The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models *

Alastair R. Hall†

North Carolina State University

and

Atsushi Inoue‡

North Carolina State University

Abstract This paper presents the limiting distribution theory for the GMM estimator when the estimation is based on a population moment condition which is subject to non–local (or fixed) misspecification. It is shown that if the parameter vector is overidentified then the weighting matrix plays a far more fundamental role than it does in the corresponding analysis for correctly specified models. Specifically, the rate of convergence of the estimator depends on the rate of convergence of the weighting matrix to its probability limit. The analysis is presented for four particular choices of weighting matrix which are commonly used in practice. In each case the limiting distribution theory is different, and also different from the limiting distribution in a correctly specified model. Statistics are proposed which allow the researcher to test hypotheses about the parameters in misspecified models.

JEL classification: C10, C32

Keywords: Misspecification, Generalized Method of Moments, Asymptotic Distribution Theory.

---

*The authors gratefully acknowledge useful conversations with Eric Ghysels, Jinyong Hahn and Lars Hansen about aspects of this work. The authors also thank participants at the Midwest Econometrics Group Meeting, Ames, IA, October, 1999, the Allied Social Science Meetings, Boston, MA, January, 2000, and the Joint Statistical Meetings, Indianapolis, IN, August 2000 and at seminar presentations at Kyoto University, Ohio State University and Waseda University. The paper has also benefitted from the comments of David Doorn and three anonymous referees.

†Corresponding author. Department of Economics, Box 8110, North Carolina State University, Raleigh, NC 27695-8110, USA. Email: alastair_hall@ncsu.edu

1 Introduction

Generalized Method of Moments (GMM) (Hansen, 1982) provides a computationally convenient approach to the estimation of nonlinear dynamic econometric models based on the type of information provided by economic theory. In consequence, the GMM estimator has been used to perform inference about the parameters of economic models in a wide variety of settings. These inference procedures build from the asymptotic theory provided by Hansen (1982) in his original paper on GMM. Specifically, Hansen establishes that if the model is correctly specified then the estimator is $\sqrt{T}$–consistent and asymptotically normally distributed. However, for some data series, the correct model specification has proved elusive and so many empirical models are misspecified. It is therefore important to develop an asymptotic theory for the GMM estimator in misspecified models.

There are two basic approaches to such an analysis. First, and by far the most common in econometrics, is to employ a local alternative analysis. Within this approach, it is assumed that the data do not satisfy the population moment condition upon which estimation is based for any finite sample size, but do so in the limit as the sample size goes to infinity. Furthermore, the data is assumed to drift towards the model used in estimation at a rate which ensures the misspecification only manifests itself in the mean of the asymptotic distribution of the estimator. All other aspects of the limiting behaviour stay the same. Therefore, the probability limit of the estimator, the asymptotic variance of the estimator and the probability limit of the covariance matrix estimator are *identical* if the model is either correctly specified or subject to local misspecification. Such an analysis is often used to provide insights into the local power properties of test statistics, and was first applied in the context of GMM by Newey (1985). However, by its very nature, this type of analysis only provides guidance on the behaviour of the estimator when the truth is just a small perturbation away from the assumed model. This weakness can be addressed by using the second approach to misspecification analysis which is based on a non–local (or fixed) alternative. Within this scenario, the nature of the misspecification remains constant throughout the sample. This approach is much rarer and to our knowledge has only been applied to GMM in the special case of the instrumental variables (IV) estimator in linear models for a particular choice of weighting matrix. In the latter context, Maasoumi and Phillips (1982) show that the combination of overidentification and misspecification causes the limiting distribution of the IV estimator to depend on the limiting distribution of the elements of the weighting matrix.[1] However to date, Maasoumi and Phillips's (1982) analysis has not been extended to the GMM estimator.[2] The lack of such a theory represents a considerable gap in our understanding of the estimator because there is no reason to suppose in practice that all misspecification is local in nature.

In this paper, we present a limiting distribution theory for the GMM estimator in non–locally misspecified models. It is shown that the combination of parameter overidentification and misspecification has two important consequences for the limiting behaviour of the GMM estimator. First, the probability limit of the GMM estimator depends on the limit of the weighting matrix. Second, the limiting distribution of the GMM estimator to depend on the limiting distribution of the elements of the weighting matrix. While both findings are to be anticipated from Maasoumi and Phillips's (1982) analysis, our results indicate that the weighting matrix plays a far more fundamental role in nonlinear dynamic models

[1]Maasoumi and Phillips (1982) only consider the case in which the weighting matrix is the inverse of the instrument cross product matrix.

[2]Gallant and White (1988) develop an asymptotic theory for a class of estimators in potentially misspecified models under very weak assumptions on the dependence structure of the data. However, this class only extends to GMM estimators of overidentified parameters in misspecified models if the weighting matrix is *fixed*; see Gallant and White (1988)[pp.11–12]. White (1994) develops a similar analysis for quasi maximum likelihood estimator (QMLE). While the QMLE can be viewed as a special case of GMM, this interpretation involves the restriction that the parameter vector is just identified. Hall (2000) uses this framework to analyze the large sample behaviour of the overidentifying restrictions test when the long run variance is estimated by a member of the class of heteroscedasticity autocorrelation consistent covariance matrix estimators.

than is revealed by the aforementioned earlier study. Specifically, it is shown that the limiting distribution of the estimator depends not only on the probability limit of the weighting matrix – as it does in correctly specified or locally misspecified models – but also on the rate of convergence of the weighting matrix to this limit. This means that there is no one single limiting distribution theory for the GMM estimator in misspecified models. Instead, the analysis must be divided into separate cases depending on large sample behaviour of the weighting matrix, W_T. In this paper, we explicitly consider four cases: (i) $W_T = W$ for all T; (ii) $T^{1/2}(W_T - W)$ converges to a normal distribution for some matrix W;[3] (iii) W_T is the inverse of a centered heteroscedasticity autocorrelation consistent covariance (HACC) matrix estimator; (iv) W_T is the inverse of an uncentred HACC estimator. To summarize the limiting behaviour that occurs, we use $\hat{\theta}_T$ denote the GMM estimator and θ_*, its probability limit. It is shown that in cases (i) and (ii), $T^{1/2}(\hat{\theta}_T - \theta_*)$ converges to a normal distribution. Furthermore, in each case the asymptotic variance is different and in neither does it equal the variance derived by Hansen (1982) for the correctly specified case. In case (iii), the limiting behaviour of the estimator depends on the rate of increase of the bandwidth, b_T. If b_T does not increase too quickly (in a sense defined below) $(T/b_T)^{1/2}(\hat{\theta}_T - \theta_*)$ converges to a normal distribution. Otherwise, $b_T^k(\hat{\theta}_T - \theta_*)$ converges to a constant where k is a constant defined below. It is shown that in case (iv) $b_T(\hat{\theta}_T - \theta_*)$ converges to a constant in most cases of practical relevance.

In practice, inference is most often based on the two–step or iterated estimator. In correctly specified models, these two estimators are asymptotically equivalent. However, our results indicate this is not the case in misspecified models. It is shown that in situations covered by Case (ii) above then the asymptotic distribution of the estimator on the i^{th} step depends on the asymptotic distributions of the estimators on all previous steps. Whereas in situations covered by Case (iii) above then this dependence only goes back as far as the second step.

[3]This case contains Maasoumi and Phillips's (1982) analysis as a special case.

As would be imagined, these results have important implications for inference about the parameter vector. Newey and West (1987a) propose Wald, likelihood ratio type and Lagrange multiplier statistics for testing the hypothesis that the parameter vector satisfies a set of nonlinear restrictions. Under the joint null hypotheses that the restrictions are satisfied and the model is correctly specified, Newey and West (1987a) show that these statistics converge to a χ^2 distribution. However, we show in Section 4 that this limiting result no longer applies if the model is misspecified even if the restrictions are true. As a result, it is possible to reject incorrectly a set of parameter restrictions using conventional statistics because of model misspecification. Therefore, we also propose two statistics which can be used to perform inference about the pseudo–parameters in misspecified models. Both have limiting χ^2 distributions under the null, and the only difference between them stems from the choice of weighting matrix.

Before we present the analysis, we wish to address a potential issue regarding the interpretation and empirical relevance of our results. A referee has argued that there is little interest in the limiting behaviour of the GMM estimator in misspecified models because the overidentifying restrictions test can be used to indicate misspecification, and once misspecification is detected then the model is rejected. Clearly, such an empirical strategy renders redundant the issue of inference in misspecified models. While this empirical strategy may be employed frequently by researchers, an inspection of the literature reveals that it is by no means universally adopted. It is possible to find a number of published studies in which inference is performed about the parameters of a misspecified model that has been estimated by GMM. Sometimes these inferences are implicit in the sense that the GMM estimates and their standard deviations (or their t statistics) are reported even though the model is rejected using the overidentifying restrictions test; examples of this practice include Cochrane (1996) and Epstein and Zin (1991). Sometimes these inferences are explicit as the following three examples illustrate. Meghir and Weber (1996)[p.1173] perform hypothesis testing on structural parameters even though Sargan's test for overidentifying restrictions reject the null with the p-value less than 1% for the transport/service marginal rate of substitution equation. Ferson and Constantinides (1991)[p.221; Table 7] discuss the sign and statistical significance of the durability parameter even when the J test reject their two-asset system. Durlauf and Maccini (1995)[p.78] discuss the significance of certain parameters in inventory models even though the overidentifying restrictions tests are significant. These citations provide clear evidence that applied researchers are interested in performing inference in misspecified models. However, to date, no statistical theory is available to guide researchers in the interpretation of the types of results described above or in the construction of suitable test statistics for those researchers who wish to perform inferences about the pseudo–parameters in misspecified models that have been estimated by GMM.

Inspection of the literature also suggests that there is a growing interest in the development of inference techniques in misspecified models. White (1982, 1994) analyses the behaviour of quasi-maximum likelihood estimators. As mentioned above, Maasoumi and Phillips (1982) study the IV estimator in linear misspecified models and Gallant and White (1988) develop a statistical theory for a class of estimators in misspecified models. In addition, Hansen and Sargent (1993), Watson (1993) and Diebold, Ohanian, and Berkowitz (1998) provide frequency-domain frameworks to evaluating misspecified dynamic economic models. Hansen and Jagannathan (1991), Hansen, Heaton, and Luttmer (1995) and Hansen and Jagannathan (1997) develop alternative measures of misspecification for financial models. Kitamura (1997) develops a framework to compare misspecified models using empirical likelihood. Our results contribute to this theoretical literature as well.

An outline of the paper is as follows. Section 2 describes the framework used to capture non–local misspecification. Section 3 presents the limiting distribution theory for GMM estimators in misspecified models. Section 4 presents the limiting behaviour of the conventional Wald, LR-type and Lagrange multiplier statistics in misspecified models, and proposes statistics for testing hypotheses about the pseudo–parameters which have limiting

χ^2 distributions under the null. Section 5 contains some concluding remarks. All proofs are relegated to a mathematical appendix.

2 Non–local misspecification within the GMM framework

To motivate the definition of misspecification and the discussion of its consequences, it is useful to define first a correctly specified model and also to summarize briefly properties of certain important statistics in this case.[4] Throughout this paper we consider the Generalized Method of Moments estimator

$$\hat{\theta}_T = \text{argmin}_{\theta\in\Theta}\, g_T(\theta)'W_T g_T(\theta) \tag{1}$$

where $g_T(\theta) = T^{-1}\sum_{t=1}^{T} f(v_t,\theta)$, $f(v_t,\theta)$ is a $q\times 1$ vector indexed by the $p\times 1$ vector θ and a vector of observed random variables v_t, and W_T is a weighting matrix. To analyze the large sample properties of this estimator, it is necessary to impose certain regularity conditions. For ease of exposition, we only highlight in the text those assumptions which are crucial to the discussion and relegate the remainder to a mathematical appendix. Following Hansen (1982), we impose the following conditions on v_t and W_T.

Assumption 1 *$\{v_t \in \mathcal{V}, t = 1, 2, \ldots\}$ is a sequence of strictly stationary and ergodic random vectors where $\mathcal{V} \subseteq \Re^s$.*

Assumption 2 *$\{W_T;\ T = 1, 2\ldots\}$ is a sequence of positive semi–definite matrices and $p\lim_{T\to\infty} W_T = W$, a positive definite matrix of constants.*

Within this framework, a correctly specified model is defined as follows.

Definition 1 Correctly specified model
The model is said to be correctly specified if there exists a unique value θ_0 in $\Theta \subset \Re^p$ such that $E[f(v_t,\theta_0)] = 0$.

[4]This framework is also employed by Hall (2000).

Notice there are two parts to Definition 1: an "orthogonality" condition, that is $E[f(v_t, \theta_0)] = 0$, and an identification condition, that is this condition only holds at θ_0. In this case, it can be shown that subject to certain regularity conditions $\hat{\theta}_T \xrightarrow{p} \theta_0$, and $T^{1/2}(\hat{\theta}_T - \theta_0)$ converges in distribution to a normal mean zero random vector; see Hansen (1982). If $q > p$ then the covariance matrix this asymptotic distribution depends on $\text{plim}_{T\to\infty} W_T$. Hansen (1982) proves that the optimal choice of W_T equals the inverse of a consistent estimator of the long run variance of the sample moment,

$$S = \lim_{T\to\infty} Var[T^{-1/2} \sum_{t=1}^{T} f(v_t, \theta_0)] \tag{2}$$

In practice this "optimal" estimator is constructed via at least a two–step procedure. On the first step, GMM estimation is performed with a sub–optimal choice of W_T to obtain a preliminary estimator of θ_0. This preliminary estimator is used to construct a consistent estimator of S, $\hat{S}_T^{-1}$ say, and then $\hat{S}_T^{-1}$ is used as the weighting matrix in the second, "optimal" GMM estimation. This process can also be iterated.

A misspecified model is defined as follows.

Definition 2 Misspecified model

A model is said to be misspecified if there is no value of θ which satisfies the orthogonality condition, that is $E[f(v_t, \theta)] = \mu(\theta)$ where $\mu : \Theta \to \Re^q$ such that $||\mu(\theta)|| > 0$ for all $\theta \in \Theta$.

Two features of this definition should be noted. First, in line with Assumption 1, $E[f(v_t, \theta)]$ is assumed constant for all t, and so our framework rules out misspecification due to structural instability. The constancy of this expectation reflects the stationarity assumption in Assumption 1, and it should be noted that additional complications may arise if this assumption is relaxed; see Gallant and White (1988). Second, the parameter vector must be overidentified, i.e. $q > p$, because if $q = p$ then there must exist some value of θ such that $E[f(v_t, \theta)] = 0$ as we now demonstrate.

Proposition 1 *Suppose that Assumptions A.1-A.6 hold and that the method of moment estimator is well defined, i.e., $p = q$ and there is a sequence $\{\hat{\theta}_T\}_{T=1}^{\infty}$ such that $\frac{1}{T}\sum_{t=1}^{T} f(v_t, \hat{\theta}_T) =$*

0 *a.s. Then there exists* $\theta_* \in \Theta$ *such that* $E[f(v_t, \theta_*)] = 0$.

Assumption A.1-A.6 are given in the mathematical appendix. While this result is trivial in one sense, it shows that misspecification in the sense of Definition 2 occurs in the method of moments framework only when the model is overidentified. Thus, we focus on the GMM estimator for overidentified moments conditions from this point on, and models are misspecified in the sense of Definition 2. This means for instance that QML and linear projections are not misspecified in our sense because they are examples of GMM estimators with $p = q$. However, the two stage least squares (2SLS) estimator does fit in our framework if the model is misspecified and there are more instruments than regressors.

By itself, Definition 2 does not imply that $\hat{\theta}_T$ has a well defined probability limit, and so we also impose the following identification condition.

Assumption 3 Identification condition for a misspecified model
There exists $\theta_*(W) \in \Theta$ *such that* $Q_0(\theta_*(W)) < Q_0(\theta)$, $\forall \theta \in \Theta \setminus \{\theta_*(W)\}$, *where* $Q_0(\theta) = E[f(v_t, \theta)]' W E[f(v_t, \theta)]$.

It should be noted that, unlike correctly specified models, there is no reason to suppose that different choices of weighting matrix lead to minimands $Q_0(\theta)$ which are minimized by the same value of θ.[5] It is for this reason that we have indexed θ_* by W. However, for most of the discussion, we can suppress the dependence of θ_* on W for notational brevity because the meaning is clear from the context. Subject to certain other regularity conditions, Hall (2000) establishes that $\hat{\theta}_T \xrightarrow{p} \theta_*$. However, he does not establish any rates of convergence for $\hat{\theta}_T$ nor any limiting distribution theory for the GMM estimator.

[5] This observation was made in the context of linear models by Maasoumi and Phillips (1982).

3 The limiting behaviour of the GMM estimator in misspecified models

Hansen (1982) establishes that if the model is correctly specified then $T^{1/2}(\hat{\theta}_T - \theta_0)$ converges in distribution to a normal random vector. For our purposes, it is useful to emphasize that this asymptotic normality holds under very weak conditions on the weighting matrix. Specifically, W_T is assumed only to be a positive semi–definite for finite T which converges in probability to a positive definite matrix of constants W. Most importantly, no assumption is needed about the rate at which W_T converges to W. In this section, it is shown that a very different picture emerges in misspecified models. The analysis is divided into two parts. To begin, we derive a generic formula for $c_T(\hat{\theta}_T - \theta_*)$ where c_T is a sequence of constants which increases with T. From this analysis, it emerges that the limiting behaviour of this statistic depends in general on that of $c_T(W_T - W)$. We then use this result to deduce the limiting behaviour of $c_T(\hat{\theta}_T - \theta_*)$ for a number of specific choices of W_T which are commonly employed in practice.

Suppose that the moment function f is differentiable. The first order conditions for GMM estimation are:

$$G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0 \tag{3}$$

where $G_T(\theta) = \partial g_T(\theta)/\partial\theta'$. If the Mean Value Theorem is used to expand $g_T(\hat{\theta}_T)$ around $g_T(\theta_*)$ then after some rearrangement (3) implies

$$c_T(\hat{\theta}_T - \theta_*) \;=\; -[G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T, \theta_*, \lambda_T)]^{-1} G_T(\hat{\theta}_T)' W_T c_T g_T(\theta_*) \tag{4}$$

where $G_T(\hat{\theta}_T, \theta_*, \lambda_T)$ is $(q \times p)$ matrix whose i^{th} row is equal to the i^{th} row of $G_T(\bar{\theta}_T^{(i)})$ where $\bar{\theta}_T^{(i)} = \lambda_T^{(i)}\theta_* + (1 - \lambda_T^{(i)})\hat{\theta}_T$ for some $0 \leq \lambda_T^{(i)} \leq 1$, and λ_T is the $q \times 1$ vector with i^{th} element $\lambda_T^{(i)}$. It is convenient to rewrite (4) as

$$c_T(\hat{\theta}_T - \theta_*) \;=\; H_{0,T}\{\, H_{1,T} \;+\; H_{2,T}\,\} \tag{5}$$

where

$$
\begin{aligned}
H_{0,T} &= -[G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T, \theta_*, \lambda_T)]^{-1} && (6)\\
H_{1,T} &= G_T(\hat{\theta}_T)' W_T (c_T/T) \sum_{t=1}^{T} [f(v_t, \theta_*) - \mu_*] && (7)\\
H_{2,T} &= G_T(\hat{\theta}_T)' W_T c_T \mu_* \; = \; H_{2,T}(1) \; + \; H_{2,T}(2) \; + \; H_{2,T}(3) \; + \; H_{2,T}(4) && (8)\\
H_{2,T}(1) &= c_T[G_T(\hat{\theta}_T) - G_T(\theta_*)]' W_T \mu_*\\
H_{2,T}(2) &= c_T[G_T(\theta_*) - G_*]' W_T \mu_*\\
H_{2,T}(3) &= G_*' c_T (W_T - W) \mu_*\\
H_{2,T}(4) &= c_T G_*' W \mu_*\\
\mu_* &= E[f(v_t, \theta_*)]\\
G_* &= E[\partial f(v_t, \theta_*)/\partial \theta'].
\end{aligned}
$$

At this stage it is useful to note two simplifications. First, the population analog to the first order conditions imply $H_{2,T}(4) = 0$. Second, $H_{2,T}(1)$ can be written as[6]

$$
\begin{aligned}
H_{2,T}(1) &= (\mu_*' W_T \otimes I_p) \mathrm{vec}\{c_T[G_T(\hat{\theta}_T) - G_T(\theta_*)]\}\\
&= (\mu_*' W_T \otimes I_p) G_T^{(2)}(\hat{\theta}_T, \theta_*, \phi_T) c_T(\hat{\theta}_T - \theta_*)\\
&= M_T c_T(\hat{\theta}_T - \theta_*), \qquad \text{say}
\end{aligned}
$$

where $G_T^{(2)}(\hat{\theta}_T, \theta_*, \phi_T)$ is the $pq \times p$ matrix whose i^{th} row is the corresponding row of $(\partial/\partial\theta')vec\{\partial f(v_t, \tilde{\theta}_T^{(i)})/\partial\theta'\}$ with $\tilde{\theta}_T^{(i)} = \phi_T^{(i)}\hat{\theta}_T + (1 - \phi_T^{(i)})\theta_*$, $0 \leq \phi_T^{(i)} \leq 1$, and ϕ_T is the $pq \times 1$ vector with i^{th} element $\phi_T^{(i)}$.

Taking advantage of these two simplifications, (5)–(8) can be used to deduce that

$$
c_T(\hat{\theta}_T - \theta_*) \; = \; [I_p \; - \; H_{0,T} M_T]^{-1} H_{0,T} \{ H_{1,T} \; + \; H_{2,T}(2) \; + \; H_{2,T}(3) \}. \tag{9}
$$

This generic equation can lead to many possible types of behaviour for $\hat{\theta}_T$ as we now demonstrate.

[6] Dhrymes (1984) Corollary 25, p.103 and the Mean Value Theorem applied to the $i - j^{th}$ element of $G_T(\hat{\theta}_T)$.

In addition to some regularity conditions, we impose the following two assumptions.

Assumption 4 *$V = \lim_{T\to\infty} Var[T^{-1/2}\sum_{t=1}^{T}(f(v_t,\theta_*)-\mu_*)]$ is a positive definite matrix of constants where $E[f(v_t,\theta_*)] = \mu_*$.*

Assumption 5 *A $p\times p$ matrix $H_* = G_*'WG_* + (\mu_*'W\otimes I_p)G_*^{(2)}$ is nonsingular where $G_* = E[\partial f(v_t,\theta_*)/\partial\theta']$ and $G_*^{(2)} = E[(\partial/\partial\theta')vec\{\partial f(v_t,\theta_*)/\partial\theta'\}]$.*

Assumption 5 guarantees that the inverse matrix in (9) is well defined in the limit.

We consider four cases: (i) $W_T = W$ for all T; (ii) $T^{1/2}(W_T - W)$ converges to a normal distribution for some matrix W; (iii) W_T is the inverse of a centred HACC estimator; (iv) W_T is the inverse of an uncentred HACC estimator.

Case (i) $W_T = W$ for all T:

The limiting distribution of the GMM estimator is given in the following theorem.

Theorem 1 *Let $W_T = W$ and suppose that Assumptions 1–5 and Assumptions A1-10 (given in the appendix) hold. In addition, let $\{\Omega_{ij}\}_{i,j=1,2}$ denote the asymptotic covariance matrix in*

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T}(f(v_t,\theta_*)-\mu_*) \\ T^{1/2}(G_T(\theta_*)-G_*)W\mu_* \end{bmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\right) \tag{10}$$

(this asymptotic normality follows from Assumptions A1-10). Then it follows that

$$T^{\frac{1}{2}}(\hat{\theta}_T - \theta_*) \xrightarrow{d} N(0,\Sigma_1)$$

where

$$\Sigma_1 = H_*^{-1}(G_*'W\Omega_{11}WG_* + G_*'W\Omega_{12} + \Omega_{21}WG_* + \Omega_{22})H_*'^{-1}$$

Theorem 1 includes the common choice of $W_T = I$ on the first step estimation. Notice that this result contains the analogous result for correctly specified models as a special case, because if $\mu_* = 0$ then the asymptotic variance reduces to

$$\Sigma_C = (G_*'WG_*)^{-1}(G_*'WVWG_*)(G_*'WG_*)^{-1} \tag{11}$$

which is the formula derived by Hansen (1982). It is remarked in footnote 1 above that Gallant and White's (1988) analysis extends to GMM estimators with fixed weighting matrices. It is easily verified that under our conditions our Theorem 1 coincides with their Theorem 5.7.[7]

Case (ii) $T^{1/2}(W_T - W)$ **converges to a normal distribution:**

We now consider the case in which the weighting matrix depends on T and is $\sqrt{T}$-asymptotically normally distributed.

Theorem 2 *Suppose that Assumptions 1–5 and Assumptions A1-9 hold. In addition, assume that*

$$\begin{pmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T}[f(v_t,\theta_*)-\mu_*] \\ T^{1/2}(G_T(\theta_*)-G_*)W\mu_* \\ T^{\frac{1}{2}}(W_T-W)\mu_* \end{pmatrix} \xrightarrow{d} N\left(0_{(p+2q)\times 1}, \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix}\right). \tag{12}$$

where the asymptotic variance covariance matrix is positive definite. [8]

Then

$$T^{1/2}(\hat{\theta}_T - \theta_*) \xrightarrow{d} N(0, \Sigma_2),$$

where $\Sigma_2 = H_*^{-1}\Omega_* H_*'^{-1}$ *and*

$$\begin{aligned} \Omega_* &= G_*'W\Omega_{11}WG_* + \Omega_{22} + G_*'\Omega_{33}G_* + G_*'W\Omega_{12} \\ &\quad + G_*'W\Omega_{13}G_* + \Omega_{21}WG_* + G_*'\Omega_{31}WG_* + \Omega_{23}G_* + G_*'\Omega_{32}. \end{aligned}$$

Theorem 2 covers two leading cases of empirical relevance: (a) $\hat{\theta}_T$ is the first step estimator calculated using the inverse of an instrument cross product matrix as the weighting matrix; (b) $\hat{\theta}_T$ is the second step estimator based on the assumption that $\{f(v_t, \theta_*) - \mu_*\}$ is a martingale difference sequence. We now discuss each of these two cases in turn.

[7] Our result can be translated into Gallant and White's (1988) notation via $H_* = \lim_{n\to\infty} A_n^*$ and $G_*'WVWG_* = \lim_{n\to\infty} B_n^*$.

[8] (12) holds under more primitive conditions on dependence and moments similar to Assumptions A9 and A10.

First, consider case (a). Let $W_T = [T^{-1}\sum_{t=1}^{T} z_t z_t']^{-1}$, where z_t is a vector of instruments, and put $M_{zz} = E[z_t z_t']$. In this case, it follows that

$$T^{1/2}(W_T - W) = -M_{zz}^{-1}\, T^{-1/2}[\sum_{t=1}^{T} z_t z_t' - M_{zz}][T^{-1}\sum_{t=1}^{T} z_t z_t']^{-1} \tag{13}$$

Therefore, this case is covered by Theorem 2 provided that $T^{-1/2}\mathrm{vech}[\sum_{t=1}^{T}(z_t z_t' - M_{zz})]$ converges to a mean zero normal distribution. Such behaviour is assumed by Maasoumi and Phillips (1982) in their analysis of the IV estimator in linear models, and Theorem 2 contains their result as a special case.

Now consider case (b). For this discussion, it is necessary to distinguish the first and second step GMM estimators. Accordingly, let $\hat{\theta}_T(i)$ denote the GMM estimator on the i^{th} step, and $p\lim_{T\to\infty}\hat{\theta}_T(i) = \theta_*(i)$. In this case, the weighting matrix is $W_T = \tilde{\Gamma}_0^{-1}$ where

$$\tilde{\Gamma}_0 = T^{-1}\sum_{t=1}^{T}[f(v_t, \hat{\theta}_T(1)) - g_T(\hat{\theta}_T(1))][f(v_t, \hat{\theta}_T(1)) - g_T(\hat{\theta}_T(1))]' \tag{14}$$

Using the same argument as (13), it can be shown that this case is covered by Theorem 2 provided $T^{1/2}vech[\tilde{\Gamma}_0 - \Gamma_0]$ converges to a mean zero normal distribution where Γ_0 is the variance of $f(v_t, \theta_*)$. However if we pursue this analysis further it reveals a fundamental dependence between the first and second step estimators. Using the Mean Value Theorem, it follows that

$$\begin{aligned} T^{1/2}vech[\tilde{\Gamma}_0 - \Gamma_0] &= T^{1/2}vech[\Gamma_0(\theta_*(1)) - \Gamma_0] \\ &\quad + \frac{\partial}{\partial\theta'}\,(vech\{\Gamma_0(\theta_*(1))\})\, T^{1/2}(\hat{\theta}_T(1) - \theta_*(1)) + o_p(1) \end{aligned} \tag{15}$$

where $\Gamma_0(\theta) = T^{-1}\sum_{t=1}^{T}[f(v_t,\theta) - g_T(\theta)][f(v_t,\theta) - g_T(\theta)]'$. Equation (5) combined with equation (15) imply the limiting distribution of $T^{1/2}(\hat{\theta}_T(2) - \theta_*(2))$ depends on the limiting distribution of $T^{1/2}(\hat{\theta}_T(1) - \theta_*(1))$ via $T^{1/2}vech[\tilde{\Gamma}_0 - \Gamma_0]$ *unless* the partial derivative matrix in (15) converges in probability to zero. In general there is no reason to expect the latter condition to hold. A similar argument holds for the iterated estimator discussed by Hansen, Heaton, and Yaron (1996). Equation (15) can be applied recursively to deduce that the limiting distribution of $T^{1/2}(\hat{\theta}_T(i) - \theta_*(i))$ depends on the limiting distributions of $\{T^{1/2}(\hat{\theta}_T(j) - \theta_*(j)), j = 1, 2 \ldots i-1\}$ in general. Theorem 2 also applies to the case where $\{f(v_t, \theta_*) - \mu_*\}$ is not a martingale difference sequence but its serial correlation is neglected when computing the weighting matrix.

To conclude the discussion of Theorem 2, we note that it also encompasses the case of correctly specified models because if $\mu_* = 0$ then $\Omega_{12} = \Omega_{21} = \Omega_{22} = 0_{q\times q}$ and the variance reduces to V_C. Furthermore, it is easily verified that if $\mu_* = 0$ then this eliminates the dependence of $T^{1/2}(\hat{\theta}_T(i) - \theta_*(i))$ on $\{T^{1/2}(\hat{\theta}_T(j) - \theta_*(j)), j = 1, 2 \ldots i-1\}$.

Case (iii) W_T is the inverse of a centred HACC estimator:

We now consider the case in which the weighting matrix is the inverse of a centred HACC matrix estimator, that is $W_T = \hat{V}_T^{-1}$ where

$$\hat{V}_T = \sum_{i=-T+1}^{T-1} \omega(i/b_T)\tilde{\Gamma}_i \tag{16}$$

and

$$\begin{aligned}
\tilde{\Gamma}_i &= T^{-1} \sum_{t=i+1}^{T} \left[f(v_t, \hat{\theta}_T(1)) - g_T(\hat{\theta}_T(1))\right] \left[f(v_{t-i}, \hat{\theta}_T(1)) - g_T(\hat{\theta}_T(1))\right]' \qquad \text{for } i \geq 0 \\
&= T^{-1} \sum_{t=-i+1}^{T} \left[f(v_{t+i}, \hat{\theta}_T(1)) - g_T(\hat{\theta}_T(1))\right] \left[f(v_t, \hat{\theta}_T(1)) - g_T(\hat{\theta}_T(1))\right]' \qquad \text{for } i < 0
\end{aligned}$$

It is assumed that the kernel, $\omega(.)$, and bandwidth, b_T, satisfy the following assumption.

Assumption 6 *(i) For all $x \in \Re$, $|\omega(x)| \leq 1$, $\omega(-x) = \omega(x)$, $\omega(0) = 1$, $\omega(x)$ is continuous at zero and for almost all $x \in \Re$, $\int_{\Re} \omega(x)^2 dx < \infty$, $\int_{\Re} \omega(x) e^{-ix\lambda} dx \geq 0$ for all $\lambda \in \Re$; (ii) $\int_{-\infty}^{\infty} \omega(x) dx = c$ where $0 < c < \infty$; (iii) $\int_{\Re} x w(x) dx < \infty$; (iv) $b_T = o(T^{1/2})$ and $b_T \to \infty$.*

Assumption 6(i) is standard in the HACC literature and satisfied by all kernels of interest in econometrics which yield positive semi–definite estimators for finite T; see Andrews (1991), Hansen (1992). Assumption 6(ii) is imposed by Hall (2000) in his analysis of the inverse of the uncentred HACC estimator. We impose it here for simplicity because we need it later in our analysis. It is easily verified that this restriction is satisfied by the Bartlett (Newey and

West, 1987b), Parzen (Gallant, 1987) and the quadratic spectral (Andrews, 1991) kernels. Assumption 6 (iii) is not typically imposed in the analysis of uncentred HACC estimators in correctly specified models, but is necessary for the analysis of centred HACC estimators in misspecified models. It can be shown that this condition is satisfied by the Bartlett, Parzen and quadratic spectral kernels. Assumption 6 (iv) gives the rate of increase of b_T used by Andrews (1991) in his proof of the consistency of $\hat{S}_T$ in correctly specified models.

For this part of the analysis, it is necessary to impose the following assumption.

Assumption 7 *Let $\hat{\theta}_T(1)$ be the first step GMM estimator, θ_* denote its probability limit and $\mu_* = \mu(\theta_*)$.*

(i)

$$\left(\frac{T}{b_T}\right)^{\frac{1}{2}} vech(\bar{V}_T - V_T) \xrightarrow{d} N(0, \Omega_V) \tag{17}$$

$$\lim_{T\to\infty} b^k(V_T - V) = C, \tag{18}$$

where Ω_V is a $q^2 \times q^2$ matrix, $k > 0$ is the characteristic exponent of the kernel ω,[9]

$$\begin{aligned}
\bar{V}_T &= \bar{\Gamma}_0 + \sum_{i=1}^{T-1} \omega(i/b_T)(\bar{\Gamma}_i + \bar{\Gamma}_i'), \\
V_T &= \Gamma_0 + \sum_{i=1}^{T-1} \omega(i/b_T)(\Gamma_i + \Gamma_i'), \\
\bar{\Gamma}_i &= T^{-1} \sum_{t=i+1}^{T} [f(v_t, \theta_*) - \mu_*][f(v_{t-i}, \theta_*) - \mu_*]', \\
C &= -\lim_{x\to 0} \left(\frac{1-\omega(x)}{|x|^k}\right) \sum_{j=-\infty}^{\infty} |j|^k \Gamma_j \;<\; \infty.
\end{aligned}$$

(ii) $\hat{\theta}_T(1) - \theta_* = O_p(T^{-1/2})$.

(iii) $\max_{1\le i\le T} \left\| \frac{1}{T}\sum_{t=1}^{i} f(v_t, \theta_*) - \mu_* \right\| = O_p(T^{-\frac{1}{2}})$.

Assumption 7(i) states that the spectral density estimator of $f(v_t, \theta_*) - \mu_*$ is asymptotically normally distributed and that the bias vanishes at an appropriate rate. More primitive assumptions can be found in the classical work of Anderson (1994), Brillinger (1975),

[9]See Anderson (1994)[Section 9.3.2] for a definition of the characteristic exponent of a kernel.

Rosenblatt (1959) and Hannan (1970). More primitive conditions are provided in the mathematical appendix (Proposition A.1). Assumption 7(ii) is satisfied if the weighting matrix on the first step satisfied the assumptions imposed in either Theorems 1 or 2. Assumption 7(iii) is satisfied if the functional central limit theorem holds for $T^{-1/2}\sum_{t=1}^{[Tr]}[f(v_t,\theta_*)-\mu_*]$ in the space of cadlag functions on the unit interval, for instance.

Theorem 3 *Let $\hat{\theta}_T(2)$ denote the second step GMM estimator based on $W_T = \hat{V}_T^{-1}$, and $\theta_*(2) = \theta_*(V^{-1})$. Also suppose that Assumptions 1, 3–7 and Assumptions A1–9 hold. Then*

$$\begin{aligned}\left(\frac{T}{b_T}\right)^{\frac{1}{2}}(\hat{\theta}_T(2)-\theta_*(2)) &\stackrel{d}{\to} N(\phi H_{**}^{-1}G_{**}'C\mu_{**},\Sigma_3) \quad \text{if } T^{1/2}/b_T^{1/2+k}\to\phi\in[0,\infty),\\ b_T^k(\hat{\theta}_T(2)-\theta_*(2)) &\stackrel{p}{\to} H_{**}^{-1}G_{**}'C\mu_{**} \quad \text{if } T^{1/2}/b_T^{1/2+k}\to\infty,\end{aligned}$$

where $\Sigma_3 = H_{**}^{-1}D'B\Omega_V B'DH_{**}'^{-1}$, $D = -(\mu_{**}'V^{-1}\otimes G_{**}'V^{-1})$, H_{**} is the matrix H_* in Assumption 5 evaluated at $\theta_* = \theta_*(2)$, $G_{**} = E[\partial f(v_t,\theta_*(2))/\partial\theta']$ and B is the selection matrix defined by $vec(S_*) = B\, vech(S_*)$.

It is interesting to contrast this result with the correctly specified case. If the model is correctly specified then $\hat{\theta}_T(2)$ converges at rate $T^{1/2}$ to a normal distribution. However, if the model is misspecified and $W_T = \hat{V}_T^{-1}$ then $H_{2,T}(2)$ becomes dominant in the limiting expression, and this causes the rate of convergence to be $(T/b_T)^{1/2}$ in the first case and b_T^k in the second case. In the second case the bias of the HACC estimator becomes dominant in the limiting expression of the GMM estimator, and this causes the degenerate limiting behaviour. The difference in the rates of convergence also means that Theorem 3 does not contain the appropriate result for the correctly specified model as a special case. This contrasts with Theorems 1 and 2 above.

This rate of convergence has important consequences for the iterated GMM estimator discussed in Hansen, Heaton, and Yaron (1996). To consider consequences, it is useful to recall first that it is shown above that if $W_T = \tilde{\Gamma}_0^{-1}$ then the asymptotic distribution of $\hat{\theta}_T(i)$ depends on the asymptotic distribution of the estimator on all previous steps. If $W_T = \hat{V}_T^{-1}$

then a similar picture emerges except that this time the dependence only goes back as far as the second step. The reason is that Assumption 7(ii) restricts the first step estimator to converge faster than the second step estimator, and so $\sqrt{T/b_T}(\hat{\theta}_T(1) - \theta_*(1)) \xrightarrow{p} 0$. However, using a similar Mean Value Theorem based argument to (15), it can be shown that the limiting distribution of $\sqrt{T/b_T} vech[V_T - V]$ on step 3 depends on the distribution of $\hat{\theta}_T(2)$ and hence on the distribution of $\sqrt{T/b_T} vech[V_T - V]$ on step 2. This effect cumulates so that in general the limiting distribution of $\sqrt{T/b_T}(\hat{\theta}_T(i) - \theta_*(i))$ depends on the limiting distribution of $\{\sqrt{T/b_T}(\hat{\theta}_T(j) - \theta_*(j)),\ j = 2, 3, \ldots j-1\}$.

Case (iv) W_T is the inverse of a uncentred HACC estimator:

Finally, we consider the case in which the weighting matrix is the inverse of an uncentred HACC matrix estimator, that is $\hat{S}_T^{-1}$ where

$$\hat{S}_T = \sum_{i=-T+1}^{T+1} \omega(i/b_T)\hat{\Gamma}_i \tag{19}$$

and

$$\begin{aligned} \hat{\Gamma}_i &= T^{-1} \sum_{t=i+1}^{T} f(v_t, \hat{\theta}_T(1)) f(v_{t-i}, \hat{\theta}_T(1))' \qquad \text{for } i \geq 0 \\ &= T^{-1} \sum_{t=-i+1}^{T} f(v_{t+j}, \hat{\theta}_T(1)) f(v_t, \hat{\theta}_T(1))' \qquad \text{for } i < 0 \end{aligned}$$

In correctly specified models, $\hat{S}_T^{-1}$ is a valid weighting matrix because it is positive semi–definite for finite T by construction and converges in probability to the inverse of the long run variance which is positive definite by assumption. However, the latter property does not extend to misspecified models. Hall (2000) shows that $\hat{S}_T^{-1}$ converges in probability to

$$S_* = V^{-1} - \frac{1}{\mu_*' V^{-1} \mu_*} V^{-1} \mu_* \mu_*' V^{-1} \tag{20}$$

where $\mu_* = E[f(v_t, \theta_*(1))]$. Hall (2000) shows that S_* has rank equal to $q-1$ and hence is singular. The form of this limit has an important implication for the GMM estimator based on $W_T = \hat{S}_T^{-1}$. The population analog to the minimand is

$$Q_0^{(2)}(\theta) = E[f(v_t, \theta)]' S_* E[f(v_t, \theta)] \tag{21}$$

Using (20) it can be seen that $Q_0^{(2)}(\theta)$ attains its minimum possible value of zero at $\theta = \theta_*(1)$. For this minimum to be unique, there must be no other value of θ which generates a value of $\mu(\theta)$ in the nullspace of S_*. Therefore we impose the condition

Assumption 8 *$S_* E[f(v,\theta)] \neq 0$ for any $\theta \in \Theta \setminus \{\theta_*(1)\}$.*

This restriction serves as the identification condition in the following result.

Theorem 4 *Let $\hat{\theta}_T(2)$ denote the second step estimator based on $W_T = \hat{S}_T^{-1}$, and suppose that Assumptions 1, 3–8 and Assumptions A1–9 hold. Then (i) $\hat{\theta}_T(2) \xrightarrow{p} \theta_*(1)$; (ii) plim $b_T[\hat{\theta}_T(2) - \theta_*(1)] = C_*$ for some $C_* \neq 0$ if $G_*' V^{-1}\mu_* \neq 0$; (iii) $T^{1/2}(\hat{\theta}_T(2) - \theta_*(1)) = O_p(1)$ if $G_*' V^{-1}\mu_* = 0$; where $G_* = E[\partial f(v_t, \theta_*(1))/\partial\theta']$ and $\mu_* = \mu(\theta_*(1))$.*

Theorem 4 (i) indicates that the second step GMM estimator converges to the same probability limit as the first step estimator. This property would be exhibited in correctly specified models, but does not typically hold in misspecified models. Its occurrence here results directly from the singularity of $W = S_*$, and, more specifically, the nature of the null space of S_*. Theorem 4 (ii) and (iii) show that the rate of convergence depends on the value of $G_*' V^{-1}\mu_*$. This latter quantity can be recognized as the population analog to the first order conditions based on $W_T = \hat{V}_T^{-1}$, and it enters the analysis via the first order conditions of the estimation. When $W_T = \hat{S}_T^{-1}$ then the population analog to the first order conditions are given by

$$G(\theta)' S_* \mu(\theta) \;=\; G(\theta)' V^{-1}\mu(\theta) \;-\; \frac{1}{\mu_*' V^{-1}\mu_*} G(\theta)' V^{-1}\mu_* \mu_*' V^{-1}\mu(\theta) \;=\; 0 \qquad (22)$$

Notice that $\theta = \theta_*(1)$ is always a solution, because in that case (22) reduces to

$$G_*' V^{-1}\mu_* \;-\; G_*' V^{-1}\mu_* \;=\; 0 \qquad (23)$$

The difference between parts (ii) and (iii) of the theorem is that in part (iii) the left hand side of (23) reduces to $0 - 0$. In practice, this coincidence only occurs if the first step weighting matrix is proportional to V^{-1}, and so is likely to be rare. Therefore, Theorem 4(ii) is likely to be the most generally applicable result.

4 Inference about the parameters in misspecified models

As remarked in the introduction, it may be of interest to test hypotheses about the parameters even in misspecified models. In this section we discuss the implications of our results for this issue and also propose certain statistics which can be used to test hypotheses about the parameters.

Consider testing for the null hypothesis

$$\mathrm{H}_0 : a\left(\theta_*(W)\right) = 0$$

against the alternative

$$\mathrm{H}_1 : a\left(\theta_*(W)\right) \neq 0,$$

where

Assumption 9 *$a : \Theta \rightarrow \Re^r$ is continuously differentiable and $A_* = \partial a\left(\theta_*(W)\right)/\partial\theta'$ has rank r.*

Notice that the dependence of θ_* on W translates to the null and alternative hypotheses. Two referees argue that this dependence makes the null hypothesis hard to interpret. We do not dispute this point. However, as pointed out in the introduction, a number of published papers do report the results from inference based on conventional test statistics even though there is clear evidence that the model is misspecified. In the first part of this section, we consider the behaviour of these conventional test statistics if the model is in fact misspecified, and thereby provide guidance on the interpretation of these types of results. At the end of the section, we briefly consider alternative methods for testing the null given above because, at the very least, researchers are evidently interested in performing these types of inferences in misspecified models. It is left to potential users to demonstrate that the null hypothesis can be interpreted meaningfully for the particular case in hand.

Let $\hat{\theta}_T$ denote the unconstrained GMM estimator and $\bar{\theta}_T$ denote the constrained GMM estimator:

$$\bar{\theta}_T = \operatorname{argmin}_{\theta\in\Theta} g_T(\theta)' W_T g_T(\theta)$$
$$\text{subject to } a(\theta) = 0.$$

Define a Wald statistic, a LM statistic, and a LR-like statistic by

$$\begin{aligned}
Wald_T &= Ta(\hat{\theta}_T)'\{\hat{A}_T(\hat{G}_T' W_T \hat{G}_T)^{-1}\hat{A}_T'\}^{-1} a(\hat{\theta}_T), \\
LM_T &= T g_T(\bar{\theta}_T)' W_T \bar{G}_T(\bar{G}_T' W_T \bar{G}_T)^{-1} \bar{G}_T' W_T g_T(\bar{\theta}_T), \\
LR_T &= T(g_T(\bar{\theta}_T)' W_T g_T(\bar{\theta}_T) - g_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T)),
\end{aligned}$$

where $\hat{A}_T = \partial a(\hat{\theta}_T)/\partial\theta'$, $\hat{G}_T = (1/T)\sum_{t=1}^T \partial f(v_t, \hat{\theta}_T)/\partial\theta'$, and $\bar{G}_T = (1/T)\sum_{t=1}^T \partial f(v_t, \bar{\theta}_T)/\partial\theta'$.

In practice, inference about the parameters is typically based on the two step or iterated estimator. In view of the comments following Theorems 2 and 3, we confine our attention to the two–step estimator. We consider the limiting behaviour of the tests statistics above when the weighting matrix is based on three different choices of long run covariance matrix estimator. The first choice is the covariance matrix estimator constructed under a martingale difference assumption. The second and third choices are the centred and uncentred versions of the HACC estimator.

We begin with the martingale difference case. As remarked above, this scenario is covered by Case (ii) (*c.f.* Theorem 2).

Theorem 5 *Let $W_T = \tilde{\Gamma}_0^{-1}$ and suppose that Assumptions 1–5, 9, and Assumptions A1–9 hold. In addition, suppose that $\{f(v_t, \theta_*) - \mu_*\}$ is a mean-zero iid or martingale difference sequence and that (12) holds. Then under the null hypothesis,*

$$\begin{aligned}
Wald_T &\xrightarrow{d} \sum_{i=1}^{r} \lambda_i^W u_i^2, \\
LM_T &\xrightarrow{d} \sum_{i=1}^{p} \lambda_i^{LM} u_i^2, \\
LR_T &\xrightarrow{d} \sum_{i=1}^{p} \lambda_i^{LR} u_i^2,
\end{aligned}$$

where $\{\lambda_i^W\}$ *are the eigenvalues of the* $r \times r$ *matrix*

$$\Sigma_2^{\frac{1}{2}} A_*'(A_*(G_*'WG_*)^{-1}A_*')^{-1}A_*\Sigma_2^{\frac{1}{2}},$$

$\{\lambda_i^{LM}\}$ *are the eigenvalues of the* $p \times p$ *matrix*

$$\Sigma_4^{\frac{1}{2}}(G_*'WG_*)^{-1}\Sigma_4^{\frac{1}{2}},$$

$$\begin{aligned}\Sigma_4 \;=\; & (G_*'W - H_*M_*)\Omega_{11}(G_*'W - H_*M_*)' + (I_p - H_*M_*)(\Omega_{22} + \Omega_{23} + \Omega_{32} + \Omega_{33})(I_p - H_*M_*)' \\ & +(G_*'W - H_*M_*)(\Omega_{12} + \Omega_{13})(I_p - H_*M_*)' + (I_p - H_*M_*)(\Omega_{21} + \Omega_{31})(G_*'W - H_*M_*)',\end{aligned}$$

$$M_* \;=\; H_*^{-1} - H_*^{-1}A_*'(A_*H_*^{-1}A_*')^{-1}A_*H_*^{-1},$$

$\{\lambda_i^{LR}\}$ *are the eigenvalues of the* $p \times p$ *matrix*

$$\Omega_*^{\frac{1}{2}} H_*^{-1}A_*'(A_*H_*^{-1}A')^{-1}A_*H_*^{-1}G^{*'}WG^*(H_*^{-1}A_*'(A_*H_*^{-1}A')^{-1}A_*H_*^{-1})'\Omega_*^{\frac{1}{2}},$$

Σ_2 *and* Ω_* *are defined in Theorem 2, and* u_i *are iid standard normal random variables.*

Theorem 5 reveals that the limiting distributions of the three statistics are mixtures of iid χ^2 distributions, but that the mixtures are different in general for each statistic. This contrasts with their behaviour if the model is correctly specified, *i.e.* $\mu_* = 0$. In the latter case, Newey and West (1987b) show that the three statistics are asymptotically equivalent and converge to a χ_r^2 distribution. The differences in misspecified models are due to the dependence of the asymptotic distributions of the unrestricted and restricted estimators upon the asymptotic distributions of the derivative matrix and weighting matrix.

We now consider the case in which $W_T = \hat{V}_T^{-1}$ (Case (iii) and Theorem 3 above).

Theorem 6 *Let* $W_T = \hat{V}_T^{-1}$. *Suppose that Assumptions 1, 3–7, 9 and Assumptions A1–9 hold and that the null hypothesis is true. If* $T^{1/2}/b_T^{1/2+k} \to \phi \in [0, \infty)$,

$$\begin{aligned}\frac{1}{b_T}Wald_T &\xrightarrow{d} \sum_{i=1}^{r} \lambda_i^W u_i^2, \\ \frac{1}{b_T}LM_T &\xrightarrow{d} \sum_{i=1}^{p} \lambda_i^{LM} u_i^2, \\ \frac{1}{b_T}LR_T &\xrightarrow{d} \sum_{i=1}^{p} \lambda_i^{LR} u_i^2,\end{aligned}$$

where $\{\lambda_i^W\}$ *are the eigenvalues of the* $r \times r$ *matrix*

$$\Sigma_3^{\frac{1}{2}} A'_{**}(A_{**}(G'_{**}WG_{**})^{-1}A'_{**})^{-1}A_{**}\Sigma_3^{\frac{1}{2}},$$

$\{\lambda_i^{LM}\}$ *are the eigenvalues of the* $p \times p$ *matrix*

$$\Sigma_3^{\frac{1}{2}} H_*(I_p - H_{**}M_{**})'(G'_{**}WG_{**})^{-1}(I_p - H_{**}M_{**})H'_*\Sigma_3^{\frac{1}{2}},$$

$$M_{**} = H_{**}^{-1} - H_{**}^{-1}A'_{**}(A_{**}H_{**}^{-1}A'_{**})^{-1}A_{**}H_{**}^{-1},$$

$\{\lambda_i^{LR}\}$ *are the eigenvalues of the* $p \times p$ *matrix*

$$\Omega_V^{\frac{1}{2}} B'DH_{**}^{-1}A'_{**}(A_{**}H_{**}^{-1}A'_{**})^{-1}A_{**}H_{**}^{-1}G'_{**}WG_{**}(H_{**}^{-1}A'_{**}(A_{**}H_{**}^{-1}A'_{**})^{-1}A_{**}H_{**}^{-1})'D'B\Omega_V^{\frac{1}{2}},$$

$u_i \sim N(-\phi\Sigma_3^{-\frac{1}{2}}H_{**}^{-1}G'_{**}C\mu_{**}, 1)$ *is iid,* Σ_3 *and* Ω_V *are defined in Theorem 3 and Assumption 7, respectively. If* $T^{1/2}/b_T^{1/2+k} \to \infty$, $\frac{b_T^{2k}}{T}Wald_T \xrightarrow{p} \kappa_W$, $\frac{b_T^{2k}}{T}LM_T \xrightarrow{p} \kappa_{LM}$ *and* $\frac{b_T^{2k}}{T}LR_T \xrightarrow{p} \kappa_{LR}$ *where* κ_W, κ_{LM} *and* κ_{LR} *are some positive constants.*

From Theorem 6 it can be seen that the three test statistics diverge to positive infinity even under the null hypothesis regardless of the rate of increase of the bandwidth.

We consider the case in which $W_T = \hat{S}_T^{-1}$ (Case (iv) and Theorem 4 above). For brevity we concentrate on the most general case in which $G_*V^{-1}\mu_* \neq 0$.

Theorem 7 *Let* $W_T = \hat{S}_T^{-1}$, *and suppose that* $G_*V^{-1}\mu_* \neq 0$. *Suppose that Assumptions 1, 3-9 and Assumptions A1-10 hold. Then under the null hypothesis,* $(b_T^2/T)Wald_T \xrightarrow{p} c_W$, $(b_T^2/T)LM_T \xrightarrow{p} c_{LM}$ *and* $(b_T^2/T)LR_T \xrightarrow{p} c_{LR}$ *where* $c_W > 0$, $c_{LM} > 0$ *and* $c_{LR} > 0$.

Theorem 7 implies that all three test statistics diverge to positive infinity even under the null hypothesis. This behaviour arises because both the restricted and unrestricted GMM estimators converge at a rate slower than $\sqrt{T}$.

Theorems 5, 6 and 7 have the following two implications. First, traditional hypothesis testing is invalid under misspecification. Even when the test statistic has a well-defined limiting distribution, the test does not have correct size. Second, in the case where HACC estimators are used, it is impossible to perform inference about the parameters which is robust to model misspecification. The reason is that the rate of convergence of the test statistics depends on whether the model is correctly specified or not. Note that this conclusion holds regardless of whether the HACC estimator is uncentred or centred. In contrast, it is possible to perform inference in the situations covered by Theorems 1 and 2 as we now demonstrate. However, we remind the reader that if the model is misspecified then the null hypothesis depends on W and so the tests presented below are for different null hypotheses.

We consider two approaches to performing this inference. The first approach is to use the usual statistics but calculate the appropriate critical value associated with the appropriate mixture distribution given in Theorem 5. These critical values can be simulated as follows. Let $\hat{\lambda}$ be the eigenvalues of a consistent estimator of the appropriate matrix given in Theorem 5.[10]

The $100\alpha\%$ critical value can be calculated as the $100(1-\alpha)^{th}$ percentile of the simulated distribution of $\sum_{i=1}^{p} \hat{\lambda}_i z_i^2$ where $z = (z_1, z_2, \ldots z_p) \sim N(0, I_p)$.[11] The second approach to inference involves using different statistics as a basis for inference. We consider Cases (i) and (ii) in turn. If $W_T = W$ (Case (i)) then H_0 can be tested using the statistic

$$Q_1 \;=\; Ta(\hat{\theta}_T)'[A(\hat{\theta}_T)\hat{V}_1 A(\hat{\theta}_T)']^{-1}a(\hat{\theta}_T) \tag{24}$$

where $\hat{V}_1$ is a consistent estimator of V_1 (given in Theorem 1). If $T^{1/2}(W_T - W)$ converges to a normal distribution (Case (ii)) then H_0 can be tested using the statistic

$$Q_2 \;=\; Ta(\hat{\theta}_T)'[A(\hat{\theta}_T)\hat{V}_2 A(\hat{\theta}_T)']^{-1}a(\hat{\theta}_T) \tag{25}$$

where $\hat{V}_2$ is a consistent estimator of $V_2 = H_*^{-1}\Sigma_* H_*^{-1}$ (defined in Theorem 2). The following

[10]It follows from Tyler (1983)[Lemma 2.2] that if $\hat{M} \xrightarrow{p} M$ then $\lambda_i(\hat{M}) \xrightarrow{p} \lambda_i(M)$ where $\lambda_i(.)$ denotes the i^{th} ordered eigenvalue of the matrix in parentheses.

[11]Note that is the model is correctly specified then $\hat{\lambda}_i \xrightarrow{p} 1$ for $i = 1, 2 \ldots p$ and so this method yields the appropriate critical value. Note that $\hat{\lambda}$

corollary to Theorems 1 and 2 gives the limiting distributions of these two statistics under H_0.

Corollary 1 *(i) If the conditions of Theorem 1 hold,* $\hat{V}_1 \xrightarrow{p} V_1$, $W_T = W_1$, *Assumption 9 holds for* $W = W_1$ *and* $a(\theta_*(W_1)) = 0$ *then* $Q_1 \xrightarrow{d} \chi^2_r$. *(ii) If the conditions of Theorem 2 hold,* $\hat{V}_2 \xrightarrow{p} V_2$, $W_T \xrightarrow{p} W_2$, *Assumption 9 holds at* $W = W_2$ *and* $a(\theta_*(W_2)) = 0$ *then* $Q_2 \xrightarrow{d} \chi^2_r$.

The matrices $\hat{V}_i$ can be constructed using consistent estimators of the components of V_i. The derivative matrix G_* can be consistently estimated by $G_T(\hat{\theta}_T)$ and the covariance matrices can be estimated using an HACC applied to the appropriate partial sums. There is no theoretical reason to prefer either version of the test as there is no theory regarding the optimal choice of weighting matrix in misspecified models. Since Q_1 involves $W_T = W$, this statistic is likely to be computationally more convenient. However, if the moment condition involves an instrument vector then it may be desirable to set W_T equal to the inverse of the instrument cross product matrix and use Q_2. Either choice is likely to be prefered to $W_T = \tilde{\Gamma}_0^{-1}$ in equation (14) because the latter induces a dependence on the asymptotic distribution of the first step estimator. It is left to future work to provide further guidance on the choice between Q_1 and Q_2, and also on the interpretation of this type of hypothesis in misspecified models.

5 Concluding Remarks

In this paper, we present a limiting distribution theory for GMM estimators when estimation is based on a population moment condition which exhibits non–local (or fixed) misspecification. It is shown that if the parameter vector is overidentified then the weighting matrix plays a far more fundamental role in the analysis than would be the case in correctly specified or locally misspecified models. Specifically, the probability limit of the estimator depends on the probability limit of the weighting matrix and the rate of convergence of the estimator depends on the rate of convergence of the weighting matrix to its limit. The latter means that there is no single distribution theory for GMM estimators in misspecified models. Rather the form of the limiting distribution has to be derived on a case by case basis. In this paper, we explicitly consider four cases: (i) $W_T = W$ for all T; (ii) $T^{1/2}(W_T - W)$ converges to a normal distribution for some matrix W; (iii) W_T is the inverse of a centered HACC estimator; (iv) W_T is the inverse of an uncentred HACC estimator. It is shown that $T^{1/2}(\hat{\theta}_T - \theta_*)$ converges to a limiting normal distribution in cases (i) and (ii). However, in case (iii) the limiting behaviour depends on the rate of increase of the bandwidth. If the bandwidth does not increase too quickly then $(T/b_T)^{1/2}(\hat{\theta}_T - \theta_*)$ converges to a limiting normal distribution; otherwise, $b_T^k((\hat{\theta}_T - \theta_*)$ converges to a degenerate limiting distribution. It is shown that in Case (iv) $b_T(\hat{\theta}_T - \theta_*)$ converges in probability to a constant in most cases of practical relevance.

In practice, inference is most often based on the two–step or iterated estimator. In correctly specified models, these two estimators are asymptotically equivalent. However, our results indicate this is not the case in misspecified models. It is shown that in situations covered by Case (ii) above then the asymptotic distribution of the estimator on the i^{th} step depends on the asymptotic distributions of the estimators on all previous steps. Whereas in situations covered by Case (iii) above then this dependence only goes back as far as the second step.

It is interesting to contrast these results with others in the literature. First, as mentioned, none of the distributions equal the ones derived by Hansen (1982) for the correctly specified case or Newey (1985) for the locally misspecified case. Second, the array of limiting behaviour described is different from that derived by either Gallant and White (1988) and White (1994) in their investigations of certain estimators in nonlinear models or Maasoumi and Phillips (1982) in their investigation of the IV estimator in linear models. All three studies find $T^{1/2}(\hat{\theta}_T - \theta_*)$ converges to a normal distribution. Although none of these studies impose stationarity, it is convenient to do so here for the purposes of comparison.

Gallant and White's (1988) framework does not cover the general case of GMM estimation in misspecified models. However, as they note, it does cover GMM with a fixed weighting matrix. Therefore, their results correspond to our case (i) above. White (1994) develops a limiting distribution theory for the quasi maximum likelihood estimator (QMLE) in non–locally misspecified models. While QMLE can be viewed as a GMM estimator, this interpretation involves the restriction that $p = q$. In this case, our stationarity assumption would imply $\mu_* = 0$ (see Proposition 1 in Section 2). It is easily shown that this restriction implies $T^{1/2}(\hat{\theta}_T - \theta_*)$ converges to a normal result and thus accords with White's (1994) results. Finally, Maasoumi and Phillips' (1982) analysis can be considered as a special case of our case (ii) above.

In the introduction it is noted that there is a growing interest in performing inference within misspecified models. Our results imply that if the parameter vector is overidentified then it is inappropriate to use conventional statistics derived under the assumption that the model is correctly specified. Therefore, we use our results to propose two new statistics which can be used to test hypotheses about the pseudo–parameters. Both these statistics are computationally convenient and have limiting chi–squared distributions under the null.

Mathematical Appendix

We first list the additional assumptions for the Lemmas and Theorems presented in the paper.

Assumption A.1 $f : \mathcal{V} \times \Theta \rightarrow \Re^q$.

Assumption A.2 Θ *is a compact set.*

Assumption A.3 $f(.,\theta)$ *is measurable for each* $\theta \in \Theta$ *and* $f(v,.)$ *is continuous on* Θ *for every* $v \in \mathcal{V}$.

Assumption A.4 $f(v_t, \theta)$ *satisfies the Uniform Weak Law of Large Numbers on* Θ.

Assumption A.5 $f(v,\theta)$ *is twice continuously differentiable with respect to* θ *on* $int(\Theta)$, *and* $\partial f(.,\theta)/\partial\theta^{'}$ *and* $\partial/\partial\theta' vec\{\partial f(.,\theta)/\partial\theta'\}$ *are measurable on* $\mathcal{V}$ *for each* $\theta \in int(\Theta)$.

Assumption A.6 *There exists a measurable function* $b(v)$ *such that* $|f_i(v,\theta)| < b(v)$, $|\partial f_i(v,\theta)/\partial\theta_j| < b(v)$ *and* $|\partial^2 f_i(v,\theta)/\partial\theta_j\partial\theta_k| < b(v)$ *in a neighbourhood of* θ_*, *for all* $i = 1, 2, ..., q,\ j, k = 1, 2, ..., p$ *and* $E[b(v)^2] < D$, *a finite constant; there exist constants* D, $\delta > 0$ *and* $r \geq 1$ *such that* $E[\{\sup_i |f_i(v,\theta_*)|\}^{4(r+\delta)}\}] < D$.

Assumption A.7 v_t *is an* α*–mixing sequence with size* $-3r/(r-1)$, $r > 1$.

Assumption A.8 θ_* *is an interior point of* Θ.

Assumption A.9 *(i)*

$$X_t = \begin{bmatrix} f(v_t, \theta_*) \\ vec\{\,(\partial/\partial\theta')f(v_t,\theta_*)\,\} \\ vec[\,(\partial/\partial\theta')vec\{\,(\partial/\partial\theta')f(v_t,\theta_*)\,\}\,] \end{bmatrix}$$

is an α*–mixing process with size* $-3r/(r-1)$ *for* $r > 1$ *and* $E[\| X_t \|^{4r}] < \infty$; *(ii)* $E[\sup_{\theta\in\Theta} \| (\partial^2/\partial\theta\partial\theta^{'})f_i(v_t,\theta) \|] < \infty$.

Assumption A.10

$$\lim_{T\rightarrow\infty} \frac{1}{T} Var\left[\sum_{t=1}^{T} \begin{pmatrix} f(v_t,\theta_*) \\ \mathrm{vec}\{\,(\partial/\partial\theta')f(v_t,\theta_*)\,\} \end{pmatrix}\right]$$

exists and is positive definite.

Assumptions A5 to A6 guarantee the Uniform Weak Law of Large Numbers for g_T and G_T among other things. Assumptions A7 to A10 are for the asymptotic normality of the normalized sum of f and its derivatives. More primitive conditions for the Central Limit Theorem and the Uniform Weak Law of Large Numbers can be found in *inter alia* Wooldridge (1994).

The following is a vector generalization of Theorem 9.4.1 and Theorem 9.3.3 of Anderson (1994).

Proposition A.1. Suppose that

1. $$f(v_t, \theta_*) \ = \ \mu_* + \sum_{s=-\infty}^{\infty} \Theta_j \varepsilon_{t-s} \tag{26}$$

   where $\sum_{s=-\infty}^{\infty} \|\Theta_j\| < \infty$, $\{\varepsilon_t\}$ is a sequence of iid random vectors with finite fourth moments.

2. $\omega(0) = 1$, $\omega(x) = 0\ \forall |x| > 1$, $\omega(x) = \omega(-x)$ and $|\omega(x)| \leq M$ for $\forall x$ and some $M > 0$, $\omega(x)$ is continuous in x, and its characteristic exponent, k, satisfies $k \geq 1$.

3. Let k denote the characteristic exponent of ω. Then $\sum_{j=-\infty}^{\infty} |j|^k \|\Gamma_j\| < \infty$ where Γ_s is the jth population autocovariance of $f(v_t, \theta_*)$.

4. $b_T \to \infty$ as $T \to \infty$ and $b_T^k / T \to 0$.

Then Assumption 7(i) holds.

Remark. The assumption of linear processes is not necessary but simplifies the exposition. More primitive conditions can be given in terms of stationarity, dependence and moments (Rosenblatt, 1959).

Proof of Proposition 1.

Suppose to the contrary that the model is misspecified, i.e., $\|\mu(\theta)\| > 0$ for all $\theta \in \Theta$. By assumption

$$\frac{1}{T} \sum_{t=1}^{T} f(v_t, \hat{\theta}_T) = 0. \tag{27}$$

Since Θ is compact by Assumption A.2 and $\mu(\theta)$ is continuous in θ by Assumptions A.5 and A.6, the Weierstrass theorem implies that there is $\delta > 0$ such that

$$\delta \leq \|\mu(\theta)\| \tag{28}$$

for all $\theta \in \Theta$.

$$\left\| \frac{1}{T} \sum_{t=1}^{T} f(v_t, \hat{\theta}_T) - \mu(\hat{\theta}_T) \right\| \leq \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^{T} f(v_t, \theta) - \mu(\theta) \right\|. \tag{29}$$

It follows from (27), (28) and (29) that

$$\delta \leq \|\mu(\hat{\theta}_T)\| \leq \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^{T} f(v_t, \theta) - \mu(\theta) \right\|. \tag{30}$$

However, the right-hand side of (30) is $o_p(1)$ by Assumption A.4. A contradiction.

Proofs for the results in Section 3

Note: we suppress the dependence of $\hat{\theta}_T$ and θ_* upon W_T and W respectively for notational brevity.

We use the following results in the proofs of Theorems 1, 2 and 3. Under Assumptions 1, 2, 4–6 and Assumptions A.1–6, it follows from standard arguments (*e.g.* Newey and McFadden's (1994) Theorem 2.1 and Wooldridge's (1994) Theorem 7.1) that

$$\hat{\theta}_T \xrightarrow{p} \theta_*. \tag{31}$$

where θ_* is defined by Assumption 5 for the appropriate choice of W.

It follows from Assumptions 1, 4–6, Assumptions A.1–9, and equation (31) that

$$I_p - H_{0,T} M_T \xrightarrow{p} I_p + (G_*' W G_*)^{-1} (\mu_*' W \otimes I_p) G_*^{(2)}) = (G_*' W G_*)^{-1} H_*.$$

Thus, by Assumption 7, we have

$$(I_p - H_{0,T} M_T)^{-1} H_{0,T} \xrightarrow{p} -H_*^{-1}. \tag{32}$$

Proof of Theorem 1. Put $c_T = T^{1/2}$. It follows from Assumption 1, 2, 4–5, Assumptions A.1–A.3, A.5–A.6, A.8–A.9, the conditions of the theorem and $W_T = W$ that

$$H_{1,T} + H_{2,T}(2) + H_{2,T}(3) = H_{1,T} + H_{2,T}(2) + o_p(1) \xrightarrow{d} N(0, G_*' W \Omega_{11} W G_* + G_*' W \Omega_{12} + \Omega_{21} W G_* + \Omega_{22}). \tag{33}$$

Therefore, the desired conclusion follows from (32) and (33).

Proof of Theorem 2. Similarly, with $c_T = T^{1/2}$,

$$H_{1,T} + H_{2,T}(2) + H_{2,T}(3) \xrightarrow{d} N(0, \Omega_*), \tag{34}$$

where

$$\begin{aligned}\Omega_* &= G_*' W \Omega_{11} W G_* + \Omega_{22} + G_*' \Omega_{33} G_* + G_*' W \Omega_{12} \\ &\quad + G_*' W \Omega_{13} G_* + \Omega_{21} W G_* + G_*' \Omega_{31} W G_* + \Omega_{23} G_* + G_*' \Omega_{32}.\end{aligned}$$

Thus, the desired conclusion follows from (32) and (34).

Proof of Theorem 3. Put $c_T = (T/b_T)^{1/2}$. Since $(1/(b_T T^{1/2})) \sum_{t=1}^T [f(v_t, \hat{\theta}_T) - \mu_*] = o_p(1)$, $(T/b_T)^{1/2}(W_T - W)$ becomes dominant in the limiting expression. If we show that

$$\left(\frac{T}{b_T}\right)^{\frac{1}{2}} (\hat{V}_T - \bar{V}_T) = o_p(1), \tag{35}$$

then the first result will follow from Assumption 7 and the delta method. Let

$$\tilde{V}_T = \tilde{\Gamma}_0(\theta_*) + \sum_{i=1}^{T-1} \omega(i/b_T)(\tilde{\Gamma}_i(\theta_*) + \tilde{\Gamma}_i(\theta_*)'),$$

where

$$\tilde{\Gamma}_i(\theta) = \frac{1}{T} \sum_{t=i+1}^{T} (f(v_t, \theta) - g_T(\theta))(f(v_{t-i}, \theta) - g_T(\theta))'.$$

First, we shall show that $(T/b_T)^{1/2}(\hat{V}_T - \tilde{V}_T) = o_p(1)$, and next we shall show that $(T/b_T)^{1/2}(\tilde{V}_T - \bar{V}_T) = o_p(1)$. The (j, k) element of the $q \times q$ matrix

$$\hat{V}_T - \tilde{V}_T = \tilde{\Gamma}_0(\hat{\theta}_T(1)) - \tilde{\Gamma}_0(\theta_*) + \sum_{i=1}^{T-1} \omega(i/b_T)(\tilde{\Gamma}_i(\hat{\theta}_T(1)) + \tilde{\Gamma}_i(\hat{\theta}_T(1))' - \tilde{\Gamma}_i(\theta_*) - \tilde{\Gamma}_i(\theta_*)')$$

is

$$\begin{aligned}\hat{V}_{T,jk} - \tilde{V}_{T,jk} &= \tilde{\Gamma}_{0,jk}(\hat{\theta}_T) - \tilde{\Gamma}_{0,jk}(\theta_*) \\ &\quad + \sum_{i=1}^{T-1} \omega(i/b_T)(\tilde{\Gamma}_{i,jk}(\hat{\theta}_T(1)) + \tilde{\Gamma}_{i,kj}(\hat{\theta}_T(1)) - \tilde{\Gamma}_{i,jk}(\theta_*) - \tilde{\Gamma}_{i,kj}(\theta_*)) \\ &= \left\{ \frac{\partial}{\partial \theta} \tilde{\Gamma}_{0,jk}(\bar{\theta}_T) + \sum_{i=1}^{T-1} \omega(i/b_T) \left(\frac{\partial}{\partial \theta} \tilde{\Gamma}_{i,jk}(\bar{\theta}_T) + \frac{\partial}{\partial \theta} \tilde{\Gamma}_{i,kj}(\hat{\theta}_T) \right) \right\}' (\hat{\theta}_T(1) - \theta_*),\end{aligned}$$

where $\bar{\theta}_T$ is a point between $\hat{\theta}_T(1)$ and θ_* and $\tilde{\Gamma}_{i,jk}(\theta)$ is the (j,k) element of $\tilde{\Gamma}_i(\theta)$. Notice that

$$\begin{aligned}\frac{\partial}{\partial\theta}\tilde{\Gamma}_{jk}(\hat{\theta}_T) &= \frac{1}{T}\sum_{t=1}^{T}(f_{j,\theta}(v_t,\bar{\theta}_T)-g_{j,T,\theta}(\bar{\theta}_T))(f_k(v_{t-i},\bar{\theta}_T)-g_{k,T}(\bar{\theta}_T)) \\ &\quad +\frac{1}{T}\sum_{t=1}^{T}(f_j(v_t,\bar{\theta}_T)-g_{j,T}(\bar{\theta}_T))(f_{k,\theta}(v_{t-i},\bar{\theta}_T)-g_{k,T,\theta}(\bar{\theta}_T)),\end{aligned}$$

where $f_{j,\theta}(v_t,\theta)=(\partial/\partial\theta)f_j(v_t,\theta)$ and $g_{j,T,\theta}(\theta)=(1/T)\sum_{t=1}^{T}f_{j,\theta}(v_t,\theta)$, and so

$$\frac{\partial}{\partial\theta}\tilde{\Gamma}_{0,jk}(\bar{\theta}_T)+\sum_{i=1}^{T-1}\omega(i/b_T)\left(\frac{\partial}{\partial\theta}\tilde{\Gamma}_{i,jk}(\bar{\theta}_T)+\frac{\partial}{\partial\theta}\tilde{\Gamma}_{i,kj}(\bar{\theta}_T)\right)$$

can be viewed as another HACC estimator of zero-mean process $\{f_\theta(v_t,\bar{\theta}_T)-g_{T,\theta}(\bar{\theta}_T)\}$ and thus is well-behaved under our assumptions. Since $\hat{\theta}_T(1)-\theta_*=O_p(T^{-1/2})$,

$$\hat{V}_{T,jk}-\tilde{V}_{T,jk}=O_p(T^{-1/2}),$$

and thus

$$\left(\frac{T}{b_T}\right)^{\frac{1}{2}}(\hat{V}_T-\tilde{V}_T)=O_p\left(\frac{1}{b_T}\right)=o_p(1). \tag{36}$$

The (j,k) element of $\tilde{\Gamma}_i(\theta_*)$ is

$$\begin{aligned}\tilde{\Gamma}_{i,jk}(\theta_*) &= \frac{1}{T}\sum_{t=i+1}^{T}(f_j(v_t,\theta_*)-g_{T,j}(\theta_*))(f_k(v_{t-i},\theta_*)-g_{T,k}(\theta_*)) \\ &= \frac{1}{T}\sum_{t=i+1}^{T}(f_j(v_t,\theta_*)-\mu_{*,j}+\mu_{*,j}-g_{T,j}(\theta_*))(f_k(v_{t-i},\theta_*)-\mu_{*,k}+\mu_{*,k}-g_{T,k}(\theta_*)) \\ &= \frac{1}{T}\sum_{t=i+1}^{T}(f_j(v_t,\theta_*)-\mu_{*,j})(f_k(v_t,\theta_*)-\mu_{*,k}) \qquad (37) \\ &\quad +(\mu_{*,j}-g_{T,j}(\theta_*))\frac{1}{T}\sum_{t=i+1}^{T}(f_k(v_{t-i},\theta_*)-\mu_{*,k}) \qquad (38) \\ &\quad +(\mu_{*,k}-g_{T,k}(\theta_*))\frac{1}{T}\sum_{t=i+1}^{T}(f_j(v_t,\theta_*)-\mu_{*,j}) \qquad (39) \\ &\quad +\frac{T-i}{T}(\mu_{*,k}-g_{T,k}(\theta_*))(\mu_{*,j}-g_{T,j}(\theta_*)). \qquad (40)\end{aligned}$$

The first term on the right-hand side (37) corresponds to the (j,k) element of $\bar{\Gamma}_i$. The second and third terms (38) and (39) are $O_p(T^{-1})$ uniform in i by assumption. The fourth term (40) is $O_p(T^{-1})$. By the definition of $\tilde{V}_T$ and $\bar{V}_T$, it follows that

$$\tilde{V}_T = \bar{V}_T + O_p\left(\frac{b_T}{T}\right)$$

and thus

$$\left(\frac{T}{b_T}\right)^{1/2}(\tilde{V}_T - \bar{V}_T) = O_p\left(\left(\frac{b_T}{T}\right)^{\frac{1}{2}}\right) = o_p(1). \tag{41}$$

Therefore (35) follows from (36) and (41). Since

$$\begin{aligned} d\text{vec}(G'_{**}V^{-1}\mu_{**}) &= -\text{vec}(G'_{**}V^{-1}dVV^{-1}\mu_{**}) \\ &= -(\mu'_{**}V^{-1} \otimes G'_{**}V^{-1})\text{vec}(dV) \end{aligned} \tag{42}$$

it follows from Theorem 3 and Theorem 2 in Magnus and Neudecker (1991)[p.151 and p.30, respectively] that

$$\frac{\partial G'_{**}V^{-1}\mu_{**}}{\partial vec(V)} = -(\mu'_{**}V^{-1} \otimes G'_{**}V^{-1}). \tag{43}$$

The first result follows from (35), Assumption 7 and the delta method. If $T^{1/2}/b_T^{1/2+k} \to \infty$ in the second case, then it follows from the above argument that

$$b_T^k(\hat{V}_T - \bar{V}_T) = o_p(1). \tag{44}$$

The second result follows from (44) and Assumption 7.

Proof of Theorem 4. Part (i) follows directly from Assumption 8 and standard proofs of convergence in probability (*e.g.* Newey and McFadden's (1994) Theorem 2.1 and Wooldridge's (1994) Theorem 7.1). Now consider part (ii). For simplicity set $\theta_* = \theta_*(1)$. Notice that $S_*\mu_* = 0$ implies that $p\lim_{T\to\infty} M_T = 0$. If this restriction and $c_T = b_T$ are substituted into (9) then we obtain

$$b_T(\hat{\theta}_T - \theta_*) \; = \; -\left[G_T(\hat{\theta}_T)'\hat{S}_T^{-1}G_T(\hat{\theta}_T, \theta_*, \lambda_T)\right]^{-1}G_T(\hat{\theta}_T)'\hat{S}_T^{-1}b_T g_T(\theta_*) \; + \; o_p(1) \tag{45}$$

Under the conditions of the theorem, it follows that

$$[G_T(\hat{\theta}'_T)\hat{S}_T^{-1}G_T(\hat{\theta}_T, \theta_*, \lambda_T)]^{-1}G_T(\hat{\theta}_T)' \xrightarrow{p} [G'_*S_*G_*]^{-1}G'_* \; = \; O(1) \tag{46}$$

where $G_* = E[\partial f(v_t, \theta_*)/\partial\theta']$. Notice that Assumption 8 implies $G_*' S_* G_*$ is full rank and so nonsingular. Now consider $h_T = \hat{S}_T^{-1} b_T g_T(\theta_*)$. It is convenient to rewrite h_T as

$$h_T \;=\; \hat{S}_T^{-1}(b_T/T^{1/2})\left[T^{-1/2}\sum_{t=1}^{T}(f(v_t,\theta_*) \;-\; \mu_*) \;+\; T^{1/2}\mu_*\right] \tag{47}$$

Under our conditions $T^{-1/2}\sum_{t=1}^{T}(f(v_t,\theta_*) - \mu_*) = O_p(1)$ and so the first term on the right hand side of (47) is $o_p(1)$. Now consider $b_T\hat{S}_T^{-1}\mu_*$. It is convenient to rewrite this term as,

$$b_T\hat{S}_T^{-1}\mu_* \;=\; b_T S_T^{-1}\mu_* \;+\; b_T[\hat{S}_T^{-1} \;-\; S_T^{-1}]\mu_* \tag{48}$$

where $S_T = V + B_T\mu_*\mu_*'$ and $B_T = \sum_{-T+1}^{T-1}\omega(i/b_T)$. From Hall (2000)[Theorem 1] it follows that

$$b_T S_T^{-1}\mu_* \;=\; \frac{b_T}{1 + B_T\mu_*' V^{-1}\mu_*}V^{-1}\mu_* \tag{49}$$

Also by similar arguments to the proof of Hall's (2000)[Theorem 2], it follows that $b_T[\hat{S}_T^{-1} - S_T^{-1}]\mu_* = o_p(1)$. Therefore, if we combine this result with (45)–(49) then it follows that

$$b_T(\hat{\theta}_T - \theta_*) \;=\; -[G_*' S_* G_*]^{-1} G_*' \frac{b_T}{1 + B_T\mu_*' V^{-1}\mu_*}V^{-1}\mu_* \;+\; o_p(1) \tag{50}$$

and so

$$b_T(\hat{\theta}_T - \theta_*) \;\xrightarrow{p}\; -\frac{\beta}{\mu_*' V^{-1}\mu_*}[G_*' S_* G_*]^{-1} G_*' V^{-1}\mu_* \;=\; C_*$$

where $\beta = \lim_{T\to\infty} b_T/B_T > 0$ by Assumption 6(ii).

Finally consider part (iii). Clearly if $G_*' V^{-1}\mu_* = 0$ then $C_* = 0$. In this case, $\hat{\theta}_T$ converges faster than b_T. Consider $T^{1/2}(\hat{\theta}_T - \theta_*)$. If we repeat the same sequence of steps as in the proof of part (ii), then it can be shown that

$$\begin{aligned} T^{1/2}(\hat{\theta}_T - \theta_*) &= -[G_*' S_* G_*]^{-1} G_*' S_* T^{-1/2}\sum_{t=1}^{T}(f(v_t,\theta_*) \;-\; \mu_*) \\ &= -\,[G_*' S_* G_*]^{-1} G_*' T^{1/2}(\hat{S}_T^{-1} - S_T^{-1})\mu_* \;+\; o_p(1) \end{aligned} \tag{51}$$

Under our assumptions, $[G_*' S_* G_*]^{-1} G_*' = O_p(1)$. Therefore we focus on $T^{1/2}(\hat{S}_T^{-1} - S_T^{-1})\mu_*$. It is convenient to write

$$T^{1/2}(\hat{S}_T^{-1} - S_T^{-1})\mu_* \;=\; \hat{S}_T^{-1}T^{1/2}(S_T - \hat{S}_T)S_T^{-1}\mu_* \tag{52}$$

and then consider the terms on the right hand side of (52) in turn. From Hall (2000)[Theorem 1] we have $\hat{S}_T^{-1} = O_p(1)$, and from (49) $S_T^{-1}\mu_* = O(b_T^{-1})$. Therefore $T^{1/2}(\hat{S}_T^{-1} - S_T^{-1})\mu_* = O_p(1)$, and hence $T^{1/2}(\hat{\theta}_T - \theta_*) = O_p(1)$, if $S_T - \hat{S}_T = O_p(b_T/T^{1/2})$. The latter is established in the following lemma.[12]

Lemma A.1

Under the Assumptions of Theorem 3, $S_T - \hat{S}_T = O_p(b_T/T^{1/2})$.

Proof: First notice that

$$\hat{\Gamma}_0 = \tilde{\Gamma}_0 + g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))'$$

and

$$\begin{aligned}
\hat{\Gamma}_i &= \tilde{\Gamma}_i + \frac{1}{T}\sum_{t=i+1}^{T} f(v_t,\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' + g_T(\hat{\theta}_T(1))\frac{1}{T}\sum_{t=i+1}^{T} f(v_{t-i},\hat{\theta}_T(1))' - \frac{T-i}{T}g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' \\
&= \tilde{\Gamma}_i + \frac{1}{T}\sum_{t=1}^{T} f(v_t,\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' + g_T(\hat{\theta}_T(1))\frac{1}{T}\sum_{t=1}^{T} f(v_t,\hat{\theta}_T(1))' - g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' \\
&\quad -\frac{1}{T}\sum_{t=1}^{i} f(v_t,\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' - g_T(\hat{\theta}_T(1))\frac{1}{T}\sum_{t=T-i+1}^{T} f(v_t,\hat{\theta}_T(1))' + \frac{i}{T}g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' \\
&= \tilde{\Gamma}_i + g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' \\
&\quad -\frac{1}{T}\sum_{t=1}^{i} f(v_t,\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' - g_T(\hat{\theta}_T(1))\frac{1}{T}\sum_{t=T-i+1}^{T} f(v_t,\hat{\theta}_T(1))' + \frac{i}{T}g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))',
\end{aligned}$$

for $i > 0$. Therefore, it follows that

$$\begin{aligned}
\hat{S}_T &= \hat{V}_T + \sum_{i=-T+1}^{T-1} \omega(i/b_T)g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' \\
&\quad +2\sum_{i=1}^{T-1}\omega(i/b_T)\{-\frac{1}{T}\sum_{t=1}^{i} f(v_t,\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' - g_T(\hat{\theta}_T(1))\frac{1}{T}\sum_{t=T-i+1}^{T} f(v_t,\hat{\theta}_T(1))' \\
&\quad +\frac{i}{T}g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))'\}.
\end{aligned}$$

If this result and $S_T = V + B_T\mu_*\mu_*^{'}$ are substituted into $\hat{S}_T - S_T$ then we obtain

$$\hat{S}_T - S_T \;=\; D_{1,T} \;+\; D_{2,T} \;+\; D_{3,T} \tag{53}$$

[12]This result is separated out because it is of interest in its own right.

where

$$\begin{aligned}
D_{1,T} &= \hat{V}_T \;-\; V \\
D_{2,T} &= \sum_{i=-T+1}^{T-1} \omega(i/b_T)\left[g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' \;-\; \mu_*\mu_*^{'}\right] \\
D_{3,T} &= 2\sum_{i=1}^{T-1} \omega(i/b_T)\{-\frac{1}{T}\sum_{t=1}^{i} f(v_t,\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' - g_T(\hat{\theta}_T(1))\frac{1}{T}\sum_{t=T-i+1}^{T} f(v_t,\hat{\theta}_T(1))' \\
&\quad +\frac{i}{T}g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))'\}
\end{aligned}$$

By Assumption 7 (i) and the proof of theorem 3, it follows that $D_{1,T} = O_p(\sqrt{b_T/T})$. By Assumption 6(ii) $\sum_{i=-T+1}^{T-1} \omega(i/b_T) = O(b_T)$, and under our assumptions

$$T^{1/2}\left[g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' - \mu_*\mu_*^{'}\right] = T^{-1/2}\sum_{t=1}^{T}[f(v_t,\hat{\theta}_T(1)) - \mu_*]\mu_*^{'} \;+\; o_p(1) \;=\; O_p(1)$$

and so $D_{2,T} = O_p(b_T/T^{1/2})$. So now consider $D_{3,T}$. First notice that

$$D_{3,T} = 2\sum_{i=1}^{T-1} \omega(i/b_T)\left[E_{1,T}(i) + E_{2,T}(i) + E_{3,T}(i)\right],$$

where

$$\begin{aligned}
E_{1,T}(i) &= T^{-1}\sum_{t=1}^{i} f(v_t,\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' \;-\; \frac{i}{T}\mu_*\mu_*^{'} \\
E_{2,T}(i) &= -g_T(\hat{\theta}_T(1))\frac{1}{T}\sum_{t=T-i+1}^{T} f(v_t,\hat{\theta}_T(1))' - \frac{i}{T}\mu_*\mu_*^{'} \\
E_{3,T}(i) &= \frac{i}{T}g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))'
\end{aligned}$$

Now, with some rearrangement, we have

$$T^{1/2}E_{1,T}(i) \;=\; T^{-1/2}\sum_{t=1}^{i}[f(v_t,\hat{\theta}_T(1)) - \mu_*]\mu_*^{'} \;+\; o_p(1)$$

and so under our assumptions $T^{1/2}E_{1,T}(i) = O_p(1)$. A similar analysis implies $T^{1/2}E_{2,T}(i) = O_p(1)$. Therefore, from Assumption 8(ii), $2\sum_{i=1}^{T-1} \omega(i/b_T)\left[E_{1,T}(i) + E_{2,T}(i)\right] = O_p(b_T/T^{1/2})$. Finally, consider $2\sum_{i=1}^{T-1} \omega(i/b_T)E_{3,T}(i) = 2E_{3,T}$. Under our assumptions, we have $g_T(\hat{\theta}_T(1))g_T(\hat{\theta}_T(1))' =$

$O_p(1)$ and

$$\begin{aligned}
\sum_{i=1}^{T-1} \frac{i}{T}\omega\left(\frac{i}{b_T}\right) &= \int_{1/T}^{1} [Tr]\omega\left(\frac{[Tr]}{b_T}\right) dr \\
&= \frac{b_T^2}{T}\int_{1/b_T}^{T/b_T} x\omega(x)dx \\
&= O\left(\frac{b_T^2}{T}\right).
\end{aligned}$$

Therefore $E_{3,T} = O_p(b_T^2/T)$. Combining these results we have $D_{3,T} = O_p(b_T/T^{1/2})$, and so

$$\begin{aligned}
\hat{S}_T - S_T &= D_{1,T} + D_{2,T} + D_{3,T} \\
&= O_p(\sqrt{b_T/T}) + O_p(b_T/T^{1/2}) + O_p(b_T/T^{1/2}) = O_p(b_T/T^{1/2})
\end{aligned}$$

Proofs of the results in Section 4

Note: we suppress the dependence of $\hat{\theta}_T$ and θ_* upon W_T and W respectively for notational brevity.

We shall use the following results in the proofs of Theorems 5, 6 and 7. By the standard argument, one can show that the constrained GMM estimator is consistent under the null hypothesis. Define a Lagrangian for the constrained GMM estimation by

$$L(\theta, \lambda) = -g_T(\theta)'W_T g_T(\theta) - \lambda' a(\theta)$$

where λ is a r-dimensional vector of Lagrange multipliers. The constrained GMM estimator $\bar{\theta}_T$ and the Lagrange multiplier $\bar{\theta}_T$ satisfy the following first-order necessary condition:

$$\begin{bmatrix} -G_T(\bar{\theta}_T)'W_T g_T(\bar{\theta}_T) - \bar{A}_T'\bar{\lambda}_T \\ -a(\bar{\theta}_T) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where $\bar{A}_T = \partial a(\bar{\theta}_T)/\partial\theta'$. Define

$$\bar{H}_{1,T}, \bar{H}_{2,T}, \bar{H}_{2,T}(1), \bar{H}_{2,T}(2), \bar{H}_{2,T}(3), \bar{H}_{3,T}, \bar{M}_T$$

by

$$H_{1,T}, H_{2,T}, H_{2,T}(1), H_{2,T}(2), H_{2,T}(3), H_{3,T}, M_T$$

with $\hat{\theta}_T$ replaced by $\bar{\theta}_T$, respectively. Applying the Mean Value Theorem to the first-order condition yields

$$\left[\begin{array}{c} -\bar{H}_{1,T} - \bar{H}_{2,T}(2) - \bar{H}_{2,T}(3) \\ 0 \end{array}\right] - \left[\begin{array}{cc} \bar{M}_T - \bar{H}_{0,T}^{-1} & \frac{\partial a(\tilde{\theta}_T)}{\partial \theta} \\ \frac{\partial a(\tilde{\theta}_T)}{\partial \theta'} & 0 \end{array}\right] \left[\begin{array}{c} \sqrt{T}(\bar{\theta}_T - \theta_*) \\ \sqrt{T}\bar{\lambda}_T \end{array}\right] = o_p(1),$$

where $\tilde{\theta}_T$ is a point between $\bar{\theta}_T$ and θ_*. By the uniform convergence,

$$\left[\begin{array}{c} -\bar{H}_{1,T} - \bar{H}_{2,T}(2) - \bar{H}_{2,T}(3) \\ 0 \end{array}\right] - \left[\begin{array}{cc} H_* & A_*' \\ A_* & 0 \end{array}\right] \left[\begin{array}{c} \sqrt{T}(\bar{\theta}_T - \theta_*) \\ \sqrt{T}\bar{\lambda}_T \end{array}\right] = o_p(1).$$

Thus,

$$\left[\begin{array}{c} \sqrt{T}(\bar{\theta}_T - \theta_*) \\ \sqrt{T}\bar{\lambda}_T \end{array}\right] = \left[\begin{array}{cc} H_* & A_*' \\ A_* & 0 \end{array}\right]^{-1} \left[\begin{array}{c} -\bar{H}_{1,T} - \bar{H}_{2,T}(2) - \bar{H}_{2,T}(3) \\ 0 \end{array}\right] + o_p(1).$$

Solving for $\sqrt{T}(\bar{\theta}_T - \theta_*)$ yields

$$\begin{aligned} \sqrt{T}(\bar{\theta}_T - \theta_*) &= -[H_*^{-1} - H_*^{-1}A_*'(A_*H_*^{-1}A_*')^{-1}A_*H_*^{-1}][\bar{H}_{1,T} + \bar{H}_{2,T}(2) + \bar{H}_{2,T}(3)] + o_p(1) \\ &= -M_*[\bar{H}_{1,T} + \bar{H}_{2,T}(2) + \bar{H}_{2,T}(3)] + o_p(1). \qquad (54) \end{aligned}$$

Combining the expression for $\sqrt{T}(\hat{\theta}_T - \theta_*)$ and that for $\sqrt{T}(\bar{\theta} - \theta_*)$ gives

$$\sqrt{T}(\hat{\theta}_T - \bar{\theta}_T) = -H_*^{-1}A_*'(A_*H_*^{-1}A_*')^{-1}A_*H_*^{-1}[H_{1,T} + H_{2,T}(2) + \bar{H}_{2,T}(3)] + o_p(1). \quad (55)$$

Proof of Theorem 5.

First, we consider the Wald test. Since

$$Wald_T = \sqrt{T}(\hat{\theta}_T - \theta_*)'\hat{A}_T'\{\hat{A}_T[G_T(\hat{\theta}_T)'W_TG_T(\hat{\theta}_T)]^{-1}\hat{A}_T)'\}^{-1}\hat{A}_T\sqrt{T}(\hat{\theta}_T - \theta_*),$$

$$\hat{A}_T'\{\hat{A}_T[G_T(\hat{\theta}_T)'W_TG_T(\hat{\theta}_T)]^{-1}\hat{A}_T'\}^{-1}\hat{A}_T \xrightarrow{p} A_*'(A_*(G_*'WG_*)^{-1}A_*')^{-1}A_*, \qquad (56)$$

and

$$\sqrt{T}(\hat{\theta}_T - \theta_*) \xrightarrow{d} N(0, \Sigma_2)$$

by Theorem 2, the desired conclusion immediately follows.

Second, we consider the Lagrange multiplier test. It follows from the Mean Value Theorem and the first-order condition in population that

$$
\begin{aligned}
& \bar{G}_T' W_T \sqrt{T} g_T(\bar{\theta}_T) \\
= \; & \bar{G}_T' W_T [\sqrt{T}(g_T(\theta_*) - \mu_*) + \sqrt{T}\mu_* + G_T(\tilde{\theta}_T)\sqrt{T}(\bar{\theta}_T - \theta_*)] + o_p(1) \\
= \; & G_*' W \sqrt{T}(g_T(\theta_*) - \mu_*) + \sqrt{T}(\bar{G}_T - G_T(\theta_*))' W_T \mu_* + \sqrt{T}(G_T(\theta_*) - G_*)' W_T \mu_* \\
& + G_*' \sqrt{T}(W_T - W)\mu_* + G_*' W G_* \sqrt{T}(\bar{\theta}_T - \theta_*) + o_p(1), \qquad (57)
\end{aligned}
$$

where $\tilde{\theta}_T$ is a point between $\bar{\theta}_T$ and θ_*. As in the derivation of (9), we obtain

$$
\sqrt{T}(\bar{G}_T - G_T(\theta_*))' W_T \mu_* = [(\mu_*' W \otimes I_p) G_*^{(2)}] \sqrt{T}(\bar{\theta}_T - \theta_*) + o_p(1). \qquad (58)
$$

Substituting (58) into (57) produces

$$
\begin{aligned}
& \bar{G}_T' W_T \sqrt{T} g_T(\bar{\theta}_T) \\
= \; & G_*' W \sqrt{T}(g_T(\theta_*) - \mu_*) + \sqrt{T}(G_T(\theta_*) - G_*)' W_T \mu_* + G_*' \sqrt{T}(W_T - W)\mu_* \\
& + [G_*' W G_* + (\mu_*' W \otimes I_p) G_*^{(2)}] \sqrt{T}(\bar{\theta}_T - \theta_*) \\
= \; & G_*' W \sqrt{T}(g_T(\theta_*) - \mu_*) + \sqrt{T}(G_T(\theta_*) - G_*)' W_T \mu_* + G_*' \sqrt{T}(W_T - W)\mu_* \\
& + H_* \sqrt{T}(\bar{\theta}_T - \theta_*). \qquad (59)
\end{aligned}
$$

It follows from (54) that

$$
\sqrt{T}(\bar{\theta}_T - \theta_*) = -M_*[\sqrt{T}(g_T(\theta_*) - \mu_*) + \sqrt{T}(G_T(\theta_*) - G_*)' W \mu_* + G_*'(W_T - W)\mu_*]. \qquad (60)
$$

Substituting (60) into (59) yields

$$
\begin{aligned}
\bar{G}_T' W_T \sqrt{T} g_T(\bar{\theta}_T) \; = \; & (G_*' W - H_* M_*) \sqrt{T}(g_T(\theta_*) - \mu_*) + (I_p - H_* M_*) \sqrt{T}(G_T(\theta_*) - G_*)' W_T \mu_* \\
& + (I_p - H_* M_*) \sqrt{T} G_*' (W_T - W)\mu_*. \qquad (61)
\end{aligned}
$$

The desired conclusion follows from (12), (61) and the definition of the LM statistic.

Lastly, we consider the Likelihood Ratio-like test. Since

$$
\begin{aligned}
LR_T \; = \; & \sqrt{T}(\bar{\theta}_T - \hat{\theta}_T)' \hat{G}_T' W_T \sqrt{T} g_T(\hat{\theta}_T) + \sqrt{T}(\bar{\theta}_T - \hat{\theta}_T)' \hat{G}_T' W_T \hat{G}_T \sqrt{T}(\bar{\theta}_T - \hat{\theta}_T) + o_p(1) \\
= \; & \sqrt{T}(\bar{\theta}_T - \hat{\theta}_T)' \hat{G}_T' W_T \hat{G}_T \sqrt{T}(\bar{\theta}_T - \hat{\theta}_T) + o_p(1)
\end{aligned}
$$

where the second equality follows from the first-order condition for the unconstrained GMM estimator, the desired conclusion follows from (55).

Proof of Theorem 6.

The proof of Theorem 6 is analogous to that of Theorem 5 except that Theorem 3 is applied and thus is omitted. If $T^{1/2}/b_T^{1/2+k} \to \infty$, we have

$$\begin{aligned}
\frac{b_T^{2k}}{T} Wald_T &\xrightarrow{p} \mu_{**}' C G_{**} H_{**}^{-1} A_{**}' (A_{**}(G_{**}' W G_{**})^{-1} A_{**}')^{-1} A_{**} H_{**}'^{-1} G_{**}' C \mu_{**} \\
&= \kappa_W, \\
\frac{b_T^{2k}}{T} LM_T &\xrightarrow{p} \mu_{**}' C G_{**} (I - H_{**} M_{**})' G_{**}' W G_{**} (I - H_{**} M_{**}) G_{**}' C \mu_{**} \\
&= \kappa_{LM}, \\
\frac{b_T^{2k}}{T} LR_T &\xrightarrow{p} \mu_{**}' C G_{**} H_{**}^{-1} A_{**}' (A_{**} H_{**}^{-1} A_{**}')^{-1} A_{**} H_{**}^{-1} G_{**}' W G_{**} \\
&\quad \times (H_{**}^{-1} A_{**}' (A_{**} H_{**}^{-1} A_{**}')^{-1} A_{**} H_{**}^{-1})' G_{**} C \mu_{**} \\
&= \kappa_{LR}.
\end{aligned}$$

Proof of Theorem 7.

It follows from (50) and (56) that

$$\frac{b_T^2}{T} Wald_T = C_*' A_*' (A_* (G_*' W G_*)^{-1} A_*')^{-1} A_* C_* + o_p(1).$$

Since

$$\bar{H}_{1,T} = G_T(\bar{\theta}_T)' W_T T^{-\frac{1}{2}} \sum_{t=1}^{T} [f(v_t, \theta_*) - \mu_*] = O_p(1),$$
$$\bar{H}_{2,T}(2) = T^{1/2} [G_T(\theta_* - G_*)' W_T) \mu_*] = O_p(1),$$

and

$$\begin{aligned}
\frac{b_T}{\sqrt{T}} \bar{H}_{2,T}(3) &= G_*' b_T (W_T - W) \mu_* \\
&= G_*' b_T (\hat{S}_T^{-1} - S_T^{-1} + S_T^{-1} - S_*) \mu_* \\
&= G_*' b_T (\hat{S}_T^{-1} - S_T^{-1}) \mu_* + G_*' b_T S_T^{-1} \mu_* \\
&= G_*' \hat{S}_T^{-1} (S_T - \hat{S}_T) b_T S_T^{-1} \mu_* + G_*' b_T S_T^{-1} \mu_* \\
&= \frac{\beta}{\mu_*' V^{-1} \mu_*} G_*' V^{-1} \mu_* + o_p(1),
\end{aligned}$$

where the latter follows from the arguments following equation (52), Lemma A.1 and $G_*' S_* \mu_* = 0$, it follows that

$$\frac{b_T}{\sqrt{T}}\{\bar{H}_{1,T} + \bar{H}_{2,T}(2) + \bar{H}_{2,T}(3)\} = \frac{\beta}{\mu_*' V^{-1}\mu_*} G_*' V^{-1}\mu_* + o_p(1). \tag{62}$$

By (54), (55) and (62), we obtain

$$\begin{aligned}
\frac{b_T}{\sqrt{T}}(\bar{\theta}_T - \theta_*) &= -\frac{\beta M_* G_*' V^{-1}\mu_*}{\mu_*' V^{-1}\mu_*} + o_p(1), && (63)\\
\frac{b_T}{\sqrt{T}}(\hat{\theta}_T - \bar{\theta}_T) &= -\frac{\beta}{\mu_*' V^{-1}\mu_*} H_*^{-1} A_*' (A_* H_*^{-1} A_*')^{-1} A_* H_*^{-1} G_*' V^{-1}\mu_* + o_p(1). && (64)
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{b_T^2}{T} LM_T &= \left(\frac{\beta}{\mu_*' V^{-1}\mu_*}\right)^2 \mu_*' V^{-1} G_* M_*' (G_*' S_* G_*) M_* G_*' V^{-1}\mu_* + o_p(1),\\
\frac{b_T^2}{T} LR_T &= \left(\frac{\beta}{\mu_*' V^{-1}\mu_*}\right)^2 \mu_*' V^{-1} G_*' H_*^{-1} A_*' (A_* H_*^{-1} A_*')^{-1} A_* H_*^{-1}\\
&\quad \times G_*' S_* G_* (H_*^{-1} A_*' (A_* H_*^{-1} A_*')^{-1} A_* H_*^{-1})' G_*' V^{-1}\mu_* + o_p(1).
\end{aligned}$$

References

Anderson, T. W. (1994). The Statistical Analysis of Time Series. Wiley, New York, NY, U. S. A.

Andrews, D. W. K. (1991). Heteroscedasticity and autocorrelation consistent covariance matrix estimation , Econometrica, 59: 817–858.

Brillinger, D. R. (1975). Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, Inc, New York.

Cochrane, J. H. (1996). A cross–sectional test of an investment asset pricing model , Journal of Political Economy, 104: 572–621.

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

Diebold, F. X., Ohanian, L. E., and Berkowitz, J. (1998). Dynamic equilibrium economies: a framework for comparing models and data , Review of Economic Studies, 65: 433–451.

Durlauf, S. N., and Maccini, L. J. (1995). Measuring noise in inventory models , Journal of Monetary Economics, 36: 65–89.

Epstein, L. G., and Zin, S. E. (1991). Substitution, risk aversion, and the temporal behaviour of consumption and asset returns: an empirical analysis , Journal of Political Economy, 99: 263–286.

Ferson, W. E., and Constantinides, G. M. (1991). Habit persistence and durability in aggregate consumption , Journal of Financial Economics, 29: 199–240.

Gallant, A. R. (1987). Nonlinear Statistical Models. Wiley, New York.

Gallant, A. R., and White, H. (1988). A Unified Theory of Estimation and Inference in Nonlinear Models. Basil Blackwell, Oxford.

Hall, A. R. (2000). Covariance matrix estimation and the power of the overidentifying restrictions test , Econometrica, 68: 1517–1528.

Hannan, E. J. (1970). Multiple Time Series. Wiley, New York.

Hansen, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes , Econometrica, 60: 967–972.

Hansen, L. P. (1982). Large sample properties of Generalized Method of Moments estimators , Econometrica, 50: 1029–1054.

Hansen, L. P., Heaton, J., and Luttmer, E. G. J. (1995). Econometric Evaluation of Asset Pricing Models , Review of Financial Studies, 8: 237–274.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators , Journal of Business and Economic Statistics, 14: 262–280.

Hansen, L. P., and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies , Journal of Political Economy, 99: 225–262.

——— (1997). Assessing specification errors in stochastic discounter factor models , Journal of Finance, 52: 557–590.

Hansen, L. P., and Sargent, T. J. (1993). Seasonality and approximation errors in rational expectations models , Journal of Econometrics, 55: 21–55.

Kitamura, Y. (1997). Comparing misspecified dynamic econometric models using nonparametric likelihood , Discussion paper, Department of Economics, University of Minnesota.

Maasoumi, E., and Phillips, P. C. B. (1982). On the behavior of inconsistent instrumental variable estimators , Journal of Econometrics, 19: 183–201.

Magnus, J. R., and Neudecker, H. (1991). Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York, NY.

Meghir, C., and Weber, G. (1996). Intertemporal nonseparability or borrowing restrictions? A disaggregate analysis using a U. S. consumption panel , Econometrica, 64: 1151–1181.

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.

Newey, W. K., and West, K. (1987a). Hypothesis testing with efficient method of moments testing , International Economic Review, 28: 777–787.

Newey, W. K., and West, K. D. (1987b). A simple positive semi–definite heteroscedasticity and autocorrelation consistent covariance matrix , Econometrica, 55: 703–708.

Rosenblatt, M. (1959). Statistical Analysis of Stochastic Processes with Stationary Residuals , in U. Grenander (ed.), Probability and Statistics, pp. 246–275. Wiley, New York.

Tyler, D. E. (1983). The asymptotic distribution of principal component roots under local alternatives to miltiple roots , Annals of Statistics, 11: 1232–1242.

Watson, M. W. (1993). Measures of fit for calibrated models , Journal of Political Economy, 101: 1011–1041.

White, H. (1982). Maximum likelihood estimation of misspecified models , Econometrica, 50: 1–25.

——— (1994). Estimation, Inference and Specification Analysis. Cambridge University Press, New York, NY, U. S. A.

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.