are bounded on Θ_j uniformly in n because of Assumption 12(iii), and since $d_j(\cdot)$ and its first two partial derivatives are continuous and hence uniformly continuous on compact sets, it follows from White (1984, Proposition 2.16) that

$$
\begin{aligned}
\frac{\partial^2 Q_n^j(\omega, \theta^j, \tau^j)}{\partial \theta^j\, \partial \theta^{j\prime}} - \frac{\partial^2 \bar{Q}_n^j(\theta^j, \tau^j)}{\partial \theta^j\, \partial \theta^{j\prime}} &\overset{as}{\to} 0 \text{ uniformly on } \Theta_j \times \Upsilon_j, \\
\frac{\partial^2 Q_n^j(\omega, \theta^j, \tau^j)}{\partial \tau^j\, \partial \theta^{j\prime}} - \frac{\partial^2 \bar{Q}_n^j(\theta^j, \tau^j)}{\partial \tau^j\, \partial \theta^{j\prime}} &\overset{as}{\to} 0 \text{ uniformly on } \Theta_j \times \Upsilon_j.
\end{aligned}
$$

Moreover, $\partial^2 \bar{Q}_n^j(\theta^j, \tau^j)/\partial\theta^j\partial\theta^{j\prime}$ and $\partial^2 \bar{Q}_n^j(\theta^j, \tau^j)/\partial\tau^j\partial\theta^{j\prime}$ are continuous on $\Theta_j \times \Upsilon_j$ uniformly in n.

Using the preceding results with $\ddot{\theta}_n^j - \bar{\theta}_n^j \overset{as}{\to} 0$ and $\ddot{\tau}_n^j - \bar{\tau}_n^j \overset{as}{\to} 0$ because $\hat{\theta}_n^j - \bar{\theta}_n^j \overset{as}{\to} 0$ and $\hat{\tau}_n^j - \bar{\tau}_n^j \overset{as}{\to} 0$, it follows from Domowitz and White (1982, Theorem 2.3) that

$$
\begin{aligned}
\frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial \theta^j\, \partial \theta^{j\prime}} - \frac{\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^j\, \partial \theta^{j\prime}} &\overset{as}{\to} 0, \\
\frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial \theta^j\, \partial \tau^{j\prime}} - \frac{\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^j \partial \tau^{j\prime}} &\overset{as}{\to} 0,
\end{aligned}
$$

where $\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^j\partial\theta^{j\prime}$ and $\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^j\partial\tau^{j\prime}$ are both O(1). Hence, Assumption 14 (ii) implies that $\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)/\partial\theta^j\partial\theta^{j\prime}$ is nonsingular for n sufficiently large almost surely and $A_{1jn}^{-1} = O(1)$. Thus, from (A.11) we obtain

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_n^j - \bar{\theta}_n^j) \\
&\quad = -\left(\frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial \theta^j\, \partial \theta^{j\prime}}\right)^{-1} \Bigg\{\sqrt{n}\frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^j} \\
&\qquad + \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial \theta^j\, \partial \tau^{j\prime}} \sqrt{n}(\hat{\tau}_n^j - \bar{\tau}_n^j)\Bigg\} \\
&\quad = -A_{1jn}^{-1}\Bigg\{\sqrt{n}\frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^j} + \frac{\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^j\, \partial \tau^{j\prime}} \sqrt{n}(\hat{\tau}_n^j - \bar{\tau}_n^j)\Bigg\} \\
&\qquad + \left(A_{1jn}^{-1} - \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)^{-1}}{\partial \theta^j\, \partial \theta^{j\prime}}\right)\sqrt{n}\frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^j} \\
&\qquad + \left(A_{1jn}^{-1}\frac{\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^j \partial \tau^{j\prime}} - \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)^{-1}}{\partial \theta^j\, \partial \theta^{j\prime}} \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial \theta^j \partial \tau^{j\prime}}\right)\sqrt{n}(\hat{\tau}_n^j - \bar{\tau}_n^j).
\end{aligned}
$$

But, from $A_{1jn} = O(1)$ uniformly positive definite and $A_{1jn}^{-1} = O(1)$, we have

$$A_{1jn}^{-1} - \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)^{-1}}{\partial\theta^j\,\partial\theta^{j\prime}} \xrightarrow{as} 0,$$

$$A_{1jn}^{-1}\frac{\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^j\partial\tau^{j\prime}} - \frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)^{-1}}{\partial\theta^j\,\partial\theta^{j\prime}}\frac{\partial^2 Q_n^j(\omega, \ddot{\theta}_n^j, \ddot{\tau}_n^j)}{\partial\theta^j\partial\tau^{j\prime}} \xrightarrow{as} 0,$$

using White (1984, Proposition 2.16) and $\partial^2\bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^j\partial\tau^{j\prime} = O(1)$. Thus, *provided* $\sqrt{n}\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^j = O_P(1)$ and $\sqrt{n}(\hat{\tau}_n^j - \bar{\tau}_n^j) = O_P(1)$, we obtain

$$\sqrt{n}(\hat{\theta}_n^j - \bar{\theta}_n^j) = -A_{1jn}^{-1}\left\{\sqrt{n}\frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^j} + \frac{\partial^2\bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^j\partial\tau^{j\prime}}\sqrt{n}(\hat{\tau}_n^j - \bar{\tau}_n^j)\right\} + o_P(1). \quad \text{(A.13)}$$

Consider now the first term inside the braces. From (A.12) we obtain

$$\begin{aligned}\frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^{j\prime}} &= \frac{\partial d_j\{M_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}}\left(\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial\theta^{j\prime}} - E\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial\theta^{j\prime}}\right)\\ &\quad+\frac{\partial d_j\{M_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}}E\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial\theta^{j\prime}}\\ &\quad+\frac{\partial d_j\{M_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial\theta^{j\prime}}. \end{aligned}\qquad \text{(A.14)}$$

From a Taylor expansion around $(EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j)$ we have

$$\begin{aligned}\frac{\partial d_j\{M_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}} &= \frac{\partial d_j\{EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}}\\ &\quad+\{M_n^j(\omega, \bar{\theta}_n^j) - EM_n^j(\omega, \bar{\theta}_n^j)\}'\,\frac{\partial^2 d_j\{\ddot{M}_n^j, \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^j\partial m^{j\prime}},\\ \frac{\partial d_j\{M_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial\theta^{j\prime}} &= \frac{\partial d_j\{EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial\theta^{j\prime}}\\ &\quad+\{M_n^j(\omega, \bar{\theta}_n^j) - EM_n^j(\omega, \bar{\theta}_n^j)\}'\,\frac{\partial^2 d_j(\ddot{M}_n^j, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial m^j\partial\theta^{j\prime}},\end{aligned}$$

where $EM_n^j(\omega, \bar{\theta}_n^j) \leq \ddot{M}_n^j \leq M_n^j(\omega, \bar{\theta}_n^j)$. Thus, multiplying (A.14) by $\sqrt{n}$, we obtain

$$\begin{aligned}\sqrt{n}\,\frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^{j\prime}} &= \frac{\partial d_j\{M_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}}\sqrt{n}\left(\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial\theta^{j\prime}} - E\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial\theta^{j\prime}}\right)\\ &\quad+\sqrt{n}\{M_n^j(\omega, \bar{\theta}_n^j) - EM_n^j(\omega, \bar{\theta}_n^j)\}'\\ &\quad\times\left(\frac{\partial^2 d_j(\ddot{M}_n^j, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial m^j\partial m^{j\prime}}E\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial\theta^{j\prime}} + \frac{\partial^2 d_j(\ddot{M}_n^j, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial m^j\,\partial\theta^{j\prime}}\right),\end{aligned}\qquad \text{(A.15)}$$

because $\partial\bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^{j\prime} = 0$, which follows from Assumption 14(i), i.e.

$$\frac{\partial d_j\{EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}}E\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial\theta^{j\prime}} + \frac{\partial d_j\{EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial\theta^{j\prime}} = 0.$$

But $M_n^j(\omega, \bar{\theta}_n^j) - EM_n^j(\omega, \bar{\theta}_n^j) \xrightarrow{as} 0$ from uniform convergence established earlier. Hence, because $EM_n^j\ (\omega, \bar{\theta}_n^j) = O(1)$, $\bar{\theta}_n^j \in \Theta_j$ compact, $\bar{\tau}_n^j \in \Upsilon_j$ compact, we have

$$\frac{\partial d_j\{M_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}} - \frac{\partial d_j\{EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^{j\prime}} \xrightarrow{as} 0,$$

where the second term is $O(1)$. Moreover, since $\ddot{M}_n^j - EM_n^j(\omega, \bar{\theta}_n^j) \xrightarrow{as} 0$, we have

$$\frac{\partial^2 d_j(\ddot{M}_n^j, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial m^j\, \partial m^{j\prime}} - \frac{\partial^2 d_j\{EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^j\, \partial m^{j\prime}} \xrightarrow{as} 0,$$
$$\frac{\partial^2 d_j(\ddot{M}_n^j, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial m^j\, \partial \theta^{j\prime}} - \frac{\partial^2 d_j\{EM_n^j(\omega, \bar{\theta}_n^j), \bar{\theta}_n^j, \bar{\tau}_n^j\}}{\partial m^j\, \partial \theta^{j\prime}} \xrightarrow{as} 0,$$

where the second terms are $O(1)$. Since $E\partial M_n^j(\omega, \bar{\theta}_n^j)/\partial\theta^{j\prime} = O(1)$ from Assumption 12(iii), (A.15) becomes

$$\begin{aligned}\sqrt{n}\, \frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial \theta^{j\prime}} &= \frac{\partial \bar{d}_{jn}}{\partial m^{j\prime}} \sqrt{n}\left(\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial \theta^{j\prime}} - E\,\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial \theta^{j\prime}}\right) \\ &\quad + \sqrt{n}\left\{M_n^j(\omega, \bar{\theta}_n^j) - EM_n^j(\omega, \bar{\theta}_n^j)\right\}' \\ &\quad \times \left(\frac{\partial^2 \bar{d}_{jn}}{\partial m^j \partial m^{j\prime}} E\,\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)}{\partial \theta^{j\prime}} + \frac{\partial^2 \bar{d}_{jn}}{\partial m^j\, \partial \theta^{j\prime}}\right) + o_P(1), \qquad \text{(A.16)}\end{aligned}$$

using the notation defined in the text, *provided* $\sqrt{n}\{M_n^j(\omega, \bar{\theta}_n^j) - EM_n^j(\omega, \bar{\theta}_n^j)\}$ and $\sqrt{n}\{\partial M_n^j(\omega, \bar{\theta}_n^j)/\partial\theta^{j\prime} - E\partial M_n^j(\omega, \bar{\theta}_n^j)/\partial\theta^{j\prime})\}$ are both $O_P(1)$. The first $O_P(1)$ condition follows from Assumption A1, 9 and 12 by the argument used by Gallant and White (1988, pp. 85–86) for proving that $\sqrt{n}(\psi_n^o - \bar{\psi}_n^o) = O_P(1)$. The second $O_P(1)$ condition follows similarly. Note that (A.16) implies that $\sqrt{n}\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^{j\prime} = O_P(1)$, as required for deriving (A.13) because $E\partial M_n^j(\omega, \bar{\theta}_n^j)/\partial\theta^{j\prime}$, $\partial\bar{d}_{jn}/\partial m^j$, $\partial^2\bar{d}_{jn}/\partial m^j\partial m^{j\prime}$, and $\partial^2\bar{d}_{jn}/\partial m^j\partial\theta^{j\prime}$ are all $O(1)$ as mentioned earlier.

Collecting results, as in Gallant and White (1988, p. 75) let

$$Y_{1nt}^j = \left(E\,\frac{\partial M_n^j(\omega, \bar{\theta}_n^j)'}{\partial\theta^j} \cdot \frac{\partial^2 \bar{d}_{jn}}{\partial m^j\, \partial m^{j\prime}} + \frac{\partial^2 \bar{d}_{jn}}{\partial \theta^j\, \partial m^{j\prime}}\right)\tilde{m}_{nt}^j + \frac{\partial \tilde{m}_{nt}^{j\prime}}{\partial\theta^j}\frac{\partial \bar{d}_{jn}}{\partial m^j}$$

where $\tilde{m}_{nt}^j = m_t^j(\omega, \bar{\theta}_n^j) - Em_t^j(\omega, \bar{\theta}_n^j)$. Hence, from (A.16) we have

$$\sqrt{n}\frac{\partial Q_n^j(\omega, \bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^j} = \frac{1}{\sqrt{n}}\sum_{t=1}^n Y_{1nt}^j + o_P(1),$$

where $EY_{1nt}^j = 0$. Thus, from (A.13) and Assumption 13, we obtain

$$\begin{aligned}\begin{pmatrix}\sqrt{n}(\hat{\theta}_n^j - \bar{\theta}_n^j) \\ \sqrt{n}(\hat{\tau}_n^j - \bar{\tau}_n^j)\end{pmatrix} &= -\begin{pmatrix} A_{1jn}^{-1} & -A_{1jn}^{-1}\partial^2\bar{Q}_n^j(\bar{\theta}_n^j\bar{\tau}_n^j)/\partial\theta^j\partial\tau^{j\prime}A_{2jn}^{+} \\ 0 & A_{2jn}^{+}\end{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^n\begin{pmatrix}Y_{1nt}^j \\ Y_{2nt}^j\end{pmatrix} \\ &\quad + o_P(1), \qquad \text{(A.17)}\end{aligned}$$

which is in the form of Assumption 10. Now recall that $\partial^2 \bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)/\partial\theta^j \partial\tau^{j\prime} = O(1)$ and $A_{1jn}^{-1} = O(1)$, as shown earlier, and $A_{2jn}^+ = O(1)$ by Assumption 13. Hence $A_{jn}^+ = O(1)$ as required. Second, $E(Y_{nt}^j) = 0$ where $Y_{nt}^j = (Y_{1nt}^{j\prime}, Y_{2nt}^{j\prime})'$. The r-integrability of Y_{nt}^j uniformly in n, t follows from Assumptions 12 and 13, and the boundedness of the nonstochastic matrices appearing in Y_{nt}^j. Finally, $\{Y_{nt}^j\}$ is near-epoch dependent on $\{X_t\}$ of size -1 because of Assumptions 12 and 13. This establishes Assumption 10.

Step 3: Computation of σ_n^2. From Theorem 2, we have $\sigma_n^2 = R_n' V_n R_n$ where $R_n^{j\prime} = (\partial\bar{d}_{jn}/\partial m^{j\prime}, -\partial\bar{Q}_n^j(\bar{\gamma}_n^j)/\partial\gamma^{j\prime} A_{jn}^+)$,

$$\frac{\partial\bar{Q}_n^j(\bar{\gamma}_n^j)}{\partial\gamma^{j\prime}} = \left(\frac{\partial\bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\theta^{j\prime}}, \frac{\partial\bar{Q}_n^j(\bar{\theta}_n^j, \bar{\tau}_n^j)}{\partial\tau^{j\prime}}\right) = \left(0, \frac{\partial\bar{d}_{jn}}{\partial\tau^{j\prime}}\right),$$

and A_{jn}^+ is the upper block triangular matrix in (A.17). Thus, matrix algebra gives $R_n^{j\prime} = (\partial\bar{d}_{jn}/\partial m^{j\prime}, 0, -\partial\bar{d}_{jn}/\partial\tau^{j\prime} A_{2jn}^+)$. Let $V_{nt} = (V_{nt}^{1\prime}, V_{nt}^{2\prime})'$, where $V_{nt}^j = (\tilde{m}_{nt}^{j\prime}, Y_{2nt}^{j\prime})'$. Hence, we obtain

$$\sigma_n^2 = \operatorname{Var}\left(\frac{\partial\bar{d}_{1n}}{\partial m^{1\prime}}, -\frac{\partial\bar{d}_{1n}}{\partial\tau^{1\prime}} A_{21n}^+, -\frac{\partial\bar{d}_{2n}}{\partial m^{2\prime}}, \frac{\partial\bar{d}_{2n}}{\partial\tau^{2\prime}} A_{22n}^+\right) \frac{1}{\sqrt{n}} \sum_{t=1}^n V_{nt},$$

where the last term has a covariance matrix V_n with uniform rank $s > 0$, as assumed in Theorem 3 and required by Theorem 2. The desired result follows.

Proof of Theorem 4. We prove three properties, and then the result.

Step 1. The first property is that $\hat{R}_n - R_n \stackrel{P}{\rightarrow} 0$. From the uniform strong convergence of $M_n^j(\omega, \gamma^j)$ and $\partial M_n^j(\omega, \gamma^j)/\partial\gamma^{j\prime}$ to $EM_n^j(\omega, \gamma^j)$ and $E\partial M_n^j(\omega, \gamma^j)/\partial\gamma^{j\prime}$, and from the equicontinuity of the two limit functions (see the proof of Theorem 2), it follows from Domowitz and White (1982, Theorem 2.3) that $M_n^j(\omega, \hat{\gamma}_n^j) - EM_n^j(\omega, \bar{\gamma}_n^j) \stackrel{as}{\rightarrow} 0$ and $\partial M_n^j(\omega, \hat{\gamma}_n^j)/\partial\gamma^{j\prime} - E\partial M_n^j(\omega, \bar{\gamma}_n^j)/\partial\gamma^{j\prime} \stackrel{as}{\rightarrow} 0$. Since $EM_n^j(\omega, \gamma_n^j)$ and $E\partial M_n^j(\omega, \gamma_n^j)/\partial\gamma^{j\prime}$ are both $O(1)$, and continuous functions are uniformly continuous on compact sets, it follows that $\partial\hat{d}_{jn}/\partial m^j - \partial\bar{d}_{jn}/\partial m^j \stackrel{as}{\rightarrow} 0$ and $\partial\hat{d}_{jn}/\partial\gamma^j - \partial\bar{d}_{jn}/\partial\gamma^j \stackrel{as}{\rightarrow} 0$, where $\partial\bar{d}_{jn}/\partial m^j$ and $\partial\bar{d}_{jn}/\partial\gamma^j$ are both $O(1)$. Using expressions (A.6) and (A.7), it follows from White (1984, Proposition 2.16) that $\partial Q_n^j(\omega, \hat{\gamma}_n^j)/\partial\gamma^j - \partial\bar{Q}_n^j(\omega, \bar{\gamma}_n^j)/\partial\gamma^j \stackrel{as}{\rightarrow} 0$. Since $\hat{A}_{jn}^+ - A_{jn}^+ \stackrel{P}{\rightarrow} 0$ and $A_{jn}^+ = O(1)$ by assumption, we obtain the desired property.

Step 2. The second property is that $\hat{V}_n - (V_n + \Lambda_n) \stackrel{P}{\rightarrow} 0$. Its proof follows the proof of Theorem 5.6 in Gallant and White (1988). First, we have $V_n - \ddot{V}_n \rightarrow 0$ where

$$\ddot{V}_n = \frac{w_{n0}}{n} \sum_{t=1}^n E(V_{nt} V_{nt}') + \frac{1}{n} \sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+1}^n \{E(V_{nt} V_{n,t-\tau}') + E(V_{n,t-\tau} V_{n,t}')\},$$

and $V_{nt} = (V_{nt}^{1\prime}, V_{nt}^{2\prime})'$ is the tth element of the sum (8). To see this, note that from (18) we have

$$V_{nt}^j = \begin{pmatrix} m_t^j(\omega, \bar{\gamma}_n^j) \\ \delta_t^j(\omega, \bar{\gamma}_n^j) \end{pmatrix} - E\begin{pmatrix} m_t^j(\omega, \bar{\gamma}_n^j) \\ \delta_t^j(\omega, \bar{\gamma}_n^j) \end{pmatrix}.$$

Since $V_n = \text{Var}\,(1/\sqrt{n})\sum_{t=1}^{n} V_{nt}$ it follows from Gallant and White (1988, Lemma 6.6) and Assumptions 1, 9, 15(iii)–(iv) and 16(iii)–(iv) that $V_n - \ddot{V}_n \to 0$.

Second, we have $V_n^* - (\ddot{V} + \Lambda_n) \xrightarrow{P} 0$ where

$$V_n^* = \frac{w_{n0}}{n}\sum_{t=1}^{n} U_{nt}U_{nt}' + \frac{1}{n}\sum_{\tau=1}^{m_n} w_{n\tau}\sum_{t=\tau+1}^{n}(U_{nt}U_{n,t-\tau}' + U_{n,t-\tau}U_{n,t}'), \tag{A.18}$$

$U_{nt} = (U_{nt}^{1\prime}, U_{nt}^{2\prime})'$, and $U_{nt}^{j\prime} = (m_t^j(\omega, \bar{\gamma}_n^j)', \delta_t^j(\omega, \bar{\gamma}_n^j)')$. To see this, note that

$$\begin{aligned}\ddot{V}_n + \Lambda_n = {} & \frac{w_{n0}}{n}\sum_{t=1}^{n} E(U_{nt}U_{nt}') \\ & + \frac{1}{n}\sum_{\tau=1}^{m_n} w_{n\tau}\sum_{t=\tau+1}^{n}\{E(U_{nt}U_{n,t-\tau}') + E(U_{n,t-\tau}U_{n,t}')\},\end{aligned} \tag{A.19}$$

because of (22) and $EU_{nt} = \mu_{nt}$. Consider the leading term of the difference between (A.18) and (A.19). From Assumptions 15(iii)–(iv) and 16(iii)–(iv), we have by Gallant and White (1988, Corollary 4.3) that the elements of $\{U_{nt}U_{nt}'\}$ are r-integrable uniformly in n, t and near-epoch dependent on $\{X_t\}$ of size -1 and hence of size $-1/2$. Then, applying Gallant and White (1988, Theorem 6.2) on the elements of $\{U_{nt}U_{nt}' - E(U_{nt}U_{nt}')\}$, we obtain

$$\frac{w_{n0}}{n}\sum_{t=1}^{n}(U_{nt}U_{nt}' - E(U_{nt}U_{nt}')) \xrightarrow{P} 0,$$

because $|w_{n0}| \le \Delta$ by Assumption 18. Moreover, from Gallant and White (1988, Lemma 6.7) and Assumptions 1, 9, 15(iii)–(iv) and 16(iii)–(iv), we have

$$\frac{1}{n}\sum_{\tau=1}^{m_n} w_{n\tau}\sum_{t=\tau+1}^{n}(U_{nt}U_{n,t-\tau}' - E(U_{nt}U_{n,t-\tau}')) \xrightarrow{P} 0.$$

Hence $V_n^* - (\ddot{V}_n + \Lambda_n) \xrightarrow{P} 0$ as claimed.

Third, we have $\hat{V}_n - V_n^* \xrightarrow{P} 0$. This is proved by taking a Taylor expansion around $(\bar{\gamma}_n^1, \bar{\gamma}_n^2)$ and by using an argument identical to that used by Gallant and White (1988, p. 106 and p. 118) for proving that $\bar{\pi}_n(\tilde{\theta}_n - \theta_n^0) = o_P(1)$. It is here where we use the assumption that the elements of $\{\partial m_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ and $\{\partial\delta_t^j(\omega, \gamma^j)/\partial\gamma^{j\prime}\}$ are $2r$-dominated uniformly on Γ_j. Collecting the preceding three facts, it follows that $\hat{V}_n - (V_n + \Lambda_n) \xrightarrow{P} 0$. Note that Λ_n is positive semidefinite, which immediately follows from Assumption 18 and Gallant and White (1988, Lemma 6.5) applied to $\lambda'\mu_{nt}$ for arbitrary λ.

Step 3. The third property is that $R_n'\Lambda_n R_n \to 0$ under H_0^*. Let $\rho_{nt} = \mu_{nt} - \bar{\mu}_n$. Thus, from (22) we have

$$\Lambda_n = \frac{w_{n0}}{n}\left(n\bar{\mu}_n\bar{\mu}_n' + \bar{\mu}_n\sum_{t=1}^{n}\rho_{nt}' + \sum_{t=1}^{n}\rho_{nt}\bar{\mu}_n' + \sum_{t=1}^{n}\rho_{nt}\rho_{nt}'\right)$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{\tau=1}^{m_n} w_{n\tau}\sum_{t=\tau+1}^{n}\{2\bar{\mu}_n\bar{\mu}_n' + \bar{\mu}_n(\rho_{nt}' + \rho_{n,t-\tau}') + (\rho_{nt} + \rho_{n,t-\tau})\bar{\mu}_n' + \rho_{nt}\rho_{n,t-\tau}' \\
&+\rho_{n,t-\tau}\rho_{n,t-\tau}'\} \\
&= \bar{\mu}_n\bar{\mu}_n'\left\{w_{n0} + \frac{2}{n}\sum_{\tau=1}^{m_n} w_{n\tau}(n-\tau)\right\} \\
&\quad + \frac{\bar{\mu}_n}{n}\left\{w_{n0}\sum_{t=1}^{n}\rho_{nt}' + \sum_{\tau=1}^{m_n} w_{n\tau}\sum_{t=\tau+1}^{n}(\rho_{nt}' + \rho_{n,t-\tau}')\right\} \\
&\quad + \left\{w_{n0}\sum_{t=1}^{n}\rho_{nt} + \sum_{\tau=1}^{m_n} w_{n\tau}\sum_{t=\tau+1}^{n}(\rho_{nt} + \rho_{n,t-\tau})\right\}\frac{\bar{\mu}_n'}{n} \\
&\quad + \frac{1}{n}\left\{w_{n0}\sum_{t=1}^{n}\rho_{nt}\rho_{nt}' + \sum_{\tau=1}^{m_n} w_{n\tau}\sum_{t=\tau+1}^{n}(\rho_{nt}\rho_{n,t-\tau}' + \rho_{n,t-\tau}\rho_{nt}')\right\} \\
&= \Lambda_{1n} + \Lambda_{2n} + \Lambda_{2n}' + \Lambda_{3n}.
\end{aligned}
$$

Consider the first term and more specifically

$$
\begin{aligned}
|R_n'\Lambda_{1n}R_n| &= (R_n'\bar{\mu}_n)^2\left|w_{n0} + \frac{2}{n}\sum_{\tau=1}^{m_n} w_{n\tau}(n-\tau)\right| \\
&\le (R_n'\bar{\mu}_n)^2\Delta\left\{1 + \frac{2}{n}\sum_{\tau=1}^{m_n}(n-\tau)\right\} \\
&\le (R_n'\bar{\mu}_n)^2\Delta(1 + 2m_n).
\end{aligned}
$$

Since $R_n'\bar{\mu}_n = O(n^{-1/4})$ under H_0^* by condition (i), and $m_n = o(n^{1/4})$ by Assumption 17, it follows that $R_n'\Lambda_{1n}R_n \to 0$ under H_0^*.

Consider the second term and more specifically

$$
\begin{aligned}
|R_n'\Lambda_{2n}R_n| &\le \frac{|R_n'\bar{\mu}_n|}{n}\Delta\left\|\sum_{t=1}^{n}\rho_{nt}' + \sum_{\tau=1}^{m_n}\sum_{t=\tau+1}^{n}(\rho_{nt}' + \rho_{n,t-\tau}')\right\|\|R_n\| \\
&\le \frac{|R_n'\bar{\mu}_n|}{n}\Delta K\left\{n + 2\sum_{\tau=1}^{m_n}(n-\tau)\right\}\|R_n\| \\
&\le |R_n'\bar{\mu}_n|\Delta K(1 + 2m_n)\|R_n\|,
\end{aligned}
$$

where the second inequality follows from $\|\rho_{nt}'\| \le K \le \infty$ because of Assumptions 15(iii) and 16(iii). Since $R_n'\bar{\mu}_n = O(n^{-1/4})$ and $m_n = o(n^{1/4})$ by assumption, and $R_n = O(1)$, it follows that $R_n'\Lambda_{2n}R_n \to 0$ and $R_n'\Lambda_{2n}'R_n \to 0$ under H_0^*.

Finally, consider the last term. We have

$$
R_n'\Lambda_{3n}R_n = \frac{1}{n}\left\{w_{n_0}\sum_{t=1}^{n}(R_n'\rho_{nt})^2 + 2\sum_{\tau=1}^{m_n} w_{nt}\sum_{t=\tau+1}^{n}(R_n'\rho_{nt})(R_n'\rho_{n,t-\tau})\right\}.
$$

Hence, because $|w_{n\tau}| \leq \Delta$ and $|R_n' \rho_{nt}| \leq d_n$ by assumption, we obtain

$$|R_n' \Lambda_{3n} R_n| \leq \frac{\Delta}{n} d_n^2 \left\{ n + 2 \sum_{\tau=1}^{m_n} (n - \tau) \right\} \leq \Delta d_n^2 (1 + 2m_n).$$

Since $d_n^2 = O(n^{-1/4})$ by condition (ii) and $m_n = o(n^{1/4})$ by Assumption 17, it follows that $R_n' \Lambda_{3n} R_n \to 0$. Therefore under H_0^*, we have $R_n' \Lambda_n R_n \to 0$.

Step 4. We are now in a position to prove the theorem. From Steps 1 and 2 combined with $R_n = O(1)$, we obtain $\hat{R}_n'(\hat{V}_n - (V_n + \Lambda_n))\hat{R}_n \xrightarrow{P} 0$, i.e. $\hat{\sigma}_n^2 - \hat{R}_n' V_n \hat{R}_n - \hat{R}_n' \Lambda_n \hat{R}_n \xrightarrow{P} 0$. Because $V_n = O(1)$, $R_n = O(1)$, and $\hat{R}_n - R_n \xrightarrow{P} 0$, then $\hat{R}_n' V_n \hat{R}_n - R_n' V_n R_n \xrightarrow{P} 0$. Thus, using the definition of σ_n^2, we have $\hat{\sigma}_n^2 - \sigma_n^2 - \hat{R}_n' \Lambda_n \hat{R}_n \xrightarrow{P} 0$, i.e. $\hat{\sigma}_n^2 - \sigma_n^2 - (\hat{R}_n' \Lambda_n \hat{R}_n - R_n' \Lambda_n R_n) \xrightarrow{P} 0$ because of Step 3. Therefore the proof is complete if the term in parentheses converges in probability to zero under H_0^*.

To see the latter, we note that

$$R_n' \Lambda_n R_n = \frac{w_{n0}}{n} \sum_{t=1}^{n} \mu_{nt}' R_n R_n' \mu_{nt} + \frac{2}{n} \sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+1}^{n} \mu_{n,t}' R_n R_n' \mu_{n,t-\tau}$$

while $\hat{R}_n' \Lambda_n \hat{R}_n - R_n' \Lambda_n R_n$ is given by a similar expression with $(\hat{R}_n \hat{R}_n' - R_n R_n')$ replacing $R_n R_n'$. From Step 1 and $R_n = O(1)$ we have $(\hat{R}_n \hat{R}_n' - R_n R_n') = o_P(1) R_n R_n'$. Hence $\hat{R}_n' \Lambda_n \hat{R}_n - R_n' \Lambda_n R_n = o_P(1) R_n' \Lambda_n R_n = o_p(1)$ under H_0^* by Step 3.

Proof of Lemma 1. (i) Under H_0^*, (A.2) gives $\sqrt{n} \Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) = L_n' C_n' \sqrt{n} Z_n + o_P(1)$ since $\sqrt{n} \Delta \bar{Q}_n(\bar{\gamma}_n^1, \bar{\gamma}_n^2) = o(1)$. Since $\sqrt{n} Z_n \Rightarrow N(0, I_s)$, it follows that $\sqrt{n} \Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) = o_P(1)$ if and only if $L_n' C_n' = o(1)$, i.e. if and only if $\sigma_n^2 = o(1)$ since $\sigma_n^2 = L_n' C_n' C_n L_n$.

(ii) From (A.1) we have $\sqrt{n} \Delta Q_n(\omega, \hat{\gamma}_n^1, \hat{\gamma}_n^2) = \sqrt{n} \Delta Q_n(\omega, \bar{\gamma}_n^1, \bar{\gamma}_n^2) + o_p(1)$ since $\sqrt{n}(\hat{\gamma}_n^j - \bar{\gamma}_n^j) = O_P(1)$ and $\partial \bar{Q}_n^j(\bar{\gamma}_n^j)/\partial \gamma^{j\prime} = o(1)$ for $j = 1, 2$ by assumption. The desired result follows from Part (i).

Proof of Theorem 5. First, because model 1 is nested in model 2 according to the chosen selection criterion, we can take $\bar{Q}_n^1(\cdot) = \bar{Q}_n^2\{h_n(\cdot)\}$ in view of Assumption 4. Thus H_0^* is equivalent to $\sqrt{n}[\bar{Q}_n^2\{h_n(\bar{\gamma}_n^1)\} - \bar{Q}_n^2(\bar{\gamma}_n^2)] = o(1)$. Hence, because $\bar{\gamma}_n^2$ is the identifiably unique minimizer of $\bar{Q}_n^2(\cdot)$ on Γ_2, it is easily shown by contradiction that $h_n(\bar{\gamma}_n^1) - \bar{\gamma}_n^2 = o(1)$ under H_0^*.

Next, we show that $\sigma_n^2 = o(1)$ by verifying condition (ii) of Lemma 1. Because model 1 is nested in model 2 according to the chosen selection criterion, such a condition is equivalent to $\sqrt{n}[Q_n^2\{\omega, h_n(\bar{\gamma}_n^1)\} - Q_n^2(\omega, \bar{\gamma}_n^2)] = o_P(1)$. Taking a Taylor expansion around $\bar{\gamma}_n^2$ gives $\sqrt{n} \partial Q_n^2(\omega, \ddot{\gamma}_n^2)/\partial \gamma^{2\prime}\{h_n(\bar{\gamma}_n^1) - \bar{\gamma}_n^2)\} = o_P(1)$, where $\bar{\gamma}_n^2 \leq \ddot{\gamma}_n^2 \leq h_n(\bar{\gamma}_n^1)$. This is satisfied because of assumption (ii) and $h_n(\bar{\gamma}_n^1) - \bar{\gamma}_n^2 = o(1)$. The desired result follows.