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

Alastair R. Hall

North Carolina State University and University of Manchester[2]

and

Denis Pelletier

North Carolina State University[3]

March 6, 2007

[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. The null hypothesis is that the models have equal measures of goodness of fit; one model is preferred if its goodness of fit is statistically significantly smaller than its competitor. 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. In the latter case it is shown that the choice of weighting matrices affects the outcome of the test and thus the ranking of the models.

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. The seminal work on non-nested hypothesis testing was undertaken in the context of Maximum Likelihood estimation in the 1960's and 1970's in the statistics literature; see Cox(1961, 1962) and Atkinson (1970). There was also considerable interest within the econometrics literature in the context of discriminating between non-nested regression models in the late 1970's and 1980's; see *inter alia* Pesaran and Deaton (1978), Davidson and MacKinnon (1981) and Mizon and Richard (1986).

A characteristic of all these tests is that under the null hypothesis one model is assumed to be correct. 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.[1]

With the emergence of Generalized Method of Moments (GMM) (Hansen, 1982), there has been a natural interest in developing non–nested tests within this more general framework. Singleton (1985), Ghysels and Hall (1990) and Smith (1992) propose tests that are applicable in a variety of settings[2]. However, one feature common to all three is that under the null hypothesis

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

[2]See Ramalho and Smith (2002) for an extension of these testing procedures to generalized empirical likelihood

one of the models in question is assumed to be correctly specified. Just as in the context of QML, it would be desirable to develop tests within the GMM framework that allow the researcher to determine which of two competing models is closer to the truth.

In a recent paper, 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 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

---

estimators.

[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).

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 also 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.[4] 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.[5] It is in this case that the dependence of the outcome on the weighting matrix becomes troublesome. Notice that this dependence is present not only in the case where inference is based on first-step minimands but also when second-step minimands are used because the limiting behaviour of the second step estimator depends on that of the first step estimator which in turn depends on the first step weighting matrix.[6] We present an example below that illustrates how different choices of weighting matrices lead to different rankings of two misspecified models based on the test.

Taking these results together, it would seem that, out of the scenarios considered, the only scope for using the Rivers and Vuong's (2002) test statistic is in non-locally misspecified models

[4]Examples 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).

[5]See Hall and Inoue (2003).

[6]See Hall and Inoue (2003).

in which economic theory dictates a choice of weighting matrix. However, even then, there is a further concern. Within this scenario, it is only appropriate to compare the test statistic to the appropriate percentile of the standard normal distribution if the variance condition is satisfied or, in other words, if both models really are non-locally misspecified. Ideally, a researcher would perform a pre-test for this restriction but while this variance pre-test works in some settings, it is shown below that it does not work in our setting because there is no way to discriminate usefully between non-locally misspecified models and the combined class of correctly specified or locally misspecified models.

It should be noted that our results are confined to the specific situation in which goodness of fit is measured via GMM minimands evaluated at GMM estimators and should not be construed as a general critique of the Rivers and Vuong's (2002) testing procedures.

An outline of the paper is as follows. Section 2 presents a review of the Rivers and Vuong's (2002) statistic and is broken down into two sub-sections that cover: (i) certain preliminaries including notation, assumptions and a summary of GMM estimation; (ii) the test statistic. 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 issue of whether the variance condition can be tested in our context. Section 6 concludes. All proofs are relegated to a Mathematical Appendix and simulation results in Sections 3 and 4 illustrate our results.

2 Framework and Analysis

2.1 Preliminaries

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 constituents of these population moment conditions satisfy:[7]

Assumption 1 *The* $(r \times 1)$ *random vectors* $\{v_t; -\infty < t < \infty\}$ *form a strictly stationary ergodic process with sample space* $\mathbf{V} \subseteq \Re^r$*.*

Assumption 2 *For each* $i = 1, 2$*, the function* $f^{(i)} : \mathbf{V} \times \Theta^{(i)} \to \Re^{q_i}$*, where* $q_i < \infty$*, satisfies: (i) it is continuous on* $\Theta^{(i)}$ *for each* $v_t \in \mathbf{V}$*; (ii)* $E[f^{(i)}(v_t,\theta^{(i)})]$ *exists and is finite for every* $\theta^{(i)} \in \Theta^{(i)} \subset \Re^{p_i}$*; (iii)* $E[f^{(i)}(v_t,\theta^{(i)})]$ *is continuous on* $\Theta^{(i)}$*; (iv) the system is overidentified, i.e.* $q_i > p_i$*.*

It is 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

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

and $W_T^{(i)}$ is the weighting matrix. It is assumed that the weighting matrix satisfies:

[7] To present their analysis, Rivers and Vuong (2002) impose relatively weak assumptions on the dynamic structure of the data. For the purposes of exposition, we adopt a more restrictive framework in which the data are assumed to be generated by a strictly stationary, ergodic process. This restriction is not crucial for the qualitative results below.

Assumption 3 $W_T^{(i)}$ *is a positive semi–definite symmetric matrix which converges in probability to the positive definite matrix of constants* $W^{(i)}$.

In the standard asymptotic analysis of GMM, Assumption 3 is the only restriction on the weighting matrix. However, for the analysis of Rivers and Vuong's (2002) statistic, it is necessary to be more specific about its structure. It is assumed that the weighting matrix, $W^{(i)}$ 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:

$$T^{1/2}(\hat{\tau}_T^{(i)} - \tau_0^{(i)}) = -A_\star^{(i)} T^{-1/2} \sum_{t=1}^{T} Y_t^{(i)} + o_p(1) \tag{3}$$

for some symmetric matrix of constants $A_\star^{(i)}$ and vector $Y_t^{(i)}$; and that the weighting matrix satisfies[8]:

$$T^{1/2} \left(vech[W_T^{(i)}] - vech[W^{(i)}] \right) = \Delta^{(i)} T^{1/2} (\hat{\tau}_T^{(i)} - \tau_0^{(i)}) + o_p(1) \tag{4}$$

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.

The statement of the null and alternative hypotheses of the Rivers and Vuong's (2002) test involves the probability limits of the two estimators. To ensure these probability limits exist, we introduce the following additional assumptions.

Assumption 4 $\Theta^{(i)}$ *is a compact set for* $i = 1, 2$.

Assumption 5 $E[sup_{\theta^{(i)} \in \Theta^{(i)}} ||f^{(i)}(v_t, \theta^{(i)})||] < \infty$ *for* $i = 1, 2$.

[8]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.

Under Assumptions 1-5, it can be shown that the GMM minimands converge uniformly on $\Theta^{(i)}$ to their population analogs,[9] that is:

$$\sup_{\theta^{(i)}\in\Theta^{(i)}} |Q_T^{(i)}(\theta^{(i)}) - Q_0^{(i)}(\theta^{(i)})| \xrightarrow{p} 0 \tag{5}$$

for $i = 1, 2$, where

$$Q_0^{(i)}(\theta^{(i)}) \;=\; E[f^{(i)}(v_t, \theta^{(i)})]' W^{(i)} E[f^{(i)}(v_t, \theta^{(i)})] \tag{6}$$

Assumption 6 *Assume there exists $\theta_\star^{(i)} \in \Theta^{(i)}$ such that $Q_0^{(i)}(\theta_\star^{(i)}) < Q_0^{(i)}(\theta^{(i)})$ for all $\theta^{(i)} \in \Theta^{(i)} \setminus \{\theta_\star^{(i)}\}$.*

Assumption 6 constitutes the identification condition. Note that at this stage we have made no presumption about whether either model is correctly specified, and so the identification condition cannot be deduced from the veracity of the population moment condition as is the case in standard GMM theory for correctly specified models. Assumption 6 is a specialization of the identifiable uniqueness conditions to the stationary environment considered here and is the identification condition employed by Hall and Inoue (2003) in their analysis of the limiting distribution theory of GMM estimators in misspecified models.

The probability limits of the estimators are then as follows.[10]

Lemma 1 *If Assumptions 1-6 hold then $\hat{\theta}_T^{(i)} \xrightarrow{p} \theta_\star^{(i)}$ for $i = 1, 2$.*

In addition to the foregoing assumptions, we must also place restrictions on $\theta_\star^{(i)}$ and the derivative matrices:

$$G_T^{(i)}(\theta^{(i)}) \;=\; T^{-1}\sum_{t=1}^{T} \partial f^{(i)}(v_t, \theta^{(i)})/\partial\theta^{(i)\prime} \tag{7}$$

$$G_0^{(i)}(\theta^{(i)}) \;=\; E[\partial f^{(i)}(v_t, \theta^{(i)})/\partial\theta^{(i)\prime}] \tag{8}$$

[9] For example, see Newey and McFadden (1994) or Wooldridge (1994).

[10] For a proof under more general conditions see Rivers and Vuong (2002); for a proof under the conditions stated see Hall (2005)[Sections 3.1, 4.1, 4.4].

These restrictions are as follows.

Assumption 7 $\theta_{\star}^{(i)}$ *is an interior point of* $\Theta^{(i)}$*, for* $i = 1, 2$.

Assumption 8 *For* $i = 1, 2$ *we have: (i)* $\partial f^{(i)}(v_t, \theta^{(i)})/\partial\theta^{(i)\prime}$ *exists and is continuous on* $\Theta^{(i)}$ *for each* $v_t \in \mathbf{V}$*; (ii)* $G_0^{(i)}(\theta^{(i)})$ *exists, is finite, and is continuous on some neighborhood* $N_{\epsilon}^{(i)}$ *of* $\theta_{\star}^{(i)}$*; (iii)* $\sup_{\theta^{(i)} \in N_{\epsilon}^{(i)}} ||G_T^{(i)}(\theta^{(i)}) - G_0^{(i)}(\theta^{(i)})|| \xrightarrow{p} 0.$ [11]

To conclude this sub-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 $\theta = [\theta^{(1)\prime}, \theta^{(2)\prime}]'$, $f(v_t, \theta) = [f^{(1)}(v_t, \theta^{(1)})', f^{(2)}(v_t, \theta^{(2)})']'$ and $g_T(\theta) = [g_T^{(1)}(\theta^{(1)})', g_T^{(2)}(\theta^{(2)})']'$ for $g_T^{(i)}(\theta^{(i)}) = T^{-1}\sum_{t=1}^{T} f^{(i)}(v_t, \theta^{(i)})$. 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 $S^{-1/2} = [S^{1/2}]^{-1}$.

2.2 Rivers and Vuong's test statistic and the decision rule

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:

$$N_T \;=\; \frac{T^{1/2}\{\,Q_T^{(1)}(\hat{\theta}_T^{(1)}) \,-\, Q_T^{(2)}(\hat{\theta}_T^{(2)})\,\}}{\hat{\sigma}_T} \tag{9}$$

where $\hat{\sigma}_T^2$ is a consistent estimator of σ_0^2, the limiting variance of the numerator of (9). 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

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

[11] For any matrix A, we define $\|A\| = [tr(A'A)]^{1/2}$.

for

$$\xi_t(\theta_\star) = \begin{bmatrix} f^{(1)}(v_t, \theta_\star^{(1)}) - E[f^{(1)}(v_t, \theta_\star^{(1)})] \\ Y_t^{(1)} \\ f^{(2)}(v_t, \theta_\star^{(2)}) - E[f^{(2)}(v_t, \theta_\star^{(2)})] \\ Y_t^{(2)} \end{bmatrix}, \tag{12}$$

and

$$R_\star = \left[R_\star^{(1)\prime}, -R_\star^{(2)\prime}\right]^\prime \tag{13}$$

$$R_\star^{(i)} = \begin{bmatrix} 2W^{(i)}E[f(v_t, \theta_\star^{(i)})] \\ -A_\star^{(i)}\Delta^{(i)\prime}B_i^\prime\{E[f^{(i)}(v_t, \theta_\star^{(i)})] \otimes E[f^{(i)}(v_t, \theta_\star^{(i)})]\} \end{bmatrix} \tag{14}$$

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 (3)-(4).

The null and alternative hypotheses of the test are as follows. The null hypothesis is that: $\mathcal{M}_1$ and $\mathcal{M}_2$ are asymptotically equivalent, that is

$$Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)}) \tag{15}$$

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

$$Q_0^{(1)}(\theta_\star^{(1)}) < Q_0^{(2)}(\theta_\star^{(2)}) \tag{16}$$

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

$$Q_0^{(1)}(\theta_\star^{(1)}) > Q_0^{(2)}(\theta_\star^{(2)}) \tag{17}$$

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 $\sigma_0^2 > 0$.

3 Correct specification and local misspecification

In this section we examine the limiting behavior of N_T under H_0 when both models are correct (Section 3.1) or locally misspecified (Section 3.3).

3.1 Correctly specified models

Within our context, a model is "correctly specified" if the population moment can be set equal to zero at some parameter value, that is:

Assumption 9 $\mathcal{M}_i$ *satisfies* $E[f^{(i)}(v_t, \theta_\star^{(i)})] = 0$.

Using Assumption 9 and (10)-(14), it can be seen that, for the case under consideration here, $R_\star$ is a null matrix and hence $\sigma_0^2 = 0$. Therefore, if both models are correctly specified then the null distribution of N_T does not follow from Rivers and Vuong's (2002) analysis.[12] 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 in this sub-section 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, we assume that the weighting matrices are chosen so that $W^{(i)} = \{S^{(i)}\}^{-1}$ and $W_T^{(i)}$ depends on $\hat{\tau}_T^{(i)} = \tilde{\theta}_T^{(i)}$, a preliminary GMM estimator of $\theta_\star^{(i)}$ based on $E[f^{(i)}(v_t, \theta_\star^{(i)})] = 0$. We further assume this preliminary GMM estimation is performed 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 (3)

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

are given by

$$A_\star^{(i)} = -\Big[\, G_0^{(i)}(\theta_\star^{(i)})'\, M^{(i)}\, G_0^{(i)}(\theta_\star^{(i)}) \,\Big]^{-1} \tag{18}$$

$$Y_t^{(i)} = G_0^{(i)}(\theta_\star^{(i)})'\, M^{(i)}\, f^{(i)}(v_t, \theta_\star^{(i)}) \tag{19}$$

To define $\Delta^{(i)}$, assume that $W_T^{(i)} = \{S_T^{(i)}(\tilde{\theta}_T^{(i)})\}^{-1}$. It then follows that[13]

$$\Delta^{(i)} = L_i[\{S^{(i)}(\theta_\star^{(i)})\}^{-1} \otimes \{S^{(i)}(\theta_\star^{(i)})\}^{-1}]\Sigma^{(i)} \tag{20}$$

where

$$\Sigma^{(i)} = E\left[\left. \frac{\partial vec[S_T(\theta^{(i)})]}{\partial \theta^{(i)\prime}} \right|_{\theta^{(i)}=\theta_\star^{(i)}} \right] \tag{21}$$

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

$$\hat{\sigma}_T^2 = \hat{R}_T' \hat{V}_T \hat{R}_T \tag{22}$$

where $\hat{V}_T$ is a consistent estimator of $V_\star$ based on

$$\hat{\xi}_t = \begin{bmatrix} f^{(1)}(v_t, \hat{\theta}_T^{(1)}) \\ G_T^{(1)}(\tilde{\theta}_T^{(1)})' M_T^{(1)} f^{(1)}(v_t, \tilde{\theta}_T^{(1)}) \\ f^{(2)}(v_t, \hat{\theta}_T^{(2)}) \\ G_T^{(2)}(\tilde{\theta}_T^{(2)})' M_T^{(2)} f^{(2)}(v_t, \tilde{\theta}_T^{(2)}) \end{bmatrix} \tag{23}$$

and

$$\hat{R}_T = \begin{bmatrix} \hat{R}_T^{(1)} \\ -\hat{R}_T^{(2)} \end{bmatrix} \tag{24}$$

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

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

$$\hat{R}_T^{(i)} = \begin{bmatrix} 2W_T^{(i)} g_T^{(i)}(\hat{\theta}_T^{(i)}) \\ -\hat{A}_T^{(i)} \hat{\Delta}_T^{(i)\prime} B_i' \{g_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes g_T^{(i)}(\hat{\theta}_T^{(i)})\} \end{bmatrix} \tag{25}$$

$$\hat{A}_T^{(i)} = -[G_T^{(i)}(\tilde{\theta}_T^{(i)})' M_T^{(i)} G_T^{(i)}(\tilde{\theta}_T^{(i)})]^{-1} \tag{26}$$

$$\hat{\Delta}_T^{(i)} = L_i[\{\hat{S}^{(i)}(\tilde{\theta}_T^{(i)})\}^{-1} \otimes \{\hat{S}^{(i)}(\tilde{\theta}_T^{(i)})\}^{-1}]\hat{\Sigma}_T^{(i)} \tag{27}$$

and $\hat{\Sigma}_T^{(i)}$ is a consistent estimator of $\Sigma^{(i)}$.

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

Assumption 10 *(i)* $T^{-1/2}\sum_{t=1}^{T} f(v_t, \theta_\star) \ \xrightarrow{d}\ N\,(\,0,\, S(\theta_\star)\,)$ *where* $S(\theta_\star)$,

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

is a positive definite matrix of finite constants; (ii) $rank\{G_0^{(i)}(\theta_\star^{(i)})\} = p_i$*; (iii)* $\hat{S}_T^{(i)}(\tilde{\theta}_T^{(i)}) \xrightarrow{p} S^{(i)}(\theta_\star^{(i)})$*; (iv)* $\hat{\Sigma}_T^{(i)} \xrightarrow{p} \Sigma^{(i)}$*; (v)* $M_T^{(i)}$ *is a positive semi-definite matrix that converges in probability to* $M^{(i)}$*, a positive definite matrix of constants.*

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

Theorem 1 *Let Assumptions 1-8 and 10 hold. If* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *satisfy Assumption 9 then*

$$N_T \xrightarrow{d} \frac{n_{q_1+q_2}' C^{1/2\prime} \begin{bmatrix} I_{q_1} - P_0^{(1)} & 0 \\ 0 & -[I_{q_2} - P_0^{(2)}] \end{bmatrix} C^{1/2} n_{q_1+q_2}}{2\sqrt{n_{q_1+q_2}' C^{1/2\prime}\{I_{q_1+q_2} - P_0\}' \bar{C}\{I_{q_1+q_2} - P_0\} C^{1/2} n_{q_1+q_2}}} \tag{28}$$

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

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

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

$$
\begin{aligned}
P_0 &= \begin{bmatrix} P_0^{(1)} & 0 \\ 0 & P_0^{(2)} \end{bmatrix} \\
P_0^{(i)} &= F_0^{(i)}(\theta_\star^{(i)}) \left[F_0^{(i)}(\theta_\star^{(i)})' F_0^{(i)}(\theta_\star^{(i)}) \right]^{-1} F_0^{(i)}(\theta_\star^{(i)})', \\
F_0^{(i)}(\theta_\star^{(i)}) &= S^{(i)}(\theta_\star^{(i)})^{-1/2} G_0^{(i)}(\theta_\star^{(i)}).
\end{aligned}
$$

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. Furthermore, the actual limiting distribution is non-standard and depends on nuisance parameters such as the the long run variances and covariances of the moment conditions. In Section 3.2, we present results from a simple simulation study that demonstrate the differences between the limiting distribution in Theorem 1 and the standard normal.

3.2 Simulation results 1

In this sub-section we run some Monte Carlo simulations to illustrate the above theoretical results. The experiment will consist of two correctly specified models and demonstrate that the test statistic under the null hypothesis does not have a standard normal distribution. The data generating process employed is the following:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} \tag{29}$$

$$x_{1,t} = \gamma_1 z_{1,t} + \gamma_2 z_{2,t} + \cdots + \gamma_6 z_{6,t} + u_{1,t} \tag{30}$$

$$x_{2,t} = \alpha_1 z_{1,t} + \alpha_2 z_{2,t} + \cdots + \alpha_6 z_{6,t} + u_{2,t} \tag{31}$$

and the two models we are going to compare will be

$$y_t = \beta_1 x_{1,t} + \tilde{u}_{1,t} \tag{32}$$

$$y_t = \beta_2 x_{2,t} + \tilde{u}_{2,t} \tag{33}$$

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 (32)] while the variables $z_{4,t}$, $z_{5,t}$ and $z_{6,t}$ will be used for the second model [equation (33)].

We take the following values for the parameters so that the two models are correctly specified. In equation (29), we take $\beta_1 = \beta_2 = 0.5$. As for the parameters in $x_{1,i}$ and $x_{2,i}$ we take

$$\begin{aligned}
\gamma_1 = \gamma_2 = \gamma_3 &= 0.5 \\
\gamma_4 = \gamma_5 = \gamma_6 &= 0 \\
\alpha_1 = \alpha_2 = \alpha_3 &= 0 \\
\alpha_4 = \alpha_5 = \alpha_6 &= 0.5
\end{aligned}$$

We draw the error terms $u_{0,t}$, $u_{1,t}$ and $u_{2,t}$ independently from a $N(0,1)$. We also draw independently the instruments z_t from a $N(0,1)$. We can see that under these parameter values the two models (32) and (33) are correctly specified since the missing x_t variable in each equation is uncorrelated with the set of instruments z_t employed. We take a sample size of 1,000. The 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.

In the left panel of Figure 1, we draw an histogram for the statistic N_T using 20,000 replications. In the right panel, we draw an histogram for draws from a N(0,1) distribution using the same number of replications. We can clearly see that these two distributions are different. Although both are centered on zero, the distribution of N_T has a smaller variance. Using a Lilliefors test [see Lilliefors (1967)], we can also reject the hypothesis that the N_T's have a normal distribution and so the differences between the two distributions in Figure 1 cannot be attributed to differences in their variances.

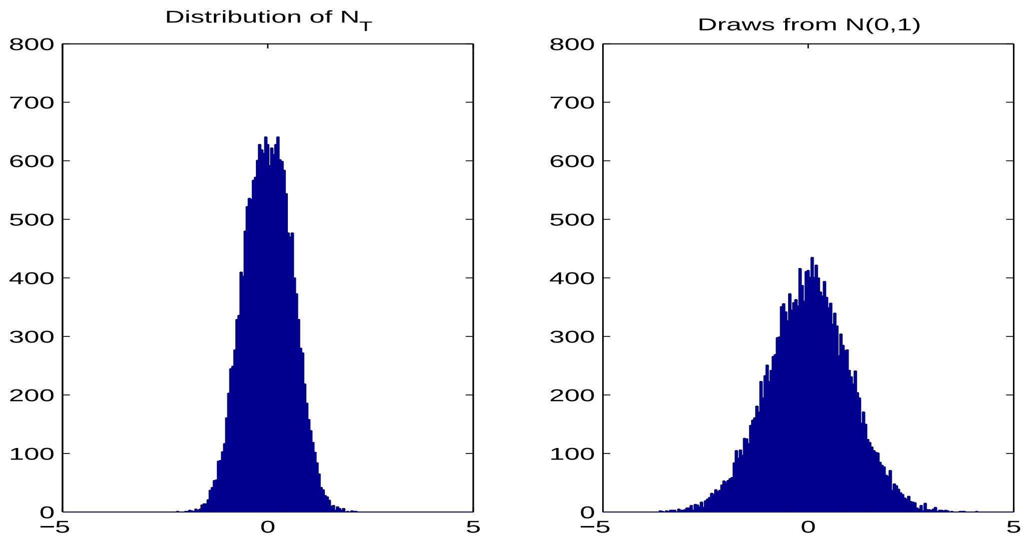

Figure 1: Correctly specified models

3.3 Locally misspecified models

Following the results in sub-sections 3.1 and 3.2, we might want to use this new limiting distribution with the test statistic N_T instead of the standard normal distribution. In this sub-section, we argue that sensible inference cannot be performed in this way. The problem is as follows. 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.[15] 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. The problem is that the limiting distribution of N_T depends on the drift parameter driving the local misspecification and this cannot be estimated

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

consistently. It follows that it is not possible to obtain the correct critical point for the Rivers and Vuong test unless both models happen to be correctly specified and the latter can never be known with certainty. This argument is substantiated below. In the remainder of this sub-section we present the limiting distribution of N_T when both models are locally misspecified and in the following sub-section provide simulation results to illustrate the dependence of the limiting distribution of N_T on the drift parameters.

Consider a scenario where the two models compared are locally misspecified models. 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 11 *$\mathcal{M}_i$ 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 $\{v_t, t = 1, \ldots, T\}$. Notice that Assumption 11 implies that the data cannot be a realization of a strictly stationary process because $E[f(v_t, \theta)]$ changes with T. This assumption is routinely invoked in the analysis of the local power of test statistics, where it is combined with the assumption that the data generation process evolves so that all sample averages converge to the same limits under the null hypothesis. For example, this framework is used to derive the local power properties of the aforementioned overidentification test; see Newey (1985) or Hall (2005, Section 5.1.3).

Given the framework in Assumption 11, 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 11 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

$$H_0 \;:\quad Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)})$$

Just as when both models are correctly specified, $\sigma_0^2 = 0$ and so the case of locally misspecified models is outside the generic conditions provided by Rivers and Vuong (2002). Therefore, we now investigate the limiting behaviour of N_T in this case. Our study of the limiting distribution theory is much like in Subsection 3.1. We assume that the weighting matrices are chosen so that $W^{(i)} = [S^{(i)}]^{-1}$ and $W_T^{(i)}$ depends on $\hat{\tau}_T^{(i)} = \tilde{\theta}_T^{(i)}$, a preliminary GMM estimator of $\theta_\star^{(i)}$ based on the moment conditions $f^{(i)}(v_t, \theta)$. We again assume that this preliminary GMM estimation is performed using a weighting matrix, $M_T^{(i)}$, that converges to a positive definite matrix of constants, $M^{(i)}$. It follows that the matrices $A_\star^{(i)}$, $\Delta^{(i)}$, $\Sigma^{(i)}$ and vector $Y_t^{(i)}$ have the same definitions [see equations (18)-(21)]. Given these definitions, the estimator $\hat{\sigma}_T^2$ remains the same as for correctly specified models [see equations (22)-(27)].

To present the limiting distribution of N_T, we impose the following regularity conditions.

Assumption 12 *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) $G_T^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{p} G_0^{(i)}$; (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

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

is a positive definite matrix of finite constants;

Comparing these assumptions with the ones in Section 3.1, the only effective difference is the mean of the limiting distribution of $T^{1/2}g_T(\theta_\star)$. The limiting distribution of N_T is given in the following theorem.

Theorem 2 *Let Assumption 12 hold. If $\mathcal{M}_1$ and $\mathcal{M}_2$ satisfy Assumption 11 then*

$$N_T \quad \xrightarrow{d} \quad \frac{(n_{q_1+q_2}+\bar{\eta})'C^{1/2'}\begin{bmatrix} I_{q_1}-P_0^{(1)} & 0 \\ 0 & -[I_{q_2}-P_0^{(2)}] \end{bmatrix} C^{1/2}(n_{q_1+q_2}+\bar{\eta})}{2\sqrt{(n_{q_1+q_2}+\bar{\eta})'C^{1/2'}\left\{I_{q_1+q_2}-P_0\right\}'\bar{C}\left\{I_{q_1+q_2}-P_0\right\}C^{1/2}(n_{q_1+q_2}+\bar{\eta})}} \tag{34}$$

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

$$\bar{\eta} \quad = \quad S^{-1/2}\begin{bmatrix} {S^{(1)}}^{1/2} & 0 \\ 0 & {S^{(2)}}^{1/2} \end{bmatrix}\begin{bmatrix} \eta^{(1)} \\ \eta^{(2)} \end{bmatrix}$$

and $C^{1/2}$, $\bar{C}$, P_0, $P_0^{(i)}$, $F_0^{(i)}$ are as defined in Theorem 1.

We omit the proof of this theorem as it is very similar to the proof of Theorem 1. It can be seen that under such local misspecification, the analysis of the limiting behavior of N_T is qualitatively similar to that presented in Theorem 1. The only difference in the limiting distribution is that $n_{q_1+q_2}$ needs to be replaced by $(n_{q_1+q_2}+\bar{\eta})$. As in the correctly specified case, this limiting distribution is not a standard normal.

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, (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.

3.4 Simulation results 2

In this section, we want to illustrate through Monte Carlo simulations the impact that local misspecification can have on the limit distribution of the N_T statistic. To this end, we consider the setup used in Section 3.2 except that we now assume that

$$\begin{bmatrix} S^{(1)}(\theta_\star^{(1)})^{-1/2} & 0 \\ 0 & S^{(2)}(\theta_\star^{(2)})^{-1/2} \end{bmatrix} E_T[u_{0,t} z_t] = T^{-1/2}\eta$$

where $z_t = [z_{1,t}, z_{2,t}, \ldots, z_{6,t}]'$. All the other aspects of the model, such as parameter values, are the same. As noted above, the null hypothesis under consideration by the Rivers and Vuong's (2002) statistic holds. In particular, $Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)}) = 0$.

This time, instead of simulating data, performing the estimation and computing the test statistic, we directly sample from the limit distribution given in equation (34). Two sets of simulations are reported in Figure 2. In the left panel, we present draws from N_T when $\eta_4 = 5$ with $\eta_i = 0$ for $i \neq 4$, that is we assume that only the second model is locally misspecified. In the right panel are draws for the case where $\eta_1 = 5$ with $\eta_i = 0$ for $i \neq 1$ so that only the first model is misspecified.

We see that the location of the test statistic can be greatly affected by local misspecification. As noted above, it is impossible to consistently estimate the drift parameters and so also impossible to simulate percentiles from the correct null hypothesis distribution for N_T when the models are locally misspecified. As an alternative, it might be natural to consider employing the test using the percentiles from the distribution in Theorem 1, that is the distribution of N_T when both models are correctly specified. However, a comparison of Figures 1 and 2 indicates that such a strategy cannot be argued to control the level of the test in any reasonable way. Hence, we conclude that no sensible inference can be performed using the N_T statistic in the case where both models are validated by the overidentifying restrictions test.

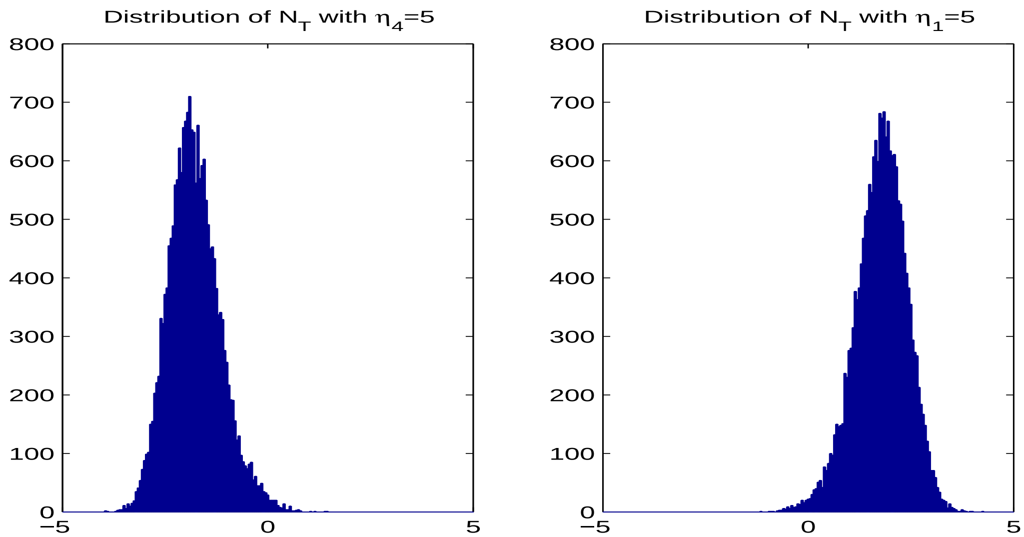

Figure 2: Locally misspecified models.

4 Non-local misspecification

The second approach to modelling misspecification in the literature is non-local (fixed) alternatives. We focus explicitly on the latter in our analysis in this section.

4.1 Misspecified models

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

Assumption 13 $\mathcal{M}_i$ *satisfies: (i)* $E[f(v_t, \theta)] = \mu(\theta)$ *where* $\mu(\theta) = [\mu^{(1)}(\theta^{(1)})', \mu^{(2)}(\theta^{(2)})']'$ *and* $\parallel \mu^{(i)}(\theta^{(i)}) \parallel \neq 0$ *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 as we now discuss.

First consider the case in which $W_T^{(i)} = I_{q_i}$. With this choice, both $A_\star^{(i)}$ and $\Delta^{(i)}$ are null matrices and so the form of σ_0^2 simplifies to:

$$\begin{aligned}\sigma_0^2 \;=\; & 4\Big\{\, \mu^{(1)}(\theta_\star^{(1)})' S^{(1)}(\theta_\star^{(1)})\mu^{(1)}(\theta_\star^{(1)}) \;+\; \mu^{(2)}(\theta_\star^{(2)})' S^{(2)}(\theta_\star^{(2)})\mu^{(2)}(\theta_\star^{(2)}) \\ & - 2\mu^{(1)}(\theta_\star^{(1)})' S^{(1,2)}(\theta_\star)\mu^{(1)}(\theta_\star^{(1)}) \,\Big\} \qquad (35)\end{aligned}$$

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

$$S(\theta_\star) \;=\; \begin{bmatrix} S^{(1)}(\theta_\star^{(1)}) & S^{(1,2)}(\theta_\star) \\ S^{(1,2)}(\theta_\star)' & S^{(2)}(\theta_\star^{(2)}) \end{bmatrix}, \qquad (36)$$

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

$$\begin{aligned}\hat{\sigma}_T^2 \;=\; & 4\Big\{\, g_T^{(1)}(\hat{\theta}_T^{(1)})' \hat{S}^{(1)}(\hat{\theta}_T^{(1)}) g_T^{(1)}(\hat{\theta}_T^{(1)}) \;+\; g_T^{(2)}(\hat{\theta}_T^{(2)})' \hat{S}^{(2)}(\hat{\theta}_T^{(2)}) g_T^{(2)}(\hat{\theta}_T^{(2)}) \\ & -2\, g_T^{(1)}(\hat{\theta}_T^{(1)})' \hat{S}^{(1,2)}(\hat{\theta}_T) g_T^{(2)}(\hat{\theta}_T^{(2)}) \,\Big\} \qquad (37)\end{aligned}$$

To analyze the behavior of the test in this case we impose:

Assumption 14 *(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 defined in (36) and 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)* $\hat{S}^{(1,2)}(\hat{\theta}_T) \xrightarrow{p} S^{(1,2)}(\theta_\star)$*.*

Hall and Inoue (2003)[Theorem 1] provide conditions under which $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_\star^{(i)})$ has a limiting normal distribution, and so Assumption 14(iii) could be replaced by these lower level assumptions.

Now consider the case in which $W_T^{(i)} = \{T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\}^{-1} = \{\hat{M}_{zz}^{(i)}\}^{-1}$ 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 (3) and (4)]. 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

$$\xi_t = \begin{bmatrix} f^{(1)}(v_t, \theta_\star^{(1)}) \\ vech\{(z_t^{(1)} z_t^{(1)\prime}) - M_{zz}^{(1)}\} \\ f^{(2)}(v_t, \theta_\star^{(2)}) \\ vech\{(z_t^{(2)} z_t^{(2)\prime}) - M_{zz}^{(2)}\} \end{bmatrix} \tag{38}$$

and $R_\star$ is defined by (13)-(14) with $A_\star^{(i)} = I_{q_i(q_i+1)/2}$ and

$$\begin{aligned} Y_t^{(i)} &= vech\{(z_t^{(i)} z_t^{(i)\prime}) - M_{zz}^{(i)}\} \\ \Delta^{(i)} &= L_i \left(\{M_{zz}^{(i)}\}^{-1} \otimes \{M_{zz}^{(i)}\}^{-1}\right) B_i \end{aligned}$$

We therefore set

$$\hat{\sigma}_T^2 = \hat{R}_T' \hat{V}_T \hat{R}_T \tag{39}$$

where $\hat{V}_T$ is a consistent estimator of $V_\star$ based on

$$\hat{\xi}_t = \begin{bmatrix} f^{(1)}(v_t, \hat{\theta}_T^{(1)}) - g_T^{(1)}(\hat{\theta}_T^{(1)}) \\ vech\{(z_t^{(1)} z_t^{(1)\prime}) - \hat{M}_{zz}^{(1)}\} \\ f^{(2)}(v_t, \hat{\theta}_T^{(2)}) - g_T^{(2)}(\hat{\theta}_T^{(2)}) \\ vech\{(z_t^{(1)} z_t^{(2)\prime}) - \hat{M}_{zz}^{(2)}\} \end{bmatrix} \tag{40}$$

and

$$\hat{R}_T = \begin{bmatrix} \hat{R}_T^{(1)} \\ -\hat{R}_T^{(2)} \end{bmatrix} \tag{41}$$

$$\hat{R}_T^{(i)} = \begin{bmatrix} 2\{\hat{M}_{zz}^{(i)}\}^{-1} g_T^{(i)}(\hat{\theta}_T^{(i)}) \\ -\hat{\Delta}_T^{(i)} B_i' \{g_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes g_T^{(i)}(\hat{\theta}_T^{(i)})\} \end{bmatrix} \tag{42}$$

$$\hat{\Delta}_T^{(i)} = L_i[\{\hat{M}_{zz}^{(i)}\}^{-1} \otimes \{\hat{M}_{zz}^{(i)}\}^{-1}] B_i \tag{43}$$

To analyze the behavior of the test in this case, we impose:

Assumption 15 *(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)[Theorem 2] provide conditions under which $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_\star^{(i)})$ has a limiting normal distribution, and so Assumption 15(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 3 *Let (i) Assumptions 1-8 hold; (ii)* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *satisfy Assumption 13; (iii)* H_0 *holds; and either: (a)* $W_T^{(i)} = I_{q_i}$*,* $\hat{\sigma}_T$ *is calculated via (37), and Assumption 14 holds; or (b):* $W_T^{(i)} = \{\hat{M}_{zz}\}^{-1}$*,* $\hat{\sigma}_T$ *is calculated via (39)-(43), and Assumption 15 holds; then* $N_T \xrightarrow{d} N(0,\,1)$

Theorem 3 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 13. 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. It might be anticipated that this problem can be avoided by using the two-step GMM estimator in which the weighting matrix equals the inverse of a consistent estimator of the long run variance of the sample moment[16]. However, this is not the case. As pointed out by Hall and Inoue (2003), the probability limit of the GMM estimator depends on the weighting matrix in misspecified models

[16]Notice that in misspecified models there is no justification for this choice of weighting matrix in terms of asymptotic efficiency. See Hall and Inoue (2003).

and so the estimate of the long run variance is dependent on the choice of first step weighting matrix even asymptotically. Therefore, even if the two-step estimator is used, the probability limit of the weighting matrix on the second step depends in general on the first step weighting matrix, and so there is the possibility that the ranking can be reversed by changing the first step weighting matrices.

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 as we demonstrate in the next sub-section.

4.2 Simulation results 3

In this sub-section we consider incorrectly specified models. We take the same setup as in Subsection 3.2 for correctly specified models but we change some parameter values so that the instruments are now correlated with the excluded explanatory variable. Our goal is to illustrate that the ranking of two incorrectly specified models can depend on the weighting matrix employed.

Since the models are incorrectly specified there is no theory for an optimal weighting matrix so we only perform a one-step estimation. We have two natural choices for the weighting matrix. The first is to take the identity matrix. The second is to use the inverse of the variance matrix of the instruments.

In this example, to obtain two incorrectly specified models we take $\gamma_4 = 0.5$ and $\alpha_i = 4$ for $i = 3, \ldots, 6$ (the value of the other γ's and α's does not change). With these values, the pairs $(x_{1,i}, z_{4,i})$ and $(x_{2,i}, z_{3,i})$ are correlated. So as to get an even more flagrant example, we also change the value of other parameters. We increase the variances of $(z_{4,t}, z_{5,t}, z_{6,t})$ to 10. With these parameter values, $\|\mu^{(1)}(\theta^{(1)})\| = \sqrt{0.5(0.5-\beta_1)^2 + (2.25-0.5\beta_1)^2} > 0$ and $\|\mu^{(2)}(\theta^{(2)})\| = \sqrt{(2.25-4\beta_2)^2 + 32(0.5-\beta_2)^2} > 0$.

For a sample size of $2,000$ and $20,000$ replications we get Figure 3 below. In the left panel, the weighting matrix for each model is the identity matrix. In the right panel, we use the inverse of the variance of the instruments. We see that by changing the weighting matrix, we can invert the ranking of the models. With the identity matrix most of the N_T's are negative (64% are below -1.28, i.e. the 10% one-sided critical value based on the standard normal distribution) while with the variance of the instruments it is the opposite (99% are above 1.28).

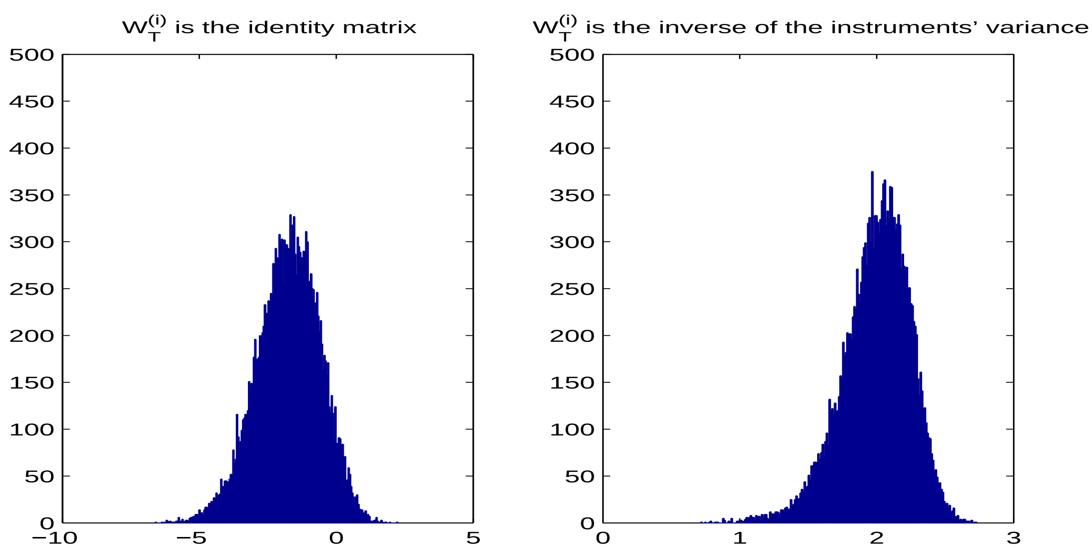

Figure 3: Misspecified models

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 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. Unfortunately, it does not seem that such a test is viable when N_T is based on GMM minimands. The problem is that we would need the limit distribution of $\hat{\sigma}_T$ under the null hypothesis that $\sigma_0^2 = 0$ and this distribution is sensitive to whether the two models are correctly specified or locally misspecified. Recall from Section 3 that if the models are correctly specified, then

$$\sqrt{T}\hat{\sigma}_T \xrightarrow{d} 2\sqrt{n'_{q_1+q_2} C^{1/2'} \left\{ I_{q_1+q_2} - P_0 \right\}' \bar{C} \left\{ I_{q_1+q_2} - P_0 \right\} C^{1/2} n_{q_1+q_2}}$$

while if the models are locally misspecified, then

$$\sqrt{T}\hat{\sigma}_T \xrightarrow{d} 2\sqrt{(n_{q_1+q_2} - \bar{\eta})' C^{1/2'} \left\{ I_{q_1+q_2} - P_0 \right\}' \bar{C} \left\{ I_{q_1+q_2} - P_0 \right\} C^{1/2} (n_{q_1+q_2} - \bar{\eta})}.$$

Therefore, just as in the case of N_T, the drift of the local misspecification is shitfting the distribution of $\sqrt{T}\hat{\sigma}_T$ away from zero. Hence, $\sqrt{T}\hat{\sigma}_T$ cannot be used to perform any sensible inference about the the hypothesis $\sigma_0^2 = 0$ for the same reasons as those discussed in Section 3.4.

Rather than focus on $\sigma_0^2 = 0$ *per se*, another possibility is to focus on the source of the zero variance in correctly specified or locally misspecified models. It can be seen from the analysis in Section 3 that the zero variance arises because the limiting average of the moment conditions is zero. The obvious way to test this condition is to use the overidentifying restrictions test. However, once again problems emerge if we attempt to use this statistic to discriminate between non-locally misspecified models ($\sigma_0^2 > 0$) and correctly specified or locally misspecified models ($\sigma_0^2 = 0$). For, supposing a nominal 5% level test is used, the probability of a significant statistic is 5% in correctly specified models, one in non-locally misspecified models but depends on the drift in locally misspecified models. Therefore, we see no way in which the overidentifying restrictions statistic can be usefully employed to test $\sigma_0^2 = 0$ versus $\sigma_0^2 > 0$.

6 Concluding remarks

Competing economic theories often lead to econometric models that are non-nested. It is, therefore, of interest to develop statistical procedures that discriminate between non-nested models. Vuong (1989) provides such procedure for models estimated by Quasi Maximum Likelihood. This test exploits the interpretation of QML estimates minimizing the Kullback Leibler metric to test which model is closer to the truth. In a recent paper, Rivers and Vuong (2002) extend Vuong (1989) in several dimensions. One of them is to allow the test to be applied to incompletely specified models such as models defined by moment conditions. Rivers and Vuong (2002) provide generic conditions under which their statistic has a standard normal limiting distribution; one such condition involves a particular variance, denoted σ_0^2, that must be non-zero for the stated result.

In this paper we analyze the limit distribution of the Rivers and Vuong's (2002) statistic in models estimated via GMM and the model comparison is based on a GMM minimand. We show that: (i) if both models are correctly specified or locally misspecified then $\sigma_0^2 = 0$ and the statistic has a non-standard limit distribution; (ii) if both models are misspecified then $\sigma_0^2 > 0$ and the statistic does converge to a limiting standard normal distribution.

At first glance, this summary might suggest that the test can be employed provided the appropriate asymptotic distribution is used but there are further complications as we have discussed in the paper and summarize now.

(a) *Correctly specified models and local misspecification:* We might want to use the statistic N_T with the asymptotic distribution derived in Theorems 1 or 2 to discriminate between two models judged correctly specified, according to the overidentification test. Unfortunately, since the overidentification test cannot discriminate between correctly specified and locally misspecified models, and we cannot estimate the drift in locally misspecified models, it would appear impossible to use these limit distributions to perform inference in practice.

(b) *Misspecified models:* Even though the limiting distribution theory is standard, some caution needs to be exercised in the use of the statistic N_T because the outcome depends on the choice of the weighting matrix. This dependence raises the possibility that the "ranking" of the model is also 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 is of interest. However, absent these economic considerations, the choice of the weighting matrix become arbitrary for in misspecified models – unlike in correctly specified models – there is no statistical theory to guide the choice of the weighting matrix. It is in this case that the dependence becomes troublesome. Monte Carlo simulations illustrate this point.

Taking these results together, it would seem that, out of the scenarios considered, the only scope for using N_T is in non-locally misspecified models in which economic theory dictates a choice of weighting matrix. However, even then, there is a further concern. Within this scenario, it is only appropriate to compare N_T to the appropriate percentile of the standard normal distribution if $\sigma_0^2 > 0$ or, in other words, if both models really are non-locally misspecified. Ideally, a researcher would perform a pre-test for this restriction but while this variance pre-test works in some settings it does not work here because there is no way to discriminate usefully between non-locally misspecified models ($\sigma_0^2 > 0$) and the combined class of correctly specified or locally misspecified models ($\sigma_0^2 = 0$).

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 3 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 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.

Therefore, our results suggest there is limited scope for using N_T based on GMM minimands to discriminate between two models estimated via GMM. In our opinion, these limitations stem from the use of GMM estimation. Unlike QML, GMM estimation can not be interpreted as minimizing the distance to the true model (or pdf thereof). However, while GMM does not have this interpretation, empirical likelihood (EL) estimation does. Kitamura (2002) proposes an extension of Vuong's (1989) methods to EL estimation of conditional moment restrictions models. Therefore, it would seem that the EL may provide a more suitable framework for the comparsion of non-nested models specified via moment restrictions.

Appendix

Proof of Theorem 1:

The test statistic N_T can be written as

$$N_T = \frac{T\left\{Q_T^{(1)}\left(\hat{\theta}_T^{(1)}\right) - Q_T^{(2)}\left(\hat{\theta}_T^{(2)}\right)\right\}}{T^{1/2}\hat{\sigma}_T}. \tag{44}$$

Analysis of the numerator in (44) is straightforward since it is simply the difference between two overidentification test statistics. Standard analysis of the overidentifying restrictions test yields[17]

$$TQ_T^{(i)}\left(\hat{\theta}_T^{(i)}\right) = \{T^{1/2}g_T^{(i)}(\theta_\star^{(i)})\}'[S^{(i)}]^{-1/2\prime}\left[I - P_0^{(i)}(\theta_\star^{(i)})\right][S^{(i)}]^{-1/2}\{T^{1/2}g_T^{(i)}(\theta_\star^{(i)})\} + o_p(1) \tag{45}$$

It follows from (45) and Assumption 9 that, dropping the dependence on $\theta_\star^{(i)}$ in places where it is obvious so as to lighten the notation,

$$T\left\{Q_T^{(1)}\left(\hat{\theta}_T^{(1)}\right) - Q_T^{(2)}\left(\hat{\theta}_T^{(2)}\right)\right\} \xrightarrow{d} n'_{q_1+q_2}C^{1/2\prime}\begin{bmatrix} I_{q_1} - P_0^{(1)} & 0 \\ 0 & -[I_{q_2} - P_0^{(2)}] \end{bmatrix} C^{1/2}n_{q_1+q_2} \tag{46}$$

Now consider the denominator of (44). 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 $T^{1/2}\hat{R}_T$. Under our assumptions, it follows from (24)-(27) that

$$\begin{aligned} T^{1/2}\hat{R}_T &= \begin{bmatrix} 2W_T^{(1)}T^{1/2}g_T^{(1)}(\hat{\theta}_T^{(1)}) \\ o_p(1) \\ -2W_T^{(2)}T^{1/2}g_T^{(2)}(\hat{\theta}_T^{(2)}) \\ o_p(1) \end{bmatrix} \\ &= 2\Gamma_V T^{1/2}g_T(\hat{\theta}_T) + o_p(1) \end{aligned} \tag{47}$$

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

where

$$\Gamma_V = \begin{bmatrix} W_T^{(1)} & 0 \\ 0 & 0 \\ 0 & -W_T^{(2)} \\ 0 & 0 \end{bmatrix},$$

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 (47) that under our assumptions:

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

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

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

Note that

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

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

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

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

$$T\hat{\sigma}^2 \xrightarrow{d} 4n'_{q_1+q_2} C^{1/2'} \{I_{q_1+q_2} - P_0\}' \bar{C} \{I_{q_1+q_2} - P_0\} C^{1/2} n_{q_1+q_2}$$

It follows that under the null hypothesis the test statistic N_T converges to

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

Proof of Theorem 3

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

$$Q_T^{(i)}(\hat{\theta}_T^{(i)}) = Q_T^{(i)}(\theta_\star^{(i)}) + \left\{ \left. \frac{\partial Q_T^{(i)}(\theta^{(i)})}{\partial \theta^{(i)}} \right|_{\theta^{(i)} = \bar{\theta}_T^{(i)}} \right\}' \left(\hat{\theta}_T^{(i)} - \theta_\star^{(i)} \right) \tag{48}$$

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

$$\Phi^{(i)}(\theta_\star^{(i)}) = 2G_0^{(i)}(\theta_\star^{(i)})' W^{(i)} E[f^{(i)}(v_t, \theta_\star^{(i)})] \tag{49}$$

Adding and subtracting $\Phi^{(i)}(\theta_\star^{(i)})(\hat{\theta}_T^{(i)} - \theta_\star^{(i)})$ in (48), we obtain

$$\begin{aligned} Q_T^{(i)}(\hat{\theta}_T^{(i)}) &= Q_T^{(i)}(\theta_\star^{(i)}) + \left\{ \Phi^{(i)}(\theta_\star^{(i)}) \right\}' \left(\hat{\theta}_T^{(i)} - \theta_\star^{(i)} \right) \\ &\quad + \left\{ \frac{\partial Q_T^{(i)}(\bar{\theta}_T^{(i)})}{\partial \theta^{(i)}} - \Phi^{(i)}(\theta_\star^{(i)}) \right\}' \left(\hat{\theta}_T^{(i)} - \theta_\star^{(i)} \right) \end{aligned} \tag{50}$$

From Assumptions 1-8, it follows that

$$\left\{ \frac{\partial Q_T^{(i)}(\bar{\theta}_T^{(i)})}{\partial \theta^{(i)}} \;-\; \Phi^{(i)}(\theta_\star^{(i)}) \right\} \;=\; o_p(1) \tag{51}$$

Equation (51) combined with $T^{1/2}\left(\hat{\theta}_T^{(i)} \;-\; \theta_\star^{(i)}\right) = O_p(1)$ yields

$$Q_T^{(i)}(\hat{\theta}_T^{(i)}) \;=\; Q_T^{(i)}(\theta_\star^{(i)}) \;+\; \left\{ \frac{\partial Q_0^{(i)}(\theta_\star^{(i)})}{\partial \theta^{(i)}} \right\}' \left(\hat{\theta}_T^{(i)} \;-\; \theta_\star^{(i)}\right) \;+\; o_p(T^{-1/2}) \tag{52}$$

Using (52), it follows that

$$\begin{aligned} T^{1/2}\left[\, Q_T^{(1)}(\hat{\theta}_T^{(1)}) \;-\; Q_T^{(2)}(\hat{\theta}_T^{(2)}) \right] \;&=\; T^{1/2}\left[\, Q_T^{(1)}(\theta_\star^{(1)}) \;-\; Q_T^{(2)}(\theta_\star^{(2)}) \right] \\ &\quad + \left\{ \Phi^{(i)}(\theta_\star^{(1)}) \right\}' \left(\hat{\theta}_T^{(1)} \;-\; \theta_\star^{(1)}\right) \\ &\quad - \left\{ \Phi^{(i)}(\theta_\star^{(2)}) \right\}' \left(\hat{\theta}_T^{(2)} \;-\; \theta_\star^{(2)}\right) \;+\; o_p(1) \end{aligned} \tag{53}$$

Finally, under H_0, we have $Q_0^{(1)}(\theta_\star^{(1)}) = Q_0^{(2)}(\theta_\star^{(2)})$ and so (53) can be written as

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\right] \;&=\; T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) - Q_0^{(1)}(\theta_\star^{(1)})\right] - T^{1/2}\left[Q_T^{(2)}(\theta_\star^{(2)}) - Q_0^{(2)}(\theta_\star^{(2)})\right] \\ &\quad + \left\{ \Phi^{(i)}(\theta_\star^{(1)}) \right\}' \left(\hat{\theta}_T^{(1)} \;-\; \theta_\star^{(1)}\right) \\ &\quad - \left\{ \Phi^{(i)}(\theta_\star^{(2)}) \right\}' \left(\hat{\theta}_T^{(2)} \;-\; \theta_\star^{(2)}\right) \;+\; o_p(1) \end{aligned} \tag{54}$$

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:

$$G_T^{(i)}(\hat{\theta}_T^{(i)})' W_T T^{-1} \sum_{t=1}^{T} f^{(i)}(v_t, \hat{\theta}_T^{(i)}) \;=\; 0 \tag{55}$$

Furthermore, the probability limits must satisfy the analogous population moment condition, that is: $\Phi^{(i)}(\theta_\star^{(i)}) = 0$. Therefore, we have

$$T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\right] \;=\; T^{1/2}\left[Q_T^{(1)}(\theta_\star^{(1)}) - Q_0^{(1)}(\theta_\star^{(1)})\right] - T^{1/2}\left[Q_T^{(2)}(\theta_\star^{(2)}) - Q_0^{(2)}(\theta_\star^{(2)})\right] + o_p(1) \tag{56}$$

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

$$\hat{h}'\hat{W}\hat{h} \,-\, h'Wh \;=\; \hat{h}'\hat{W}(\hat{h}-h) \,+\, \hat{h}'(\hat{W}-W)h \,+\, (\hat{h}-h)'Wh \tag{57}$$

Using (56) and (57), 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 (56) and (57) that

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

The result then follows immediately under the stated assumptions.

Part (b):

With $W_T^{(i)} = \{\hat{M}_{zz}^{(i)}\}^{-1}$, it follows from (56) and (57) that

$$\begin{aligned} T^{1/2}\left[Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\right] \;\; = \;\; & 2\,\mu^{(1)}(\theta_\star^{(1)})'\{M_{zz}^{(1)}\}^{-1}T^{-1/2}\sum_{t=1}^{T}[f^{(1)}(v_t,\theta_\star^{(1)}) - \mu^{(1)}(\theta_\star^{(1)})] \\ & +\mu^{(1)}(\theta_\star^{(1)})'T^{1/2}(\{\hat{M}_{zz}^{(1)}\}^{-1} - \{M_{zz}^{(1)}\}^{-1})\mu^{(1)}(\theta_\star^{(1)}) \\ & -2\,\mu^{(2)}(\theta_\star^{(2)})'\{M_{zz}^{(2)}\}^{-1}T^{-1/2}\sum_{t=1}^{T}[f^{(2)}(v_t,\theta_\star^{(2)}) - \mu^{(2)}(\theta_\star^{(2)})] \\ & -\mu^{(2)}(\theta_\star^{(2)})'T^{1/2}(\{\hat{M}_{zz}^{(2)}\}^{-1} - \{M_{zz}^{(2)}\}^{-1})\mu^{(2)}(\theta_\star^{(2)}) \\ & +o_p(1) \end{aligned} \tag{58}$$

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

$$T^{1/2}(\{\hat{M}_{zz}^{(1)}\}^{-1} - \{M_{zz}^{(1)}\}^{-1}) \;=\; -\{M_{zz}^{(1)}\}^{-1}T^{1/2}\sum_{t=1}^{T}[z_t^{(i)}z_t^{(i)\prime} - M_{zz}^{(i)}]\{\hat{M}_{zz}^{(1)}\}^{-1} \tag{59}$$

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

$$\begin{aligned} \mu^{(i)}(\theta_\star^{(i)})'T^{1/2}\left(\{\hat{M}_{zz}^{(i)}\}^{-1} - \{M_{zz}^{(i)}\}^{-1}\right)\mu^{(i)}(\theta_\star^{(i)}) \;\; = \;\; & \left[\mu^{(i)}(\theta_\star^{(i)})' \otimes \mu^{(i)}(\theta_\star^{(i)})'\right]\left[\{M_{zz}^{(i)}\}^{-1} \otimes \{M_{zz}^{(i)}\}^{-1}\right]B_i \,\times \\ & T^{-1/2}\sum_{t=1}^{T} vech[z_t^{(i)}z_t^{(i)\prime} - M_{zz}^{(i)}] + o_p(1) \end{aligned} \tag{60}$$

The result then follows from (58)-(60) and the stated assumptions.

References

Atkinson, A. C. (1970). 'A method for for discriminating between models', *Journal of the Royal Statistical Society, Series B*, 32: 211–243.

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

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.

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. (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 McFadden, D. L. (1994). 'Large sample estimation and hypothesis testing', in R. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2113–2247. Elsevier Science Publishers, Amsterdam, The Netherlands.

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.

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.

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.

Wooldridge, J. M. (1994). 'Estimation and inference for dependent processes', in R. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2641–2739. Elsevier Science Publishers, Amsterdam, The Netherlands.