Supplement to "Comparison of Misspecified Calibrated Models"*

Viktoria Hnatkovska†, Vadim Marmer†, and Yao Tang‡

February 3, 2011

S.1 Introduction

This paper contains supplemental material to Hnatkovska et al. (2011), HMT hereafter. In Section S.2, we discuss the asymptotic properties of the CMD estimators defined in Section 2 of HMT. In Section S.3, we consider model comparison tests of Sections 3 and 4 in HMT when the weight matrices used to construct the CMD criterion function are data-dependent. Section S.4 considers the model comparison testing when models are estimated and evaluated on different sets of the reduced-form parameters. In Section S.5, we discuss how to construct a confidence set for the weighting schemes favorable to one of the models. This procedure allows one to compare two models by taking into account their relative performance under various weighting schemes. Section S.6 contains the proofs of the results in HMT. Section S.7 contains the proofs of the additional results presented in Sections S.2-S.4.

Here, we would like also to make some additional remarks regarding the testing problem discussed in HMT. In our paper, we follow the approach originally developed in Vuong (1989). We believe that Vuong (1989) and similar testing problems should be discerned from the classical non-nested hypothesis testing problems (Davidson and MacKinnon, 1981; Smith, 1992). Suppose that two alternative models are non-nested and therefore cannot be both true at the same time. According to our model comparison null hypothesis, the models have equal measures of fit and, consequently, the null hypothesis implies that they are both misspecified. However, in the literature on non-nested hypothesis testing, the null

*We thank two anonymous referees and the guest co-editors for very helpful comments and suggestions.

†Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail addresses: hnatkovs@interchange.ubc.ca (Hnatkovska), vadim.marmer@ubc.ca (Marmer).

‡Department of Economics, Bowdoin College, 9700 College Station, Brunswick, Maine 04011-8497, USA. Email: ytang@bowdoin.edu.

hypothesis is that one of the models is true. Thus, the two approaches, the non-nested testing and the model comparison testing of misspecified models in the spirit of Vuong (1989), are not competing but rather complementary. The first approach can be used in a search for the true specification, while the later approach can be adopted when the econometrician believes that all alternative models are misspecified or when they all have been rejected by the overidentified restrictions or non-nested tests.

Lastly, we would like to make some comments regarding computation of the $AQLR_n$ statistic in Section 4 in HMT. In general, its computation can be quite complicated due to the need to integrate over $\mathbb{A}$ according to the distribution π. This task can be simplified in the following manner. When the CDF corresponding to π is available in the closed-form, one can draw at random from $\mathbb{A}$ according to the distribution π using a random numbers generator for uniform(0,1) distribution. Let A_j be such random draw, $j = 1, \ldots, \tau_n$, where the number of draws τ_n can depend on the sample size n. For each A_j compute $QLR_n\left(\hat{\theta}_n(A_j), \hat{\gamma}_n(A_j), A_j\right)$, and then compute

$$AQLR_n^a = \tau_n^{-1} \sum_{j=1}^{\tau_n} QLR_n\left(\hat{\theta}_n(A_j), \hat{\gamma}_n(A_j), A_j\right).$$

As $\tau_n \to \infty$, $AQLR_n^a \to AQLR_n$ in probability. In practice one should choose τ_n so that n/τ_n is close to zero, so that the effect of approximating the integral by random draws on the test would be negligible.

S.2 Asymptotic properties of the CMD estimators of structural parameters

The CMD estimator of the structural parameters are defined by (2.1) in HMT. In this section, we discuss the asymptotic properties of the CMD estimators under misspecification.

Theorem S.1 *Suppose that Assumptions 2.1, 2.2, 2.4, and 2.5 in HMT hold. Then,* $\hat{\theta}_n \to_p \theta_0$ *and* $\hat{\gamma}_n \to_p \gamma_0$.

Consider the correctly specified case: $h_0 - f(\theta_0) = 0$. In this case, Assumptions 2.1, 2.2, 2.4, 2.5, and Theorem S.1 imply that $n^{1/2}\left(\hat{\theta}_n - \theta_0\right)$ has asymptotically normal distribution with the variance matrix

$$\left(\frac{\partial f(\theta_0)'}{\partial \theta} A' A \frac{\partial f(\theta_0)}{\partial \theta'}\right)^{-1} \frac{\partial f(\theta_0)'}{\partial \theta} A' A \Lambda_0 A' A \frac{\partial f(\theta_0)}{\partial \theta'} \left(\frac{\partial f(\theta_0)'}{\partial \theta} A' A \frac{\partial f(\theta_0)}{\partial \theta'}\right)^{-1}.$$

As usual, in the correctly specified case, the efficient CMD estimator corresponds to $A_n' A_n = \hat{\Lambda}_n^{-1}$. Note, however, that when a model is misspecified, such a choice no longer leads to statistical efficiency.

Define

$$\begin{aligned}
V_{ff,0} &= F_0^{-1} \frac{\partial f(\theta_0)'}{\partial \theta} A' A \Lambda_0 A' A \frac{\partial f(\theta_0)}{\partial \theta'} F_0'^{-1}, \\
V_{fg,0} &= F_0^{-1} \frac{\partial f(\theta_0)'}{\partial \theta} A' A \Lambda_0 A' A \frac{\partial g(\gamma_0)}{\partial \gamma'} G_0'^{-1}, \\
V_{gg,0} &= G_0^{-1} \frac{\partial g(\gamma_0)'}{\partial \gamma} A' A \Lambda_0 A' A \frac{\partial g(\gamma_0)}{\partial \gamma'} G_0'^{-1}, \text{ and} \\
V_0 &= \begin{pmatrix} V_{ff,0} & V_{fg,0} \\ V_{fg,0}' & V_{gg,0} \end{pmatrix}, \qquad \text{(S.1)}
\end{aligned}$$

where the matrices F_0 and G_0 are defined in Assumption 2.5 in HMT. Note that the matrix V_0 can be singular. The following theorem describes the asymptotic distribution of the CMD estimators in the fixed weight matrix case.

Theorem S.2 *Suppose that $A_n = A$ for all $n \geq 1$. Under Assumptions 2.1, 2.2, 2.4, and 2.5 in HMT,*

$$n^{1/2} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} \to_d N\left(0_{(k+l)\times 1}, V_0\right).$$

Next, we consider the case of data-dependent weight matrices A_n. For example, suppose that $A_n' A_n \to_p \Lambda^{-1}$. While such a choice no longer optimal when the models are misspecified, the econometrician is still might be interested in using such A_n, because it assigns greater weights to the elements of h that are estimated more precisely. When the weight matrix depends on the data, we replace Assumption 2.1 with Assumption S.1 below, which assumes that the elements of $A_n' A_n$ are root-n consistent and asymptotically normal estimators of the elements of $A'A$ and can be correlated with $\hat{h}_n$. To account for the fact that the matrix $A'A$ contains duplicating elements, we introduce the following notation. Let ξ be the d-vector of the unique elements of $A'A$, and C_A be the $m^2 \times d$ selection matrix of zeros and ones such that $\operatorname{vec}(A'A) = C_A \xi$. Note that C_A is a known matrix. Also note that this representation allows for some known elements of $A'A$ to be zeros. Let $\hat{\xi}_n$ be an estimator of ξ. We can write $\operatorname{vec}(A_n' A_n) = C_A \hat{\xi}_n$.

Assumption S.1 **(a)** $n^{1/2}\left(\left(\hat{h}_n - h_0\right)', \left(\hat{\xi}_n - \xi\right)\right)' \to_d N(0, \Sigma_0)$, *where Σ_0 is a positive definite $(m+d) \times (m+d)$ matrix.*

(b) *A has full rank, and* $\operatorname{vec}(A'A) = C_A \xi$*, where* C_A *is known.*

(c) *There is* $\hat{\Sigma}_0$ *such that* $\hat{\Sigma}_0 \to_p \Sigma_0$.

This assumption is similar to condition (12) of Theorem 2 in Hall and Inoue (2003), HI hereafter. Let $V^A_{ff,0}$, $V^A_{gg,0}$, and $V^A_{fg,0}$ denote the asymptotic variance of $\hat{\theta}_n$, the asymptotic variance of $\hat{\gamma}_n$, and the asymptotic covariance of $\hat{\theta}_n$ and $\hat{\gamma}_n$ respectively:

$$\begin{aligned}
V^A_{ff,0} &= F_0^{-1} \frac{\partial f(\theta_0)'}{\partial \theta} D^A_{f,0} \Sigma_0 D^{A\prime}_{f,0} \frac{\partial f(\theta_0)}{\partial \theta'} F_0^{\prime -1}, \text{ where} \\
D^A_{f,0} &= \begin{pmatrix} A'A & \left(I_m \otimes (h_0 - f(\theta_0))'\right) C_A \end{pmatrix}, \\
V^A_{gg,0} &= G_0^{-1} \frac{\partial g(\gamma_0)'}{\partial \gamma} D^A_{g,0} \Sigma_0 D^{A\prime}_{g,0} \frac{\partial g(\gamma_0)}{\partial \gamma'} G_0^{\prime -1}, \text{ where} \\
D^A_{g,0} &= \begin{pmatrix} A'A & \left(I_m \otimes (h_0 - g(\gamma_0))'\right) C_A \end{pmatrix}; \\
V^A_{fg,0} &= F_0^{-1} \frac{\partial f(\theta_0)'}{\partial \theta} D^A_{f,0} \Sigma_0 D^{A\prime}_{g,0} \frac{\partial g(\gamma_0)}{\partial \gamma'} G_0^{\prime -1}, \text{ and} \\
V^A_0 &= \begin{pmatrix} V^A_{ff,0} & V^A_{fg,0} \\ V^{A\prime}_{fg,0} & V^A_{gg,0} \end{pmatrix}.
\end{aligned} \tag{S.2}$$

The joint asymptotic distribution of $\hat{\theta}_n$ and $\hat{\gamma}_n$ is given in the next theorem.

Theorem S.3 *Under Assumptions 2.4, 2.5 in HMT and S.1,*

$$n^{1/2} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} \to_d N\left(0_{(k+l)\times 1}, V^A_0\right).$$

The asymptotic variance of $\hat{\theta}_n$ and $\hat{\gamma}_n$ can be consistently estimated by the plug-in method, i.e. by replacing $h_0 - f(\theta_0)$ with $\hat{h}_n - f\left(\hat{\theta}_n\right)$ and so on.

As discussed in HI, Assumption S.1(a) rules out HAC-type estimators of $A'A$. To handle HAC-type estimators, one can impose a condition similar to Assumption 7 in HI. According to this assumption, $A_n'A_n$ is a consistent and asymptotically normal centered HAC estimator, however, its convergence rate is slower than $n^{-1/2}$. In this case, one can show that the CMD estimators are consistent and asymptotically normal with a slower than $n^{-1/2}$ convergence rate. The asymptotic distribution will be driven in this case only by estimation of $A'A$, as can be easily seen from the expansion in (2.8). Such an approach can also be used to extend the tests discussed in the next section to allow for HAC-type estimators $A_n'A_n$.

S.3 Model comparison with data-dependent weight matrices

In this section, we extend the tests in Sections 3 and 4 of HMT to the case when the weight matrices are data-dependent as described in Assumption S.1. The following result establishes the null asymptotic distribution of the QLR statistic in Section 3.1 of HMT for the nested models case.

Theorem S.4 *Suppose that Assumptions 2.3-2.5 in HMT and S.1 hold, and $\mathcal{G} \subset \mathcal{F}$. Then, under H_0,*

$$nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) \to_d Z'\Sigma_0^{1/2}\left(W_{g,0}^A - W_{f,0}^A\right)\Sigma_0^{1/2}Z,$$

where

$$\begin{aligned}
Z &\sim N\left(0, I_{m+d}\right), \\
W_{f,0}^A &= W_{f,0}^A(1) - W_{f,0}^A(2) - W_{f,0}^A(3) - W_{f,0}^A(4), \\
W_{f,0}^A(1) &= D_{f,0}^{A\prime}\frac{\partial f(\theta_0)}{\partial\theta'}F_0'^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}A'A\frac{\partial f(\theta_0)}{\partial\theta'}F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}D_{f,0}^A, \\
W_{f,0}^A(2) &= \begin{pmatrix} A'A & 0 \end{pmatrix}'\frac{\partial f(\theta_0)}{\partial\theta'}F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}D_{f,0}^A \\
&\quad + D_{f,0}^{A\prime}\frac{\partial f(\theta_0)}{\partial\theta'}F_0'^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}\begin{pmatrix} A'A & 0 \end{pmatrix}, \\
W_{f,0}^A(3) &= D_{f,0}^{A\prime}\frac{\partial f(\theta_0)}{\partial\theta'}F_0'^{-1}\left(M_{f,0}' + M_{f,0}\right)F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}D_{f,0}^A, \\
W_{f,0}^A(4) &= \begin{pmatrix} 0 & \left(I_m \otimes (h_0 - f(\theta_0))'\right)C_A \end{pmatrix}'\frac{\partial f(\theta_0)}{\partial\theta'}F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}D_{f,0}^A \\
&\quad + D_{f,0}^{A\prime}\frac{\partial f(\theta_0)}{\partial\theta'}F_0'^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}\begin{pmatrix} 0 & \left(I_m \otimes (h_0 - f(\theta_0))'\right)C_A \end{pmatrix}.
\end{aligned}$$

Here $D_{f,0}^A$ is defined in (S.2) and $W_{g,0}^A$ is defined similarly to $W_{f,0}^A$.

As in the fixed A case, the null asymptotic distribution is mixed χ^2. The mixing matrices $W_{f,0}^A$ and $W_{g,0}^A$ depend on unknown parameters, however, they can be consistently estimated by the plug-in method, and the critical values can be obtained by simulations as described in Section 3.1 of HMT.

The following result provides the null asymptotic distribution of the QLR statistic in Section 3.2 in the case of non-nested models.

Theorem S.5 *Suppose that Assumptions 2.3-2.5 in HMT and S.1 hold, and $\mathcal{F} \cap \mathcal{G} = \emptyset$. Then, under H_0, $n^{1/2}QLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) \to_d N\left(0, \omega_{A,0}^2\right)$, where $\omega_{A,0}$ is given by*

$$\left\|\Sigma_0^{1/2}\begin{pmatrix} 2A'A\left(f(\theta_0) - g(\gamma_0)\right) \\ C_A\left[\left(I_m \otimes (h_0 - g(\gamma_0))\right)(h_0 - g(\gamma_0)) - \left(I_m \otimes (h_0 - f(\theta_0))\right)(h_0 - f(\theta_0))\right] \end{pmatrix}\right\|.$$

Again, as in the case of fixed weight matrices, $\omega_{A,0}$ is strictly positive unless models are nested ($f(\theta_0) = g(\gamma_0)$), and the asymptotic variance $\omega_{A,0}^2$ can be consistently estimated by the plug-in method.

The averaged and sup-norm tests in Section 4 of HMT are derived under the assumption that the weight matrices A in $\mathbb{A}$ are known. This rules out many important cases where at least some A's in $\mathbb{A}$ are unknown but can be consistently estimated. Below we show how the results in Section 4 of HMT can be extended to allow for such weight matrices.

As for the previous results in this section, we replace Assumption 2.1 of HMT with Assumption S.1. The vector ξ now contains the unknown unique elements of A's in $\mathbb{A}$ The selection matrix C_A in part (b) of the assumption will be different for each A in $\mathbb{A}$. If some A consists only of known elements, its corresponding C_A is a matrix of zeros. Let

$$W_0(A) = W_{g,0}^A - W_{f,0}^A,$$

where $W_{g,0}^A$ and $W_{f,0}^A$ are defined in Theorem S.4. When f and g are nested, we have the following result.

Theorem S.6 *Suppose that Assumptions 2.3, 4.1 in HMT and S.1 hold, and $\mathcal{G} \subset \mathcal{F}$. Let $Z \sim N(0, I_{m+d})$. **(a)** Under H_0^a, $nAQLR_n \to_d Z'\Sigma_0^{1/2}(\int W_0(A)\pi(dA))\Sigma_0^{1/2}Z$. **(b)** Under H_0^s, $nSQLR_n \to_d \sup_{A\in\mathbb{A}}(Z'\Sigma_0^{1/2}W_0(A)\Sigma_0^{1/2}Z)$.*

For the non-nested case, define $s(A)$ as

$$\begin{pmatrix} 2A'A\left(f\left(\theta_0(A)\right) - g\left(\gamma_0(A)\right)\right) \\ C_A\left[\left(I_m \otimes \left(h_0 - g\left(\gamma_0(A)\right)\right)\right)\left(h_0 - g\left(\gamma_0(A)\right)\right) - \left(I_m \otimes \left(h_0 - f\left(\theta_0(A)\right)\right)\right)\left(h_0 - f\left(\theta_0(A)\right)\right)\right] \end{pmatrix}.$$

We have the following result.

Theorem S.7 *Suppose that Assumptions 2.3, 4.1 in HMT and S.1 hold, and $\mathcal{F} \cap \mathcal{G} = \emptyset$. Let $\{X(A) \in R : A \in \mathbb{A}\}$ be a mean zero Gaussian process such that the covariance of $X(A_1)$ and $X(A_2)$, $A_1, A_2 \in \mathbb{A}$, is $s(A_1)'\Sigma_0 s(A_2)$. **(a)** Under H_0^a, $n^{1/2}AQLR_n \to_d N(0, \int_{\mathbb{A}}\int_{\mathbb{A}} s(A_1)'\Sigma_0 s(A_2)\pi(dA_1)\pi(dA_2))$. **(b)** Under H_0^s, $\lim_{n\to\infty} P(n^{1/2}SQLR_n > c) \leq P(\sup_{A\in\mathbb{A}} X(A) > c)$.*

S.4 Model comparison with estimation and evaluation on different sets of reduced-form parameters

In the calibration literature, model parameters are often estimated or calibrated using one set of reduced-form characteristics, while model evaluation is conducted on another.

For example, a structural model can be estimated to match first moments, and evaluated with respect to second moments. This case is discussed in this section. It is analogous to out-of-sample model evaluation in the forecasting literature[1]. It also corresponds to the case of model comparison without lack-of-fit minimization in RV.

We find that when a model is estimated and evaluated on different sets of reduced-form parameters, the QLR statistic has asymptotically normal distribution regardless of whether f and g are nested or non-nested. The reason is that even when models are nested a bigger model does not necessarily provides a better fit, since the deep parameters are not calibrated to minimize the distance between the truth and the part of the model used for evaluation. This conclusion is in agreement with the results in Section 6 of RV.

Next, we introduce the notation and assumptions of this section. We partition $h_0 = \left(h_{1,0}', h_{2,0}'\right)'$, where $h_{1,0}$ is an m_1-vector, and $h_{2,0}$ is an m_2-vector, $m_1+m_2 = m$. Similarly, we partition $\hat{h}_n = \left(\hat{h}_{1,n}', \hat{h}_{2,n}'\right)'$, $f(\theta) = \left(f_1(\theta)', f_2(\theta)'\right)'$, and $g(\gamma) = \left(g_1(\gamma)', g_2(\gamma)\right)'$. Next, consider the weight matrices A_1 and A_2, where A_i is $m_i \times m_i$, $i = 1, 2$. At the estimation stage, the parameters are calibrated using only the first m_1 reduced-form characteristics and the weight matrix A_1:

$$\hat{\theta}_n(A_{1,n}) = \arg\min_{\theta\in\Theta} \left\| A_{1,n}\left(\hat{h}_{1,n} - f_1(\theta)\right)\right\|^2, \text{ and}$$
$$\hat{\gamma}_n(A_{1,n}) = \arg\min_{\gamma\in\Gamma} \left\| A_{1,n}\left(\hat{h}_{1,n} - g_1(\gamma)\right)\right\|^2.$$

At the evaluation stage, models are compared using the remaining m_2 reduced-form characteristics and the weight matrix A_2:

$$H_0 : \|A_2(h_{2,0} - f_2(\theta_0(A_1)))\| = \|A_2(h_{2,0} - g_2(\gamma_0(A_1)))\|. \tag{S.3}$$
$$H_f : \|A_2(h_{2,0} - f_2(\theta_0(A_1)))\| < \|A_2(h_{2,0} - g_2(\gamma_0(A_1)))\|. \tag{S.4}$$
$$H_g : \|A_2(h_{2,0} - f_2(\theta_0(A_1)))\| > \|A_2(h_{2,0} - g_2(\gamma_0(A_1)))\|. \tag{S.5}$$

We make the following assumption.

Assumption S.2 **(a)** *f_2 and g_2 are misspecified according to Definition 2.1.*

(b) *$A_{1,n} \to_p A_1$, $A_{2,n} \to_p A_2$; A_1 and A_2 have full ranks.*

(c) *Assumption 2.4 holds for A_1, f_1, and g_1.*

(d) $\frac{\partial f_2(\theta_0(A_1))'}{\partial\theta} A_2' A_2 (h_{2,0} - f_2(\theta_0(A_1))) \neq 0$; $\frac{\partial g_2(\gamma_0(A_1))'}{\partial\gamma} A_2' A_2 (h_{2,0} - g_2(\gamma_0(A_1))) \neq 0$.

[1]See, for example, West and McCracken (1998).

According to part (a) of the assumption, models are misspecified with respect to the second set of reduced-form parameters h_2. Note that the pseudo-true values of the parameters are defined with respect to A_1 and the first m_1 reduced-form characteristics. Consequently, the first-order condition (2.2) does not hold for f_2, g_2, h_2, and A_2, since $\theta_0(A_1)$ and $\gamma_0(A_1)$ are not the minimizers of the CMD criterion for the remaining m_2 reduced-form characteristics, as described in part (d).

The QLR statistic is now defined as

$$QLR_n\left(\hat{\theta}_n(A_{1,n}), \hat{\gamma}_n(A_{1,n}), A_{2,n}\right) = -\left\|A_{2,n}\left(\hat{h}_{2,n} - f_2\left(\hat{\theta}_n(A_{1,n})\right)\right)\right\|^2 + \left\|A_{2,n}\left(\hat{h}_{2,n} - g_2\left(\hat{\gamma}_n(A_{1,n})\right)\right)\right\|^2. \quad \text{(S.6)}$$

Define further

$$J_{f,0} = \left(-\frac{\partial f_2(\theta_0(A_1))}{\partial\theta'} F_{1,0}^{-1} \frac{\partial f_1(\theta_0(A_1))'}{\partial\theta} A_1' A_1 \quad I_{m_2} \right),$$
$$J_{g,0} = \left(-\frac{\partial g_2(\gamma_0(A_1))}{\gamma'} G_{1,0}^{-1} \frac{\partial g_1(\gamma_0(A_1))'}{\partial\gamma} A_1' A_1 \quad I_{m_2} \right),$$

where $F_{1,0}$ and $G_{1,0}$ are defined similarly to F_0 and G_0 in (2.6) and (2.7) respectively, but using A_1, $h_{1,0}$, f_1, and g_1. In the case of fixed weight matrices, we have the following result.

Theorem S.8 *Suppose that Assumptions 2.1 and S.2 hold, and $A_{1,n} = A_1, A_{2,n} = A_2$ for all n.*

(a) *Under H_0 in (S.3), $n^{1/2} QLR_n\left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2\right) \to_d N\left(0, \omega_{21,0}^2\right)$, where*

$$\omega_{21,0} = 2\left\|\Lambda_0^{1/2}\left(J_{g,0}' A_2' A_2\left(h_{2,0} - g_2(\gamma_0(A_1))\right) - J_{f,0}' A_2' A_2\left(h_{2,0} - f_2(\theta_0(A_1))\right)\right)\right\|.$$

(b) *Under H_f in (S.4), $n^{1/2} QLR_n\left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2\right) \to \infty$ with probability one; under the alternative H_g in (S.5), $n^{1/2} QLR_n\left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2\right) \to -\infty$ with probability one.*

As before the QLR statistic is asymptotically normal when models are non-nested. Now, however, it is asymptotically normal also in the nested case. This is because there is no minimization of the lack-of-fit functions in (S.6). Thus, when models are estimated using one set of reduced-form parameters and evaluated using another, one follows the rule regardless of whether models are nested, non-nested, or overlapping. One should reject the null of equivalent models when

$$n^{1/2}\left|QLR_n\left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2\right)\right| / \hat{\omega}_{21,n} > z_{1-\alpha/2},$$

where $\hat{\omega}_{21,n}$ is a consistent estimator of $\omega_{21,0}$. A consistent estimator of $\omega_{21,0}$ can be obtained by the plug-in method, since all the elements of $\omega_{21,0}$ can be consistently estimated. Note that, when $f_2(\theta_0(A_1)) = g_2(\gamma_0(A_1))$, which can occur if models are nested or overlapping, the columns corresponding to I_{m_2} in $J_{f,0}$ and $J_{g,0}$ do not contribute to the asymptotic variance; however, this will be reflected automatically by any consistent estimator $\hat{\omega}_{21,n}$.

When the weight matrices are data dependent, one can adjust the asymptotic variance of the QLR statistic in a manner similar to that in Theorem S.5.

S.5 Confidence sets for weighting schemes favorable to one of the models

In Section 4 of HMT, we discuss averaged and sup-norm tests that take into account models' relative performance under various choices of the weighting schemes. In this section we discuss another such approach. When all the considered models are misspecified, it is possible that model g provides a better approximation to one set of reduced-form characteristics, say h_1, and model f performs better on another set of h. In such a case, it might be of interest to see how large the weight of h_1 has to be for model g to be preferred to f overall. One can compare f and g by approximating the set of weighting schemes under which model g is preferred to model f. If this set is large, one can argue that model g is a viable alternative to f. On the other hand, if this set is very small in some sense, one can argue that g can be as good as f only under very special circumstances.

Let $\mathcal{A}_0$ be a collection of weighting schemes under which g is preferred to f:

$$\mathcal{A}_0 = \{A \in \mathbb{A} : \|A(h_0 - g(\gamma_0(A)))\| - \|A(h_0 - f(\theta_0(A)))\| \leq 0\}.$$

The confidence set (CS) for $\mathcal{A}_0$, denoted by $CS_{n,1-\alpha}$, is defined by the following condition

$$\lim_{n\to\infty} P(A \in CS_{n,1-\alpha}) \geq 1-\alpha \text{ for all } A \in \mathcal{A}_0.$$

The CS be constructed by inversion of the basic QLR test discussed in Section 3. First, given $A \in \mathbb{A}$, compute $QLR_n(A)$. Next, test $H_0 : A \in \mathcal{A}_0$ as follows: reject H_0 when $QLR_n(A) > z_{1-\alpha}\hat{\omega}_n/\sqrt{n}$, if models are non-nested. If models are nested, assuming that $\mathcal{G} \subset \mathcal{F}$, one can use the mixed χ^2 critical values as described in Section 3.1 to test H_0. If models are overlapping, one can apply the sequential procedure of Section 3.3. The confidence set $CS_{n,1-\alpha}$ is given by the collection of all A for which $H_0 : A \in \mathcal{A}_0$ cannot be rejected.

S.6 Proofs of the results in HMT

Proof of (2.8). First, note that by Theorem S.1, $\hat{\theta}_n \to_p \theta_0$. Next, applying the mean value expansion to $f\left(\hat{\theta}_n\right)$,

$$
\begin{aligned}
0 &= \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} A_n' A_n \left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right) \\
&= \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} A_n' A_n \left(\hat{h}_n - f(\theta_0) - \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'}\left(\hat{\theta}_n - \theta_0\right)\right) \\
&= \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} \left(A_n' A_n \left(\hat{h}_n - h_0\right) + \left(A_n' A_n - A'A\right)\left(h_0 - f(\theta_0)\right)\right) \\
&\quad + \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} A'A\left(h_0 - f(\theta_0)\right) - \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} A_n' A_n \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'}\left(\hat{\theta}_n - \theta_0\right),
\end{aligned}
$$

where $\widetilde{\theta}_n$ is the mean value. Next,

$$
\begin{aligned}
&\frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} A'A\left(h_0 - f(\theta_0)\right) \\
&= \left(I_k \otimes \left(h_0 - f(\theta_0)\right)' A'A\right) \operatorname{vec}\left(\frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'}\right) \\
&= \frac{\partial f(\theta_0)'}{\partial \theta} A'A\left(h_0 - f(\theta_0)\right) \\
&\quad + \left(I_k \otimes \left(h_0 - f(\theta_0)\right)' A'A\right) \frac{\partial}{\partial \theta'} \operatorname{vec}\left(\frac{\partial f\left(\overline{\theta}_n\right)}{\partial \theta'}\right)\left(\hat{\theta}_n - \theta_0\right) \\
&= M_{f,n}\left(\hat{\theta}_n - \theta_0\right),
\end{aligned}
\tag{S.7}
$$

where

$$
M_{f,n} = \left(I_k \otimes \left(h_0 - f(\theta_0)\right)' A'A\right) \frac{\partial}{\partial \theta'} \operatorname{vec}\left(\frac{\partial f\left(\overline{\theta}_n\right)}{\partial \theta'}\right), \tag{S.8}
$$

and $\overline{\theta}_n$ is the mean value. Note that the last equality in (S.7) follows from the population first-order condition (2.2). Define

$$
F_n = \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} A_n' A_n \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} - M_{f,n}. \tag{S.9}
$$

The result follows since by Theorem S.1, $M_{f,n} \to_p M_{f,0}$ and $F_n \to_p F_0$. □

Proof of Lemma 3.1. Let $h_{f,0} = f(\theta_0)$ and $h_{g,0} = g(\gamma_0)$. Then under the null of models equivalence we have $\|A(h_0 - h_{f,0})\| = \|A(h_0 - h_{g,0})\|$. However, since models are nested, $h_{g,0} \in \mathcal{F}$, and there should be some $\tilde{\theta}_0 \in \Theta$ such that $h_{g,0} = f\left(\tilde{\theta}_0\right)$ which violates Assumption 2.4 if $h_{f,0} \neq h_{g,0}$. $\square$

Proof of Theorem 3.1. In the case of the fixed weight matrix, using (2.8) the following expansion is obtained.

$$
\begin{aligned}
&n\left\|A\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 \\
&= n\left\|A\left(\hat{h}_n - f(\theta_0)\right)\right\|^2 + n^{1/2}\left(\hat{h}_n - h_0\right)' A'AW_{f,0}A'An^{1/2}\left(\hat{h}_n - h_0\right) + o_p(1). \qquad \text{(S.10)}
\end{aligned}
$$

To show (S.10), write

$$
\left\|A\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 = \left\|A\left(\hat{h}_n - f(\theta_0)\right)\right\|^2 + S_{1,n} + S_{2,n} + S_{3,n}, \qquad \text{(S.11)}
$$

where

$$
\begin{aligned}
S_{1,n} &= \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right)' A'A\left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right), \\
S_{2,n} &= -2\left(\hat{h}_n - h_0\right)' A'A\left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right), \\
S_{3,n} &= -2\left(h_0 - f(\theta_0)\right)' A'A\left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right).
\end{aligned}
$$

Now, one obtains (S.10) by expanding $f\left(\hat{\theta}_n\right)$ in $S_{1,n}$, $S_{2,n}$, and $S_{3,n}$ around $f(\theta_0)$ and using (2.8). Let $\widetilde{\theta}_n$ denote the mean value. For $S_{1,n}$ and $S_{2,n}$, we have

$$
\begin{aligned}
nS_{1,n} &= n^{1/2}\left(\hat{\theta}_n - \theta_0\right)' \frac{\partial f\left(\widetilde{\theta}_n\right)'}{\partial \theta} A'A \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} n^{1/2}\left(\hat{\theta}_n - \theta_0\right) \\
&= \left(n^{1/2}\left(\hat{h}_n - h_0\right)' A'A \frac{\partial f(\theta_0)}{\partial \theta'} F_0^{-1}\right) \frac{\partial f(\theta_0)'}{\partial \theta} A'A \frac{\partial f(\theta_0)}{\partial \theta'} \times \\
&\quad \times \left(F_0^{-1} \frac{\partial f(\theta_0)'}{\partial \theta} A'An^{1/2}\left(\hat{h}_n - h_0\right)\right) + o_p(1) \\
&= n^{1/2}\left(\hat{h}_n - h_0\right)' A'AW_{f,0}(1)\, A'An^{1/2}\left(\hat{h}_n - h_0\right) + o_p(1).
\end{aligned}
$$

$$
\begin{aligned}
nS_{2,n} &= -2n^{1/2}\left(\hat{h}_n - h_0\right)' A'A \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} n^{1/2}\left(\hat{\theta}_n - \theta_0\right) \\
&= -2n^{1/2}\left(\hat{h}_n - h_0\right)' A'A \frac{\partial f(\theta_0)}{\partial \theta'} \left(F_0^{-1} \frac{\partial f(\theta_0)'}{\partial \theta} A'An^{1/2}\left(\hat{h}_n - h_0\right)\right) + o_p(1) \\
&= -n^{1/2}\left(\hat{h}_n - h_0\right)' A'AW_{f,0}(2)\, A'An^{1/2}\left(\hat{h}_n - h_0\right) + o_p(1).
\end{aligned}
$$

In the case of $S_{3,n}$, after expanding $f\left(\hat{\theta}_n\right)$, one can apply the result in (S.7) to the term $(h_0 - f(\theta_0))' A'A\left(\partial f\left(\widetilde{\theta}_n\right)/\partial\theta'\right)$, which leads to $M_{f,0}$ in the expression for $W_{f,0}(3)$:

$$
\begin{aligned}
nS_{3,n} &= -2\left[n^{1/2}(h_0 - f(\theta_0))' A'A\frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial\theta'}\right] n^{1/2}\left(\hat{\theta}_n - \theta_0\right) \\
&= -2n^{1/2}\left(\hat{\theta}_n - \theta_0\right)' M_{f,n} n^{1/2}\left(\hat{\theta}_n - \theta_0\right) \\
&= -2\left(n^{1/2}\left(\hat{h}_n - h_0\right)' A'A\frac{\partial f(\theta_0)}{\partial\theta'}F_0^{-1}\right) M_{f,0}\left(F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}A'An^{1/2}\left(\hat{h}_n - h_0\right)\right) \\
&\quad + o_p(1) \\
&= -n^{1/2}\left(\hat{h}_n - h_0\right)' A'AW_{f,0}(3) A'An^{1/2}\left(\hat{h}_n - h_0\right) + o_p(1).
\end{aligned}
$$

An expansion similar to (S.10) is available for $n\left\|A\left(\hat{h}_n - g(\hat{\gamma}_n)\right)\right\|^2$ with f, θ, and F replaced by g, γ, and G. Hence,

$$
\begin{aligned}
nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) &= -n\left\|A\left(\hat{h}_n - f(\theta_0)\right)\right\|^2 + n\left\|A\left(\hat{h}_n - g(\gamma_0)\right)\right\|^2 \\
&\quad + n^{1/2}\left(\hat{h}_n - h_0\right)' A'A\left(W_{g,0} - W_{f,0}\right) A'An^{1/2}\left(\hat{h}_n - h_0\right) + o_p(1). \qquad \text{(S.12)}
\end{aligned}
$$

Under the null, the first summand on the right-hand side of (S.12) is zero by Lemma 3.1, and the result in part (a) of the theorem follows by Assumption 2.1(a).

Since under H_f, $\|A(h_0 - f(\theta_0))\|^2 \leq \|A(h_0 - g(\gamma_0))\|^2$, part (b) of the theorem follows from (S.12). □

Proof of Theorem 3.2. From (S.10), by adding and subtracting h_0, we obtain

$$
\left\|A\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 = \|A(h_0 - f(\theta_0))\|^2 + 2(h_0 - f(\theta_0))' A'A\left(\hat{h}_n - h_0\right) + O_p\left(n^{-1}\right),
$$

with a similar expression for $\left\|A\left(\hat{h}_n - g(\hat{\gamma}_n)\right)\right\|^2$. Hence,

$$
\begin{aligned}
QLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) &= -\|A(h_0 - f(\theta_0))\|^2 + \|A(h_0 - g(\gamma_0))\|^2 \\
&\quad + 2(f(\theta_0) - g(\gamma_0))' A'A\left(\hat{h}_n - h_0\right) + O_p\left(n^{-1}\right). \qquad \text{(S.13)}
\end{aligned}
$$

Since $\mathcal{F} \cap \mathcal{G} = \emptyset$, we have that $f(\theta_0) \neq g(\gamma_0)$, and the result follows from Assumption 2.1(a). □

Proof of Theorem 4.1. First, note that in the case of nested models for all $A \in \mathbb{A}$,

$\|A(h_0 - g(\gamma_0(A)))\|^2 \geq \|A(h_0 - f(\theta_0(A)))\|^2$, and thus, under H_0^a, we have that for all $A \in \mathbb{A}$, $\|A(h_0 - g(\gamma_0(A)))\|^2 = \|A(h_0 - f(\theta_0(A)))\|^2$.

We show next that under H_0^a, $nQLR_n\left(\hat{\theta}_n(A), \hat{\gamma}_n(A), A\right)$ converges weakly to a stochastic process indexed by A. According to Theorem (10.2) of Pollard (1990), for weak convergence one needs to show finite dimensional convergence and stochastic equicontinuity of $nQLR_n\left(\hat{\theta}_n(A), \hat{\gamma}_n(A), A\right)$ with respect to A. Finite dimensional convergence follows by the same arguments as in the proof of Theorem 3.1.

For stochastic equicontinuity, similarly to (S.10) one can show that for all $A \in \mathbb{A}$,

$$\begin{aligned}
&\left\|A\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 \\
&= \left\|A\left(\hat{h}_n - f(\theta_0)\right)\right\|^2 + \left(\hat{h}_n - h_0\right)' A'AW_{f,n}A'A\left(\hat{h}_n - h_0\right) + o_p\left(n^{-1}\right),
\end{aligned}$$

where

$$\begin{aligned}
W_{f,n} &= W_{f,n}(1) - W_{f,n}(2) - W_{f,n}(3), \\
W_{f,n}(1) &= \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n'^{-1} \frac{\partial f\left(\widetilde{\theta}_n\right)'}{\partial \theta} A'A \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta}, \\
W_{f,n}(2) &= \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n'^{-1} \frac{\partial f\left(\widetilde{\theta}_n\right)'}{\partial \theta} + \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta}, \\
W_{f,n}(3) &= \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n'^{-1}\left(M_{f,n}' + M_{f,n}\right) F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta},
\end{aligned}$$

and F_n, $M_{f,n}$ are as defined in (S.9) and (S.8) respectively. A similar expansion holds for $\left\|A\left(\hat{h}_n - g(\hat{\gamma}_n)\right)\right\|^2$, and we can write

$$\begin{aligned}
n&\left|QLR_n\left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_1\right) - QLR_n\left(\hat{\theta}_n(A_2), \hat{\gamma}_n(A_2), A_2\right)\right| \\
&\leq n\left\|\hat{h}_n - h_0\right\|^2 K_n \|A_1 - A_2\|^\delta + o_p(1), \qquad \text{(S.14)}
\end{aligned}$$

where $\delta > 0$, $K_n = O_p(1)$ and independent of $(A_1 - A_2)$. Since $W_{f,n}$ and $W_{g,n}$, where $W_{g,n}$ is defined similarly to $W_{f,n}$, are continuous in A, the reminder term in (S.14) is $o_p(1)$ uniformly in A. Stochastic equicontinuity of $nQLR_n\left(\hat{\theta}_n(A), \hat{\gamma}_n(A), A\right)$ follows from Lemma 2(a) of Andrews (1992).

The results of the theorem follow now from weak convergence by the continuous mapping theorem (CMT). □

Proof of Theorem 4.2. Convergence of finite dimensional distributions and stochastic equicontinuity can be established from (S.13). The results of the theorem will follow by the CMT. □

S.7 Proofs of the results in Sections S.2-S.4

Proof of Theorem S.1. For consistency of $\hat{\theta}_n$, it is sufficient to show uniform convergence of $\left\|A_n\left(\hat{h}_n - f(\theta)\right)\right\|^2$ to $\|A(h_0 - f(\theta))\|^2$ on Θ. The desired result will follow from Assumptions 2.4 and 2.5 by the usual argument for extremum estimators (see, for example, Theorem 2.1 in Newey and McFadden (1994)).

$$\left\|A_n\left(\hat{h}_n - f(\theta)\right)\right\|^2 - \|A(h_0 - f(\theta))\|^2 = R_{1,n} - 2R_{2,n}(\theta) + R_{3,n}(\theta), \text{ where}$$

$$\begin{aligned} R_{1,n} &= \hat{h}_n A_n' A_n \hat{h}_n - h_0' A_n' A_n h_0, \\ R_{2,n}(\theta) &= \left(\hat{h}_n - h_0\right)' A_n' A_n f(\theta) \\ R_{3,n}(\theta) &= (h_0 - f(\theta))' \left(A_n' A_n - A'A\right)(h_0 - f(\theta)). \end{aligned}$$

By Assumption 2.1(a) and 2.2, $|R_{1,n}| \to_p 0.$. Due to Assumption 2.5 (a) and (c), f is bounded on Θ (Davidson, 1994, Theorem 2.19), and therefore,

$$\begin{aligned} \sup_{\theta\in\Theta} |R_{2,n}(\theta)| &\leq \left\|A_n' A_n\right\| \left\|\left(\hat{h}_n - h_0\right)\right\| \sup_{\theta\in\Theta} \|f(\theta)\| \\ &\to_p 0, \end{aligned}$$

by Assumptions 2.1(a) and 2.2. Similarly,

$$\begin{aligned} \sup_{\theta\in\Theta} |R_{3,n}(\theta)| &\leq \left\|A_n' A_n - A'A\right\| \sup_{\theta\in\Theta} \|h_0 - f(\theta)\|^2 \\ &\leq \left\|A_n' A_n - A'A\right\| \left(\|h_0\| + \sup_{\theta\in\Theta} \|f(\theta)\|\right)^2 \\ &\to_p 0. \end{aligned}$$

The proof of $\hat{\gamma}_n \to_p \gamma_0$ is identical with f and θ replaced by g and γ respectively. □

Proof of Theorem S.2. One can expand the first-order conditions for $\hat{\gamma}_n$ similarly to that of $\hat{\theta}_n$, equation (2.8). Taking into account that $A_n = A$ for all n,

$$n^{1/2} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} = \begin{pmatrix} F_n^{-1} \frac{\partial f(\hat{\theta}_n)'}{\partial \theta} \\ G_n^{-1} \frac{\partial g(\hat{\gamma}_n)'}{\partial \gamma} \end{pmatrix} A'A n^{1/2} \left(\hat{h}_n - h_0\right),$$

where

$$G_n = \frac{\partial g(\hat{\gamma}_n)'}{\partial \gamma} A_n' A_n \frac{\partial g(\widetilde{\gamma}_n)}{\partial \gamma'} - M_{g,n},$$
$$M_{g,n} = \left(I_l \otimes (h_0 - g(\gamma_0))' A'A\right) \frac{\partial}{\partial \gamma'} \text{vec}\left(\frac{\partial g(\overline{\gamma}_n)}{\partial \gamma'}\right),$$

and $\widetilde{\gamma}_n, \overline{\gamma}_n$ are between $\hat{\gamma}_n$ and γ_0. The result follows from Theorem S.1, Assumptions 2.1(a) and 2.5(e). □

Proof of Theorem S.3. The result follows immediately from (2.8), a similar expansion for $\hat{\gamma}_n$, and the assumptions of the theorem by writing

$$\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} =$$
$$= \begin{pmatrix} F_n^{-1} \frac{\partial f(\hat{\theta}_n)'}{\partial \theta} \begin{pmatrix} A_n' A_n & I_m \otimes (h_0 - f(\theta_0))' \end{pmatrix} \\ G_n^{-1} \frac{\partial g(\hat{\gamma}_n)'}{\partial \gamma} \begin{pmatrix} A_n' A_n & I_m \otimes (h_0 - g(\gamma_0))' \end{pmatrix} \end{pmatrix} \begin{pmatrix} \hat{h}_n - h_0 \\ \text{vec}(A_n' A_n - A'A) \end{pmatrix}$$
$$= \begin{pmatrix} F_n^{-1} \frac{\partial f(\hat{\theta}_n)'}{\partial \theta} \begin{pmatrix} A_n' A_n & \left(I_m \otimes (h_0 - f(\theta_0))'\right) C_A \end{pmatrix} \\ G_n^{-1} \frac{\partial g(\hat{\gamma}_n)'}{\partial \gamma} \begin{pmatrix} A_n' A_n & \left(I_m \otimes (h_0 - g(\gamma_0))'\right) C_A \end{pmatrix} \end{pmatrix} \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}.$$

□

Proof of Theorem S.4. As in the proof of Theorem 3.1, write

$$\left\| A_n \left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right) \right\|^2 = \left\| A_n \left(\hat{h}_n - f(\theta_0)\right) \right\|^2 + S_{1,n}^A + S_{2,n}^A + S_{3,n}^A, \qquad \text{(S.15)}$$

where

$$S_{1,n}^A = \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right)' A_n' A_n \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right),$$
$$S_{2,n}^A = -2\left(\hat{h}_n - h_0\right)' A_n' A_n \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right),$$
$$S_{3,n}^A = -2\left(h_0 - f(\theta_0)\right)' A_n' A_n \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right).$$

Under the null in the nested case, $f(\theta_0) = g(\theta_0)$, and therefore $\left\| A_n \left(\hat{h}_n - f(\theta_0)\right) \right\|^2 = \left\| A_n \left(\hat{h}_n - g(\gamma_0)\right) \right\|^2$. Define

$$D_{f,n}^A = \begin{pmatrix} A_n' A_n & \left(I_m \otimes (h_0 - f(\theta_0))'\right) C_A \end{pmatrix}.$$

By expanding $f\left(\hat{\theta}_n\right)$ around $f(\theta_0)$ and using (2.8), we obtain the following expression for $S^A_{1,n}$:

$$\begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}' D^{A\prime}_{f,n} \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n'^{-1} \frac{\partial f\left(\tilde{\theta}_n\right)'}{\partial \theta} A_n' A_n \\ \times \frac{\partial f\left(\tilde{\theta}_n\right)}{\partial \theta'} F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} D^A_{f,n} \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix},$$

where $\tilde{\theta}_n$ is the mean value. Similarly, for $S^A_{2,n}$ we obtain

$$\begin{aligned} &\begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}' \begin{pmatrix} A_n' A_n \\ 0 \end{pmatrix} \frac{\partial f\left(\tilde{\theta}_n\right)}{\partial \theta'} F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} D^A_{f,n} \times \\ &\times \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix} \\ &+ \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}' D^{A\prime}_{f,n} \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n'^{-1} \frac{\partial f\left(\tilde{\theta}_n\right)'}{\partial \theta} \begin{pmatrix} A_n' A_n \\ 0 \end{pmatrix}' \\ &\times \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}. \end{aligned}$$

Next, for $S^A_{3,n}$ write

$$\begin{aligned} -2S^A_{3,n} = &\left(h_0 - f(\theta_0)\right)' A' A \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right) + \\ &+ \left(h_0 - f(\theta_0)\right)' \left(A_n' A_n - A' A\right) \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right). \end{aligned} \tag{S.16}$$

For the first summand on the right-hand side of (S.16), applying the mean-value expansion to $f\left(\hat{\theta}_n\right)$ around $f(\theta_0)$ and by (S.7), we obtain

$$\begin{aligned} &\left(h_0 - f(\theta_0)\right)' A' A \left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right) = \\ &= \left(h_0 - f(\theta_0)\right)' A' A \frac{\partial f\left(\tilde{\theta}_n\right)}{\partial \theta'} \left(\hat{\theta}_n - \theta_0\right) \\ &= \left(\hat{\theta}_n - \theta_0\right)' M_{f,n} \left(\hat{\theta}_n - \theta_0\right) \\ &= \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}' D^{A\prime}_{f,n} \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n'^{-1} M_{f,n} \end{aligned}$$

$$\times F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} D_{f,n}^A \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}.$$

For the second summand on the right-hand side of (S.16), write

$$\begin{aligned}
&(h_0 - f(\theta_0))' \left(A_n' A_n - A'A\right)\left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right) = \\
&= \left(\hat{\xi}_n - \xi\right)' C_A' \left(I_m \otimes (h_0 - f(\theta_0))\right)\left(f\left(\hat{\theta}_n\right) - f(\theta_0)\right) \\
&= \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}' \begin{pmatrix} 0 \\ C_A' \left(I_m \otimes (h_0 - f(\theta_0))\right) \end{pmatrix} \frac{\partial f\left(\tilde{\theta}_n\right)}{\partial \theta'} \\
&\quad \times F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta} D_{f,n}^A \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix}.
\end{aligned}$$

The result of the theorem follows by Assumption S.1 from the above expressions for $S_{1,n}, S_{2,n}, S_{3,n}$ and by establishing a similar expansion for $\left\|A_n\left(\hat{h}_n - g(\hat{\gamma}_n)\right)\right\|^2$. □

Proof of Theorem S.5. From (S.15) we have

$$\begin{aligned}
&\left\|A_n\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 - \left\|A(h_0 - f(\theta_0))\right\|^2 \\
&= (h_0 - f(\theta_0))' \left(A_n' A_n - A'A\right)(h_0 - f(\theta_0)) + 2(h_0 - f(\theta_0))' A_n' A_n \left(\hat{h}_n - h_0\right) \\
&\quad + O_p\left(n^{-1}\right) \\
&= \begin{pmatrix} 2(h_0 - f(\theta_0))' A_n' A_n & (h_0 - f(\theta_0))' \left(I_m \otimes (h_0 - f(\theta_0))'\right) \end{pmatrix} \\
&\quad \times \begin{pmatrix} \hat{h}_n - h_0 \\ \operatorname{vec}\left(A_n' A_n - A'A\right) \end{pmatrix} + O_p\left(n^{-1}\right) \\
&= \begin{pmatrix} 2(h_0 - f(\theta_0))' A_n' A_n & (h_0 - f(\theta_0))' \left(I_m \otimes (h_0 - f(\theta_0))'\right) C_A \end{pmatrix} \\
&\quad \times \begin{pmatrix} \hat{h}_n - h_0 \\ \hat{\xi}_n - \xi \end{pmatrix} + O_p\left(n^{-1}\right).
\end{aligned}$$

Using a similar expansion for $\left\|A_n\left(\hat{h}_n - g(\hat{\gamma}_n)\right)\right\|^2$, the result follows by Assumption S.1. □

The proof of Theorem S.6 is similar to that of Theorems 4.1 and S.4, and therefore omitted. The proof of Theorem S.7 is similar to that of Theorems 4.2 and S.5, and therefore omitted.

Proof of Theorem S.8. From (S.11), $\left\|A_2\left(\hat{h}_{2,n}-f_2\left(\hat{\theta}_n\left(A_1\right)\right)\right)\right\|^2$ can be expanded as

$$
\begin{aligned}
&\left\|A_2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)\right\|^2+2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2\left(\hat{h}_{2,n}-h_{2,0}\right) \\
&\quad-2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2 \frac{\partial f_2\left(\theta_0\left(A_1\right)\right)}{\partial \theta'}\left(\hat{\theta}_n\left(A_1\right)-\theta_0\left(A_1\right)\right)+o_p\left(n^{-1/2}\right) \\
&=\left\|A_2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)\right\|^2 \\
&\quad-2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2 \frac{\partial f_2\left(\theta_0\left(A_1\right)\right)}{\partial \theta'} F_{1,0}^{-1} \frac{\partial f_1\left(\theta_0\left(A_1\right)\right)'}{\partial \theta} A_1' A_1\left(\hat{h}_{1,n}-h_{1,0}\right) \\
&\quad+2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2\left(\hat{h}_{2,n}-h_{2,0}\right) \\
&\quad+o_p\left(n^{-1/2}\right) \\
&=\left\|A_2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)\right\|^2+2\left(h_{2,0}-f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2 J_{f,0}\left(\hat{h}_n-h_0\right) \\
&\quad+o_p\left(n^{-1/2}\right).
\end{aligned}
$$

□

References

Andrews, D. W. K., 1992. Generic uniform convergence. Econometric Theory 8 (2), 241–257.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, New York.

Davidson, R., MacKinnon, J. G., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 49 (3), 781–793.

Hall, A. R., Inoue, A., 2003. The large sample behavior of the generalized method of moments estimator in misspecified models. Journal of Econometrics 114 (2), 361–394.

Hnatkovska, V., Marmer, V., Tang, Y., 2011. Comparison of misspecified calibrated models: The minimum distance approach, Working Paper, University of British Columbia.

Newey, W. K., McFadden, D. L., 1994. Large sample estimation and hypothesis testing. In: Engle, R. F., McFadden, D. L. (Eds.), Handbook of Econometrics. Vol. IV. Elsevier, Amsterdam, Ch. 36, pp. 2111–2245.

Pollard, D., 1990. Empirical Processes: Theory and Applications. Vol. 2 of NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Mathematical Statistics, Hayward, California.

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

Smith, R. J., 1992. Non-nested tests for competing models estimated by generalized method of moments. Econometrica 60 (4), 973–980.

Vuong, Q. H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307–333.

West, K. D., McCracken, M. W., 1998. Regression-based tests of predictive ability. International Economic Review 39, 817–840.