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]

June 20, 2008

[1]We thank Atsushi Inoue, Eric Renault, Quang Vuong and Ken West for several 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

Rivers and Vuong (2002) develop a very general framework for choosing between two competing dynamic models. Within their framework, inference is based on a statistic that compares measures of goodness of fit between the two models. Under the null hypothesis, Rivers and Vuong (2002) show that their test statistic has a standard normal distribution under generic conditions that are argued to allow for a variety of estimation methods including Generalized Method of Moments (GMM). In this paper, we analyze the limiting distribution of Rivers and Vuong's (2002) statistic under the null hypothesis when inference is based on a comparison of GMM minimands evaluated at GMM estimators. It is shown that the limiting behaviour of this statistic depends on whether the models in question are correctly specified, locally misspecified or misspecified. Specifically, it is shown that: (i) if both models are correctly specified or locally misspecified then Rivers and Vuong's (2002) generic conditions are not satisfied, and the limiting distribution of the test statistic is non-standard under the null; (ii) if both models are misspecified then the generic conditions are satisfied, and so the statistic has a standard normal distribution 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. A characteristic of the early work on non-nested hypothesis testing is that under the null hypothesis one model is assumed to be correct[1]. This is clearly a viable approach to model selection but there is, of course, the chance that the test procedures indicate that either both models are correct or that neither are correct. 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). White (1982) shows that QML can be interpreted as choosing estimates to minimize the Kullback Leibler metric for the distance between the assumed probability density function (pdf) and the true pdf. Vuong (1989) exploits this interpretation to propose a test of which model is closer to the truth based on the difference of the QML's.[2]

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. The analysis covers the case in which the measure of goodness of fit is the optimand for parameter estimation and also the case in which it is not. 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

[1] See Cox(1961, 1962) and Atkinson (1970) in the context of Maximum Likelihood estimation, Pesaran and Deaton (1978), Davidson and MacKinnon (1981) and Mizon and Richard (1986) in the context of regression models.

[2] Also see Sin and White (1996) for a related information criterion approach.

used in the estimation.[3] 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; for if this variance is zero, then Rivers and Vuong (2002) show that their test statistic does not have a standard normal limiting distribution under the null.

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 because the variance mentioned in the previous paragraph is zero. We further show, in this case, that the statistic does not converge to a limiting normal distribution but to a non-standard distribution that is a function of nuisance parameters, 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. However, we argue that some caution needs to be exercised in its use because the outcome of the statistic depends on the choice of weighting matrix. This dependence raises the possibility that the "ranking" of the models is determined by the choice of weighting matrix. Whether or not this 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 weighting matrix is of interest. However, absent these economic considerations, the choice of the weighting matrix becomes arbitrary for in misspecified models – unlike in correctly specified models – there is no statistical theory to guide the choice of the weighting matrix.[4] It is in this case that the dependence of the outcome on the weighting matrix becomes troublesome.

An outline of the paper is as follows. Section 2 presents a review of the Rivers and Vuong's (2002)

[3] See Rivers and Vuong (2002)[p.3 and p.13]. The latter version of the test has been employed by Carpentier and Weaver (1997) and Nauges and Thomas (2003). For other approaches to non-nested hypothesis testing within the GMM framework, see Singleton (1985), Ghysels and Hall (1990) and Smith (1992).

[4] See Hall and Inoue (2003).

statistic in the context of GMM estimation. Section 3 analyzes the limiting distribution of the statistic under the null hypothesis in the case where both models are two correctly specified or two locally misspecified models. Section 4 studies the limiting distribution when the two models are misspecified. Section 5 considers the problem of testing whether the key variance (mentioned above) is zero, and Section 6 briefly considers some extensions. The main proofs are relegated to a Mathematical Appendix.[5]

2 Framework and Test Statistic

For $i = 1, 2$, let $f^{(i)} : \mathbf{V} \times \Theta^{(i)} \to \Re^{q_i}$, where $q_i < \infty$, $\Theta \subset \Re^p_i$ 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 parameters of both models are estimated via GMM; these estimators are defined as follows:

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

where

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

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

$$Q^{(i)}_0(\theta^{(i)}) = E[f^{(i)}(v_t, \theta^{(i)})]' W^{(i)} E[f^{(i)}(v_t, \theta^{(i)})]. \tag{3}$$

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

[5]For brevity, certain conditions are suppressed in the text. A more explicit accounting of the regularity conditions can be found in an earleir version of the paper which is available from the authors upon request.

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.[6] The null hypothesis of the test is that: $\mathcal{M}_1$ and $\mathcal{M}_2$ are asymptotically equivalent, that is

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

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

$$Q_0^{(1)}(\theta_\star^{(1)}) \;<\; Q_0^{(2)}(\theta_\star^{(2)}) \tag{6}$$

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

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

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:

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

[6] For proofs specific to GMM see: for correctly specified models, Hansen (1982)[Theorem 2.1] or Hall (2005)[Theorem 3.1]; for non-locally misspecified models, see Hall (2000)[Lemma 1] or Hall (2005)[Theorem 4.1].

for some symmetric matrix of constants $A_\star^{(i)}$ and vector $Y_t^{(i)}$; and that the weighting matrix satisfies[7]:

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

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, and are considered below on a case by case basis.

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

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

and

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

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).

To conclude this section, we introduce some additional notation. 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 $G_T(\bar{\theta}^{(j)})$ 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 $S^{-1/2} = [S^{1/2}]^{-1}$.

[7]This assumption could be relaxed to allow estimators with a rate of convergence depending on a bandwith as in Newey and West (1987). It would complicate the notation but would not qualitatively change our results. Although it should be noted that it would change the rate statements in Proposition 1 in Section 5.

3 Correctly and Locally Misspecified Models

In this section we examine the limiting behavior of N_T under H_0 when the models are correctly specified or locally misspecified. In practice, the characterization of the model as correctly or incorrectly specified is based on the outcome of the overidentifying restrictions test. Therefore, the designation "correctly specified" is more appropriately denoted as "the overidentifying restrictions test is insignificant". Now, even in the limit, an insignificant statistic can occur with non-negligible probability not only because the model is correctly specified but also because the model is locally misspecified.[8] Therefore, even in the limit, the category "the overidentifying restrictions test is insignificant" contains both correctly specified and locally misspecified models, and, in fact, both types of model satisfy the null hypothesis of the Rivers and Vuong test.

Consider a scenario where the two models compared are locally misspecified. Within the GMM framework, this is most naturally captured via a Pitman drift on the population moment conditions. We assume that the moment conditions are invalid but the size of the violation is $O(T^{-1/2})$ and so disappears at the limit, that is

Assumption 1 *$\mathcal{M}_i$ 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 $S^{(i)}(\theta_\star^{(i)}) = \lim_{T\to\infty} Var[g_T^{(i)}(\theta_\star^{(i)})]$.[9]

Given the framework in Assumption 1, 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 1 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 of the Rivers and Vuong test still holds, that is $H_0 : Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)})$.

[8] See Newey (1985) or Hall (2005, Section 5.1.3).

[9] Assumption 1 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).

Using Assumption 1 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.[10] 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.

Below 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[11]

$$\Delta^{(i)} = L_i[\{S^{(i)}(\theta_\star^{(i)})\}^{-1} \otimes \{S^{(i)}(\theta_\star^{(i)})\}^{-1}]\Sigma^{(i)} \tag{14}$$

where

$$\Sigma^{(i)} = E\left[\left. \frac{\partial vec[S_T(\theta^{(i)})]}{\partial \theta^{(i)\prime}} \right|_{\theta^{(i)}=\theta_\star^{(i)}} \right] \tag{15}$$

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.[12] Given these definitions, it is natural to set

$$\hat{\sigma}_T^2 = \hat{R}_T' \hat{V}_T \hat{R}_T \tag{16}$$

[10]The result in question is Rivers and Vuong (2002)[Theorem 3].

[11]See Dhrymes(1984)[Proposition 99, p.115; Proposition 106, p.124].

[12]The interested reader is referred to Hall (2005)[p.103] for an example.

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.[13]

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

Assumption 2 *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*

$$T^{1/2} g_T(\theta_\star) \xrightarrow{d} N\left(\begin{bmatrix} S^{(1)}(\theta_\star^{(1)})\eta^{(1)} \\ S^{(2)}(\theta_\star^{(2)})\eta^{(2)} \end{bmatrix}, S(\theta_\star)\right), \ where\ S(\theta_\star) \ = \ \begin{bmatrix} S^{(1)}(\theta_\star^{(1)}) & S^{(1,2)}(\theta_\star) \\ S^{(1,2)}(\theta_\star)' & S^{(2)}(\theta_\star^{(2)}) \end{bmatrix},$$

and $S(\theta_\star)$ *is a positive definite matrix of finite constants.*

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

Theorem 1 *Let Assumption 2 hold (and so* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *satisfy Assumption 1) then*

$$N_T \ \xrightarrow{d} \ \frac{(n_{q_1+q_2} + \bar{\eta})' C^{1/2\prime} \begin{bmatrix} I_{q_1} - P_0^{(1)} & 0 \\ 0 & -[I_{q_2} - P_0^{(2)}] \end{bmatrix} C^{1/2}(n_{q_1+q_2} + \bar{\eta})}{2\sqrt{(n_{q_1+q_2} + \bar{\eta})' C^{1/2\prime} \left\{I_{q_1+q_2} - P_0\right\}' \bar{C} \left\{I_{q_1+q_2} - P_0\right\} C^{1/2}(n_{q_1+q_2} + \bar{\eta})}} \tag{17}$$

[13]See the working paper verison of this paper for further details.

where $n_{q_1+q_2} \sim N(0, I_{q_1+q_2})$,

$$C^{1/2} = \begin{bmatrix} S^{(1)}(\theta_\star^{(1)})^{-1/2} & 0 \\ 0 & S^{(2)}(\theta_\star^{(2)})^{-1/2} \end{bmatrix} S(\theta_\star)^{1/2},$$

$$\bar{C} = \begin{bmatrix} S^{(1)}(\theta_\star^{(1)})^{-1/2} & 0 \\ 0 & -S^{(2)}(\theta_\star^{(2)})^{-1/2} \end{bmatrix} S(\theta_\star) \begin{bmatrix} S^{(1)}(\theta_\star^{(1)})^{-1/2} & 0 \\ 0 & -S^{(2)}(\theta_\star^{(2)})^{-1/2} \end{bmatrix}',$$

$$P_0 = \begin{bmatrix} P_0^{(1)} & 0 \\ 0 & P_0^{(2)} \end{bmatrix}, \qquad P_0^{(i)} = F_0^{(i)}(\theta_\star^{(i)}) \left[F_0^{(i)}(\theta_\star^{(i)})' F_0^{(i)}(\theta_\star^{(i)}) \right]^{-1} F_0^{(i)}(\theta_\star^{(i)})',$$

$$F_0^{(i)}(\theta_\star^{(i)}) = S^{(i)}(\theta_\star^{(i)})^{-1/2} G_0^{(i)}(\theta_\star^{(i)}), \qquad \bar{\eta} = S^{-1/2} \begin{bmatrix} S^{(1)^{1/2}} & 0 \\ 0 & S^{(2)^{1/2}} \end{bmatrix} \begin{bmatrix} \eta^{(1)} \\ \eta^{(2)} \end{bmatrix}.$$

It is evident from Theorem 1 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 the null hypothesis is satisfied because both models are correctly specified or locally misspecified. Furthermore, the actual limiting distribution is non-standard and depends on nuisance parameters, some of which can be consistently estimated (the long run variances and covariances of the moment conditions) and some of which cannot (the local misspecification $\bar{\eta}$).

The implication of this result is that we cannot use the test statistic N_T to discriminate between two models judged correctly specified, according to the overidentification test. This conclusion is drawn from the following logical sequence: (i) the overidentification test cannot discriminate between correctly specified and locally misspecified models with probability one even in the limit, (ii) under local misspecification the limit distribution of N_T is a function of the drift (iii) the drift cannot be consistently estimated. As a result, we see no way to simulate percentiles from the appropriate limit distribution.

To illustrate the sensitivity of N_T to the form of local alternatives, we report the results from a small simulation study. The data generating process considered is the following:

$$y_t = x_{1,t} + x_{2,t} + u_{0,t} \tag{18}$$

$$x_{1,t} = z_{1,t} + z_{2,t} + z_{3,t} + \gamma z_{4,t} + u_{1,t} \tag{19}$$

$$x_{2,t} = \alpha z_{3,t} + z_{4,t} + z_{5,t} + z_{6,t} + u_{2,t} \tag{20}$$

and the two models we compare are

$$y_t = \beta_1 x_{1,t} + \tilde{u}_{1,t} \tag{21}$$

$$y_t = \beta_2 x_{2,t} + \tilde{u}_{2,t} \tag{22}$$

i.e. we exclude one of the two explanatory variables. The variables $z_{1,t}$, $z_{2,t}$ and $z_{3,t}$ will be used as instruments for the first model [equation (21)] while the variables $z_{4,t}$, $z_{5,t}$ and $z_{6,t}$ will be used as instruments for the second model [equation (22)]. We draw the error terms u_t and the instruments z_t independently from a $N(0,1)$. We take a sample size equal to 1000, big enough so the simulated test statistics can be considered as draws from the limit distribution. The GMM estimation is done in two steps and we use an identity matrix as weighting matrix in the first step and the optimal weighting matrix in the second step.

Figure 1 contains the results for three scenarios. In the left panel, only the second model is locally misspecified ($\alpha = 10/\sqrt{T}$ and $\gamma = 0$). In the second panel, both models are correctly specified ($\alpha = \gamma = 0$). In the third panel, only the first model is locally misspecified ($\gamma = 10/\sqrt{T}$ and $\alpha = 0$). We draw an histogram for the statistic N_T using 20,000 replications for each scenario. We can clearly see that these distributions are not draws from a Normal distribution even though the null hypothesis is true for each case[14]. One conclusion that can be drawn from this exercise is that if one were to use the critical values from a $N(0,1)$ in conjunction with the statistic N_T, then the level of the test could not be controlled. Different local misspecification can move the distirbution of N_T left or right.

4 Non-local misspecification

The second approach to modelling misspecification in the literature is non-local (fixed) alternatives. Following Hall (2000) and Hall and Inoue (2003), a model is said to be misspecified in our context if there is no parameter value at which the moment condition can be set equal to zero, that is

[14]Lilliefors tests [see Lilliefors (1967)] reject the hypothesis that the N_T's have a normal distribution.

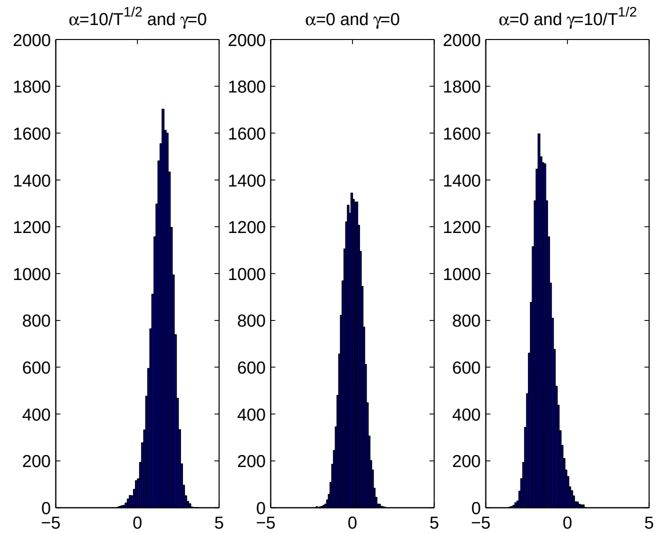

Figure 1: N_T statistic for correctly specified and locally misspecified models

Assumption 3 *$\mathcal{M}_i$ 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 $\theta^{(i)} \in \Theta^{(i)}$*

To implement the test, it is necessary to choose the weighting matrices. 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, we consider the case in which inference is based on GMM estimation with a weighting matrix that is either a matrix of constants, such as the identity matrix, or the inverse of an instrument cross product matrix. For these two cases, the construction of $\hat{\sigma}_T^2$ is different. The details are relegated to the appendix.

To analyze the behavior of the test in this case we impose one of the following two sets of assumptions:

Assumption 4 *(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 $S(\theta_\star)$ is a positive definite matrix of finite constants; (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) $\hat{S}^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{p} S^{(i)}(\theta_\star^{(i)})$ for $i = 1, 2$; (v) $\hat{S}^{(1,2)}(\hat{\theta}_T) \xrightarrow{p} S^{(1,2)}(\theta_\star)$.*

Assumption 5 *(i) $T^{-1/2}\sum_{t=1}^{T}\xi_t(\theta_\star) \xrightarrow{d} N(0, V_\star)$ where $V_\star$ is a positive semi-definite matrix; (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)* $\hat{V}_T \xrightarrow{p} V_\star$.

Hall and Inoue (2003)[Theorems 1 and 2] provide conditions under which $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_\star^{(i)})$ has a limiting normal distribution, and so Assumption 4(iii) or 5(iii) could be replaced by these lower level assumptions.

The following theorem gives the limiting distribution of N_T for these two choices of weighting matrix.

Theorem 2 *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 3; (iii)* H_0 *holds; and either: (a)* $W_T^{(i)} = I_{q_i}$ *and Assumption 4 holds; or (b):* $W_T^{(i)} = \{T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\}^{-1}$ *and Assumption 5 holds; then* $N_T \xrightarrow{d} N(0,\, 1)$.

Theorem 2 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 3. This result would appear to indicate that there is scope for using this statistic to compare two misspecified models estimated via GMM. However, 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.[15] Whether or not 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 (2006). However, absent these economic considerations, the choice of the weighting matrix and the relative ranking of the models can become arbitrary.

[15] An earlier version of this paper contains simulation evidence to illustrate this point.

5 Testing $\sigma_0^2 = 0$

As noted by Rivers and Vuong (2002), N_T only converges to a standard normal distribution under certain regularity conditions, and key amongst these conditions, is the restriction that $\sigma_0^2 > 0$. Our analysis highlights that this condition is apposite to the case where inference is based on a comparsion of GMM minimands. For on the one hand, if the null is satisfied because both models are respectively correctly specified or locally misspecifed then this variance condition fails and N_T converges to a non-standard distribution. It is further shown that this limiting distribution depends upon the drift and, as a result, it is not possible to develop satisfactory inference procedures based on N_T in this case. On the other hand, if the null is satisfied because both models are non-locally misspecified then the variance condition is satisfied and N_T has a standard normal limiting distribution.

This dichotomy creates a problem for any researcher wishing to use the test: how can he/she assess whether $\sigma_0^2 > 0$? A natural solution is 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 1, it follows that if the model is 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, $\bar{\eta}$. This means that the implementation of the test is problematic because the critical value for the asymptotically valid $100\alpha\%$ depends on $\bar{\eta}$ 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 ($\bar{\eta} = 0$) 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. 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 1 *and* $\eta^{(i)} = 0$ for $i = 1, 2$.[16] 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 1 for* $i = 1, 2$ *and Assumption 2 holds then* $lim_{T\to\infty} P(\textit{Type I} \,|\, \alpha) = 0$. *(ii) If* $\mathcal{M}_i$ *satisfy Assumption 3 for* $i = 1, 2$ *and the other assumptions of Theorem 2 hold then* $lim_{T\to\infty} P(\textit{Type II} \,|\, \alpha) = 0$.

Proposition 1 demonstrates that it is possible to develop a testing strategy that discriminates 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.

6 Extensions

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. For example, the test has been implemented in this form by Carpentier and Weaver (1997) and Nauges and Thomas (2003). An inspection of the proofs of Theorems 1 and 2 indicates that the same qualitative results go through whether the test is or is not based on the same weighting matrices as used in the estimations. In the case of two correctly specified models, the form of the limiting distribution changes from the one in Theorem 1 but it is not standard normal in general. Again, if the two models are correctly specified the

[16]This statement assumes the other conditions of Theorem 1 hold as well.

null will hold by construction and if they are locally misspecified the limit distribution will be a function of the drift. In the case of two non-locally misspecified models, the formulae for $\hat{\sigma}_T$ changes but once appropriately modified, the limiting distribution of N_T is standard normal. However, this version of the test is also subject to the same concerns raised above.

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) proposes an extension of Vuong's (1989) methods to GEL estimation of conditional moment restrictions models.[17] An immediate advantage of GEL methods is that there is no weighting matrix and thus the model ranking in unambiguous. However, GEL methods do share with GMM methods the problems highlighted above concerning with 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 case in which both models are correctly specified and considers its behaviour only under non-local alternatives. It is easily seen from his analysis that the same problems arise in the GEL setting when locally misspecification 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.[18] [19]

[17]Chen, Hong, and Shum (2007) extend Vuong's methods to the comparison of parametric models estimated by Maximum Likelihood to moment condition models estimated by EL. See Ramalho and Smith (2002) for an alternative approach to non-nested testing within the GEL framework.

[18]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.

[19]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 1 above.

Appendix

(i) Proof of Theorem 1:

It is useful to begin by defining two matrices:

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

The proof is then as follows. The test statistic N_T can be written as

$$N_T = \frac{T\left\{Q_T^{(1)}\left(\hat{\theta}_T^{(1)}\right) - Q_T^{(2)}\left(\hat{\theta}_T^{(2)}\right)\right\}}{T^{1/2}\hat{\sigma}_T}. \tag{23}$$

Analysis of the numerator in (23) is straightforward since it is simply the difference between two overidentification test statistics. Standard analysis of the overidentifying restrictions test yields[20]

$$TQ_T^{(i)}\left(\hat{\theta}_T^{(i)}\right) = \{T^{1/2}g_T^{(i)}(\theta_\star^{(i)})\}'[S^{(i)}]^{-1/2\prime}\left[I - P_0^{(i)}(\theta_\star^{(i)})\right][S^{(i)}]^{-1/2}\{T^{1/2}g_T^{(i)}(\theta_\star^{(i)})\} + o_p(1) \tag{24}$$

It follows from (24) and Assumption 2 that, dropping the dependence on $\theta_\star^{(i)}$ in places where it is obvious so as to lighten the notation,

$$T\left\{Q_T^{(1)}\left(\hat{\theta}_T^{(1)}\right) - Q_T^{(2)}\left(\hat{\theta}_T^{(2)}\right)\right\} \xrightarrow{d} (n_{q_1+q_2}+\bar{\eta})'C^{1/2\prime}\begin{bmatrix} I_{q_1} - P_0^{(1)} & 0 \\ 0 & -[I_{q_2} - P_0^{(2)}] \end{bmatrix} C^{1/2}(n_{q_1+q_2}+\bar{\eta}) \tag{25}$$

Now consider the denominator of (23). 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

$$\begin{aligned} T^{1/2}\hat{R}_T &= \left[2T^{1/2}g_T^{(1)}(\hat{\theta}_T^{(1)})'W_T^{(1)}, o_p(1), -2T^{1/2}g_T^{(2)}(\hat{\theta}_T^{(2)})'W_T^{(2)}, o_p(1)\right]' \\ &= 2\Gamma_V T^{1/2}g_T(\hat{\theta}_T) + o_p(1) \end{aligned} \tag{26}$$

[20]For example see Hall (2005)[equation (3.36),p.73].

Using a Mean Value Theorem expansion for $g_T(\hat{\theta}_T)$ and the standard asymptotic representation for $(\hat{\theta}_T - \theta_\star)$ (e.g. see Hall (2005)[equation (3.26)]), it follows from (26) that under our assumptions:

$$\begin{aligned} T^{1/2}\hat{R}_T &\stackrel{d}{\to} 2\Gamma_V \left\{ I_{q_1+q_2} - \begin{bmatrix} {S^{(1)}}^{1/2} P_0^{(1)} {S^{(1)}}^{-1/2} & 0 \\ 0 & {S^{(2)}}^{1/2} P_0^{(2)} {S^{(2)}}^{-1/2} \end{bmatrix} \right\} S^{1/2}(n_{q_1+q_2} + \bar{\eta}) \\ &= 2\Gamma_V \left\{ I_{q_1+q_2} - \begin{bmatrix} {S^{(1)}}^{1/2} & 0 \\ 0 & {S^{(2)}}^{1/2} \end{bmatrix} P_0 \begin{bmatrix} {S^{(1)}}^{-1/2} & 0 \\ 0 & {S^{(2)}}^{-1/2} \end{bmatrix} \right\} S^{1/2}(n_{q_1+q_2} + \bar{\eta}) \end{aligned}$$

Now consider $\hat{V}_T$. Under our assumptions, we have that $\hat{V}_T \stackrel{p}{\to} \Gamma_U S \Gamma_U'$. Furthermore, we have

$$\Gamma_V' \Gamma_U S \Gamma_U' \Gamma_V = \begin{bmatrix} {S^{(1)}}^{-1/2} & 0 \\ 0 & {S^{(2)}}^{-1/2} \end{bmatrix}' \bar{C} \begin{bmatrix} {S^{(1)}}^{-1/2} & 0 \\ 0 & {S^{(2)}}^{-1/2} \end{bmatrix}$$

Therefore, combining these results for constituents of the denominator, we obtain

$$T\hat{\sigma}^2 \stackrel{d}{\to} 4(n_{q_1+q_2} + \bar{\eta})' C^{1/2\prime} \{I_{q_1+q_2} - P_0\}' \bar{C} \{I_{q_1+q_2} - P_0\} C^{1/2}(n_{q_1+q_2} + \bar{\eta}) \tag{27}$$

The result then follows from (25) and (27).

(ii) Construction of $\hat{\sigma}_T^2$ in non-locally misspecified models

Case (a): $W_T^{(i)} = I_{q_i}$.

With this choice, both $A_\star^{(i)}$ and $\Delta^{(i)}$ are null matrices and so the form of σ_0^2 simplifies to:

$$\begin{aligned} \sigma_0^2 &= 4 \Big\{ \mu^{(1)}(\theta_\star^{(1)})' S^{(1)}(\theta_\star^{(1)}) \mu^{(1)}(\theta_\star^{(1)}) + \mu^{(2)}(\theta_\star^{(2)})' S^{(2)}(\theta_\star^{(2)}) \mu^{(2)}(\theta_\star^{(2)}) \\ &\qquad - 2\mu^{(1)}(\theta_\star^{(1)})' S^{(1,2)}(\theta_\star) \mu^{(1)}(\theta_\star^{(1)}) \Big\} \end{aligned} \tag{28}$$

where $S(\theta_\star) = \lim_{T\to\infty} Var[T^{1/2} g_T(\theta_\star)]$ and

$$S(\theta_\star) = \begin{bmatrix} S^{(1)}(\theta_\star^{(1)}) & S^{(1,2)}(\theta_\star) \\ S^{(1,2)}(\theta_\star)' & S^{(2)}(\theta_\star^{(2)}) \end{bmatrix}, \tag{29}$$

The obvious choice of $\hat{\sigma}_T^2$ is therefore,

$$\begin{aligned} \hat{\sigma}_T^2 &= 4 \Big\{ g_T^{(1)}(\hat{\theta}_T^{(1)})' \hat{S}^{(1)}(\hat{\theta}_T^{(1)}) g_T^{(1)}(\hat{\theta}_T^{(1)}) + g_T^{(2)}(\hat{\theta}_T^{(2)})' \hat{S}^{(2)}(\hat{\theta}_T^{(2)}) g_T^{(2)}(\hat{\theta}_T^{(2)}) \\ &\qquad - 2\, g_T^{(1)}(\hat{\theta}_T^{(1)})' \hat{S}^{(1,2)}(\hat{\theta}_T) g_T^{(2)}(\hat{\theta}_T^{(2)}) \Big\} \end{aligned} \tag{30}$$

Case (b): $W_T^{(i)} = \{T^{-1}\sum_{t=1}^T z_t^{(i)} z_t^{(i)\prime}\}^{-1} = \{\hat{M}_{zz}^{(i)}\}^{-1}$ *where* $z_t^{(i)}$ *are the instruments and* $W^{(i)} = E[z_t^{(i)} z_t^{(i)\prime}] = \{M_{zz}^{(i)}\}^{-1}$.

For this case, the weighting matrix does not depend on the previous step estimates so we treat $W_T^{(i)}$ as the nuisance parameters [see equations (8) and (9)]. Accordingly, we have $\sigma_0^2 = R_\star' V_\star R_\star$ where $V_\star = \lim_{T\to\infty} Var[T^{-1/2}\sum_{t=1}^T \xi_t]$ for $\xi_t = [a_{1,t}', b_{1,t}', a_{2,t}', b_{2,t}']$ where $a_{i,t} = f^{(i)}(v_t, \theta_\star^{(i)}) - E[f^{(i)}(v_t, \theta_\star^{(i)})]$ and $b_{i,t} = vech\{(z_t^{(1)} z_t^{(i)\prime}) - M_{zz}^{(i)}\}$ for $i = 1, 2$, and $R_\star$ is defined by (13) with $A_\star^{(i)} = I_{q_i(q_i+1)/2}$ and $Y_t^{(i)} = vech\{(z_t^{(i)} z_t^{(i)\prime}) - M_{zz}^{(i)}\}$ and $\Delta^{(i)} = L_i\left(\{M_{zz}^{(i)}\}^{-1} \otimes \{M_{zz}^{(i)}\}^{-1}\right) B_i$. We therefore set

$$\hat{\sigma}_T^2 \;=\; \hat{R}_T' \hat{V}_T \hat{R}_T \tag{31}$$

where $\hat{V}_T$ is a consistent estimator of $V_\star$ based on $\hat{\xi}_t = \left[\hat{a}_{1,t}', \hat{b}_{1,t}', \hat{a}_{2,t}', \hat{b}_{2,t}'\right]$ where $\hat{a}_{i,t} = f^{(i)}(v_t, \hat{\theta}_T^{(i)}) - g_T^{(i)}(\hat{\theta}_T^{(i)})$, $\hat{b}_{i,t} = vech\{(z_t^{(i)} z_t^{(i)\prime}) - \hat{M}_{zz}^{(i)}\}$ for $i = 1, 2$, and

$$\hat{R}_T \;=\; \begin{bmatrix} \hat{R}_T^{(1)} \\ -\hat{R}_T^{(2)} \end{bmatrix}, \quad \hat{R}_T^{(i)} \;=\; \begin{bmatrix} 2\{\hat{M}_{zz}^{(i)}\}^{-1} g_T^{(i)}(\hat{\theta}_T^{(i)}) \\ -\hat{\Delta}_T^{(i)} B_i'\{g_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes g_T^{(i)}(\hat{\theta}_T^{(i)})\} \end{bmatrix}, \quad \hat{\Delta}_T^{(i)} \;=\; L_i[\{\hat{M}_{zz}^{(i)}\}^{-1} \otimes \{\hat{M}_{zz}^{(i)}\}^{-1}] B_i$$

(iii) Proof of Theorem 2

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

$$Q_T^{(i)}(\hat{\theta}_T^{(i)}) \;=\; Q_T^{(i)}(\theta_\star^{(i)}) \;+\; \left\{ \left.\frac{\partial Q_T^{(i)}(\theta^{(i)})}{\partial \theta^{(i)}}\right|_{\theta^{(i)} = \bar{\theta}_T^{(i)}} \right\}' \left(\hat{\theta}_T^{(i)} \;-\; \theta_\star^{(i)}\right) \tag{32}$$

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

$$\Phi^{(i)}(\theta_\star^{(i)}) \;=\; 2G_0^{(i)}(\theta_\star^{(i)})' W^{(i)} E[f^{(i)}(v_t, \theta_\star^{(i)})] \tag{33}$$

It the follows from (32) that under our assumptions, we have

$$Q_T^{(i)}(\hat{\theta}_T^{(i)}) \;=\; Q_T^{(i)}(\theta_\star^{(i)}) \;+\; \left\{ \frac{\partial Q_0^{(i)}(\theta_\star^{(i)})}{\partial \theta^{(i)}} \right\}' \left(\hat{\theta}_T^{(i)} \;-\; \theta_\star^{(i)}\right) \;+\; o_p(T^{-1/2}) \tag{34}$$

and hence

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) \;-\; Q_T^{(2)}(\hat{\theta}_T^{(2)}) \right] \;&=\; T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) \;-\; Q_T^{(2)}(\theta_\star^{(2)}) \right] \\ &\quad + \left\{ \Phi^{(i)}(\theta_\star^{(1)}) \right\}' T^{1/2}\left(\hat{\theta}_T^{(1)} \;-\; \theta_\star^{(1)}\right) \\ &\quad - \left\{ \Phi^{(i)}(\theta_\star^{(2)}) \right\}' T^{1/2}\left(\hat{\theta}_T^{(2)} \;-\; \theta_\star^{(2)}\right) \;+\; o_p(1) \end{aligned} \tag{35}$$

Finally, under H_0, we have $Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)})$ and so (35) can be written as

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\right] &= T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) - Q_0^{(1)}(\theta_\star^{(1)})\right] - T^{1/2}\left[Q_T^{(2)}(\theta_\star^{(2)}) - Q_0^{(2)}(\theta_\star^{(2)})\right] \\ &\quad + \left\{ \Phi^{(i)}(\theta_\star^{(1)})\right\}' T^{1/2}\left(\hat{\theta}_T^{(1)} - \theta_\star^{(1)}\right) \\ &\quad - \left\{ \Phi^{(i)}(\theta_\star^{(2)})\right\}' T^{1/2}\left(\hat{\theta}_T^{(2)} - \theta_\star^{(2)}\right) + o_p(1) \end{aligned} \tag{36}$$

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

$$T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\right] = T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) - Q_0^{(1)}(\theta_\star^{(1)})\right] - T^{1/2}\left[Q_T^{(2)}(\theta_\star^{(2)}) - Q_0^{(2)}(\theta_\star^{(2)})\right] + o_p(1) \tag{37}$$

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

$$\hat{h}'\hat{W}\hat{h} - h'Wh = \hat{h}'\hat{W}(\hat{h} - h) + \hat{h}'(\hat{W} - W)h + (\hat{h} - h)'Wh \tag{38}$$

Using (37) and (38), we now deduce the results for the two choices of weighting matrices considered in the theorem.

Part (a): With $W_T^{(i)} = I_{q_i}$, it follows from (37) and (38) that

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

The result then follows immediately under the stated assumptions.

Part (b): With $W_T^{(i)} = \{\hat{M}_{zz}^{(i)}\}^{-1}$, it follows from (37) and (38) that

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\right] \;=\;& 2\,\mu^{(1)}(\theta_\star^{(1)})'\{M_{zz}^{(1)}\}^{-1}T^{-1/2}\sum_{t=1}^{T}[f^{(1)}(v_t,\theta_\star^{(1)}) - \mu^{(1)}(\theta_\star^{(1)})] \\ &+ \mu^{(1)}(\theta_\star^{(1)})'T^{1/2}(\{\hat{M}_{zz}^{(1)}\}^{-1} - \{M_{zz}^{(1)}\}^{-1})\mu^{(1)}(\theta_\star^{(1)}) \\ &- 2\,\mu^{(2)}(\theta_\star^{(2)})'\{M_{zz}^{(2)}\}^{-1}T^{-1/2}\sum_{t=1}^{T}[f^{(2)}(v_t,\theta_\star^{(2)}) - \mu^{(2)}(\theta_\star^{(2)})] \\ &- \mu^{(2)}(\theta_\star^{(2)})'T^{1/2}(\{\hat{M}_{zz}^{(2)}\}^{-1} - \{M_{zz}^{(2)}\}^{-1})\mu^{(2)}(\theta_\star^{(2)}) \\ &+ o_p(1) \qquad (39) \end{aligned}$$

Using $A^{-1} - B^{-1} = B^{-1}(B-A)A^{-1}$, we have

$$T^{1/2}(\{\hat{M}_{zz}^{(1)}\}^{-1} - \{M_{zz}^{(1)}\}^{-1}) \;=\; -\{M_{zz}^{(1)}\}^{-1}T^{1/2}\sum_{t=1}^{T}[z_t^{(i)}z_t^{(i)\prime} - M_{zz}^{(i)}]\{\hat{M}_{zz}^{(1)}\}^{-1} \qquad (40)$$

and then from Dhrymes (1984)[Corollary 25, p.103]

$$\begin{aligned} \mu^{(i)}(\theta_\star^{(i)})'T^{1/2}\left(\{\hat{M}_{zz}^{(i)}\}^{-1} - \{M_{zz}^{(i)}\}^{-1}\right)\mu^{(i)}(\theta_\star^{(i)}) \;=\;& \left[\mu^{(i)}(\,\theta_\star^{(i)})' \otimes \mu^{(i)}(\theta_\star^{(i)})'\right]\left[\{M_{zz}^{(i)}\}^{-1} \otimes \{M_{zz}^{(i)}\}^{-1}\right] \times \\ & B_i T^{-1/2}\sum_{t=1}^{T} vech[z_t^{(i)}z_t^{(i)\prime} - M_{zz}^{(i)}] + o_p(1) \qquad (41) \end{aligned}$$

The result then follows from (39)-(41) and the stated assumptions.

(iv) 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. Pötscher (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

Atkinson, A. C. (1970). 'A method for discriminating between models', *Journal of the Royal Statistical Society, Series B*, 32: 323–353.

Carpentier, A., and Weaver, R. D. (1997). 'Damage control productivity: why econometrics matters', *American Journal of Agricultural Economics*, 79: 47–61.

Chen, X., Hong, H., and Shum, M. (2007). 'Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models', *Journal of Econometrics*, 141: 109–140.

Cox, D. (1961). 'Tests of separate families of hypotheses', in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, Berkeley, CA, U. S. A.

——— (1962). 'Further results on tests of separate families of hypotheses', *Journal of the Royal Statistical Society, Series B*, 24: 406–424.

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

Dhrymes, P. J. (1984). *Mathematics for Econometrics*. Springer-Verlag, New York, NY, U. S. A., second edn.

Dridi, R., Guay, A., and Renault, E. (2006). 'Indirect inference and calibration of dynamic general equilibrium models', Forthcoming in Journal of Econometrics.

Ghysels, E., and Hall, A. R. (1990). 'Testing non–nested Euler conditions with quadrature–based methods of approximation', *Journal of Econometrics*, 46: 273–308.

Hall, A. R. (2000). 'Covariance matrix estimation and the power of the overidentifying restrictions test', *Econometrica*, 68: 1517–1527.

——— (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. (1982). 'Large sample properties of Generalized Method of Moments estimators', *Econometrica*, 50: 1029–1054.

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 Dunamic 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.

Lilliefors, H. W. (1967). 'On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown', *Journal of the American Statistical Association* , 62(318): 399–402.

Mizon, G. E., and Richard, J. F. (1986). 'The encompassing principle and its application to testing non–nested hypotheses', *Econometrica*, 54: 657–678.

Nauges, C., and Thomas, A. (2003). 'Long-run study of residential water consumption', *Environmental and Resource Economics*, 26: 25–43.

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

Newey, W. K., and West, K. D. (1987). 'A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica*, 55: 703–708.

Pesaran, M. H., and Deaton, A. (1978). 'Testing non–nested nonlinear regression models', *Econometrica*, 46: 677–694.

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

Ramalho, J. J., and Smith, R. J. (2002). 'Generalized empirical likelihood non-nested tests', *Journal of Econometrics*, 107: 99–125.

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

Sin, C.-Y., and White, H. (1996). 'Information criteria for selecting possibly misspecified parametric models', *Journal of Econometrics*, 71: 207–225.

Singleton, K. J. (1985). 'Testing specifications of economic agents' intertemporal optimum problems in the prescence of alternative models', *Journal of Econometrics*, 30: 391–413.

Smith, R. J. (1992). 'Non–nested tests for competing models estimated by Generalized Method of Moments', *Econometrica*, 60: 973–980.

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

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

White, H. (1982). 'Maximum likelihood in misspecified models', *Econometrica*, 50: 1–25.