Econometrica, Vol. 68, No. 6 (November, 2000), 1517–1527

COVARIANCE MATRIX ESTIMATION AND THE POWER OF THE OVERIDENTIFYING RESTRICTIONS TEST

BY ALASTAIR R. HALL[1]

1. INTRODUCTION

GENERALIZED METHOD OF MOMENTS (GMM) (Hansen (1982)) has been widely applied to estimate the parameters of economic and statistical models based on time series data.[2] In many cases of interest, the parameter vector is overidentified and so inference is based on the two-step estimator proposed by Hansen (1982). The second step of this procedure requires the construction of a positive semidefinite consistent estimator of the long run covariance matrix of the sample moment. Unless the economic model implies a certain form for this covariance matrix, it is desirable to use an estimator that is consistent under the weakest possible conditions. This requirement has motivated the development of the class of heteroscedasticity and autocorrelation consistent covariance (HACC) matrix estimators that are now routinely used by practitioners in the calculation of the two-step GMM estimator. In applications, it is customary to assume the model is correctly specified during the estimation, and only to assess the specification after the second-step using a statistic such as the overidentifying restrictions test. One important consequence of this methodology is that the HACC estimator is calculated under the assumption that the model is correctly specified.

In this paper, we examine the implications of model misspecification for the HACC covariance matrix estimator and the overidentifying restrictions test. It is shown that the HACC estimator is asymptotically equivalent to the sum of two matrices: one of these matrices is nonsingular and $O(1)$; the other is of rank one and $O(b_T)$ where b_T is the bandwidth used in the HACC estimator. It is shown that this structure implies the inverse of the HACC estimator converges to a singular matrix as $b_T \to \infty$ with T, and this limiting matrix has rank $q-1$ where q is the dimension of the moment condition. It is shown that this limiting behavior translates into an overidentifying restrictions test that is $O_p(T/b_T)$. In contrast, it is shown that the overidentifying restrictions test is consistent and $O_p(T)$ if the covariance matrix estimator is consistent under both null and alternative. This can be achieved by constructing the HACC estimator using the sample moment in mean deviation form.[3]

[1]I am grateful for useful conversations about this work with Peter Burridge, Atsushi Inoue, and Ada Wossink, and for the comments from seminar participants at the Departments of Economics at Cornell University, North Carolina State University, Pennsylvania State University, and the University of Arizona. This paper has also greatly benefited from the comments of Don Andrews and three anonymous referees. Part of this work was undertaken while the author was a visiting research fellow at the Department of Economics at the University of Birmingham, UK, and this support is gratefully acknowledged.

[2]For example, see the citations in the review articles by Hall (1993) and Ogaki (1993).

[3]Since this research was undertaken, Don Andrews has drawn my attention to footnote 3 in his paper (Andrews (1999)) on moment selection in which it is recommended that variances be calculated using the data in mean deviation form. However, as he also pointed out to me, he provides neither a citation nor analysis to support the recommendation in his paper.

An overview of the paper is as follows. Section 2 presents a brief review of the two-step GMM estimator and HACC estimators. Section 3 examines the properties of these estimators and the overidentifying restrictions test when the model is misspecified. All proofs and certain other technical details are relegated to an Appendix.

2. TWO-STEP GMM ESTIMATION IN CORRECTLY SPECIFIED MODELS

Consider the case in which a model implies the $p \times 1$ vector of parameters, θ_0, and a set of observed variables, v_t, satisfy the $q \times 1$ population moment condition

$$E[f(v_t, \theta_0)] = 0. \tag{1}$$

The GMM estimator of θ_0 based on (1) is defined to be

$$\hat{\theta}_T = \text{Argmin}_{\theta \in \Theta} g_T(\theta)' W_T g_T(\theta) \tag{2}$$

where $g_T(\theta) = T^{-1}\sum_{t=1}^{T} f(v_t, \theta)$ and W_T is a positive semidefinite weighting matrix that converges in probability to a positive definite matrix of constants. If $q > p$, then the choice of W_T becomes important because it determines the asymptotic covariance matrix of $\hat{\theta}_T$. In these circumstances, Hansen (1982) proves that the optimal choice of weighting matrix converges in probability to S^{-1} where

$$S = \lim_{T \to \infty} \text{var}[T^{1/2} g_T(\theta_0)]. \tag{3}$$

This limit can be achieved by setting $W_T = \hat{S}_T^{-1}$ where $\hat{S}_T$ is a positive semidefinite consistent estimator of S. Since S typically depends on θ_0, the use of this optimal weighting matrix necessitates a two step estimation procedure. In this paper, we consider the case in which the first step estimator, $\hat{\theta}_T(1)$, is used to construct the following HACC estimator:

$$\hat{S}_T = \sum_{i=-T+1}^{T-1} \omega(i/b_T)\hat{\Gamma}_i \tag{4}$$

where

$$\hat{\Gamma}_i = T^{-1} \sum_{t=i+1}^{T} f(v_t, \hat{\theta}_T(1)) f(v_{t-i}, \hat{\theta}_T(1))' \quad \text{for } i \geq 0,$$

$$= T^{-1} \sum_{t=-i+1}^{T} f(v_{t+i}, \hat{\theta}_T(1)) f(v_t, \hat{\theta}_T(1))' \quad \text{for } i < 0,$$

and the kernel, $\omega(\cdot)$, and bandwidth, b_T, satisfy certain restrictions given below. The second step (or optimal) estimator, $\hat{\theta}_T(2)$, is then calculated using $W_T = \hat{S}_T^{-1}$ in (2).

Up to this point, the model is assumed to be correctly specified. It is only after the second step estimation that credence is given to the possibility of misspecification, and

the overidentifying restrictions test is applied to assess the validity of (1). This statistic is conveniently calculated as T times the second-step GMM minimand evaluated at $\hat{\theta}_T(2)$, namely

$$J_T = T g_T(\hat{\theta}_T(2))' \hat{S}_T^{-1} g_T(\hat{\theta}_T(2)). \tag{5}$$

Hansen (1982) shows that if (1) holds, then J_T converges to a χ^2_{q-p} distribution. This sequence of events has the important consequence that in many applications the HACC estimator used in the second step estimation, and hence the overidentifying restrictions test, is calculated under the assumption that (1) holds.

3. TWO STEP GMM ESTIMATION IN MISSPECIFIED MODELS

In this section, we consider the case where the original population moment condition in (1) is invalid. Certain assumptions are required to facilitate the analysis. However, for brevity, we only highlight in the text those assumptions that are relevant to the exposition and relegate the rest to the mathematical Appendix.

Following Hansen (1982), we impose the following condition on v_t.

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

If the original population moment condition in (1) is invalid, then there must be no value of θ at which the population moment condition is satisfied. So, if we define

$$E[f(v_t, \theta)] = \mu(\theta), \tag{6}$$

then model misspecification is captured by the following assumption.

ASSUMPTION 2: *$\mu: \Theta \to \Re^q$ such that $\| \mu(\theta) \| > 0$ for all $\theta \in \Theta$.*

To deduce the impact of this misspecification on the HACC estimator and the overidentifying restrictions test, it is necessary to begin with the first step estimator. Let W_T denote the weighting matrix for the first step estimation. For completeness, we formally impose the conditions described in the previous section.

ASSUMPTION 3: *W_T is a positive semidefinite matrix that converges in probability to the positive definite matrix of constants W.*

So, the first step estimator is defined as

$$\hat{\theta}_T(1) = \text{Argmin}_{\theta \in \Theta} g_T(\theta)' W_T g_T(\theta) \tag{7}$$

where $\Theta \subset \Re^p$. To establish that $\hat{\theta}_T(1)$ converges to a probability limit, it is necessary to impose an identification condition.[4]

[4]Assumption 4 is similar to the "identifiable uniqueness" condition that underpins Gallant and White's analysis of misspecified models; see Gallant and White (1988, p. 19).

ASSUMPTION 4: *There exists* $\theta_* \in \Theta$ *such that* $Q_0(\theta_*) < Q_0(\theta), \forall \theta \in \Theta \setminus \{\theta_*\}$, *where* $Q_0(\theta) = E[f(v_t, \theta)]' W E[f(v_t, \theta)]$.

Notice that in general θ_* depends on W, although we suppress this dependence for ease of notation. Once this identification condition is imposed, we can appeal to a combination of Newey and McFadden's (1994) Theorem 2.1 and Wooldridge's (1994) Theorem 7.1 to deduce the following result.[5]

LEMMA 1: *If Assumptions* 1–4, *and Assumptions A*.1–*A*.4 *in the Appendix hold*, *then* $\hat{\theta}_T(1) \xrightarrow{p} \theta_*$.

After the first step estimation, it is the behavior of the population moment condition at θ_* that becomes important. Consequently, we introduce the notation

$$E[f(v_t, \theta_*)] = \mu_*. \tag{8}$$

Notice that Assumption 2 implies $\mu_* \neq 0$.

We now turn to the impact on misspecification on both the HACC estimator and its inverse because it is the latter that appears in the overidentifying restrictions test. For this analysis, it is necessary to impose certain additional restrictions. First we limit attention to HACC estimators whose kernels and bandwidths satisfy the following properties.

ASSUMPTION 5: (*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*) $b_T = o(T^{1/2})$ *and* $b_T \to \infty$; (*iii*) $\int_{\Re} \omega(x) dx = c$ *where* $0 < c < \infty$.

It can be verified that Assumption 5 is satisfied by the Bartlett (Newey and West (1987)), Parzen (Gallant (1987)), or the quadratic spectral (Andrews (1991)) kernels.

It is also necessary to assume the following.

ASSUMPTION 6: $V = \lim_{T \to \infty} \text{var}[T^{-1/2} \sum_{t=1}^{T} (f(v_t, \theta_*) - \mu_*)]$ *is a positive definite matrix of constants*.

We can now appeal to Gallant and White's (1988) Theorem 6.8 to deduce the following lemma.[6]

LEMMA 2: *If Assumptions* 1–6 *and A*.1–*A*.8 *in the mathematical Appendix hold*, *then* $\hat{S}_T = V + M_T + o_p(1)$ *where* $M_T = B_T \mu_* \mu_*'$, $B_T = \sum_{i=-T+1}^{T-1} \omega(i/b_T)$, *and hence* $B_T/b_T = c + o(1)$ *for c defined in Assumption* 5(*iii*).

[5]Our identification condition is different from Wooldridge's but his proof is easily adapted to take account of this difference. Note that Wooldridge's result is based on Newey and McFadden's (1994) Theorem 2.1 and these authors employ an analogous identification context to Assumption 4 albeit in the context of correctly specified models.

[6]Strictly, Gallant and White (1988) only prove their result for $b_T = o(T^{1/4})$ and kernels of the form $\omega(x) = 0$ for $x > 1$. However, it is straightforward to extend the result to $b_T = o(T^{1/2})$ and kernels satisfying Assumption 5(i) using Andrews's (1991) Theorem 1.

Lemma 1 indicates that the large sample behavior of $\hat{S}_T$ is identical to the behavior of $V + M_T$. The matrix V is positive definite and $O(1)$; the matrix M_T has rank equal to one and increases at rate b_T. We now consider the implications of Lemma 2 for $\hat{S}_T^{-1}$.

THEOREM 1: *If Assumptions* 1–6 *and A*.1–*A*.8 *hold then*:
(i)

$$\hat{S}_T^{-1} \xrightarrow{p} S_* = V^{-1} - \frac{1}{\mu_*' V^{-1} \mu_*} V^{-1} \mu_* \mu_*' V^{-1};$$

(ii) S_* *is a positive semidefinite matrix with rank equal to* $q - 1$ *and a nullspace spanned by* μ_*.

We now consider how this behavior impacts the properties of the overidentifying restrictions test.

THEOREM 2: *If Assumptions* 1–6 *and A*.1–*A*.8 *hold, then* $J_T = O_p(T/b_T)$.

This result states that the overidentifying restrictions test can not increase faster than rate T/b_T. By itself, Theorem 2 does not imply J_T is a consistent test; however this is in fact the case.[7]

Inspection of the proof reveals that this dependence on b_T stems directly from the invalid assumption that the model is correctly specified. Therefore it seems intuitively reasonable that a more powerful test will be obtained if the HACC matrix is constructed in such a way that it is consistent regardless of whether or not the population moment condition is correct. This can be achieved by using the estimator

$$\hat{V}_T = \sum_{i=-T+1}^{T-1} \omega(i/b_T) \tilde{\Gamma}_i \tag{9}$$

where

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

For the rest of the paper, we consider the properties of the GMM estimator and overidentifying restrictions test when this "centered" version of the HACC estimator is used instead of the "uncentered" version, $\hat{S}_T$. Therefore we now consider the second step GMM estimator

$$\tilde{\theta}_T(2) = \text{Argmin}_{\theta \in \Theta} g_T(\theta)' \hat{V}_T^{-1} g_T(\theta) \tag{10}$$

[7]In an earlier version of this paper, it is shown that $(b_T/T)J_T \xrightarrow{p} a$ where a is a finite positive constant, and hence that the test is consistent. However, this analysis was dropped for brevity on the advice of the editor and referees.

and its associated overidentifying restrictions test

$$\tilde{J}_T = Tg_T(\tilde{\theta}_T(2))'\hat{V}_T^{-1}g_T(\tilde{\theta}_T(2)). \tag{11}$$

The following theorem establishes the properties of $\hat{V}_T$, $\tilde{\theta}_T(2)$, and $\tilde{J}_T$. However, before these results can be presented it is necessary to introduce the following identification condition.

ASSUMPTION 7: *There exists* $\theta_{**} \in \Theta$ *such that* $\tilde{Q}_0(\theta_{**}) < \tilde{Q}_0(\theta), \forall \theta \in \Theta \setminus \{\theta_{**}\}$ *and* $\tilde{Q}_0(\theta) = E[f(v_t, \theta)]'V^{-1}E[f(v_t, \theta)]$.

THEOREM 3: *If Assumptions* 1–7 *and A*.1–*A*.9 *hold, then*: (i) $\hat{V}_T \xrightarrow{p} V$; (ii) $\tilde{\theta}_T(2) \xrightarrow{p} \theta_{**}$; (*iii*) $T^{-1}\tilde{J}_T \xrightarrow{p} \epsilon$ *for some finite positive constant* ϵ.

Theorem 3(iii) implies that $\tilde{J}_T$ is consistent and increases at rate T, which, from Theorem 2, is faster than J_T. Therefore $\tilde{J}_T$ is more powerful than J_T in large samples.

To conclude this paper, we present the results from a simulation study designed to illustrate the extent to which our asymptotic results manifest themselves in finite samples.[8] Data were generated from the following model:

$$y_t = x_t + \gamma z_{1,t} + u_t,$$
$$x_t = z_{1,t} + z_{2,t} + e_t,$$

for $t = 1, 2 \ldots T$ where $(z_{1,t}, z_{2,t}, u_t, e_t)' \sim IN(0, \Sigma)$ and Σ is the symmetric matrix with *i-j*th element σ_{ij} whose nonzero upper triangular elements are given by $\sigma_{ii} = 1$ for $i = 1, 2 \ldots 4$, $\sigma_{12} = \sigma_{34} = 0.5$. GMM estimation of a scalar parameter θ is performed under the assumption that the following population moment condition holds:

$$E[f(v_t, \theta)] = E[z_t(y_t - x_t\theta)] = 0 \tag{12}$$

where $z_t = (z_{1,t}, z_{2,t})'$. Notice that if $\gamma = 0$, then (12) holds at $\theta = 1$; otherwise the model is misspecified. Simulation results reported in an earlier version of this paper indicate that both $\|g_T(\hat{\theta}_T(1))\|$ and $|\hat{\theta}_T(1) - 1|$ increase monotonically with γ over the range of values considered. On each replication, both J_T and $\tilde{J}_T$ are calculated using a HACC estimator with the Bartlett kernel and a bandwidth chosen via the data based method proposed by Newey and West (1994). To present this method, it is notationally convenient to consider the generic case in which it is desired to estimate the long run variance $\lim_{T\to\infty} \text{var}[T^{-1/2}\sum_{t=1}^{T} d_t]$ for some random vector d_t.

Newey and West's (1994) *bandwidth selection method*[9]

1. Construct the scalar random variable $h_t = w' * d_t$ where w is a prespecified vector of constants discussed below.
2. Construct $\hat{\sigma}_j = (T-1)^{-1}\sum_{t=j+2}^{T} h_t h_{t-j}$ for $j = 0, 1, \ldots n$.
3. Calculate $\hat{s}^{(\nu)} = 2\sum_{j=1}^{n} j\hat{\sigma}_j$ and $\hat{s}^{(0)} = \hat{\sigma}_0 + 2\sum_{j=1}^{n}\hat{\sigma}_j$.
4. Calculate $\hat{\gamma} = 1.1447[\{\hat{s}^{(\nu)}/\hat{s}^{(0)}\}^2]^{1/3}$.
5. Set $b_T = \text{int}\{\hat{\gamma}T^{1/3}\}$.

[8] All calculations are performed using MATLAB.

[9] It should be noted that Newey and West (1994) actually recommend the use of a prewhitened and recolored covariance matrix estimator. However, this step is omitted in our design to highlight more clearly the impact of misspecification on HACC estimators.

TABLE I

SUMMARY STATISTICS FOR BANDWIDTH AND OVERIDENTIFYING RESTRICTIONS TEST AT $T = 300$

c	γ	Med($\hat{b}_T$)	Med(J_T)	Pow(J_T)	Med($\tilde{b}_T$)	Med($\tilde{J}_T$)	Pow($\tilde{J}_T$)
4	0.000	4	0.476	0.051	4	0.481	0.061
	0.125	4	1.310	0.186	4	1.353	0.213
	0.250	4	4.736	0.600	4	5.296	0.631
	0.375	5	9.474	0.923	4	12.016	0.934
	0.500	6	13.744	0.996	4	20.721	0.997
	10.00	10	21.661	1.000	4	110.900	1.000
12	0.00	11	0.501	0.046	12	0.518	0.078
	0.125	11	1.341	0.171	12	1.473	0.241
	0.250	12	4.300	0.563	12	5.702	0.651
	0.375	15	7.389	0.905	11	12.867	0.940
	0.500	18	9.228	0.993	11	22.059	0.997
	10.00	24	11.370	1.000	12	116.227	1.000

Notes: Med(J) denotes the median of the statistic J; Pow(J) denotes the power of the test based on J with nominal size 0.05.

If the HACC is calculated with uncentered data, then $d_t = f(v_t, \hat{\theta}_T(1))$ and the selected bandwidth is denoted $\hat{b}_T$. If the HACC is calculated with centered data, then $d_t = f(v_t, \hat{\theta}_T(1)) - g_T(\hat{\theta}_T(1))$, and the selected bandwidth is denoted $\tilde{b}_T$. To implement this approach, it is necessary to specify w and n. We set $w = (1, -1)', n = c\text{int}\{(T/100)^{2/9}\}$, and consider two values for c, that is $c = 4, 12$.[10]

For brevity, we only report results in Table I for $\gamma = 0.0, 0.125, 0.25, 0.375, 0.5, 10.00$ and $T = 300$, but summarize corresponding results for $T = 1000$ in the text. All calculations are based on 10,000 replications. First consider the size properties of the test (i.e. $\gamma = 0$). At $T = 300$, the empirical sizes of J_T and $\tilde{J}_T$ tend to be closer to their nominal value of 0.05 with $c = 4$ than with $c = 12$. Of the two statistics, J_T exhibits empirical size closer to the nominal value as would be expected since it is based on the HACC estimator, which exploits the information in (1). By $T = 1000$, the empirical sizes of J_T with $c = 4, 12$ and $\tilde{J}_T$ with $c = 4$ are within two simulation standard errors of the nominal value; the corresponding quantity for $\tilde{J}_T$ with $c = 12$ is slightly more than three simulation standard errors from the nominal value. Now consider the power properties of the test (i.e. $\gamma > 0$). Three features stand out. First, $\tilde{J}_T$ is more powerful than J_T and sometimes by as much as approximately 10%. Second, the median of $\hat{b}_T$ increases with the degree of misspecification (i.e. γ) but the median of $\tilde{b}_T$ is unaffected. The difference is even more striking at $T = 1000$: when $\gamma = 10$, the medians of $\hat{b}_T$ with $c = 4$ and $c = 12$ are respectively 16 and 36, whereas the corresponding medians of $\tilde{b}_T$ are 4 and 12 respectively. Third, the median of $\tilde{J}_T$ increases far more rapidly than the median of J_T. Again, the contrast is more striking at $T = 1000$. For $\gamma = 10$, the median of $\tilde{J}_T$ is 359.637 ($c = 4$) and 363.018

[10] The values of n are chosen to mimic those employed in Newey and West's (1994) simulation study. The choice of w requires a little more explanation. It turns out that within our simulation design the moment condition approximately takes the form $(k, -k)'$. Therefore, if we set $w = (1, 1)'$ —which is similar to the choice used by Newey and West (1994) in their simulations—then h_t is essentially the same process regardless of whether the data are centered or not. Since this is just an artifact of the simulation design, we set $w = (1, -1)$ in order to illustrate the potential impact of misspecification on Newey and West's method procedure for calculating the bandwidth.

($c = 12$). The corresponding figures for J_T are 51.346 and 25.755. This evidence suggests the use of Newey and West's bandwidth selection method without the mean correction can lead to relatively large values of the bandwidth, and hence exacerbate the problems caused by the failure to account for misspecification in the construction of the HACC estimator.

Dept. of Economics, North Carolina State University, Box 8110, Raleigh, NC 27695-8110, U.S.A.; alastair_hall@ncsu.edu

Manuscript received April, 1998; final revision received August, 1999.

Mathematical Appendix

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

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

Assumption A.2: *Θ is a compact set*.

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

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

Assumption A.5: *$f(v, \theta)$ is continuously differentiable with respect to θ on $\text{int}(\Theta)$ and this differential $\partial f(\cdot, \theta)/\partial \theta'$ is measurable on $\mathscr{V}$ for each $\theta \in \text{int}(\Theta)$.*

Assumption A.6: *There exists a measurable function $b(v)$ such that $|f_i(v, \theta)| < b(v), |\partial f_i(v, \theta)/\partial \theta_j| < b(v)$ for all $i, j = 1, 2 \ldots q$ 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: *θ_{**} is an interior point of Θ.*

More primitive conditions for the Central Limit Theorem and the Uniform Weak Law of Large Numbers can be found in inter alia Woolridge (1994).

Proof of Theorem 1: The proof of part (i) is broken down into two parts. We first show that $\hat{S}_T^{-1} = S_T^{-1} + o_p(1)$ where $S_T = V + B_T \mu_* \mu_*'$.[11] We then use the form of S_T^{-1} to deduce the result stated in the theorem.

The proof is based on two matrix results, which for convenience are stated first. Let $\|A\| = \sqrt{\text{tr}(A'A)}$, and $\lambda_{\max}(A)$, $\lambda_{\min}(A)$ denote the largest and smallest eigenvalues of A respectively.

[11] I am greatly indebted to an anonymous referee for suggesting the strategy used to prove this statement.

It can be shown that for $q \times q$ symmetric matrices A and B:[12]

$$\|A\| \leq q\lambda_{\max}(A), \tag{13}$$

$$|\lambda_{\min}(A) - \lambda_{\min}(B)| \leq \|A - B\|. \tag{14}$$

To begin the proof, note that for any T, we have[13]

$$S_T^{-1} = V^{-1} - \frac{B_T}{1 + B_T \mu_*' V^{-1} \mu_*} V^{-1} \mu_* \mu_*' V^{-1} \tag{15}$$

and so $\|S_T^{-1}\| = O(1)$. Now, Lemma 2 implies that $\|\hat{S}_T - S_T\| = o_p(1)$. Since V is positive definite, it follows that $\lambda_{\min}(S_T) > 0$; see Magnus and Neudecker (1991, p. 208). These two properties combined with (14) imply that $\lim_{T \to \infty} P[\lambda_{\min}(\hat{S}_T) > 0] = 1$, and so

$$\|\hat{S}_T^{-1}\| \leq q\lambda_{\max}(\hat{S}_T^{-1}) = q\{\lambda_{\min}(\hat{S}_T)\}^{-1} = O_p(1). \tag{16}$$

Therefore,

$$\begin{aligned} \|\hat{S}_T^{-1} - S_T^{-1}\| &= \|\hat{S}_T^{-1}(S_T - \hat{S}_T)S_T^{-1}\| \leq \|\hat{S}_T^{-1}\| \, \|S_T - \hat{S}_T\| \, \|S_T^{-1}\| \\ &= O_p(1)o_p(1)O(1) = o_p(1) \end{aligned}$$

and so we have established that $\hat{S}_T^{-1} = S_T^{-1} + o_p(1)$. To complete the proof of part (i) note that as $T \to \infty$, we have $B_T/(1 + B_T \mu_*' V^{-1} \mu_*) \xrightarrow{p} 1/(\mu_*' V^{-1} \mu_*)$. The desired result then follows directly from (15).

The proof of part (ii) follows immediately from

$$S_* = V^{-1/2'} \left[I_q - x(x'x)^{-1} x' \right] V^{-1/2} \tag{17}$$

where $V^{-1/2}$ is the $(q \times q)$ matrix that satisfies $V^{-1} = V^{-1/2'} V^{-1/2}$ and $x = V^{-1/2} \mu_*$.

PROOF OF THEOREM 2:[14] Let the minimand on the second step GMM estimation be $Q_T(\theta)$. By definition $Q_T(\hat{\theta}_T(2)) \leq Q_T(\theta_*)$, and so it is sufficient to prove that $TQ_T(\theta_*) = O_p(T/b_T)$. By the Cauchy-Schwarz inequality and Lemma 2, we have

$$TQ_T(\theta_*) \leq |T/B_T| \, |B_T Q_T(\theta_*)| \leq |T/b_T| \|b_T/B_T\| B_T Q_T(\theta_*)|. \tag{18}$$

Since $T/b_T = O(T/b_T)$ we need to show that the remaining two terms in (18) are $O_p(1)$. We first show that $b_T/B_T = O(1)$. Let $c_T = B_T/b_T$. From Assumption 5(iii) it follows that $c_T = c + o(1)$ where $0 < c < \infty$. Therefore $b_T/B_T = c_T^{-1} = (c + o(1))^{-1} = c^{-1} + o(1) = O(1)$ because c is a finite positive constant. We now show that $B_T Q_T(\theta_*) = O_p(1)$. Since

$$B_T Q_T(\theta_*) = B_T^{1/2} g_T(\theta_*)' \hat{S}_T^{-1} B_T^{1/2} g_T(\theta_*),$$

we first consider $B_T^{1/2} g_T(\theta_*)$. By definition, we have

$$B_T^{1/2} g_T(\theta_*) = B_T^{1/2} \mu_* + (B_T/T)^{1/2} T^{-1/2} \sum_{t=1}^{T} [f(v_t, \theta_*) - \mu_*]. \tag{19}$$

Under our conditions $T^{-1/2} \Sigma_{t=1}^{T} [f(v_t, \theta_*) - \mu_*] = O_p(1)$, and so it follows from Assumption 5(ii)–(iii) and (19) that $B_T^{1/2} g_T(\theta_*) = B_T^{1/2} \mu_* + o_p(1)$. Therefore, it follows that

$$B_T Q_T(\theta_*) = B_T \mu_*' \hat{S}_T^{-1} \mu_* + o_p(1) \tag{20}$$

$$= B_T \mu_*' S_T^{-1} \mu_* + B_T \mu_*' (\hat{S}_T^{-1} - S_T^{-1}) \mu_* + o_p(1). \tag{21}$$

[12] Equation (13) follows from Magnus and Neudecker (1991, Theorem 14, p. 211). Equation (14) follows from Golub and van Loan (1989, Corollary 8.1.3, p. 411).

[13] For example, see Morrison (1976, p. 69).

[14] I am greatly indebted to two anonymous referees whose comments greatly shortened this proof.

Using (15) it can be shown that

$$B_T \mu_*' S_T^{-1} \mu_* = \frac{B_T \mu_*' V^{-1} \mu_*}{1 + B_T \mu_*' V^{-1} \mu_*} = n_{1,T}, \quad \text{say.}$$

Assumption 5(iii) and Lemma 2 together imply that $B_T \to \infty$ as $T \to \infty$ and so $\lim_{T \to \infty} n_{1,T} = 1$. Now consider the second term in (21), that is

$$B_T \mu_*' (\hat{S}_T^{-1} - S_T^{-1}) \mu_* = B_T \mu_*' \hat{S}_T^{-1} (S_T - \hat{S}_T) S_T^{-1} \mu_* = n_{2,T}, \quad \text{say.}$$

Using (16) and Lemma 2, it follows that $n_{2,T} = O_p(1) o_p(1) B_T S_T^{-1} \mu_*$. Now $B_T S_T^{-1} \mu_* = [B_T/(1 + B_T \mu_*' V^{-1} \mu_*)] V^{-1} \mu_* = O(1)$ and so $n_{2,T} = o_p(1)$. Therefore, we have

$$B_T Q_T(\theta_*) = n_{1,T} + n_{2,T} + o_p(1) = n_{1,T} + o_p(1) = O_p(1)$$

and so the desired result follows from (18).

PROOF OF THEOREM 3: *Part* (*i*): To simplify the presentation let $\hat{f}_t = f(v_t, \hat{\theta}_T(1))$, $\bar{f} = g_T(\hat{\theta}_T(1))$, and $\bar{f}_i = (T-i)^{-1} \sum_{t=i+1}^{T} f(v_t, \hat{\theta}_T(1))$. Using these definitions, it can be shown that

$$\tilde{\Gamma}_i = \tilde{\Gamma}_i^0 - A_{i,T} + C_{i,T} + C_{i,T}' \tag{22}$$

where

$$A_{i,T} = \frac{T-i}{T} [\bar{f} - \mu_*][\bar{f} - \mu_*]' \quad \text{and} \quad C_{i,T} = \frac{T-i}{T} [\bar{f}_i - \mu_*][\bar{f} - \mu_*]'.$$

Using similar arguments to Andrews' (1991) proof of this Theorem 1, it can be shown that

$$\sum_{i=-T+1}^{T-1} \omega(i/b_T) \tilde{\Gamma}_i^0 \xrightarrow{p} V. \tag{23}$$

Therefore, the desired result will follow if

$$A = \sum_{i=-T+1}^{T-1} \omega(i/b_T) A_{i,T} \xrightarrow{p} 0, \tag{24}$$

$$C = \sum_{i=-T+1}^{T-1} \omega(i/b_T) C_{i,T} \xrightarrow{p} 0. \tag{25}$$

Equations (24) and (25) can be deduced from the following properties that are implied by the conditions of the theorem:

$$(a) \qquad \sum_{i=-T+1}^{T-1} \omega(i/b_T) \frac{T-i}{T} = O(b_T);$$

$$(b) \qquad \bar{f} - \mu_* = O_p(T^{-1/2});$$

$$(c) \qquad \bar{f}_i - \mu_* = O_p(T^{-1/2}), \quad \text{for all } i = 1, 2 \ldots b_T.$$

It is easily shown that (*a*)–(*c*) plus $b_T = o(T^{1/2})$ and $b_T \to \infty$ imply (24) and (25).

Part (*ii*): This follows directly from Wooldridge's (1994) Theorem 7.1.

Part (*iii*): From parts (i) and (ii) it follows that

$$T^{-1} \tilde{J}_T = \mu_{**}' V^{-1} \mu_{**} + o_p(1).$$

Assumptions 2 and 6 together imply that $\epsilon = \mu_{**}' V^{-1} \mu_{**} > 0$.

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–564.

GALLANT, A. R. (1987): *Nonlinear Statistical Models*. New York, NY: Wiley.

GALLANT, A. R., AND H. WHITE (1988): *A Unified Theory of Estimation and Inference in Nonlinear Models*. Oxford, UK: Basil Blackwell.

GOLUB, G. H., AND C. F. VAN LOAN (1989): *Matrix Computations*, Second Edn. Baltimore, MD: Johns Hopkins Press.

HALL, A. R. (1993): "Some Aspects of Generalized Method of Moments Estimation," in *Handbook of Statistics*, Vol. 11, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod. Amsterdam, The Netherlands: Elsevier Science Publishers, pp. 393–417.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

MAGNUS, J. R., AND H. NEUDECKER (1991): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York, NY: Wiley.

MORRISON, D. F. (1976): *Multivariate Statistical Methods*, Second Edn. Tokyo, Japan: McGraw-Hill.

NEWEY, W. K., AND D. L. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. L. McFadden. Amsterdam, The Netherlands: Elsevier Science Publishers, pp. 2113–2247.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple Positive Semi-definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

OGAKI, M. (1993): "Generalized Method of Moments: Econometric Applications," in *Handbook of Statistics*, Vol. 11, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod. Amsterdam, The Netherlands: Elsevier Science Publishers, pp. 455–488.

WOOLDRIDGE, J. M. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. L. McFadden. Amsterdam, The Netherlands: Elsevier Science Publishers, pp. 2641–2739.