Non-Nested Testing in Models Estimated via Generalized Method of Moments[1]

Alastair R. Hall

University of Manchester[2]

and

Denis Pelletier

North Carolina State University[3]

May 3, 2010

[1]We thank Atsushi Inoue, Eric Renault, Quang Vuong, Ken West, a co-editor and two anonymous referees for helpful comments.

[2]Corresponding author: Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK. Email: alastair.hall@manchester.ac.uk

[3]Department of Economics, Box 8110, North Carolina State University, Raleigh, NC 27695-8110, USA. Email: denis_pelletier@ncsu.edu

Abstract

We analyze the limiting distribution of Rivers and Vuong's (2002) statistic for choosing between two competing dynamic models based on a comparison of GMM minimands. It is shown that: (i) if both models are misspecified then the statistic has a standard normal distribution under the null hypothesis of equal fit but the ranking could be determined by the choice of the weighting matrix; (ii) if both models are correctly specified or locally misspecified then the limiting distribution of the test statistic is non-standard under the null.

JEL classification: C10, C32

Keywords: Generalized Method of Moments, Non-nested Hypothesis Testing, Model Selection.

1 Introduction

Competing economics theories often lead to econometric models that are non-nested in the sense that one model is not obtained as a special case of the other. It is, therefore, of interest to develop statistical procedures that discriminate between non-nested models. In these circumstances, it may be considered attractive to have some method that allows the researcher to determine which - if either - of the two models is closer to the truth in some sense. Vuong (1989) provides such a test for models estimated by Quasi Maximum Likelihood (QML). More recently, Rivers and Vuong (2002) extend Vuong's (1989) approach to provide a very general framework for the comparison of two competing dynamic models. In this more general context, inference is based on a test statistic that compares measures of goodness of fit for the two models; one model is preferred if its goodness of fit is statistically significantly smaller than its competitor. Rivers and Vuong (2002) provide generic conditions under which the statistic has a limiting standard normal distribution under the null hypothesis that both models are "equally good", a concept that is defined below. These generic conditions are very general and it is argued that they cover the situation in which the competing models are estimated via GMM and then compared using either the GMM minimands employed in the estimations or GMM type minimands that are different from those used in the estimation.[1] In spite of this seeming generality, Rivers and Vuong (2002) show that the aforementioned distributional result rests crucially on the assumption that a certain variance is non-zero.

In this paper, we investigate whether these generic conditions in fact cover GMM estimators and minimands. It turns out that the analysis depends crucially on whether the models in question are correctly specified, locally misspecified or non-locally misspecified. It is shown that if both models are correctly specified or locally misspecified then Rivers and Vuong's (2002) generic conditions are not satisfied and the statistic does not converge to a limiting normal distribution but to a non-standard distribution that is a function of nuisance parameters, some of which may not be consistently estimable. However, if both models are non-locally misspecified then the generic conditions are satisfied and the Rivers and Vuong's (2002) statistic does converge to the limiting standard normal distribution. The latter result indicates that there is scope for using the Rivers and Vuong statistic to compare two misspecified models estimated via GMM although caution may need to be exercised in its use because the outcome can depend on the choice of weighting matrix. Thus our results reveal important limitations to the use of GMM minimands for model selection in this way, which in turn contribute to the wider literature on

[1] See Rivers and Vuong (2002)[p.3 and p.13].

model selection via goodness of fit in econometrics; see *e.g.* Rivers and Vuong (2002), Marcellino and Rossi (2008) and Kitamura (2002).

2 Framework and Analysis

For $i = 1, 2$, let $f^{(i)} : \mathbf{V} \times \Theta^{(i)} \rightarrow \Re^{q_i}$, where $q_i < \infty$, $\Theta^{(i)} \subset \Re^{p_i}$ and let $\{v_t\}$ be a stationary and ergodic sequence of $d-$ dimensional random vectors in $\mathbf{V}$. Suppose it is desired to compare two models denoted $\mathcal{M}_1$ and $\mathcal{M}_2$, and that each implies a population moment condition as follows:

$$\begin{aligned} \mathcal{M}_1 \;&\Rightarrow\; E[f^{(1)}(v_t, \bar{\theta}^{(1)})] \;=\; 0 \qquad \text{for a unique } \bar{\theta}^{(1)} \in \Theta^{(1)}, \\ \mathcal{M}_2 \;&\Rightarrow\; E[f^{(2)}(v_t, \bar{\theta}^{(2)})] \;=\; 0 \qquad \text{for a unique } \bar{\theta}^{(2)} \in \Theta^{(2)}. \end{aligned}$$

It is assumed that the models are non-nested, *i.e.* the moment conditions of one model are not a subset or sub-case of another model. It is also assumed that the parameters of both models are estimated via GMM; these estimators are defined as follows:

$$\hat{\theta}_T^{(i)} \;=\; argmin_{\theta^{(i)} \in \Theta^{(i)}} \, Q_T^{(i)}(\theta^{(i)}), \qquad \text{for } i = 1, 2 \tag{1}$$

where

$$Q_T^{(i)}(\theta^{(i)}) = \left\{ T^{-1} \sum_{t=1}^{T} f^{(i)}(v_t, \theta^{(i)}) \right\}' W_T^{(i)} \left\{ T^{-1} \sum_{t=1}^{T} f^{(i)}(v_t, \theta^{(i)}) \right\} \tag{2}$$

and $W_T^{(i)}$ is the weighting matrix. The population analog of the GMM minimands is for $i = 1, 2$,

$$Q_0^{(i)}(\theta^{(i)}) = E[f^{(i)}(v_t, \theta^{(i)})]' W^{(i)} E[f^{(i)}(v_t, \theta^{(i)})]. \tag{3}$$

where $W^{(i)} = plim_{T \rightarrow \infty} W_T^{(i)}$.

Rivers and Vuong (2002) introduce a very general framework that includes the cases where the metric of model comparison either involves the minimands employed in the estimation or some other measure of goodness of fit. We consider the case in which the metric involves the GMM minimands and so the test statistic is:

$$N_T \;=\; \frac{T^{1/2} \{ \, Q_T^{(1)}(\hat{\theta}_T^{(1)}) \;-\; Q_T^{(2)}(\hat{\theta}_T^{(2)}) \, \}}{\hat{\sigma}_T} \tag{4}$$

where $\hat{\sigma}_T^2$ is a consistent estimator of σ_0^2, the limiting variance of the numerator of (4). This variance and its estimator are discussed below.

To present the null and alternative hypotheses of the test, we must introduce notation for the probability limits of $\hat{\theta}_T^{(1)}$ and $\hat{\theta}_T^{(2)}$. Accordingly, we define $plim_{T\to\infty}\hat{\theta}_T^{(i)} = \theta_*^{(i)}$ for $i = 1, 2$. This convergence result can be established under certain regularity conditions which are omitted for brevity here as they are now standard in the literature. The null hypothesis of the test is that: $\mathcal{M}_1$ and $\mathcal{M}_2$ are asymptotically equivalent, that is

$$H_0 \ : \ Q_0^{(1)}(\theta_\star^{(1)}) \ = \ Q_0^{(2)}(\theta_\star^{(2)}). \tag{5}$$

There are two alternative hypotheses of interest: $\mathcal{M}_1$ is asymptotically better than $\mathcal{M}_2$, that is

$$H_1^{(a)} \ : \ Q_0^{(1)}(\theta_\star^{(1)}) \ < \ Q_0^{(2)}(\theta_\star^{(2)}) \tag{6}$$

and $\mathcal{M}_2$ is asymptotically better than $\mathcal{M}_1$, that is

$$H_1^{(b)} \ : \ Q_0^{(1)}(\theta_\star^{(1)}) \ > \ Q_0^{(2)}(\theta_\star^{(2)}). \tag{7}$$

Rivers and Vuong (2002) present regularity conditions under which N_T converges to a standard normal distribution under H_0. For the purposes of our subsequent analysis, it is useful to highlight just one of these conditions, namely $\sigma_0^2 > 0$.

Apart from the standard assumption that the weighting matrix $W_T^{(i)}$ is positive semi-definite and converges in probability to a positive definite limit, it is assumed that $W^{(i)}$ depends on a vector of nuisance parameters $\tau_0^{(i)}$ and that $\hat{\tau}_T^{(i)}$ is an estimator of $\tau_0^{(i)}$. So that we have, with an obvious abuse of notation, $W^{(i)} = W^{(i)}(\tau_0^{(i)})$ and $W_T^{(i)} = W_T^{(i)}(\hat{\tau}_T^{(i)})$. It is assumed that the nuisance parameters satisfy:

$$T^{1/2}(\hat{\tau}_T^{(i)} \ - \ \tau_0^{(i)}) \ = \ -A_\star^{(i)}\, T^{-1/2}\sum_{t=1}^{T} Y_t^{(i)} \ + \ o_p(1) \tag{8}$$

for some symmetric matrix of constants $A_\star^{(i)}$ and vector $Y_t^{(i)}$; and that the weighting matrix satisfies

$$T^{1/2}\left(vech[W_T^{(i)}] - vech[W^{(i)}]\right) \ = \ \Delta^{(i)} T^{1/2}(\hat{\tau}_T^{(i)} \ - \ \tau_0^{(i)}) \ + \ o_p(1) \tag{9}$$

for some matrix of constants $\Delta^{(i)}$. The definitions of $A_\star^{(i)}$, $Y_t^{(i)}$ and $\Delta^{(i)}$ depend on the choice of weighting matrix.

Within our framework of GMM minimands with stationary processes, σ_0^2 has the following form:

$$\sigma_0^2 \ = \ R_\star^{'} V_\star R_\star \tag{10}$$

where

$$V_\star \ = \ \lim_{T\to\infty} \frac{1}{T} Var\left[\sum_{t=1}^{T} \xi_t(\theta_\star)\right] \tag{11}$$

for

$$\xi_t(\theta_\star) \;=\; \left[f^{(1)}(v_t,\theta_\star^{(1)})' - E[f^{(1)}(v_t,\theta_\star^{(1)})'], Y_t^{(1)'}, f^{(2)}(v_t,\theta_\star^{(2)})' - E[f^{(2)}(v_t,\theta_\star^{(2)})'], Y_t^{(2)'}\right]' \quad (12)$$

and

$$R_\star \;=\; \left[R_\star^{(1)'}, -R_\star^{(2)'}\right]', \qquad R_\star^{(i)} \;=\; \left[\begin{array}{c} 2W^{(i)}E[f(v_t,\theta_\star^{(i)})] \\ -A_\star^{(i)}\Delta^{(i)\prime}B_i'\{E[f^{(i)}(v_t,\theta_\star^{(i)})]\otimes E[f^{(i)}(v_t,\theta_\star^{(i)})]\} \end{array}\right] \quad (13)$$

where B_i is the $q_i^2 \times q_i(q_i+1)/2$ matrix such that $vec(W^{(i)}) = B_i vech(W^{(i)})$, and $A_\star^{(i)}$, $Y_t^{(i)}$ and $\Delta^{(i)}$ are defined implicitly in (8)-(9).

Some additional notation will be useful. On occasion, it is convenient to combine the parameters and moment functions from both models into one vector and so we define $\theta = [\theta^{(1)\prime}, \theta^{(2)\prime}]'$, $f(v_t,\theta) = [f^{(1)}(v_t,\theta^{(1)})', f^{(2)}(v_t,\theta^{(2)})']'$, $g_T(\theta) = [g_T^{(1)}(\theta^{(1)})', g_T^{(2)}(\theta^{(2)})']'$ for $g_T^{(i)}(\theta^{(i)}) = T^{-1}\sum_{t=1}^T f^{(i)}(v_t,\theta^{(i)})$, $G_T^{(i)}(\theta^{(i)}) = T^{-1}\sum_{t=1}^T \partial f^{(i)}(v_t,\theta^{(i)})/\partial\theta^{(i)\prime}$, $G_0^{(i)}(\theta^{(i)}) = E[\partial f^{(i)}(v_t,\theta^{(i)})/\partial\theta^{(i)\prime}]$ and $G_T^{(i)}(\theta_1,\theta_2,\lambda)$ is the $(q_i \times p_i)$ matrix whose j^{th} row is the corresponding row of $G_T^{(i)}(\bar\theta^{(j)})$ where $\bar\theta^{(j)} = \lambda_j\theta_1 + (1-\lambda_j)\theta_2$ for some $0 \le \lambda_j \le 1$, and λ is the $(q \times 1)$ vector with j^{th} element λ_j. Finally, we denote the Choleski decomposition of a matrix S by $S^{1/2}$ such that $S = S^{1/2}S^{1/2\prime}$ and we denote the inverse of $S^{1/2}$ by $S^{-1/2} \equiv [S^{1/2}]^{-1}$.

We now consider the limiting behaviour of N_T under H_0. To facilitate this discussion, it is useful to relate the H_0 in (5) to the scenarios regarding the population moment condition employed in the analysis of model specification tests within the GMM framework. There are three such scenarios: (i) the moment condition is correctly specified, $E[f^{(i)}(v_t,\theta_*^{(i)})] = 0$; (ii) the moment condition is locally misspecified, $E[f^{(i)}(v_t,\theta_*^{(i)})] = T^{-1/2}\eta^{(i)}$, $\eta^{(i)} \neq 0$; (iii) the moment condition is misspecified, $E[f^{(i)}(v_t,\theta^{(i)})] \neq 0$ $\forall\,\theta^{(i)} \in \Theta^{(i)}$, the so-called fixed or non-local alternative. Notice that H_0 in (5) can hold under any of (i)-(iii).[2]

For pedagogic convenience, we begin with the case of non-local misspecification, that is

Assumption 1 *$\mathcal{M}_i$ satisfies: (i) $E[f(v_t,\theta)] = \mu(\theta)$ where $\mu(\theta) = [\mu^{(1)}(\theta^{(1)})', \mu^{(2)}(\theta^{(2)})']'$ and $\| \mu^{(i)}(\theta^{(i)}) \| \neq 0$ for all $\theta^{(i)} \in \Theta^{(i)}$*

[2] As a consequence, while the scenarios (i), (ii) and (iii) represent respectively the null, local alternative and fixed alternative for the overidentifying restrictions test, these scenarios do not fulfil the same roles for the Rivers and Vuong statistics considered here. However, since (i)-(iii) are routinely applied to analyze model specification tests within the GMM framework, we believe it is of interest - and natural - to analyze the behaviour of model selection tests within the GMM framework under the same scenarios.

To implement the estimations and test, it is necessary to choose the weighting matrices. From the results in Hall and Inoue (2003), since the models are misspecified, there is no advantage to employing a weighting matrix that converges to the inverse of the long run variance of the sample moment condition and hence to employing iterated GMM estimation. Therefore, in the following we only study the case where the weighting matrix is a matrix of constants, such as the identity matrix.[3] To analyze the behavior of the test in this case we impose the following set of assumptions:

Assumption 2 *(i)* $T^{-1/2}\sum_{t=1}^{T}\{\, f(v_t,\theta_\star) - E[f(v_t,\theta_\star)]\,\} \xrightarrow{d} N\,(\,0,\, S(\theta_\star)\,)$ *where*

$$S(\theta_\star) \;=\; \begin{bmatrix} S^{(1)}(\theta_\star^{(1)}) & S^{(1,2)}(\theta_\star) \\ S^{(1,2)}(\theta_\star)' & S^{(2)}(\theta_\star^{(2)}) \end{bmatrix},$$

is a positive definite matrix of finite constants that is partitioned conformably with $f(.)$*; (ii)* $rank\{G_0^{(i)}(\theta_\star^{(i)})\} = p_i$*; (iii)* $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_\star^{(i)}) = O_p(1)$*; (iv)* $\hat{S}^{(i)}(\hat{\theta}_T^{(i)})$ *is a* $q_i \times q_i$ *matrix satisfying* $\hat{S}^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{p} S^{(i)}(\theta_\star^{(i)})$ *for* $i = 1, 2$*; (v)* $\hat{S}^{(1,2)}(\hat{\theta}_T)$ *is a* $q_1 \times q_2$ *matrix satisfying* $\hat{S}^{(1,2)}(\hat{\theta}_T) \xrightarrow{p} S^{(1,2)}(\theta_\star)$*.*

Hall and Inoue (2003)[Theorem 1] provides conditions under which $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_\star^{(i)})$ has a limiting normal distribution, and so Assumption 2(iii) could be replaced by these lower level assumptions. The following theorem gives the limiting distribution of N_T for this choice of weighting matrix.

Theorem 1 *Let (i)* $\{v_t\}$*,* $f(.)^{(i)}$*,* $\theta_*^{(i)}$ *and* $\Theta_*^{(i)}$ *satisfy Hall (2005) Assumptions 3.1, 3.2, 3.8-3.10, 4.2, and 4.3 hold; (ii)* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *satisfy Assumption 1; (iii)* H_0 *holds; the weighting matrix is* $W_T^{(i)} = I_{q_i}$ *and Assumption 2 holds; then* $N_T \xrightarrow{d} N(0,\, 1)$*.*

Theorem 1 confirms the results of Rivers and Vuong (2002) in that the statistic N_T has a limiting standard normal distribution under the null hypothesis if both models are misspecified in the sense of Assumption 1. This result would appear to indicate that there is scope for using this statistic to compare two misspecified models estimated via GMM. However, in our opinion, some caution needs to be exercised in its use as we now explain. The null hypothesis involves the population analog to the minimands. These minimands depend on the weighting matrices and also the probability limits of the estimators. In general, the relative magnitudes of the minimands, $Q_0^{(i)}(\theta_\star^{(i)})$, are sensitive to the choice of weighting matrices, and so the relative ranking can be reversed by changing the weighting matrices. Whether or not

[3] In an earlier version of this paper, we consider the case in which the weighting matrix is an instrument cross product matrix. The limiting distribution of N_T presented in Theorem 1 below is the same although the construction of an appropriate $\hat{\sigma}_T$ does depend on the choice of weighting matrix; see Hall and Pelletier (2007).

this dependence on the weighting matrix is a weakness depends on the setting. In some cases, economic theory dictates an appropriate choice of weighting matrix and so only the outcome with this choice of the weighting matrix is of interest. Examples in this vein are the assessment of specification errors in asset pricing models, e.g. see Hansen and Jagannathan (1997), or dynamic stochastic equilibrium models, e.g. see Dridi, Guay, and Renault (2007). However, absent these economic considerations, the choice of the weighting matrix and the relative ranking of the models can become arbitrary since there is no optimal weighting matrix in this context.

We now consider the case in which the population moment conditions are correctly specified or locally misspecified.

Assumption 3 *$\mathcal{M}_i$ satisfies $S^{(i)}(\theta_\star^{(i)})^{-1/2}E_T[f^{(i)}(v_t,\theta_\star^{(i)})] = T^{-1/2}\eta^{(i)}$ where $\eta^{(i)}$ is a vector of finite constants.*

The operator $E_T[.]$ denotes expectations with respect to the joint probability distribution of $\{v_t, t = 1,\ldots,T\}$ and $S^{(i)}(\theta_\star^{(i)}) = \lim_{T\to\infty} Var[g_T^{(i)}(\theta_\star^{(i)})]$.[4] Notice that if $\eta^{(i)} = 0$ then model i is correctly specified. Rather than just consider the case of correct specification, we allow for local misspecification as well because the latter casts light on the finite sample behaviour of the test statistic when the degree of misspecification is small. Given the framework in Assumption 3, we must modify the definition of the population minimands as follows $Q_0^{(i)}(\theta^{(i)}) = \lim_{T\to\infty} E_T\left[f^{(i)}(v_t,\theta^{(i)})\right]' W^{(i)} \lim_{T\to\infty} E_T\left[f^{(i)}(v_t,\theta^{(i)})\right]$. Notice that Assumption 3 implies $Q_0^{(i)}(\theta_\star^{(i)}) = 0$ for both models. Therefore, although the models are not correctly specified, the local nature of this misspecification implies that the null hypothesis in (5) still holds, that is $H_0 : Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)})$.

Using Assumption 3 and (10)-(13), it can be seen that, for the case under consideration here, $R_\star$ is a null vector and hence $\sigma_0^2 = 0$. Therefore, if both models are either correctly specified or locally misspecified then the null distribution of N_T does not follow from Rivers and Vuong's (2002) analysis (see their Theorem 3). We note that Rivers and Vuong (2002)[Section 6] provide generic conditions under which the test does not have a limiting standard normal distribution because $\sigma_0^2 = 0$. An inspection of these conditions indicates that they include the case covered here although this is not noted in their discussion of the results.

[4]Assumption 3 implies that the data cannot be a realization of a strictly stationary process, unless $\eta^{(i)} = 0$, because $E[f(v_t,\theta)]$ changes with T. Instead the process can be viewed as a perturbation of a stationary process; see Newey (1985) or Hall (2005, Section 5.1.3).

Below, for completeness, we present the appropriate limiting distribution theory for the test statistic in this case. To do so, it is necessary to be more specific about the construction of $\hat{\sigma}_T$, and hence the weighting matrices employed. Since both models are assumed correctly specified or at most locally misspecified, we assume that the weighting matrices are chosen so that $W^{(i)} = \{S^{(i)}\}^{-1}$ and $W_T^{(i)}$ depends on $\hat{\tau}_T^{(i)}$, a preliminary GMM estimator of $\theta_\star^{(i)}$ using a weighting matrix, $M_T^{(i)}$, that converges to a positive definite matrix of constants, $M^{(i)}$. In this case, it follows that the matrix $A_\star^{(i)}$ and vector $Y_t^{(i)}$ in (8) are given by $A_\star^{(i)} = -\left[G_0^{(i)}(\theta_\star^{(i)})'\,M^{(i)}\,G_0^{(i)}(\theta_\star^{(i)})\right]^{-1}$ and $Y_t^{(i)} = G_0^{(i)}(\theta_\star^{(i)})'\,M^{(i)}\,f^{(i)}(v_t,\theta_\star^{(i)})$. To define $\Delta^{(i)}$, assume that $W_T^{(i)} = \{S_T^{(i)}(\hat{\tau}_T^{(i)})\}^{-1}$. It then follows that $\Delta^{(i)} = L_i[\{S^{(i)}(\theta_\star^{(i)})\}^{-1} \otimes \{S^{(i)}(\theta_\star^{(i)})\}^{-1}]\Sigma^{(i)}$ where $\Sigma^{(i)} = E\left[\partial vec[S_T(\theta^{(i)})]\,\partial\theta^{(i)\prime}|_{\theta^{(i)}=\theta_\star^{(i)}}\right]$ and L_i is a $q_i(q_i+1)/2 \times q_i^2$ selection matrix such that $vech[W^{(i)}] = L_i vec[W^{(i)}]$. The exact form of $\Sigma^{(i)}$ depends on the choice of covariance matrix estimator. We leave that unspecified and only impose high level assumptions on $\Sigma^{(i)}$ below. Given these definitions, it is natural to set

$$\hat{\sigma}_T^2 \;=\; \hat{R}_T'\hat{V}_T\hat{R}_T \tag{14}$$

where $\hat{R}_T$ and $\hat{V}_T$ are consistent estimators of $R_\star$ and $V_\star$ constructed using the obvious sample analogs to the population quantities that make up these matrices.[5]

To present the limiting distribution of N_T, it is necessary to impose the following additional regularity conditions.

Assumption 4 *The observed data are assumed to be a realization from a stochastic process $\{v_t; t = 1,2,\ldots\}$ which satisfies the following conditions: (i) $\hat{\theta}_T^{(i)} \xrightarrow{p} \theta_\star^{(i)}$; (ii) $g_T^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{p} 0$; (iii) $G_T^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{p} G_0^{(i)}$, $G_T^{(i)}(\hat{\theta}_T^{(i)},\theta_*^{(i)},\lambda_T) \xrightarrow{p} G_0^{(i)}$; (iv) $W_T^{(i)} \xrightarrow{p} [S^{(i)}(\theta_\star^{(i)})]^{-1}$, a positive definite matrix; (v) the limit distribution of the moment conditions satisfies*

$$T^{1/2}g_T(\theta_\star) \xrightarrow{d} N\left(\begin{bmatrix} S^{(1)}(\theta_\star^{(1)})\eta^{(1)} \\ S^{(2)}(\theta_\star^{(2)})\eta^{(2)} \end{bmatrix},\ S(\theta_\star)\right),$$

where $S(\theta_\star)$ is a positive definite matrix of finite constants; (vi) equations (8) and (9) hold with the definitions given in the sentences preceding (14) and $\Sigma^{(i)}$ is a matrix of finite constants.

The limiting distribution of N_T is given in the following theorem.

[5]See Hall and Pelletier (2007) for further details.

Theorem 2 *Let Assumptions 3 and 4 hold for* $i = 1, 2$ *then* $N_T \xrightarrow{d} \xi_1/\xi_2$ *where* ξ_1 *and* ξ_2 *are two random variables defined in the appendix whose distributions depend on* q_1*,* q_2*, certain population moments of the data and the drift parameters* $\{\eta^{(i)}\}$.

It is evident from Theorem 2 that N_T does not have a limiting standard normal distribution in the case where it is used to compare two models via their GMM minimands and both models are either correctly specified or locally misspecified. The result for local misspecification reveals that if the models are misspecified but the degree of misspecification is small then the behaviour of the statistic in moderate sized samples is more like the non-standard distribution encountered in the correctly specified case than the standard normal distribution.

Given the sensitivity of the limiting behaviour of N_T to the specification of the model or more generally the value of σ_0^2, it is desirable to implement some formal pre-test of $\sigma_0^2 = 0$ against $\sigma_0^2 > 0$. Such a test is developed by Vuong (1989) in the context of QML and is suggested by Rivers and Vuong (2002) albeit in the context of their very general framework. We now consider the implementation of such a test in our context. The obvious test statistic is $T\hat{\sigma}_T^2$. From the proof of Theorem 2, it follows that if the models are locally misspecified then $T\hat{\sigma}_T^2$ converges in distribution to a mixture of non-central chi-squareds with the noncentrality parameters depending on the drift parameter, η. This means that the implementation of the test is problematic because the critical value for the asymptotically valid $100\alpha\%$ depends on η which is itself not estimable consistently.

One solution is to adopt a decision rule based on the limiting distribution of $T\hat{\sigma}_T^2$ in the case where the models are both correctly specified ($\eta = 0$) because in this case

$$T\hat{\sigma}_T^2 \xrightarrow{d} \sum_{i=1}^{q_1+q_2} w_i z_i^2$$

where $\{z_i\}$ are a sequence of i.i.d. standard normal random variables and $\{w_i\}$ are the eigenvalues of $C^{1/2\prime}\{I_{q_1+q_2} - P_0\}'\bar{C}\{I_{q_1+q_2} - P_0\}C^{1/2} = D$, say. The various elements composing D are defined in the proof of Theorem 2 and can be consistently estimated using sample analogs. In the simulation of the critical values, the weights are replaced by the eigenvalues of a consistent estimator of D. To consider the properties of such a strategy, define $\gamma(\alpha)$ to be the value such that $P(\zeta > \gamma(\alpha)) = \alpha$ where $\zeta \sim \sum_{i=1}^{q_1+q_2} w_i z_i^2$. Consider the decision rule

$$DR(\alpha) : \text{ reject } H_0 : \sigma_0^2 \text{ in favour of } H_1 : \sigma_0^2 > 0 \text{ if } T\hat{\sigma}_T^2 > \gamma(\alpha)$$

Let $P(\text{Type I} \,|\, \alpha)$ and $P(\text{Type II} \,|\, \alpha)$ denote the probabilities of Type I and Type II errors respectively associated with the decision rule $DR(\alpha)$.

Clearly if $DR(\alpha)$ is implemented with a fixed α then it only yields a test satisfying $lim_{T\to\infty}P(\text{Type I} \,|\, \alpha) = \alpha$ in the case where $\mathcal{M}_i$ satisfy Assumption 3 *and* $\eta^{(i)} = 0$ for $i = 1, 2$.[6] However, if the decision rule is implemented with a value of α that tends to zero as $T \to \infty$ then this problem is mitigated in the limit as the following proposition demonstrates.

Proposition 1 *Let $\alpha = \tilde{\alpha}_T$ with $\tilde{\alpha}_T \to 0$ as $T \to \infty$ and $ln[\tilde{\alpha}_T] = o(T)$. (i) If $\mathcal{M}_i$ satisfy Assumption 3 for $i = 1, 2$ and Assumption 4 holds then $lim_{T\to\infty}P(\text{Type I} \,|\, \alpha) = 0$. (ii) If $\mathcal{M}_i$ satisfy Assumption 1 for $i = 1, 2$ and the other assumptions of Theorem 1 hold then $lim_{T\to\infty}P(\text{Type II} \,|\, \alpha) = 0$.*

Proposition 1 demonstrates that it is possible to develop a testing strategy that disciminates between states of the world in which $\sigma_0^2 = 0$ (correctly specified and locally misspecified models) and $\sigma_0^2 > 0$ (non-locally misspecified models) with probability one in the limit, and thus provides a justification in the limit for the use of this test as a pre-test.

Our theoretical results suggest the following decision tree. We can start by testing "$\sigma^2 = 0$" using the decision rule in Proposition 1. If this hypothesis is rejected then N_T can be used with a decision rule based on the standard normal distribution to test H_0 in (5) (subject to the caveats above regarding the choice of weighting matrix mentioned above). If "$\sigma^2 = 0$" cannot be rejected then it indicates that both models are either correctly specified or locally misspecified, in which case H_0 in (5) holds by construction since $Q_0^{(i)}(\theta_\star^{(i)}) = 0$ for $i = 1, 2$. We note that N_T is consistent against the alternatives in (6) and (7) for which $\hat{\sigma}_T^2$ converges in probability to a positive constant, and this class includes the case in which one model is correctly specified or locally misspecified and the other model is non-locally misspecified. Thus if either (6) or (7) hold then this decision tree leads to the model with the smaller population minimand with probability one in the limit.

A referee has argued that in the case where the pre-test is insignificant, the statistic N_T can be compared to a critical value obtained by simulating under the null hypothesis that both models are correctly specified.[7] Such an approach offers the potential to inform with respect to the hypothesis that one model is correctly specified and the other is locally misspecified, which is different than H_0 in (5) but nevertheless of interest. However some exploratory simulations indicate that the marginal information in the test

[6]This statement assumes the other conditions of Theorem 2 hold as well.

[7]If $\sigma_0^2 = 0$, then inference can be based on $T^{1/2}$ time the numerator of N_T, *i.e.* T times the difference between the minimands; see Vuong (1989) for a similar approach in QMLE.

is very limited. The reason is that in the GMM context N_T is based on the difference of the overidentifying restrictions tests from each model, each of which is a test of the individual model specification. Thus, the numerator of N_T only adds to the researcher's ability to discriminate between a model that is correctly specified and one that is locally misspecified if it is significant but both overidentifying restrictions test statistics are insignificant. Our simulations indicate that it does not happen often, the intuition being that the power of both (the numerator of) N_T and the overidentifying restrictions test statistic are driven by the same drift, and similar magnitudes of drift make both statistics significant. We conjecture this finding may be more generally true.

It may be desired to use a different weighting matrix in the GMM minimands used to measure the distance between the two models than the ones used in the estimation of the parameters. An inspection of the proofs of Theorems 1 and 2 indicates that the same qualitative results go through if the test is not based on the same weighting matrices as used in the estimations.

There has been a growing interest in the estimation of moment-models via Generalized Empirical Likelihood (GEL) (Smith (1997)) and it is reasonable to wonder if GEL suffers similar deficiencies to GMM for the kind of inference problem described here. Kitamura(2000, 2002) and Shi (2010) propose extensions of Vuong's (1989) methods to GEL estimation of conditional moment restrictions models and to models defined by moment inequalities, respectively. An immediate advantage of GEL methods is that there is no weighting matrix and thus the model ranking is unambiguous. However, GEL methods do share with GMM methods the problems highlighted above concerning the comparison of two correctly specified or locally misspecified models. Kitamura (2000) develops an analogous test for $\sigma_0^2 = 0$ within his setting. However, he concentrates on the cases in which both models are correctly specified or both are "overlapping" misspecified models with a common pseudo true measure, and considers the test's behaviour only under non-local alternatives for which $\sigma_0^2 \neq 0$. It is easily seen from his analysis that the same problems arise in the GEL setting when local misspecification of the population moment condition occurs. Proposition 1 above can easily be extended to cover the GEL case and thus provides a justification for the pre-test in the GEL setting as well.[8] [9]

[8] To match Kitamura's setting in which $\hat{\sigma}_T^2$ is calculated with a HAC estimator, we must set $ln[\alpha_T] = o(T/b_T)$ where b_T is the bandwidth used in the HAC estimator.

[9] Kitamura (2000)[p.12] does observe that his test of σ_0^2 has power against certain local alternatives but does not relate these possibilities back to the moment conditions as in our framework nor explore its implications further as done in Theorem 2 above.

Appendix

(i) Proof of Theorem 1

Applying the Mean Value Theorem to $Q_T^{(i)}(\theta^{(i)})$ around $\theta^{(i)} = \theta_\star^{(i)}$, we obtain

$$Q_T^{(i)}(\hat{\theta}_T^{(i)}) \;=\; Q_T^{(i)}(\theta_\star^{(i)}) \;+\; \left\{ \frac{\partial Q_T^{(i)}(\theta^{(i)})}{\partial \theta^{(i)}}\bigg|_{\theta^{(i)}=\bar{\theta}_T^{(i)}} \right\}' \left(\hat{\theta}_T^{(i)} \;-\; \theta_\star^{(i)} \right) \tag{15}$$

where $\bar{\theta}_T^{(i)} = \lambda_T \theta_\star^{(i)} \;+\; (1-\lambda_T)\hat{\theta}_T^{(i)}$ for some $0 \leq \lambda_T \leq 1$. Now define

$$\Phi^{(i)}(\theta_\star^{(i)}) \;=\; 2G_0^{(i)}(\theta_\star^{(i)})' W^{(i)} E[f^{(i)}(v_t, \theta_\star^{(i)})] \tag{16}$$

It the follows from (15) that under our assumptions, we have

$$Q_T^{(i)}(\hat{\theta}_T^{(i)}) \;=\; Q_T^{(i)}(\theta_\star^{(i)}) \;+\; \left\{ \frac{\partial Q_0^{(i)}(\theta_\star^{(i)})}{\partial \theta^{(i)}} \right\}' \left(\hat{\theta}_T^{(i)} \;-\; \theta_\star^{(i)} \right) \;+\; o_p(T^{-1/2}) \tag{17}$$

and hence

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) \;-\; Q_T^{(2)}(\hat{\theta}_T^{(2)}) \right] \;&=\; T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) \;-\; Q_T^{(2)}(\theta_\star^{(2)}) \right] \\ &\quad + \left\{ \Phi^{(i)}(\theta_\star^{(1)}) \right\}' T^{1/2}\left(\hat{\theta}_T^{(1)} \;-\; \theta_\star^{(1)} \right) \\ &\quad - \left\{ \Phi^{(i)}(\theta_\star^{(2)}) \right\}' T^{1/2}\left(\hat{\theta}_T^{(2)} \;-\; \theta_\star^{(2)} \right) \;+\; o_p(1) \end{aligned} \tag{18}$$

Finally, under H_0, we have $Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)})$ and so (18) can be written as

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)}) \right] \;&=\; T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) - Q_0^{(1)}(\theta_\star^{(1)}) \right] - T^{1/2}\left[Q_T^{(2)}(\theta_\star^{(2)}) - Q_0^{(2)}(\theta_\star^{(2)}) \right] \\ &\quad + \left\{ \Phi^{(i)}(\theta_\star^{(1)}) \right\}' T^{1/2}\left(\hat{\theta}_T^{(1)} \;-\; \theta_\star^{(1)} \right) \\ &\quad - \left\{ \Phi^{(i)}(\theta_\star^{(2)}) \right\}' T^{1/2}\left(\hat{\theta}_T^{(2)} \;-\; \theta_\star^{(2)} \right) \;+\; o_p(1) \end{aligned} \tag{19}$$

This equation simplifies further. Under our assumptions, the GMM estimator can be obtained by solving the first order conditions associated with the minimization in (1), that is: $G_T^{(i)}(\hat{\theta}_T^{(i)})' W_T T^{-1} \sum_{t=1}^T f^{(i)}(v_t, \hat{\theta}_T^{(i)}) = 0$. Furthermore, the probability limits must satisfy the analogous population moment condition, that is: $\Phi^{(i)}(\theta_\star^{(i)}) = 0$. Therefore, we have

$$T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)}) \right] \;=\; T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) - Q_0^{(1)}(\theta_\star^{(1)}) \right] - T^{1/2}\left[Q_T^{(2)}(\theta_\star^{(2)}) - Q_0^{(2)}(\theta_\star^{(2)}) \right] + o_p(1) \tag{20}$$

Notice that $Q_T^{(i)}(.)$ and $Q_0^{(i)}(.)$ have the generic structures $\hat{h}'\hat{W}\hat{h}$ and $h'Wh$ respectively, and that

$$\hat{h}'\hat{W}\hat{h} \,-\, h'Wh \;=\; \hat{h}'\hat{W}(\hat{h}-h) \,+\, \hat{h}'(\hat{W}-W)h \,+\, (\hat{h}-h)'Wh \tag{21}$$

With the choice $W_T^{(i)} = I_{q_i}$ for the weighting matrix, it follows from (20) and (21) that

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\right] \;&=\; 2\Bigg\{\mu^{(1)}(\theta_\star^{(1)})'T^{-1/2}\sum_{t=1}^{T}[f^{(1)}(v_t,\theta_\star^{(1)}) - \mu^{(1)}(\theta_\star^{(1)})] \\ &\quad + \mu^{(2)}(\theta_\star^{(2)})'T^{-1/2}\sum_{t=1}^{T}[f^{(2)}(v_t,\theta_\star^{(2)}) - \mu^{(2)}(\theta_\star^{(2)})]\Bigg\} + o_p(1) \end{aligned}$$

The result then follows immediately under the stated assumptions.

(ii) Theorem 2: definitions and proof

Definitions: $\xi_1 = d'diag\left(I_{q_1} - P_0^{(1)}, -[I_{q_2} - P_0^{(2)}]\right)d$, $d = C^{1/2}(n_{q_1+q_2} + \bar{\eta})$,
$C^{1/2} = diag\left(S^{(1)}(\theta_\star^{(1)})^{-1/2}, S^{(2)}(\theta_\star^{(2)})^{-1/2}\right)S(\theta_\star)^{1/2}$, $n_{q_1+q_2} \sim N(0, I_{q_1+q_2})$,
$P_0^{(i)} \;=\; F_0^{(i)}(\theta_\star^{(i)})\left[F_0^{(i)}(\theta_\star^{(i)})'F_0^{(i)}(\theta_\star^{(i)})\right]^{-1}F_0^{(i)}(\theta_\star^{(i)})'$, $F_0^{(i)}(\theta_\star^{(i)}) \;=\; S^{(i)}(\theta_\star^{(i)})^{-1/2}G_0^{(i)}(\theta_\star^{(i)})$,
$\xi_2 = 2\sqrt{d'\{I_{q_1+q_2} - P_0\}'\bar{C}\{I_{q_1+q_2} - P_0\}d}$, $\bar{C} \;=\; A\,S(\theta_\star)\,A'$,
$A = diag\left(S^{(1)}(\theta_\star^{(1)})^{-1/2},\, -S^{(2)}(\theta_\star^{(2)})^{-1/2}\right)$, $P_0 \;=\; diag\left(P_0^{(1)},\, P_0^{(2)}\right)$.

Sketch Proof:

The test statistic N_T can be written as

$$N_T \;=\; \frac{T\left\{Q_T^{(1)}\left(\hat{\theta}_T^{(1)}\right) - Q_T^{(2)}\left(\hat{\theta}_T^{(2)}\right)\right\}}{T^{1/2}\hat{\sigma}_T}. \tag{22}$$

Standard analysis of the overidentifying restrictions test yields (*e.g.* see Hall (2005)[equation (3.36),p.73])

$$TQ_T^{(i)}\left(\hat{\theta}_T^{(i)}\right) \;=\; \{T^{1/2}g_T^{(i)}(\theta_\star^{(i)})\}'[S^{(i)}]^{-1/2'}\left[I - P_0^{(i)}(\theta_\star^{(i)})\right][S^{(i)}]^{-1/2}\{T^{1/2}g_T^{(i)}(\theta_\star^{(i)})\} + o_p(1) \tag{23}$$

Now consider the denominator of (22). It is most convenient to study $T\hat{\sigma}_T^2 = T^{1/2}\hat{R}_T'\,\hat{V}_T\,T^{1/2}\hat{R}_T$. First consider $T^{1/2}\hat{R}_T$, the sample analog of $R_\star$. Under our assumptions, it follows that

$$\begin{aligned} T^{1/2}\hat{R}_T \;&=\; \left[2T^{1/2}g_T^{(1)}(\hat{\theta}_T^{(1)})'W_T^{(1)}, o_p(1), -2T^{1/2}g_T^{(2)}(\hat{\theta}_T^{(2)})'W_T^{(2)}, o_p(1)\right]' \\ &=\; 2\Gamma_V T^{1/2}g_T(\hat{\theta}_T) + o_p(1) \end{aligned} \tag{24}$$

where Γ_V is defined below. The result the follows by using a Mean Value Theorem expansion for $g_T(\hat{\theta}_T)$, the standard asymptotic representation for $(\hat{\theta}_T - \theta_\star)$ (e.g. see Hall (2005)[equation (3.26)]) and

$\hat{V}_T \xrightarrow{p} \Gamma_U S \Gamma_U'$. The matrices Γ_V and Γ_U are:

$$\Gamma_V = \begin{bmatrix} W_T^{(1)} & 0 & 0 & 0 \\ 0 & 0 & -W_T^{(2)} & 0 \end{bmatrix}', \Gamma_U = \begin{bmatrix} I_{q_1} & \{G_0^{(1)'} M^{(1)}\}' & 0 & 0 \\ 0 & 0 & I_{q_2} & \{G_0^{(2)'} M^{(2)}\}' \end{bmatrix}$$

(iii) Proof of Proposition 1

Since D is positive semi-definite, $w_i \geq 0$; without loss of generality, we assume the eigenvalues are ordered so that $w_i \geq w_{i+1}$. Let n_w denote the number of non-zero eigenvalues of D. It follows that

$$w_{n_w} \sum_{i=1}^{n_w} z_i^2 \leq \sum_{i=1}^{q_1+q_2} w_i z_i^2 = \sum_{i=1}^{n_w} w_i z_i^2 \leq w_1 \sum_{i=1}^{n_w} z_i^2$$

and so $w_{n_w} c(\alpha) \leq \gamma(\alpha) \leq w_1 c(\alpha)$ where $c(\alpha)$ is the $1 - \alpha$ quantile of the $\chi^2_{n_w}$ distribution. Potscher (1983, Theorem 5.8) establishes that for $\lim_{T\to\infty} \kappa_T = \infty$, $c(\alpha_T) = o(\kappa_T)$ if and only if $ln[\alpha_T] = o(\kappa_T)$. The proof is completed by noting that under our assumptions if the models are correctly specified or locally misspecified then $T\hat{\sigma}_T^2 = O_p(1)$ and if the models are non-locally misspecified then $\hat{\sigma}_T^2 \xrightarrow{p} \sigma_0^2 > 0$ and hence $T\hat{\sigma}_T^2$ diverges at rate T.

References

Dridi, R., Guay, A., and Renault, E. (2007). 'Indirect inference and calibration of dynamic general equilibrium models', *Journal of Econometrics*, 136: 397–430.

Hall, A., and Pelletier, D. (2007). 'Non-Nested Testing in Models Estimated via Generalized Method of Moments', Economics Working Papers 11, Department of Economics, North Carolina State University.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press, Oxford, U. K.

Hall, A. R., and Inoue, A. (2003). 'The large sample behaviour of the Generalized Method of Moments estimator in misspecified models', *Journal of Econometrics*, 114: 361–394.

Hansen, L. P., and Jagannathan, R. (1997). 'Assessing specification errors in stochastic discount factor models', *Journal of Finance*, 52(2): 557–590.

Kitamura, Y. (2000). 'Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood', University of Pennsylvania.

Kitamura, Y. (2002). 'A likelihood-based approach to the analysis of a class of nested and non-nested models', Discussion paper, Department of Economics, University of Pennsylvania, Philadelphia, PA, USA.

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

Newey, W. K. (1985). 'Generalized Method of Moments specification testing', *Journal of Econometrics*, 29: 229–256.

Potscher, B. M. (1983). 'Order estimation in ARMA models by lagrangian multiplier tests', *Annals of Statistics*, 11: 872–885.

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

Shi, X. (2010). 'Model Selection Tests for Nonnested Moment Inequality Models', Yale University.

Smith, R. J. (1997). 'Alternative semi-parametric likelihood approaches to generalized method of moments estimation', *Economics Journal*, 107: 503–519.

Vuong, Q. (1989). 'Likelihood ratio tests for model selection and non–nested hypotheses', *Econometrica*, 57: 307–334.