Econometrica, Vol. 71, No. 4 (July, 2003), 1121–1159

THE TIME SERIES AND CROSS-SECTION ASYMPTOTICS OF DYNAMIC PANEL DATA ESTIMATORS

BY JAVIER ALVAREZ AND MANUEL ARELLANO[1]

In this paper we derive the asymptotic properties of within groups (WG), GMM, and LIML estimators for an autoregressive model with random effects when both T and N tend to infinity. GMM and LIML are consistent and asymptotically equivalent to the WG estimator. When $T/N \to 0$ the fixed T results for GMM and LIML remain valid, but WG, although consistent, has an asymptotic bias in its asymptotic distribution. When T/N tends to a positive constant, the WG, GMM, and LIML estimators exhibit negative asymptotic biases of order $1/T$, $1/N$, and $1/(2N-T)$, respectively. In addition, the crude GMM estimator that neglects the autocorrelation in first differenced errors is inconsistent as $T/N \to c > 0$, despite being consistent for fixed T. Finally, we discuss the properties of a random effects pseudo MLE with unrestricted initial conditions when both T and N tend to infinity.

KEYWORDS: Autoregressive models, random effects, panel data, within-groups, generalized method of moments, maximum likelihood, double asymptotics.

1. INTRODUCTION

IN A REGRESSION MODEL for panel data containing lags of the dependent variable, the within-groups (WG) estimator can be severely downward biased when the time series (T) is short regardless of the cross-sectional size of the panel (N). This has been a well known fact since the Monte Carlo simulations reported by Nerlove (1967, 1971) and the exact calculation of the bias for the first-order autoregressive model derived by Nickell (1981). Moreover, Anderson and Hsiao (1981) showed the sensitivity of maximum likelihood estimators to alternative assumptions about initial conditions and asymptotic plans. As a result, they proposed to estimate their model in first-differences by instrumental variables using either the dependent variable lagged two periods or its first-differences as instruments. Anderson and Hsiao argued that the advantage of these estimators was that they were consistent whatever the form of the initial conditions and whether T or N or both were tending to infinity. Inconsistency for fixed T as N tends to infinity has been regarded as an undesirable property since in most micro panels T is small while N is large. Subsequently, Holtz-Eakin, Newey, and Rosen

[1] We thank Gary Chamberlain, Jan Magnus, Enrique Sentana, Neil Shephard, a co-editor, three anonymous referees, and seminar audiences at CEMFI, IFS, London, and CentER, Tilburg, for useful comments. An earlier version of this paper was presented at the International Conference on Panel Data, Göteborg, June 1998, and at the Econometric Society European Meeting in Berlin, September 1998. All remaining errors are our own. The first author acknowledges research funding from a CEMFI Ph.D. scholarship, from the Spanish DGI, Grant BEC2002-03097, and from Generalitat Valenciana, Grant CTIDIB/2002/175.

(1988) and Arellano and Bond (1991) proposed GMM estimators that used all the available lags at each period as instruments for the equations in first differences, hence relying on a number of orthogonality conditions that grew at the rate of $T(T-1)/2$. These estimates were shown to be consistent for fixed T, and the simulations reported by Arellano and Bond suggested significant efficiency gains of the GMM estimates relative to those of the Anderson-Hsiao type. However, applied econometricians have tended to use in practice less than the total number of instruments available when that number (which depends on T) was judged to be not sufficiently small relative to the cross-sectional sample size. This practice reflects a concern with the small sample properties of GMM estimators, which have been shown not to be free from bias either, as reported in, for example, Kiviet (1995) or Alonso-Borrego and Arellano (1999). This concern led Alonso-Borrego and Arellano to consider symmetrically normalized GMM estimators of the LIML type, which in simulations exhibited less bias but more dispersion than conventional GMM.

In this paper we show that further insight into the relative merits of dynamic panel data estimators can be obtained by allowing both N and T to tend to infinity and studying their behavior for alternative relative rates of increase for T and N. Our analysis is motivated by the increasing availability of micropanels in which the value of T is not negligible relative to N (such as the household incomes panel in the US (PSID), or the balance-sheet-based company panels that are available in many countries). Thus this paper does not belong to the recent literature on country or regional macropanels (which has focused on models with unit roots, or models with more general forms of heterogeneity as, for example, in Pesaran and Smith (1995), and Canova and Marcet (1995)), although some of our results may be also relevant in that context. The importance of the results in this paper is that they lead to a reassessment of alternative panel data estimators for autoregressive models existing in the literature.

Specifically, we establish the asymptotic properties of WG, GMM, and LIML estimators for a first-order autoregressive model with individual effects when both N and T tend to infinity. We show that the three estimators are consistent when $T/N \to c$ for $0 < c \leq 2$. The basic intuition behind this result is that, contrary to the structural equation setting where too many instruments produces overfitting and undesirable closeness to the OLS coefficients (cf. Kunitomo (1980), Morimune (1983), or Bekker (1994), who show that 2SLS is inconsistent as the number of instruments tends to infinity), here a large number of instruments is associated with larger values of T, and in such a case closeness to OLS (the WG estimator) becomes increasingly desirable since the "endogeneity bias" tends to zero as T tends to infinity. Nevertheless, WG, GMM, and LIML exhibit a bias term in their asymptotic distributions; the biases are of orders $1/T, 1/N$, and $1/(2N-T)$, respectively. Provided $T < N$, the asymptotic GMM bias is always smaller than the WG bias, and the LIML bias is smaller than the other two. When $T = N$ the expressions for the three asymptotic biases are all equal. When $T/N \to 0$ the fixed T results for GMM and LIML remain valid. Conversely, the asymptotic bias in the WG estimator only disappears when $N/T \to 0$.

Some other results emerge from this setting. The three estimators are asymptotically normal and have the same asymptotic variance, given by the large T variance of least-squares when the individual effects are known. Another interesting result is that a crude GMM estimator that neglects the first-difference structure of the errors is inconsistent as T tends to infinity, while it would only be asymptotically inefficient for fixed T as N tends to infinity. The intuition here is again that with an increasingly large number of instruments the instrumental variables estimates will approach the OLS estimates in first differences, which cannot be consistent as $T \to \infty$.

Finally, we consider a random effects (pseudo) maximum likelihood estimator (RML) that leaves the mean and variance of initial conditions unrestricted but enforces time series homoskedasticity. For fixed T, RML is more efficient but less robust than GMM or LIML, since unlike the latter RML requires homoskedasticity for consistency. However, as both T and N tend to infinity RML becomes robust to time series heteroskedasticity, and its asymptotic variance coincides with those of GMM and LIML. The difference is that unlike GMM or LIML, RML does not exhibit an asymptotic bias, because it does not entail incidental parameters in the N or T dimensions.

The paper is organized as follows. Section 2 presents the model and the estimators. In Section 3 we establish the asymptotic properties of WG, GMM, and LIML estimators, and provide some discussion of the implications of the results. We also show the inconsistency of the crude GMM estimator in first-differences, and discuss the properties of the RML estimator in the large T and N context. Section 4 reports some Monte Carlo simulations to evaluate the accuracy of the approximations. Finally, Section 5 contains some concluding remarks and plans for future work.

2. THE MODEL AND THE ESTIMATORS

The Model. We consider an autoregressive process for panel data given by

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it} \qquad (t = 1, \dots, T;\ i = 1, \dots, N), \tag{1}$$

where $|\alpha| < 1$ and v_{it} has zero mean given $\eta_i, y_{i0}, \dots, y_{it-1}$. For notational convenience we assume that y_{i0} is also observed. Moreover, for the presentation of the estimators below, it is convenient to introduce the notation $x_{it} = y_{it-1}$ and write model (1) in the form:

$$y_i = \alpha x_i + \eta_i \iota_T + v_i \tag{2}$$

where $y_i = (y_{i1}, \dots, y_{iT})'$, $x_i = (x_{i1}, \dots, x_{iT})'$, ι_T is a $T \times 1$ vector of ones, and $v_i = (v_{i1}, \dots, v_{iT})'$.

The Within-Groups Estimator. The within-groups or covariance estimator is given by

$$\hat{\alpha}_{WG} = \frac{\sum_{i=1}^N x_i' Q_T y_i}{\sum_{i=1}^N x_i' Q_T x_i} \tag{3}$$

where $Q_T = I_T - \iota_T \iota_T'/T$ is the WG operator of order T.

The WG estimator can also be written as OLS in orthogonal deviations (cf. Arellano and Bover (1995)). The forward orthogonal deviations operator A is the $(T-1)\times T$ upper triangular matrix such that $A'A = Q_T$ and $AA' = I_{T-1}$. Thus, if $\mathrm{var}(v_i) = \sigma^2 I_T$, the $(T-1)\times 1$ vector of errors in orthogonal deviations $v_i^* = Av_i$ also has $\mathrm{var}(v_i^*) = \sigma^2 I_{T-1}$.[2] Notice that since $A\iota_T = 0$, in the equation in orthogonal deviations the individual effects are differenced out:

$$y_i^* = \alpha x_i^* + v_i^*, \tag{4}$$

and letting $x^* = (x_1^{*'}, \ldots, x_N^{*'})'$ and $y^* = (y_1^{*'}, \ldots, y_N^{*'})'$ we have

$$\hat{\alpha}_{WG} = \frac{x^{*'} y^*}{x^{*'} x^*}. \tag{5}$$

The GMM Estimator. For any value of $T, E(x_{it}^* v_{it}^*) \neq 0$ and as a consequence $\hat{\alpha}_{WG}$ is inconsistent for fixed T as N tends to infinity. However,

$$E(z_{it} v_{it}^*) = 0 \qquad (t = 1, \ldots, T-1) \tag{6}$$

where $z_{it} = (x_{i1}, \ldots, x_{it})'$, and therefore GMM estimators of α based on such moment conditions will be consistent for fixed T (cf. Arellano and Bond (1991) and Arellano and Bover (1995)). In (6) there are $q = T(T-1)/2$ orthogonality conditions that can be written as

$$E(Z_i' v_i^*) = 0$$

where Z_i is a $(T-1)\times q$ block diagonal matrix whose tth block is z_{it}'. Moreover, provided v_{it} has constant variance σ^2 given $\eta_i, y_{i0}, \ldots, y_{it-1}$,

$$E(Z_i' v_i^* v_i^{*\prime} Z_i) = \sigma^2 E(Z_i' Z_i), \tag{7}$$

in which case an asymptotically efficient GMM estimator of α relative to the moment conditions in (6) is given by

$$\hat{\alpha}_{GMM} = \frac{x^{*\prime} Z(Z'Z)^{-1} Z' y^*}{x^{*\prime} Z(Z'Z)^{-1} Z' x^*} \tag{8}$$

where $Z = (Z_1', \ldots, Z_N')'$. This is the GMM estimator whose properties we analyze in this paper. A computationally useful alternative expression for $\hat{\alpha}_{GMM}$ is

$$\hat{\alpha}_{GMM} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} M_t y_t^*}{\sum_{t=1}^{T-1} x_t^{*\prime} M_t x_t^*}, \tag{9}$$

[2] The vector v_i^* has elements of the form

$$v_{it}^* = c_t \left[v_{it} - \frac{1}{(T-t)} (v_{it+1} + \cdots + v_{iT}) \right] \qquad (t = 1, \ldots, T-1)$$

with $c_t^2 = (T-t)/(T-t+1)$.

where x_t^* and y_t^* are the $N \times 1$ vectors whose ith elements are x_{it}^* and y_{it}^*, respectively, $M_t = Z_t(Z_t'Z_t)^{-1}Z_t'$ and Z_t is the $N \times t$ matrix whose ith row is z_{it}'.[3]

Notice that invertibility of $Z'Z$ requires that $N \geq T-1$, but the projections involved remain well defined in any case. We shall retain the condition that $N \geq T-1$ for simplicity, and because the GMM estimator is motivated in a situation in which T is smaller than N. Nevertheless, it is straightforward to extend the results in this paper to allow for any combination of values of T and N by considering a generalized formulation of (9) using $M_t = Z_t(Z_t'Z_t)^+Z_t'$ where $(Z_t'Z_t)^+$ denotes the Moore-Penrose inverse of $Z_t'Z_t$. In this way, $M_t = Z_t(Z_t'Z_t)^{-1}Z_t'$ if $t < N$ and $M_t = I_N$ if $t \geq N$. Thus, the contributions of terms with $t > N$ to the GMM formula coincide with the corresponding terms for WG.

Finally, $\hat{\alpha}_{GMM}$ can also be written using the equations in first differences as opposed to orthogonal deviations (cf. Arellano and Bover (1995)). In such case:

$$(10) \qquad \hat{\alpha}_{GMM} = \frac{\Delta x' Z[Z'(I_N \otimes H)Z]^{-1}Z'\Delta y}{\Delta x' Z[Z'(I_N \otimes H)Z]^{-1}Z'\Delta x}$$

where Δx and Δy are $(T-1)N \times 1$ vectors of the variables in first differences, and H is a $(T-1) \times (T-1)$ matrix whose diagonal elements are equal to two, the elements in the first subdiagonal are equal to minus one, and the remaining elements are equal to zero.

As shown by Ahn and Schmidt (1995), the orthogonality conditions in (6) are not the only restrictions on the data second-order moments implied by conditional mean independence and homoskedasticity of v_{it}, but these are the only ones that remain valid in the absence of homoskedasticity or lack of correlation between v_{it} and η_i.

The LIML Estimator. The "limited information maximum likelihood" (LIML) analog estimator solves the following problem:

$$(11) \qquad \hat{\alpha}_{LIML} = \arg\min_{a} \frac{(y^* - ax^*)'Z(Z'Z)^{-1}Z'(y^* - ax^*)}{(y^* - ax^*)'(y^* - ax^*)}.$$

It is a symmetrically normalized estimator of the kind considered by Alonso-Borrego and Arellano (1999), and it is asymptotically equivalent to the GMM estimator for fixed T as $N \to \infty$. It can also be regarded as a "continuously updated" GMM estimator in the terminology of Hansen, Heaton, and Yaron (1996). That is, instead of keeping σ^2 fixed in the weighting matrix of the GMM criterion, it is continuously updated by making it a function of the argument in the estimating criterion. It does not correspond to any meaningful maximum

[3] In our notation, for the sake of simplicity, the individual-specific matrix Z_i is distinguished from the period-specific matrix Z_t by the choice of index. Similarly, v_i and v_t denote individual $(T \times 1)$ and period specific $(N \times 1)$ vectors, respectively. Throughout, the index i is the only one used to denote individuals, whereas time periods are denoted by t and other indices.

likelihood estimator; it is only a LIML analog estimator in the sense of the minimax instrumental-variable interpretation given by Sargan (1958) to the original LIML estimator.[4]

We can write down a simple explicit expression for $\hat{\alpha}_{LIML}$ by noticing that the minimized criterion in (11) is the minimum generalized characteristic root $\hat{\ell}$ of the polynomial equation:

$$\det[\ell(W^{*\prime}W^*) - W^{*\prime}Z(Z'Z)^{-1}Z'W^*] = 0 \qquad (12)$$

where $W^* = (y^* : x^*)$. As $N \to \infty$ for fixed T, $\hat{\ell} \xrightarrow{p} 0$ since the population projection matrix is singular.

Now the first order conditions for (11) are

$$(1, -a)[W^{*\prime}Z(Z'Z)^{-1}Z'W^* - \hat{\ell}W^{*\prime}W^*]\begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0 \qquad (13)$$

from which we obtain

$$\hat{\alpha}_{LIML} = \frac{x^{*\prime}Z(Z'Z)^{-1}Z'y^* - \hat{\ell}(x^{*\prime}y^*)}{x^{*\prime}Z(Z'Z)^{-1}Z'x^* - \hat{\ell}(x^{*\prime}x^*)}. \qquad (14)$$

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

Assumptions. In this section we derive the asymptotic properties of the previous estimators when both T and N tend to infinity under the following assumptions:

ASSUMPTION A1: $\{v_{it}\}$ $(t = 1, \ldots, T;\ i = 1, \ldots, N)$ *are i.i.d. across time and individuals and independent of* η_i *and* y_{i0}, *with* $E(v_{it}) = 0, \mathrm{var}(v_{it}) = \sigma^2$, *and finite moments up to fourth order.*

ASSUMPTION A2: *The initial observations satisfy*

$$y_{i0} = \frac{\eta_i}{1-\alpha} + w_{i0} \qquad (i = 1, \ldots, N),$$

where w_{i0} *is independent of* η_i *and i.i.d. with the steady state distribution of the homogeneous process, so that* $w_{i0} = \sum_{j=0}^{\infty} \alpha^j v_{i(-j)}$.

ASSUMPTION A3: η_i *are i.i.d. across individuals with* $E(\eta_i) = 0, \mathrm{var}(\eta_i) = \sigma_\eta^2$, *and finite fourth order moment.*

While these assumptions will be used in deriving the asymptotic properties of the estimators, the estimators themselves do not rely on the specification of initial conditions.

[4] We nevertheless prefer to keep the LIML label to refer to these estimators, since much of their motivation draws on the finite sample literature for LIML in the instrumental variable context.

3.1. *The WG Estimator*

We first consider the covariance or WG estimator defined in (3) and (5):

$$\hat{\alpha}_{WG} - \alpha = \frac{x^{*'}v^*}{x^{*'}x^*}. \tag{15}$$

The results collected in the following lemma are useful in establishing the asymptotic properties of the WG estimator. All proofs are in the Appendix.

LEMMA 1: *Under Assumptions A1 and A2,*

$$E(x^{*'}v^*) = -N\frac{\sigma^2}{(1-\alpha)}\left[1-\frac{1}{T}\left(\frac{1-\alpha^T}{1-\alpha}\right)\right]. \tag{16}$$

Moreover, as $T \to \infty$, regardless of whether N is fixed or tends to infinity,

$$\operatorname{var}\left(\frac{x^{*'}v^*}{(NT)^{1/2}}\right) \to \frac{\sigma^4}{(1-\alpha^2)}, \tag{17}$$

$$\frac{1}{NT}(x^{*'}x^*) \xrightarrow{p} \frac{\sigma^2}{(1-\alpha^2)}. \tag{18}$$

It is well known that $\hat{\alpha}_{WG}$ is consistent as $T \to \infty$ regardless of the asymptotic behavior of N (cf. Anderson and Hsiao (1981), or Nickell (1981)). Indeed, in view of (16) and (17), $(x^{*'}v^*)/NT$ converges to zero in mean square, which implies that $\operatorname{plim}(x^{*'}v^*/NT) = 0$. Together with (18), this implies that

$$\hat{\alpha}_{WG} \xrightarrow{p} \alpha \quad \text{as} \quad T \to \infty. \tag{19}$$

We now turn to consider asymptotic normality. The result is contained in the following theorem.

THEOREM 1 (Asymptotic normality of the WG estimator): *Let conditions A1 and A2 hold. Then, as $T \to \infty$, regardless of whether N is fixed or tends to infinity*:

$$(NT)^{-1/2}[(x^{*'}v^*) - E(x^{*'}v^*)] \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4}{(1-\alpha^2)}\right). \tag{20}$$

Moreover, provided $N/T^3 \to 0$,

$$\sqrt{NT}\left[\hat{\alpha}_{WG} - \left(\alpha - \frac{1}{T}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2). \tag{21}$$

The implication of Theorem 1 is that even if the covariance estimator is always consistent provided $T \to \infty$, its asymptotic distribution may contain an asymptotic bias term when $N \to \infty$, depending on the relative rates of increase of T and N. If $\lim(N/T) = 0$ (which includes N fixed) there is no asymptotic bias:

$$\sqrt{NT}(\hat{\alpha}_{WG} - \alpha) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2) \tag{22}$$

but if $\lim(N/T) > 0$, the bias term in expression (21) must be kept.

Of these two situations, the second is more relevant here since we wish to compare WG estimates with GMM and LIML estimates in environments in which T is not larger than N, and for such datasets the assumption $N/T \to 0$ is not very realistic. Notice that (21) has been obtained under the assumption that $N/T^3 \to 0$. The asymptotic bias will contain additional terms for lower relative rates of increase of T. For example, if $\lim(N/T^3) \neq 0$ but $N/T^5 \to 0$, the bias will include a T^2 term as the one shown in the proof to Theorem 1.

The result in Theorem 1 has been independently found by Hahn and Kuersteiner (2002) under slightly more general conditions. Hahn and Kuersteiner (2002)'s paper has a different focus since they are concerned with the development of a bias-corrected estimator when both N and T are large.

3.2. *The GMM Estimator*

We now turn to consider the GMM estimator defined in (8), (9), or (10):

$$\hat{\alpha}_{GMM} - \alpha = \frac{x^{*\prime} M v^*}{x^{*\prime} M x^*} \tag{23}$$

where $M = Z(Z'Z)^{-1}Z'$. As before, some preliminary results are collected in a Lemma.

LEMMA 2: *Under Assumptions A*1, *A*2, *and A*3,

$$E(x^{*\prime} M v^*) = -T\frac{\sigma^2}{(1-\alpha)}\left[1 - \frac{1}{T(1-\alpha)}\sum_{t=1}^{T}\frac{(1-\alpha^t)}{t}\right]. \tag{24}$$

Moreover, as both N and T tend to infinity, provided $(\log T)^2/N \to 0$,

$$\operatorname{var}\left(\frac{x^{*\prime} M v^*}{(NT)^{1/2}}\right) \to \frac{\sigma^4}{(1-\alpha^2)}, \tag{25}$$

$$\frac{1}{NT}(x^{*\prime} M x^*) \xrightarrow{p} \frac{\sigma^2}{1-\alpha^2}, \tag{26}$$

and provided $T/N \to c, 0 \leq c < \infty$,

$$\frac{1}{NT}(v^{*\prime} M v^*) \xrightarrow{p} \sigma^2 \frac{c}{2}. \tag{27}$$

The condition $(\log T)^2/N \to 0$ provides a limit on how slow N can tend to infinity relative to T. Since the GMM estimator is motivated in environments with T smaller than N, this is not an unreasonable assumption. It would be certainly satisfied if $T/N \to c$ for $0 \leq c < \infty$. Given these results, we can consider the consistency and asymptotic normality of $\hat{\alpha}_{GMM}$ in the following Theorem.

THEOREM 2 (Consistency and asymptotic normality of the GMM estimator): *Let conditions* $A1$, $A2$, *and* $A3$ *hold. Then as both* N *and* T *tend to infinity, provided* $(\log T)^2/N \to 0$, $\hat{\alpha}_{GMM}$ *is consistent for* α:

$$\hat{\alpha}_{GMM} \xrightarrow{p} \alpha. \tag{28}$$

Moreover, provided $T/N \to c, 0 \leq c < \infty$,

$$\sqrt{NT}\left[\hat{\alpha}_{GMM} - \left(\alpha - \frac{1}{N}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2). \tag{29}$$

When $T \to \infty$, the number of the GMM orthogonality conditions $q = T(T-1)/2$ also tends to infinity. In spite of this fact, the theorem shows that $\hat{\alpha}_{GMM}$ remains consistent. This is in contrast to the situation in the structural equation setting where the two-stage least squares estimator has been shown to be inconsistent when both the number of instruments and the sample size tend to infinity, while their ratio tends to a positive constant (cf. Kunitomo (1980), Morimune (1983), and Bekker (1994)). The intuition for the consistency of $\hat{\alpha}_{GMM}$ is that in our context as T tends to infinity, the "endogeneity bias" tends to zero, and so closeness of $\hat{\alpha}_{GMM}$ to $\hat{\alpha}_{WG}$ for larger values of T becomes a desirable property of the GMM estimator.

The theorem also shows that as $T \to \infty, \hat{\alpha}_{GMM}$ is asymptotically normal but unless $\lim(T/N) = 0$, it exhibits a bias term in its asymptotic distribution. When $0 < \lim(T/N) < \infty$, Theorems 1 and 2 provide a clean comparison between the GMM and WG estimators. Namely, they have the same asymptotic variance, and a similar expression for their (negative) asymptotic biases, which nevertheless differ in their orders of magnitude: $(1+\alpha)/N$ for GMM and $(1+\alpha)/T$ for WG.

For a class of linear GMM problems, Koenker and Machado (1999) found that $q^3/N \to 0$ was a sufficient condition for the validity of conventional asymptotic inference about heteroskedasticity-robust GMM estimators, where q is the number of moment conditions. It is interesting to notice that in our case if $T/N \to 0$ as $N \to \infty$ the fixed T homoskedastic asymptotic inferences about $\hat{\alpha}_{GMM}$ are valid. Since here $q = T(T-1)/2$, it turns out that there is a much tighter condition for the validity of standard fixed T inferences in the homoskedastic dynamic panel data context.

3.3. *The LIML Estimator*

The LIML estimator defined in (14) can be written as

$$\hat{\alpha}_{LIML} - \alpha = \frac{x^{*\prime} M v^* - \hat{\ell}(x^{*\prime} v^*)}{x^{*\prime} M x^* - \hat{\ell}(x^{*\prime} x^*)}. \tag{30}$$

The limit in probability of $\hat{\ell}$ is given in the following lemma.

LEMMA 3: *Under Assumptions A*1, *A*2, *and A*3 *as both N and T tend to infinity, and* $T/N \to c, 0 \leq c \leq 2$,

$$\hat{\ell} \xrightarrow{p} \frac{c}{2}. \tag{31}$$

An implication of this result is that $c \leq 2$ is a necessary condition for the consistency of the LIML estimator. In effect, for a given d, under the assumptions of Lemma 3

$$\frac{(y^* - dx^*)' M (y^* - dx^*)}{(y^* - dx^*)'(y^* - dx^*)} \xrightarrow{p} \frac{(d-\alpha)^2 + (1-\alpha^2)c/2}{(d-\alpha)^2 + (1-\alpha^2)}. \tag{32}$$

Provided $c \leq 2$, the limiting criterion is minimized at $d = \alpha$, taking the value $c/2$. If, on the contrary, $c > 2$, the limiting criterion can be reduced for any $d > \alpha$, tending to one as $d \to \pm\infty$. However, the condition $\lim(T/N) \leq 2$ should not be regarded as a restrictive assumption since the LIML estimator is motivated in a situation in which T is smaller than N, and the actual formula we consider is only defined for $(T-1)/N \leq 1$. The following theorem considers consistency and asymptotic normality of $\hat{\alpha}_{LIML}$.

THEOREM 3 (Consistency and asymptotic normality of the LIML estimator): *Let conditions A*1, *A*2, *and A*3 *hold. Then as both N and T tend to infinity, provided* $T/N \to c, 0 \leq c \leq 2$, $\hat{\alpha}_{LIML}$ *is consistent for* α:

$$\hat{\alpha}_{LIML} \xrightarrow{p} \alpha. \tag{33}$$

Moreover,

$$\sqrt{NT}\left[\hat{\alpha}_{LIML} - \left(\alpha - \frac{1}{(2N-T)}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2). \tag{34}$$

The theorem shows that like GMM, the LIML estimator is consistent despite $T \to \infty$ and $T/N \to c$. Also, $\hat{\alpha}_{LIML}$ is asymptotically normal with the same asymptotic variance as the GMM and WG estimates. Unless $T/N \to 0$, it has a (negative) asymptotic bias with a similar expression as the asymptotic biases of WG and GMM, but again differing in its order of magnitude: $(1+\alpha)/T$ for WG, $(1+\alpha)/N$ for GMM, and $(1+\alpha)/(2N-T)$ for LIML. Therefore, provided $T < N$, the LIML asymptotic bias is the smallest of the three.

3.4. *The Crude GMM Estimator in First Differences*

We noticed in equation (10) that the asymptotically efficient GMM estimator could also be written using the moment conditions in first differences as opposed to orthogonal deviations. In such a case, however, the optimal weighting matrix becomes $[Z'(I_N \otimes H)Z]^{-1}$ instead of $(Z'Z)^{-1}$ in order to take into account the serial correlation in the errors in first-differences. In this section we consider the crude IV or GMM estimator in first differences that uses $(Z'Z)^{-1}$ as the weighting matrix:

$$\hat{\alpha}_{CIV} = \frac{\Delta x' Z (Z'Z)^{-1} Z' \Delta y}{\Delta x' Z (Z'Z)^{-1} Z' \Delta x}. \tag{35}$$

For fixed T as N tends to infinity, this estimator is asymptotically inefficient relative to $\hat{\alpha}_{GMM}$, but it is still consistent and asymptotically normal, and as such it may be regarded as a computationally simpler alternative to $\hat{\alpha}_{GMM}$ (for example, Holtz-Eakin, Newey, and Rosen (1988) use CIV estimators as their one-step GMM estimates). However, the results in the previous sections suggest that, since the "endogeneity bias" in first differences does not tend to zero as $T \to \infty$, there may be more fundamental differences between $\hat{\alpha}_{CIV}$ and $\hat{\alpha}_{GMM}$ when both T and N tend to infinity. We address this issue in the following theorem.

THEOREM 4 (Inconsistency of the crude GMM estimator in first differences): *Let conditions* A1, A2, *and* A3 *hold. Then as both* N *and* T *tend to infinity, provided* $T/N \to c, 0 \leq c < \infty$,

$$\frac{1}{NT}(\Delta x' M \Delta v) \xrightarrow{p} -\sigma^2 \frac{c}{2}, \tag{36}$$

$$\frac{1}{NT}(\Delta x' M \Delta x) \xrightarrow{p} \sigma^2 \left(\frac{c}{2} + \frac{1-\alpha}{1+\alpha} \right), \tag{37}$$

and

$$\hat{\alpha}_{CIV} \xrightarrow{p} \alpha - \frac{(1+\alpha)}{2} \left(\frac{c}{2-(1+\alpha)(2-c)/2} \right). \tag{38}$$

The crude GMM estimator is therefore inconsistent when $T \to \infty$ unless $c = 0$. Moreover, the bias may be qualitatively relevant. In a squared panel ($c = 1$) the biases will be enormous, but even in a panel whose cross-sectional size is ten times the time series dimension ($c = 0.1$) the biases are substantial (some numerical calculations of the bias are reported in the next section). Notice that at $c = 2$, the bias of $\hat{\alpha}_{CIV}$ coincides with that of the OLS regression in first differences. This result further illustrates the shortcomings of large N, fixed T asymptotics in evaluating the relative merits of the estimators. In effect, according to the fixed T approximations, in the comparison between $\hat{\alpha}_{GMM}$ and $\hat{\alpha}_{CIV}$ there is only a second order difference in precision, whereas when $T/N \to c > 0$, $\hat{\alpha}_{GMM}$ is still consistent but $\hat{\alpha}_{CIV}$ is not.

3.5. *The Random Effects ML Estimator*

In this section we discuss the random effects ML estimator $\hat{\alpha}_{RML}$ when $(\eta_i, y_{i0}, \ldots, y_{iT})$ are jointly normally distributed and Assumption A1 is satisfied. Let the covariance matrix of η_i and y_{i0} be

$$\operatorname{var}\begin{pmatrix} \eta_i \\ y_{i0} \end{pmatrix} = \begin{pmatrix} \sigma_\eta^2 & \gamma_{\eta 0} \\ \gamma_{\eta 0} & \gamma_{00} \end{pmatrix}. \tag{39}$$

Under Assumption A2 $\gamma_{\eta 0} = \sigma_\eta^2/(1-\alpha)$ and $\gamma_{00} = [\sigma_\eta^2/(1-\alpha)^2] + \sigma^2/(1-\alpha^2)$, but here $\gamma_{\eta 0}$ and γ_{00} are free parameters. Thus $\hat{\alpha}_{RML}$ is also the conditional MLE given y_{i0}. As a result, it will be robust to alternative initial conditions when T is small, and yet the likelihood in this case does not depend on parameters whose number grows with T or N, so that no asymptotic biases will occur when both N and T tend to infinity. From the point of view of the large N, fixed T asymptotics, RML is more efficient but less robust than GMM or LIML, since contrary to the latter RML requires time series homoskedasticity for consistency. However, as both T and N tend to infinity RML turns out to be robust to heteroskedasticity, but unlike GMM or LIML it does not exhibit an asymptotic bias. This is, therefore, another instance in which the N and T asymptotics suggests a reassessment of the relative merits of competing estimators.[5]

As shown in the Appendix, the log Gaussian density of $(y_{i1}, \ldots, y_{iT})$ given y_{i0} can be written as

$$\log f(y_{i1}, \ldots, y_{iT} \mid y_{i0}) = -\frac{(T-1)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y_i^* - \alpha x_i^*)'(y_i^* - \alpha x_i^*) - \frac{1}{2}\log\omega^2 - \frac{1}{2\omega^2}(\bar{y}_i - \alpha\bar{x}_i - \varphi y_{i0})^2 \tag{40}$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^T y_{it}$, $\bar{x}_i = T^{-1}\sum_{t=1}^T x_{it}$, and (φ, ω^2) are the linear projection coefficients of $(\bar{y}_i - \alpha\bar{x}_i)$ on y_{i0} given by $\varphi = \gamma_{\eta 0}/\gamma_{00}$, and $\omega^2 = \sigma_\eta^2 - \varphi^2\gamma_{00} + \sigma^2/T$. Hence, by concentrating φ, ω^2, and σ^2 out of the log likelihood, the RML estimator can be expressed as

$$\hat{\alpha}_{RML} = \arg\min_a \left\{ \log[(y^* - ax^*)'(y^* - ax^*)] + \frac{1}{(T-1)}\log[(\bar{y} - a\bar{x})'S_0(\bar{y} - a\bar{x})] \right\} \tag{41}$$

where $S_0 = I_N - y_0 y_0'/(y_0' y_0)$, $y_0 = (y_{10}, \ldots, y_{N0})'$, $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_N)'$, and $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)'$.[6] Consistency and asymptotic normality of $\hat{\alpha}_{RML}$ is considered in the following theorem. The data are not assumed to be normally distributed, so we regard $\hat{\alpha}_{RML}$ as a pseudo ML estimator.

THEOREM 5 (Consistency and asymptotic normality of the RML estimator): *Let conditions A*1, *A*2, *and A*3 *hold. Then as both N and T tend to infinity,* $\hat{\alpha}_{RML}$ *is consistent for* α:

$$\hat{\alpha}_{RML} \xrightarrow{p} \alpha. \tag{42}$$

[5] We thank Gary Chamberlain for suggesting we consider the RML estimator in this context.

[6] The estimator in (41) does not restrict σ_η^2 to be nonnegative. We may obtain ML estimates of α that enforce $\sigma_\eta^2 \geq 0$, from an alternative concentrated likelihood that is only a function of α and σ_η^2/σ^2. In such case, a boundary solution at $\sigma_\eta^2 = 0$ may occur. This problem was discussed by Maddala (1971).

Moreover, provided $0 \leq \lim(N/T) < \infty$,

$$\sqrt{NT}(\hat{\alpha}_{RML} - \alpha) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2). \tag{43}$$

Obviously, the RML estimator is also consistent and asymptotically normal for fixed T as $N \to \infty$ under the stated conditions, but in such a case the asymptotic variance will take a different expression.

This estimator and a generalized least squares estimator of the same model were considered by Blundell and Smith (1991) and have been discussed further by Blundell and Bond (1998) (in their formulation the model is not transformed into orthogonal deviations together with an average equation as we do).[7]

4. MONTE CARLO EVIDENCE

In this section we report some Monte Carlo simulations of the estimators discussed above for various combinations of values of N and T. We wish to assess the accuracy of the asymptotic approximations derived in Section 3. Various simulation exercises for dynamic panel data estimators have already been conducted in other work, but since the existing results typically concentrate on small values of T, they do not provide the type of evidence required here (an exception is the recent Monte Carlo analysis in Judson and Owen (1997)).

In Table I we report medians, interquartile ranges, and median absolute errors of the WG, GMM, LIML, CIV, and RML estimators for $\alpha = 0.2, 0.5$, and 0.8, and for $N = 100$ with $T^o = 10$, 25, and 50, where $T^o = T + 1$ (the actual number of time series observations in the data). Similar experiments with $N = 50$ are reported in Table II. For all cases we conducted 1000 replications from the model specified in Sections 2 and 3 under normality with $\sigma^2 = 1$ and $\sigma_\eta^2 = 0$. While the exact distribution of the WG estimator is invariant to both σ_η^2 and σ^2, the distributions of the other estimators are only invariant to (σ_η^2/σ^2). Their dependence on σ_η^2, however, vanishes as T tends to infinity, and for the values of T that we consider here, the effect of changing σ_η^2 on the results turned out to be small (as can be seen from Tables A1–A4 in the Appendix, which contain the results for $\sigma_\eta^2 = 0.2$ and 1).

In Table III we calculate and subtract from the value of α the asymptotic biases of the estimates, using the theoretical results in Section 3 (RML is not reported because it has no asymptotic bias). A comparison of those figures with the Monte Carlo medians in Tables I and II reveals that the asymptotic biases provide a very accurate approximation to the finite sample median biases of all the estimators in our experiments. It is interesting to notice that the bias of the GMM estimator is always smaller than the WG bias (even in a squared panel with $T^o = 50$ and $N = 50$), and that the bias of LIML is in turn smaller than the GMM bias. It is also noticeable that the GMM bias changes with N, and the

[7] The problem with the GLS estimates of α and φ based on preliminary estimates of ω^2 and σ^2 is that they are only consistent if based on consistent estimates of ω^2 and σ^2, and they are only asymptotically equivalent to ML if based on asymptotically efficient estimates of ω^2 and σ^2.

TABLE I

MEDIANS, INTERQUARTILE RANGES, AND MEDIAN ABSOLUTE ERRORS OF THE ESTIMATORS ($N = 100$)[a]

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML
								$T^o = 10$							
median	0.065	0.188	0.196	0.139	0.202	0.318	0.481	0.493	0.384	0.500	0.554	0.763	0.792	0.514	0.799
iqr	0.047	0.056	0.057	0.074	0.056	0.048	0.060	0.061	0.083	0.058	0.044	0.069	0.074	0.124	0.073
mae	0.135	0.030	0.029	0.062	0.028	0.182	0.032	0.031	0.116	0.029	0.246	0.046	0.037	0.286	0.036
								$T^o = 25$							
median	0.149	0.187	0.193	0.048	0.199	0.434	0.483	0.492	0.235	0.500	0.714	0.774	0.790	0.281	0.799
iqr	0.026	0.028	0.029	0.040	0.028	0.025	0.028	0.029	0.045	0.028	0.021	0.027	0.029	0.061	0.024
mae	0.051	0.017	0.015	0.152	0.014	0.065	0.019	0.015	0.265	0.014	0.086	0.025	0.015	0.519	0.012
								$T^o = 50$							
median	0.175	0.188	0.192	−0.068	0.199	0.468	0.485	0.491	0.077	0.499	0.760	0.779	0.789	0.112	0.799
iqr	0.019	0.019	0.020	0.026	0.019	0.017	0.018	0.019	0.029	0.018	0.014	0.015	0.017	0.036	0.014
mae	0.025	0.014	0.011	0.268	0.009	0.032	0.015	0.012	0.423	0.009	0.040	0.020	0.012	0.688	0.007

[a] $\sigma_\eta^2 = 0, \sigma^2 = 1$, 1000 replications, iqr is the 75th–25th interquartile range; mae denotes the median absolute error.

TABLE II

MEDIANS, INTERQUARTILE RANGES, AND MEDIAN ABSOLUTE ERRORS OF THE ESTIMATORS ($N = 50$)[a]

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML
							$T^o = 10$								
median	0.063	0.176	0.191	0.084	0.201	0.317	0.462	0.486	0.292	0.499	0.556	0.729	0.781	0.358	0.793
iqr	0.068	0.079	0.081	0.101	0.078	0.067	0.083	0.086	0.119	0.082	0.060	0.096	0.111	0.157	0.093
mae	0.136	0.042	0.041	0.116	0.039	0.183	0.049	0.044	0.207	0.041	0.244	0.074	0.058	0.442	0.048
							$T^o = 25$								
median	0.149	0.178	0.187	−0.065	0.200	0.436	0.470	0.484	0.081	0.502	0.714	0.756	0.780	0.117	0.800
iqr	0.039	0.041	0.043	0.049	0.042	0.038	0.040	0.044	0.058	0.041	0.029	0.037	0.043	0.070	0.034
mae	0.050	0.027	0.023	0.265	0.021	0.064	0.031	0.024	0.419	0.020	0.086	0.044	0.025	0.683	0.017
							$T^o = 50$								
median	0.176	0.178	0.180	−0.222	0.200	0.468	0.471	0.475	−0.093	0.500	0.760	0.764	0.770	−0.015	0.799
iqr	0.027	0.027	0.029	0.028	0.027	0.024	0.025	0.028	0.033	0.025	0.019	0.021	0.026	0.037	0.020
mae	0.025	0.023	0.021	0.422	0.014	0.031	0.029	0.025	0.593	0.012	0.040	0.036	0.030	0.815	0.010

[a] $\sigma_\eta^2 = 0$, $\sigma^2 = 1$, 1000 replications, iqr is the 75th–25th interquartile range; mae denotes the median absolute error.

TABLE III

ASYMPTOTIC BIASES OF THE ESTIMATES[a]

	$\alpha=0.2$				$\alpha=0.5$				$\alpha=0.8$			
	WG	GMM	LIML	CIV	WG	GMM	LIML	CIV	WG	GMM	LIML	CIV
					$N=100$							
$T^o=10$	0.067	0.188	0.194	0.137	0.333	0.485	0.492	0.381	0.600	0.782	0.791	0.512
$T^o=25$	0.150	0.188	0.193	0.047	0.437	0.485	0.491	0.235	0.725	0.782	0.790	0.281
$T^o=50$	0.175	0.188	0.192	−0.069	0.469	0.485	0.490	0.076	0.763	0.782	0.788	0.112
					$N=50$							
$T^o=10$	0.067	0.176	0.187	0.081	0.333	0.470	0.483	0.287	0.600	0.764	0.780	0.352
$T^o=25$	0.150	0.176	0.184	−0.065	0.437	0.470	0.480	0.081	0.725	0.764	0.776	0.116
$T^o=50$	0.175	0.176	0.176	−0.224	0.469	0.470	0.471	−0.095	0.763	0.764	0.765	−0.015

[a]For WG the figures show $\alpha-(1+\alpha)/T$; for GMM, $\alpha-(1+\alpha)/N$; for LIML, $\alpha-(1+\alpha)/(2N-T)$, and for CIV, $\alpha-\frac{(1+\alpha)}{2}\left(\frac{c}{2-(1+\alpha)(2-c)/2}\right)$, where $c=T/N$.

LIML bias changes with both N and T^o as expected. The tables also provide an assessment of the CIV bias. Notice that even with $T^o=10$ the biases of the CIV estimator are substantial. In fact, except for $\alpha=0.2$ and 0.5 with $T^o=10$ and $N=100$, they are always larger than the WG bias! Finally, as expected, RML is virtually median unbiased in all experiments.

Turning to dispersion, LIML always has a larger interquartile range than GMM, but the difference between the two is very small (although less so with $\alpha=0.8$ and $N=50$). WG has the smallest interquartile range. The differences with GMM, LIML, and RML are noticeable when $T^o=10$, but become small with $T^o=25$ or 50. The large T asymptotic interquartile range (that is, $1.349[(1-\alpha^2)/NT]^{1/2}$) does not approximate well the GMM or LIML interquartile ranges for $T^o=10$, but becomes a reasonable approximation when $T^o=25$ or 50, specially for the smaller values of α. Concerning CIV, this estimator always has the largest dispersion, which suggests that in addition to biases there are substantial efficiency losses in using the crude GMM estimator.

Finally, concerning median absolute errors, RML is the estimator that performs best in all the experiments. Among the others, LIML is always the estimator with the smallest median absolute error in the experiments with $\sigma_\eta^2=0$ (Tables I and II), followed by GMM, WG, and CIV, except for three cases in which the mae of CIV is smaller than that of WG. Nevertheless, the ranking is less obvious in the experiments with $\sigma_\eta^2>0$. When $N=100$, GMM outperforms LIML in terms of mae on three occasions (Tables A1 and A2), and with $N=50, T^o=50, \sigma_\eta^2=1$, WG has the smallest mae followed by GMM, LIML, and CIV (Table A4).

5. CONCLUSIONS

In this paper we show that in autoregressive panel data models, the GMM and LIML estimators that use all the available lags at each period as instruments are consistent and asymptotically normal when both N and T tend to infinity.

They attain the same asymptotic variance as the least squares estimator when the individual effects are known and $T \to \infty$. In addition, we establish that when T/N tends to a positive constant the WG, GMM, and LIML estimators are asymptotically biased with negative asymptotic biases of order $1/T$, $1/N$, and $1/(2N-T)$, respectively. When $T/N \to 0$ the fixed T results for GMM and LIML remain valid. Conversely, the asymptotic bias in the WG estimator only disappears when $N/T \to 0$. We also show that the crude GMM estimator that neglects the autocorrelation in the first differenced errors is inconsistent as $T/N \to c > 0$, despite being consistent for fixed T. Finally, we consider a random effects MLE that leaves the mean and variance of initial conditions unrestricted but enforces time series homoskedasticity; this estimator has no asymptotic bias because it does not entail incidental parameters in the N or T dimensions, and it becomes robust to heteroskedasticity as T tends to infinity. The results of some Monte Carlo simulations for data with $T^o = 10$, 25, 50 and $N = 50$, 100 suggest that the asymptotic approximations are a reliable guidance for the sampling distributions of the estimators.

Our results highlight the importance of understanding the properties of panel data estimators as the time series information accumulates even for micropanels with moderate values of T: In a fixed T framework, GMM and LIML are asymptotically equivalent, but as T increases LIML exhibits a smaller asymptotic bias than GMM. Moreover, for fixed T the IV estimators in orthogonal-deviations and first-differences are both consistent, whereas as T increases the former remains consistent but the latter is inconsistent.

In future work we plan to extend the current results in three directions. Firstly, we would like to relax the initial conditions and homoskedasticity assumptions. A second natural extension is to study the properties of "two-step" GMM estimators. These estimators use weighting matrices that remain consistent estimates of the covariance of the moments under heteroskedasticity. Finally, we plan to consider the properties of estimators that allow for time dummies when T is not fixed.

Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, 03071 Alicante, Spain; alvarez@merlin.fae.ua.es; http://merlin.fae.ua.es/alvarez/

and

CEMFI, Casado del Alisal 5, 28014 Madrid, Spain; arellano@cemfi.es; http://www.cemfi.es/~arellano/

Manuscript received December, 1998; final revision received October, 2002.

APPENDIX

A.1. *Within-Groups*

Lemma 1

PROOF OF (16): Firstly, note that

$$E(x^{*'} v^*) = E\left(\sum_{i=1}^{N} x_i' Q_T v_i\right) = N E(x_i' Q_T v_i). \tag{A1}$$

Next, since $E(x_i'v_i) = TE(y_{it-1}v_{it}) = 0$, we have

$$E(x_i'Q_T v_i) = E(x_i'v_i) - \frac{1}{T}\iota_T' E(v_i x_i')\iota_T = -\frac{\sigma^2}{T}\iota_T' C_T \iota_T \tag{A2}$$

where $E(v_i x_i') = \sigma^2 C_T$. Notice that the (t,s)th element of C_T is $\sigma^2\alpha^{(s-t-1)}$ for $t < s$, and zero otherwise. Adding up the elements of this matrix the result follows.

PROOF OF (17) AND (18): Let w_{it} be the homogeneous AR(1) process

$$w_{it} = y_{it} - \frac{\eta_i}{(1-\alpha)} \tag{A3}$$

with corresponding vectors $w_i = (w_{i1}, \ldots, w_{iT})'$ and $w_{i(-1)} = (w_{i0}, \ldots, w_{i(T-1)})'$, so that $w_i^* = Aw_i = y_i^*$, $w_{i(-1)}^* = Aw_{i(-1)} = x_i^*$. Moreover, letting $\bar{v}_i = T^{-1}v_i'\iota_T$, etc., we have:

$$(NT)^{-1/2}x^{*'}v^* = (NT)^{-1/2}\sum_i\sum_t v_{it}w_{it-1} - (T/N)^{1/2}\sum_i \bar{v}_i\bar{w}_{i(-1)}. \tag{A4}$$

Firstly

$$\operatorname{var}\left(\frac{1}{\sqrt{NT}}\sum_i\sum_t v_{it}w_{it-1}\right) = \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_t v_{it}w_{it-1}\right) \to \sum_{j=-\infty}^{\infty}\gamma_j = \gamma_0 \tag{A5}$$

$$= E(v_{it}^2 w_{it-1}^2) = E(v_{it}^2)E(w_{it-1}^2) = \frac{\sigma^4}{1-\alpha^2}$$

where $\gamma_j = E(v_{it}w_{it-1}v_{i(t-j)}w_{i(t-1-j)}) = 0$ for $j \neq 0$ (cf. Anderson (1971)).

Next

$$\operatorname{var}\left(\left(\frac{T}{N}\right)^{1/2}\sum_i \bar{v}_i\bar{w}_{i(-1)}\right) = \frac{T}{N}\operatorname{var}\left(\sum_i \bar{v}_i\bar{w}_{i(-1)}\right) = T\operatorname{var}(\bar{v}_i\bar{w}_{i(-1)}) = O\left(\frac{1}{T}\right) \tag{A6}$$

since $\operatorname{var}(\bar{v}_i\bar{w}_{i(-1)}) = O(T^{-2})$ as justified below.

Therefore, (17) follows:

$$\operatorname{var}\left(\frac{1}{\sqrt{NT}}x^{*'}v^*\right) = \frac{\sigma^4}{1-\alpha^2} + O\left(\frac{1}{T}\right).$$

To prove (18) we establish convergence in mean square. Let us write

$$\frac{x^{*\prime}x^*}{NT} = \frac{1}{NT}\sum_i w_{i(-1)}'Qw_{i(-1)} = \frac{1}{N}\sum_i\left(\frac{1}{T}\sum_t w_{i(t-1)}^2 - \bar{w}_{i(-1)}^2\right) \tag{A7}$$

$$E\left(\frac{x^{*\prime}x^*}{NT}\right) = E(w_{i(t-1)}^2) - E(\bar{w}_{i(-1)}^2) = E(w_{i(t-1)}^2) - O(T^{-1}) \tag{A8}$$

$$= \frac{\sigma^2}{1-\alpha^2} - \frac{1}{T}\frac{\sigma^2}{(1-\alpha^2)}\left[\frac{(1+\alpha)}{(1-\alpha)} - \frac{1}{T}\frac{2\alpha(1-\alpha^T)}{(1-\alpha)^2}\right].$$

The explicit expression (A8) will be of later use. We have used the fact that $E(\bar{w}_{i(-1)}^2) = T^{-2}\iota_T' E(w_i w_i')\iota_T$ and the (t,s) element of $E(w_i w_i')$ is $\sigma^2\alpha^{|t-s|}/(1-\alpha^2)$. Thus,

$$E\left(\frac{x^{*\prime}x^*}{NT}\right) \to \frac{\sigma^2}{1-\alpha^2}.$$

Moreover,

$$\operatorname{var}\left(\frac{x^{*\prime}x^*}{NT}\right) = \frac{1}{N}\operatorname{var}\left(\frac{1}{T}\sum_t w_{i(t-1)}^2 - \bar{w}_{i(-1)}^2\right) \to 0 \tag{A9}$$

since $\operatorname{var}((1/T)\sum_t w_{i(t-1)}^2) = O(T^{-1})$ and $\operatorname{var}(\bar{w}_{i(-1)}^2) = O(T^{-2})$.

Finally, let us justify the assertion that $\text{var}(\bar{w}^2_{i(-1)})$ and $\text{var}(\bar{v}_i\bar{w}_{i(-1)})$ are $O(T^{-2})$. Given our assumptions, these two variances are finite for any T. As $T\to\infty$, $\sqrt{T}\bar{w}_{i(-1)} \xrightarrow{d} \xi_i \sim \mathcal{N}(0, \sigma^2/(1-\alpha)^2)$, which implies $\text{var}(T\bar{w}^2_{i(-1)}) \to \text{var}(\xi_i^2)$ and $\text{var}(\bar{w}^2_{i(-1)}) = O(T^{-2})$. Similarly, $\sqrt{T}(\bar{v}_i, \bar{w}_{i(-1)}) \xrightarrow{d} (\zeta_i, \xi_i)$, which are jointly normally distributed with a singular covariance matrix, so that $\text{var}(T\bar{v}_i\bar{w}_{i(-1)}) \to \text{var}(\zeta_i\xi_i)$ and $\text{var}(\bar{v}_i\bar{w}_{i(-1)}) = O(T^{-2})$.

Theorem 1

PROOF OF (20): In view of (16) we have

$$\mu_{NT} = E[(NT)^{-1/2}x^{*'}v^*] = -\left(\frac{N}{T}\right)^{1/2}\frac{\sigma^2}{(1-\alpha)} + \frac{N^{1/2}}{T^{3/2}}\frac{\sigma^2(1-\alpha^T)}{(1-\alpha)^2}. \tag{A10}$$

Subtracting μ_{NT} from (A4),

$$(NT)^{-1/2}(x^{*'}v^*) - \mu_{NT} = (NT)^{-1/2}\sum_i\sum_t v_{it}w_{it-1} - R_{NT} \tag{A11}$$

where

$$R_{NT} = (T/N)^{1/2}\sum_i \bar{v}_i\bar{w}_{i(-1)} + \mu_{NT}.$$

Clearly, $E(R_{NT}) = 0$ and in view of (A6) $\lim_{T\to\infty}\text{var}(R_{NT}) = 0$, which suffices to establish that R_{NT} is $o_p(1)$.

Finally, from a standard central limit theorem for autoregressive processes (cf. Anderson (1971, Ch. 5, Theorem 5.5.7), and Anderson (1978)):

$$(NT)^{-1/2}\sum_i\sum_t v_{it}w_{it-1} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4}{(1-\alpha^2)}\right). \tag{A12}$$

Since R_{NT} is $o_p(1)$, also

$$(NT)^{-1/2}(x^{*'}v^*) - \mu_{NT} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4}{(1-\alpha^2)}\right),$$

which establishes the first result of the theorem.

PROOF OF (21): In view of (18), by Cramer's theorem we have

$$\left(\frac{x^{*'}x^*}{NT}\right)^{-1}[(NT)^{-1/2}(x^{*'}v^*) - \mu_{NT}] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2)$$

or

$$\sqrt{NT}(\hat{\alpha}_{WG} - \alpha) - \left[E\left(\frac{x^{*'}x^*}{NT}\right)\right]^{-1}\mu_{NT} - R^o_{NT} \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2) \tag{A13}$$

where

$$R^o_{NT} = \left\{\left(\frac{x^{*'}x^*}{NT}\right)^{-1} - \left[E\left(\frac{x^{*'}x^*}{NT}\right)\right]^{-1}\right\}\mu_{NT}.$$

Because of (A9) we have

$$\frac{x^{*'}x^*}{NT} - E\left(\frac{x^{*'}x^*}{NT}\right) = O_p(1/\sqrt{NT}), \tag{A14}$$

and by the delta method also

$$\left(\frac{x^{*\prime}x^{*}}{NT}\right)^{-1}-\left[E\left(\frac{x^{*\prime}x^{*}}{NT}\right)\right]^{-1}=O_{p}(1/\sqrt{NT}).$$

Thus, since $\mu_{NT}=O(\sqrt{N/T})$ the term R_{NT}^{o} is $o_{p}(1)$ as $T\to\infty$ regardless of N.

Moreover, a second order expansion of the inverse of the expected value of $(x^{*\prime}x^{*})/NT$ given in (A8) gives

$$\left[E\left(\frac{x^{*\prime}x^{*}}{NT}\right)\right]^{-1}=\frac{(1-\alpha^{2})}{\sigma^{2}}\left[1+\frac{1}{T}\frac{(1+\alpha)}{(1-\alpha)}\right]+O(T^{-2}) \tag{A15}$$

and therefore

$$\left[E\left(\frac{x^{*\prime}x^{*}}{NT}\right)\right]^{-1}\mu_{NT}=-\left(\frac{N}{T}\right)^{1/2}(1+\alpha)-\left(\frac{N}{T^{3}}\right)^{1/2}\frac{(1+\alpha)(\alpha+\alpha^{T})}{(1-\alpha)}+O\left[\left(\frac{N}{T^{5}}\right)^{1/2}\right]. \tag{A16}$$

The second result of the theorem follows from noticing that when $N/T^{3}\to 0$ the second term of the right-hand side in the expression above is also negligible.

A.2. *GMM*

The following two preliminary lemmae collect the basic building blocks that will be used in proving Lemma 2 and Theorem 2.

LEMMA C1: *Let κ_3 and κ_4 be the third and fourth-order cumulants of v_{it}. Also let d_t and d_s be $N\times 1$ vectors containing the diagonal elements of M_t and M_s, respectively, so that $\operatorname{tr}(M_t)=d_t'\iota=t$, $\operatorname{tr}(M_s)=d_s'\iota=s$, and $d_t'd_s\leq\min(t,s)$. Then, under Assumption A1, for $l\geq r\geq t$, $p\geq q\geq s$, and $t\geq s$:*

$$\operatorname{cov}(v_l'M_tv_r,v_p'M_sv_q)=\begin{cases}2\sigma^{4}s+\kappa_{4}E(d_t'd_s)\leq(2\sigma^{4}+\kappa_{4})s & \text{if } l=r=p=q,\\ \kappa_{3}E(d_t'M_sv_q) & \text{if } l=r=p\neq q<t,\\ \sigma^{4}s & \text{if } l=p\neq r=q,\\ 0 & \text{otherwise,}\end{cases} \tag{A17}$$

where $|E(d_t'M_sv_q)|\leq(st)^{1/2}\sigma$.

PROOF: We begin by showing that the following holds:

$$\operatorname{cov}_t(v_l'M_tv_r,v_p'M_sv_q)=\begin{cases}2\sigma^{4}s+\kappa_{4}d_t'd_s & \text{if } l=r=p=q,\\ \kappa_{3}d_t'M_sv_q & \text{if } l=r=p\neq q<t,\\ \sigma^{4}s & \text{if } l=p\neq r=q,\\ 0 & \text{otherwise,}\end{cases} \tag{A18}$$

where $E_t(.)$ denotes an expectation conditional on η_i and $\{v_{i(t-j)}\}_{j=1}^{\infty}$.

We have

$$\operatorname{cov}_t(v_l'M_tv_r,v_p'M_sv_q)=E_t(v_l'M_tv_rv_p'M_sv_q)-E_t(v_l'M_tv_r)E_t(v_p'M_sv_q).$$

If $p<t$, then $v_p'M_sv_q$ is constant and the covariance vanishes.

The conditional mean terms are given by:

$$E_t(v_l'M_tv_r)=\operatorname{tr}[M_tE_t(v_rv_l')]=\begin{cases}\sigma^{2}\operatorname{tr}(M_t)=\sigma^{2}t & \text{if } l=r,\\ 0 & \text{if } l\neq r,\end{cases} \tag{A19}$$

$$E_t(v_p'M_sv_q)=\operatorname{tr}[M_sE_t(v_qv_p')]=\begin{cases}\sigma^{2}\operatorname{tr}(M_s)=\sigma^{2}s & \text{if } p=q,\\ 0 & \text{if } p\neq q.\end{cases} \tag{A20}$$

As for the leading term we have

$$E_t(v_l'M_tv_rv_p'M_sv_q)=\begin{cases}E_t(v_l'M_tv_lv_l'M_sv_l) & \text{if } l=r=p=q,\\ E_t(v_l'M_tv_lv_l')M_sv_q & \text{if } l=r=p\neq q<t,\\ \operatorname{tr}[M_tE_t(v_rv_r')M_sE_t(v_lv_l')] & \text{if } l=p\neq r=q,\\ 0 & \text{otherwise.}\end{cases} \tag{A21}$$

Firstly, note that in view of the form of the mean terms above there is only a nonzero mean-product subtraction in covariances with $l=r=p=q$.

For the first type of nonzero terms,

$$\begin{aligned}E_t(v_l'M_tv_lv_l'M_sv_l)&=\sum_i\sum_j\sum_k\sum_\ell m_{ij}^{(t)}m_{k\ell}^{(s)}E_t(v_{il}v_{jl}v_{kl}v_{\ell l})\\&=(3\sigma^4+\kappa_4)d_t'd_s+\sigma^4\sum_i\sum_{k\neq i}m_{ii}^{(t)}m_{kk}^{(s)}+2\sigma^4\sum_i\sum_{j\neq i}m_{ij}^{(t)}m_{ij}^{(s)}\\&=\kappa_4d_t'd_s+\sigma^4\operatorname{tr}(M_t)\operatorname{tr}(M_s)+2\sigma^4\operatorname{tr}(M_tM_s)\\&=\kappa_4d_t'd_s+\sigma^4ts+2\sigma^4s,\end{aligned} \tag{A22}$$

where $m_{ij}^{(t)}$ and $m_{k\ell}^{(s)}$ denote elements of M_t and M_s, respectively. Subtracting the product of means, the result follows.

For the second type,

$$E_t(v_l'M_tv_lv_l')M_sv_q=\kappa_3d_t'M_sv_q, \tag{A23}$$

and finally

$$\operatorname{tr}[M_tE_t(v_rv_r')M_sE_t(v_lv_l')]=\sigma^4\operatorname{tr}(M_tM_s)=\sigma^4s. \tag{A24}$$

Given (A18), the unconditional covariances follow from

$$\operatorname{cov}(v_l'M_tv_r,v_p'M_sv_q)=E[\operatorname{cov}_t(v_l'M_tv_r,v_p'M_sv_q)]+\operatorname{cov}[E_t(v_l'M_tv_r),E_t(v_p'M_sv_q)]$$

and the fact that the second term vanishes.

To prove $[E(d_t'M_sv_q)]^2\le ts\sigma^2$ notice that

$$(d_t'M_sv_q)^2\le(d_t'M_sd_t)(v_q'M_sv_q)\le d_t'd_t(v_q'M_sv_q)\le t(v_q'M_sv_q), \tag{A25}$$

hence

$$E[(d_t'M_sv_q)^2]\le tE(v_q'M_sv_q)=ts\sigma^2. \tag{A26}$$

Finally, since $E[(d_t'M_sv_q)^2]=\operatorname{var}(d_t'M_sv_q)+[E(d_t'M_sv_q)]^2$, it turns out that

$$[E(d_t'M_sv_q)]^2\le ts\sigma^2. \tag{A27}$$

Q.E.D.

LEMMA C2: *Under Assumptions A1, A2, and A3, as $T\to\infty$ regardless of whether $N\to\infty$ or is fixed we have*

$$\frac{1}{NT}\sum_{t=1}^{T-1}w_{t-1}'M_tw_{t-1}\xrightarrow{m.s.}\frac{\sigma^2}{1-\alpha^2}. \tag{A28}$$

PROOF: Let $w_{t-1} = y_{t-1} - \mu$ where $\mu = (\mu_1, \ldots, \mu_N)'$ and $\mu_i = \eta_i/(1-\alpha)$ with variance σ_μ^2. Also let μ_t^* be the $N \times 1$ vector of errors of the population linear projection of μ on Z_t:

$$\mu_t^* = \mu - Z_t \gamma_t \tag{A29}$$

where $\gamma_t = [E(z_{it} z_{it}')]^{-1} E(z_{it}\mu_i)$. Also letting V_t be the $t \times t$ autoregressive matrix whose (j,k) element is given by $\alpha^{|j-k|}/(1-\alpha^2)$, and $\lambda = \sigma_\mu^2/\sigma^2$:

$$\gamma_t = \lambda(\lambda \iota_t \iota_t' + V_t)^{-1}\iota_t = \frac{\lambda}{1+\lambda \iota_t' V_t^{-1}\iota_t} V_t^{-1}\iota_t. \tag{A30}$$

Hence, using the triangular decomposition of V_t^{-1}, we obtain the following expression for the ith component of μ_t^*:

$$\begin{aligned} \mu_{it}^* &= \mu_i - z_{it}'\gamma_t = \mu_i(1-\iota_t'\gamma_t) - (w_{i0}, \ldots, w_{i(t-1)})\gamma_t \\ &= \frac{1}{1+\lambda\iota_t'V_t^{-1}\iota_t}\left[\mu_i - \lambda(w_{i0}, \ldots, w_{i(t-1)})V_t^{-1}\iota_t\right] \\ &= \frac{1}{1+\lambda\iota_t'V_t^{-1}\iota_t}\left[\mu_i - \lambda(1-\alpha^2)w_{i0} - \lambda(1-\alpha)(v_{i1} + \cdots + v_{i(t-1)})\right] \end{aligned} \tag{A31}$$

where

$$\iota_t' V_t^{-1}\iota_t = 1 - \alpha^2 + (t-1)(1-\alpha)^2. \tag{A32}$$

Note that $\iota_t'V_t^{-1}\iota_t = O(t)$ and that μ_{it}^* is a linear combination of $(t+1)$ independent random variables. Thus,

$$E(\mu_{it}^{*2}) = \frac{\sigma_\mu^2}{1+\lambda[1-\alpha^2+(t-1)(1-\alpha)^2]} = O\left(\frac{1}{t}\right). \tag{A33}$$

Moreover, given the existence of fourth-order moments of μ_i and v_{it}, we also have

$$E(\mu_{it}^{*4}) = O\left(\frac{1}{t^2}\right). \tag{A34}$$

Now consider the decomposition:

$$\begin{aligned} w_{t-1}'M_t w_{t-1} &= w_{t-1}'w_{t-1} - w_{t-1}'(I_N - M_t)w_{t-1} \\ &= w_{t-1}'w_{t-1} - \mu_t^{*\prime}(I_N - M_t)\mu_t^*, \end{aligned} \tag{A35}$$

where the second equality follows from the fact that $w_{t-1} = y_{t-1} - Z_t\gamma_t - \mu_t^*$ and $(I_N - M_t)\times (y_{t-1} - Z_t\gamma_t) = 0$.

Therefore,

$$\frac{1}{N(T-1)}\sum_{t=1}^{T-1} E(w_{t-1}'M_t w_{t-1}) = E(w_{i(t-1)}^2) + \frac{1}{N(T-1)}\sum_{t=1}^{T-1} E[\mu_t^{*\prime}(I_N - M_t)\mu_t^*]. \tag{A36}$$

Moreover,

$$\mu_t^{*\prime}(I_N - M_t)\mu_t^* \le \lambda_{\max}(I_N - M_t)(\mu_t^{*\prime}\mu_t^*), \tag{A37}$$

where $\lambda_{\max}(I_N - M_t)$ denotes the maximum eigenvalue of $(I_N - M_t)$, which is equal to 1 because $(I_N - M_t)$ is idempotent,[8] so that

$$\frac{1}{NT}\sum_{t=1}^{T-1} E[\mu_t^{*\prime}(I_N - M_t)\mu_t^*] \le \frac{1}{NT}\sum_{t=1}^{T-1} E(\mu_t^{*\prime}\mu_t^*) = \frac{1}{T}\sum_{t=1}^{T-1} E(\mu_{it}^{*2}) = \frac{1}{T}O(\log T) \to 0. \tag{A38}$$

[8] In a case in which $N < T-1$ and $M_t = Z_t(Z_t'Z_t)^+ Z_t'$, for values of t such that $t > N$ we have $M_t = I_N$ so that $\mu_t^{*\prime}(I_N - M_t)\mu_t^* = 0$.

Hence

$$\text{(A39)} \qquad \frac{1}{NT}\sum_{t=1}^{T-1} E(w'_{t-1}M_t w_{t-1}) \to E(w^2_{i(t-1)}) = \frac{\sigma^2}{1-\alpha^2}.$$

To establish mean-square convergence, we now show that the variances of $(NT)^{-1}\sum_{t=1}^{T-1} w'_{t-1}w_{t-1}$ and $(NT)^{-1}\sum_{t=1}^{T-1}\mu_t^{*\prime}\mu_t^*$ tend to zero. First,

$$\text{(A40)} \qquad \operatorname{var}\left(\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}w_{t-1}\right) = \frac{1}{N}\operatorname{var}\left(\frac{1}{T}\sum_{t=1}^{T-1} w^2_{i(t-1)}\right) = \frac{1}{N}O\left(\frac{1}{T}\right) \to 0.$$

Finally,

$$\text{(A41)} \qquad \begin{aligned} \operatorname{var}\left(\frac{1}{NT}\sum_{t=1}^{T-1}\mu_t^{*\prime}\mu_t^*\right) &= \frac{1}{N}\operatorname{var}\left(\frac{1}{T}\sum_{t=1}^{T-1}\mu_{it}^{*2}\right) \\ &= \frac{1}{N}\left[\frac{1}{T^2}\sum_t \operatorname{var}(\mu_{it}^{*2}) + \frac{2}{T^2}\sum_s\sum_{t>s}\operatorname{cov}(\mu_{it}^{*2},\mu_{is}^{*2})\right] \\ &\le \frac{1}{N}\left[\frac{1}{T^2}\sum_t O\left(\frac{1}{t^2}\right) + \frac{2}{T^2}\sum_s\sum_{t>s} O\left(\frac{1}{t}\right)O\left(\frac{1}{s}\right)\right] \to 0. \end{aligned} \qquad \textit{Q.E.D.}$$

Lemma 2

We shall use the decomposition:

$$\text{(A42)} \qquad \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*\prime}M_t v_t^* = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w'_{t-1}M_t v_t^* - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t\tilde{v}'_{tT}M_t v_t^*$$

where

$$\text{(A43)} \qquad x_t^* = \psi_t w_{t-1} - c_t\tilde{v}_{tT},$$

$$\text{(A44)} \qquad \psi_t = c_t\left(1 - \frac{\alpha\phi_{T-t}}{T-t}\right),$$

$$\text{(A45)} \qquad \tilde{v}_{tT} = \frac{1}{T-t}(\phi_{T-t}v_t + \cdots + \phi_1 v_{T-1}),$$

and $\phi_j = (1-\alpha^j)/(1-\alpha)$, so that $E_t(x_{it}^*) = \psi_t w_{i(t-1)}$.

Proof of (24): Only the second term in the right-hand side of (A42) has nonzero mean, so that

$$\text{(A46)} \qquad E(x^{*\prime}Mv^*) = -\sum_{t=1}^{T-1} E(c_t\tilde{v}'_{tT}M_t v_t^*)$$

and

$$\text{(A47)} \qquad \begin{aligned} E(c_t\tilde{v}'_{tT}M_t v_t^*) &= \frac{1}{(T-t+1)}E\left[\left(v_t - \frac{(v_{t+1}+\cdots+v_T)}{(T-t)}\right)' M_t(\phi_{T-t}v_t + \cdots + \phi_1 v_{T-1})\right] \\ &= \frac{t\sigma^2}{(T-t+1)}\left(\phi_{T-t} - \frac{\phi_{T-t-1}+\cdots+\phi_1}{T-t}\right) \\ &= \frac{\sigma^2}{(1-\alpha)}\frac{t}{(T-t+1)}\left(\frac{\phi_{T-t}}{T-t} - \alpha^{T-t}\right) = \frac{\sigma^2 t}{(1-\alpha)}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right), \end{aligned}$$

where we have used the fact that $E(v'_{t+j}M_t v_{t+k})$ equals $t\sigma^2$ for $j=k$ and zero for $j\neq k$, and also that $\phi_1+\cdots+\phi_{j-1}=(j-\phi_j)/(1-\alpha)$ and $\phi_j=\phi_{j-1}+\alpha^{j-1}$. Finally, adding terms and changing the index of the sum to $s=T-t+1$, we get

$$\text{(A48)}\qquad \begin{aligned} E(x^{*\prime}Mv^*) &= -\frac{\sigma^2}{(1-\alpha)}\sum_{s=2}^{T}(T-s+1)\left(\frac{\phi_{s-1}}{s-1}-\frac{\phi_s}{s}\right)\\ &= -\frac{\sigma^2}{(1-\alpha)}\left(\sum_{s=1}^{T-1}(T-s)\frac{\phi_s}{s}-\sum_{s=2}^{T}(T-s+1)\frac{\phi_s}{s}\right)\\ &= -\frac{\sigma^2}{(1-\alpha)}\left(\sum_{s=1}^{T-1}(T-s)\frac{\phi_s}{s}-\sum_{s=2}^{T}(T-s)\frac{\phi_s}{s}-\sum_{s=2}^{T}\frac{\phi_s}{s}\right)\\ &= -\frac{\sigma^2}{(1-\alpha)}\left((T-1)-\sum_{s=2}^{T}\frac{\phi_s}{s}\right)=-\frac{T\sigma^2}{(1-\alpha)}\left[1-\frac{1}{T(1-\alpha)}\sum_{t=1}^{T}\frac{(1-\alpha^t)}{t}\right]. \end{aligned}$$

PROOF OF (25): Letting $\bar{v}_{tT}=(v_t+\cdots+v_T)/(T-t+1)$, we further decompose the two terms of (A42) as follows:[9]

$$\text{(A49)}\qquad \frac{1}{\sqrt{NT}}x^{*\prime}Mv^*=\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}w'_{t-1}M_t v_t-\Upsilon_{11NT}-\Upsilon_{12NT}\right)-(\Upsilon_{21NT}-\Upsilon_{22NT})$$

where

$$\text{(A50)}\qquad \Upsilon_{11NT}=\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}w'_{t-1}M_t\bar{v}_{tT},$$

$$\text{(A51)}\qquad \Upsilon_{12NT}=\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\frac{c_t\alpha\phi_{T-t}}{(T-t)}w'_{t-1}M_t v_t^*,$$

$$\text{(A52)}\qquad \Upsilon_{21NT}=\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\bar{v}'_{tT}M_t v_t,$$

$$\text{(A53)}\qquad \Upsilon_{22NT}=\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\bar{v}'_{tT}M_t\bar{v}_{tT}.$$

The variance of the leading term satisfies:

$$\text{(A54)}\qquad \operatorname{var}\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}w'_{t-1}M_t v_t\right)=\frac{1}{NT}\sum_{t=1}^{T-1}\operatorname{var}(w'_{t-1}M_t v_t)=\frac{\sigma^2}{NT}\sum_{t=1}^{T-1}E(w'_{t-1}M_t w_{t-1}).$$

This is so because for $t>s$ $\operatorname{cov}(w'_{t-1}M_t v_t, w'_{s-1}M_s v_s)=E[w'_{t-1}M_t E_t(v_t)w'_{s-1}M_s v_s]=0$. Thus, in view of Lemma C2 we have

$$\text{(A55)}\qquad \operatorname{var}\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}w'_{t-1}M_t v_t\right)\to\frac{\sigma^4}{1-\alpha^2}.$$

To prove that this is also the limit of $\operatorname{var}[(NT)^{-1/2}x^{*\prime}Mv^*]$, we now show that the variances of Υ_{11NT}, Υ_{12NT}, Υ_{21NT}, and Υ_{22NT} tend to zero. Firstly, note that

$$\text{(A56)}\qquad \operatorname{var}(\Upsilon_{11NT})=\frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}E(w'_{t-1}M_t\bar{v}_{tT}\bar{v}'_{sT}M_s w_{s-1}).$$

[9] Note that $v_t^*=(v_t-\bar{v}_{tT})/c_t$.

For $t \geq s$,

$$E(w'_{t-1}M_t\bar{v}_{tT}\bar{v}'_{sT}M_s w_{s-1}) = E[w'_{t-1}M_t E_t(\bar{v}_{tT}\bar{v}'_{sT})M_s w_{s-1}] = \frac{\sigma^2 E(w'_{t-1}M_s w_{s-1})}{(T-s+1)}$$
$$= \frac{\sigma^2}{(T-s+1)} E[E_s(w'_{t-1})M_s w_{s-1}]$$
$$= \frac{\sigma^2}{(T-s+1)} \alpha^{t-s} E(w'_{s-1}M_s w_{s-1}). \tag{A57}$$

Also,

$$E(w'_{s-1}M_s w_{s-1}) \leq E(w'_0 M_1 w_0) \quad \text{for} \quad s \geq 1. \tag{A58}$$

Therefore,

$$\operatorname{var}(\Upsilon_{11NT}) \leq \frac{\sigma^2}{NT} E(w'_0 M_1 w_0)\left[\left(\frac{1}{T} + \frac{1}{T-1} + \cdots + \frac{1}{2}\right) + \frac{2}{T}(\alpha + \cdots + \alpha^{T-2}) + \frac{2}{T-1}(\alpha + \cdots + \alpha^{T-3}) + \cdots + \frac{2}{3}\alpha\right]$$
$$= \frac{\sigma^2}{NT}\frac{E(w'_0 M_1 w_0)}{(1-\alpha)}\left[(1+\alpha)\left(\frac{1}{2} + \cdots + \frac{1}{T}\right) - 2\left(\frac{\alpha}{2} + \cdots + \frac{\alpha^{T-1}}{T}\right)\right]$$
$$\leq \frac{\sigma^2}{T(1-\alpha)}[E(w_{i0}^2) + E(\mu_{i1}^{*2})]\left[(1+\alpha)\left(\frac{1}{2} + \cdots + \frac{1}{T}\right) - 2\left(\frac{\alpha}{2} + \cdots + \frac{\alpha^{T-1}}{T}\right)\right]$$
$$\to 0. \tag{A59}$$

Next,

$$\operatorname{var}(\Upsilon_{12NT}) = \frac{1}{NT}\operatorname{var}\left(\sum_{t=1}^{T-1}\frac{c_t \alpha \phi_{T-t}}{(T-t)} w'_{t-1}M_t v_t^*\right)$$
$$= \frac{1}{NT}\sum_{t=1}^{T-1}\frac{\alpha^2\phi_{T-t}^2}{(T-t)(T-t+1)}\operatorname{var}(w'_{t-1}M_t v_t^*)$$
$$= \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\frac{\alpha^2\phi_{T-t}^2}{(T-t)(T-t+1)} E(w'_{t-1}M_t w_{t-1})$$
$$\leq \frac{\sigma^2 E(w'_0 M_1 w_0)}{N}\frac{1}{T}\sum_{t=1}^{T-1}\frac{\alpha^2\phi_{T-t}^2}{(T-t)(T-t+1)} \to 0. \tag{A60}$$

This is so because

$$\operatorname{var}(w'_{t-1}M_t v_t^*) = E(w'_{t-1}M_t v_t^* v_t^{*\prime} M_t w_{t-1})$$
$$= E[w'_{t-1}M_t E_t(v_t^* v_t^{*\prime})M_t w_{t-1}] = \sigma^2 E(w'_{t-1}M_t w_{t-1}), \tag{A61}$$

and the covariance terms are zero. In effect, for $t > s$,

$$\operatorname{cov}(w'_{t-1}M_t v_t^*, w'_{s-1}M_s v_s^*) = E(w'_{t-1}M_t v_t^* v_s^{*\prime} M_s w_{s-1})$$
$$= E[w'_{t-1}M_t E_t(v_t^* v_s^{*\prime})M_s w_{s-1}] = 0 \tag{A62}$$

since $E_t(v_t^* v_s^{*\prime}) = 0$.

Turning to consider the variance of Υ_{21NT}, in view of Lemma C1, the only nonzero terms are as follows:

$$\text{(A63)} \qquad \operatorname{var}(\Upsilon_{21NT}) = \frac{1}{NT}\operatorname{var}\left[\sum_{t=1}^{T-1}\frac{1}{(T-t)}v_t'M_t(\phi_{T-t}v_t+\cdots+\phi_1 v_{T-1})\right] = a_{0NT}+a_{1NT}$$

where

$$\text{(A64)} \qquad a_{0NT} = \frac{1}{NT}\sum_{t=1}^{T-1}\frac{1}{(T-t)^2}\left[\phi_{T-t}^2\operatorname{var}(v_t'M_tv_t)+\cdots+\phi_1^2\operatorname{var}(v_t'M_tv_{T-1})\right]$$
$$= \frac{1}{NT}\sum_{t=1}^{T-1}\frac{1}{(T-t)^2}\left\{\phi_{T-t}^2[2\sigma^4 t+\kappa_4 E(d_t'd_t)]+(\phi_{T-t-1}^2+\cdots+\phi_1^2)\sigma^4 t\right\}$$

and

$$\text{(A65)} \qquad a_{1NT} = \frac{2}{NT}\sum_{t=1}^{T-2}\left[\frac{\phi_{T-t-1}^2\operatorname{cov}(v_t'M_tv_{t+1}, v_{t+1}'M_{t+1}v_{t+1})}{(T-t)(T-t-1)}\right.$$
$$\left. +\cdots+\frac{\phi_1^2\operatorname{cov}(v_t'M_tv_{T-1}, v_{T-1}'M_{T-1}v_{T-1})}{(T-t)}\right]$$
$$= \frac{2}{NT}\sum_{t=1}^{T-2}\left[\frac{\phi_{T-t-1}^2\kappa_3 E(d_{t+1}'M_tv_t)}{(T-t)(T-t-1)}+\cdots+\frac{\phi_1^2\kappa_3E(d_{T-1}'M_tv_t)}{(T-t)}\right].$$

Using Lemma C1 and the fact that $\phi_j^2 < 1/(1-\alpha)^2$ for all j,

$$\text{(A66)} \qquad a_{0NT} \le \frac{1}{NT}\sum_{t=1}^{T-1}\frac{t}{(T-t)^2}\left[\phi_{T-t}^2(2\sigma^4+\kappa_4)+(\phi_{T-t-1}^2+\cdots+\phi_1^2)\sigma^4\right]$$
$$\le \frac{1}{(1-\alpha)^2}\frac{1}{NT}\sum_{t=1}^{T-1}\frac{t}{(T-t)^2}\left[(2\sigma^4+\kappa_4)+(T-t-1)\sigma^4\right]$$
$$= \frac{\sigma^4}{(1-\alpha)^2}\frac{1}{NT}\sum_{t=1}^{T-1}\frac{t}{(T-t)}+\frac{(\sigma^4+\kappa_4)}{(1-\alpha)^2}\frac{1}{NT}\sum_{t=1}^{T-1}\frac{t}{(T-t)^2} = O\left(\frac{\log T}{N}\right).$$

The latter follows from

$$\text{(A67)} \qquad \sum_{t=1}^{T-1}\frac{t}{(T-t)} = \sum_{s=1}^{T-1}\frac{T-s}{s} = T\sum_{s=2}^{T-1}\frac{1}{s}+1 = O(T\log T).$$

Moreover, in view of the triangle inequality, and the fact that $|E(d_{t+j}'M_tv_t)| \le (t+j)\sigma$:

$$\text{(A68)} \qquad |a_{1NT}| \le \frac{2|\kappa_3|\sigma}{(1-\alpha)^2}\frac{1}{NT}\sum_{t=1}^{T-2}\frac{1}{(T-t)}\left(\frac{(t+1)}{(T-t-1)}+\cdots+\frac{(T-1)}{1}\right)$$
$$\le \frac{2|\kappa_3|\sigma}{(1-\alpha)^2}\frac{1}{NT}\left(\frac{1}{2}+\cdots+\frac{1}{T-1}\right)\left[T\left(\frac{1}{2}+\cdots+\frac{1}{T-1}\right)+1\right] = O\left(\frac{(\log T)^2}{N}\right).$$

Finally, we must consider the term Υ_{22NT}. We begin by establishing the order of magnitude of $\operatorname{var}(\bar{v}_{tT}'M_t\tilde{v}_{tT})$. Let $\tilde{v}_{itT}$ and $\bar{v}_{itT}$ denote the ith elements of $\tilde{v}_{tT}$ and $\bar{v}_{tT}$, respectively, and let

$$\text{(A69)} \qquad r_{itT} = (T-t)^{1/2}\left(c_t^2\tilde{v}_{itT}-\frac{1}{(1-\alpha)}\bar{v}_{itT}\right)$$
$$\equiv -\frac{c_t^2}{(T-t)^{1/2}}\frac{(\alpha^{T-t}v_{it}+\cdots+\alpha v_{i(T-1)}+v_{iT})}{(1-\alpha)}$$

so that $\mathrm{var}(r_{itT}) = O[1/(T-t)]$. Now let us consider the order of magnitude of the variances of the terms in the expression

$$(T-t)c_t^2\tilde{v}_{itT}\bar{v}_{itT} = \frac{1}{(1-\alpha)}(T-t)\bar{v}_{itT}^2 + \sqrt{T-t}\bar{v}_{itT}r_{itT}. \tag{A70}$$

Clearly, $\mathrm{var}[(T-t)\bar{v}_{itT}^2] = O(1)$. Moreover, since $\mathrm{var}(\sqrt{T-t}\bar{v}_{itT}) = O(1)$ we have that $\mathrm{var}(\sqrt{T-t}\bar{v}_{itT}r_{itT}) = o(1)$. Therefore, $\mathrm{var}((T-t)c_t^2\tilde{v}_{itT}\bar{v}_{itT}) = O(1)$ and also

$$\mathrm{var}(\tilde{v}_{itT}\bar{v}_{itT}) = O\left(\frac{1}{(T-t)^2}\right). \tag{A71}$$

Hence, since the elements of $\bar{v}_{tT}$ and $\tilde{v}_{tT}$ are iid,

$$\mathrm{var}(\bar{v}_{tT}'\tilde{v}_{tT}) = \mathrm{var}(\bar{v}_{1tT}\tilde{v}_{1tT}) + \cdots + \mathrm{var}(\bar{v}_{NtT}\tilde{v}_{NtT}) = O\left(\frac{N}{(T-t)^2}\right) \tag{A72}$$

and also $\mathrm{var}_t(\bar{v}_{tT}'M_t\tilde{v}_{tT}) = O[t/(T-t)^2]$ so that[10]

$$\mathrm{var}(\bar{v}_{tT}'M_t\tilde{v}_{tT}) = O\left(\frac{t}{(T-t)^2}\right). \tag{A73}$$

The variance of Υ_{22NT} is given by

$$\mathrm{var}(\Upsilon_{22NT}) = \frac{1}{NT}\mathrm{var}\left(\sum_{t=1}^{T-1}\bar{v}_{tT}'M_t\tilde{v}_{tT}\right) = b_{0NT} + b_{1NT} \tag{A74}$$

where

$$b_{0NT} = \frac{1}{NT}\sum_{t=1}^{T-1}\mathrm{var}(\bar{v}_{tT}'M_t\tilde{v}_{tT}) = \frac{1}{NT}O\left(\sum_{t=1}^{T-1}\frac{t}{(T-t)^2}\right) = O\left(\frac{1}{N}\right) \tag{A75}$$

and

$$b_{1NT} = \frac{2}{NT}\sum_s\sum_{t>s}\mathrm{cov}(\bar{v}_{tT}'M_t\tilde{v}_{tT}, \bar{v}_{sT}'M_s\tilde{v}_{sT}). \tag{A76}$$

Now, since $|\mathrm{cov}(\bar{v}_{tT}'M_t\tilde{v}_{tT}, \bar{v}_{sT}'M_s\tilde{v}_{sT})| \le \mathrm{var}(\bar{v}_{tT}'M_t\tilde{v}_{tT})^{1/2}\mathrm{var}(\bar{v}_{sT}'M_s\tilde{v}_{sT})^{1/2}$,

$$\begin{aligned} |b_{1NT}| &\le \frac{2}{NT}\sum_s\sum_{t>s}|\mathrm{cov}(\bar{v}_{tT}'M_t\tilde{v}_{tT}, \bar{v}_{sT}'M_s\tilde{v}_{sT})| \\ &\le \frac{2}{NT}\sum_s\sum_{t>s}O\left(\frac{\sqrt{t}}{T-t}\right)O\left(\frac{\sqrt{s}}{T-s}\right) \\ &\le \frac{2}{NT}\sum_t O\left(\frac{t}{T-t}\right)\sum_s O\left(\frac{1}{T-s}\right) = O\left(\frac{(\log T)^2}{N}\right). \end{aligned} \tag{A77}$$

Provided $(\log T)^2/N \to 0$, b_{1NT} also converges to zero and it follows that $\mathrm{var}(\Upsilon_{22NT}) \to 0$.

[10] This follows from the fact that $M_t = C_t'\Lambda_t C_t$ where C_t is an orthogonal matrix and

$$\Lambda_t = \begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix}.$$

PROOF OF (26): Using the decomposition in (A43) we have

$$\text{(A78)}\qquad \frac{1}{NT}(x^{*\prime}Mx^*)=\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w_{t-1}'M_t w_{t-1}-\frac{2}{NT}\sum_{t=1}^{T-1}c_t\psi_t w_{t-1}'M_t\tilde{v}_{tT}+\frac{1}{NT}\sum_{t=1}^{T-1}c_t^2\tilde{v}_{tT}'M_t\tilde{v}_{tT}.$$

The first term of the right-hand side converges in probability to $\sigma^2/(1-\alpha^2)$ in view of Lemma C2 and the fact that $\psi_t^2=1-O[1/(T-t)]$. The second term has zero mean and its variance is shown to tend to zero using similar arguments as used for Υ_{11NT}. The third term is analogous to Υ_{22NT}. Its mean is given by

$$\begin{aligned}\text{(A79)}\qquad E\left(\frac{1}{NT}\sum_{t=1}^{T-1}c_t^2\tilde{v}_{tT}'M_t\tilde{v}_{tT}\right)&=\frac{1}{NT}\sum_{t=1}^{T-1}c_t^2E\{\mathrm{tr}[M_tE_t(\tilde{v}_{tT}\tilde{v}_{tT}')]\}\\&=\frac{1}{NT}\sum_{t=1}^{T-1}c_t^2tE(\tilde{v}_{itT}^2)\\&=\frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\frac{t}{(T-t)(T-t+1)}(\phi_{T-t}^2+\cdots+\phi_1^2)\\&=O\left(\frac{\log T}{N}\right),\end{aligned}$$

so that it converges to zero as long as $\log T/N\to 0$. Finally, its variance is shown to tend to zero in the same way as done for Υ_{22NT}.

PROOF OF (27): Using the fact that $E_t(v_t^*v_t^{*\prime})=\sigma^2 I_N$ and $\mathrm{tr}(M_t)=t$, we have

$$\text{(A80)}\qquad E\left(\frac{1}{NT}v^{*\prime}Mv^*\right)=\frac{1}{NT}\sum_{t=1}^{T-1}E\{\mathrm{tr}[M_tE_t(v_t^*v_t^{*\prime})]\}=\frac{\sigma^2}{NT}\sum_{t=1}^{T-1}t\to\sigma^2\frac{c}{2}.$$

Moreover, using $v_t^*=(v_t-\bar{v}_{tT})/c_t$,

$$\text{(A81)}\qquad \frac{1}{NT}v^{*\prime}Mv^*=\frac{1}{NT}\sum_{t=1}^{T-1}c_t^{-2}v_t'M_tv_t-\frac{2}{NT}\sum_{t=1}^{T-1}c_t^{-2}\bar{v}_{tT}'M_tv_t+\frac{1}{NT}\sum_{t=1}^{T-1}c_t^{-2}\bar{v}_{tT}'M_t\bar{v}_{tT}.$$

In view of Lemma C1 $\mathrm{var}(v_t'M_tv_t)\le(2\sigma^4+\kappa_4)t$ and $\mathrm{cov}(v_t'M_tv_t,v_s'M_sv_s)=0$ for $t\neq s$. Hence, the variance of the first term satisfies

$$\begin{aligned}\text{(A82)}\qquad \mathrm{var}\left(\frac{1}{NT}\sum_{t=1}^{T-1}c_t^{-2}v_t'M_tv_t\right)&=\frac{1}{N^2T^2}\sum_{t=1}^{T-1}c_t^{-4}\mathrm{var}(v_t'M_tv_t)\\&\le(2\sigma^4+\kappa_4)\frac{1}{N^2T^2}\sum_{t=1}^{T-1}t\left(1+\frac{1}{T-t}\right)^2\to 0.\end{aligned}$$

The second and third terms of the right-hand side of (A81) are analogous to Υ_{21NT} and Υ_{22NT}, and their variances are shown to tend to zero using similar arguments as used for those terms.

Theorem 2

PROOF OF (28): Consistency follows directly from Lemma 2: From (24) and (25) $(x^{*'}Mv^*)/NT$ converges to zero in mean square, and therefore also in probability, whereas from (26) $(x^{*'}Mx^*)/NT$ is bounded in probability.

PROOF OF (29): In the right-hand side of expression (A49), $-(\Upsilon_{21NT}-\Upsilon_{22NT})$ is the only term that has nonzero mean, which is given by (24) scaled by $(NT)^{-1/2}$. Since we showed that the variances of Υ_{11NT}, Υ_{12NT}, Υ_{21NT}, and Υ_{22NT} tend to zero, it follows that

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*\prime} M_t v_t^* - \mu_{NT}^{\dagger} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' M_t v_t + o_p(1) \tag{A83}$$

$$= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' v_t - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}'(I_N - M_t)v_t + o_p(1) \tag{A84}$$

where

$$\mu_{NT}^{\dagger} = (NT)^{-1/2} E(x^{*\prime} M v^*) = -\left(\frac{T}{N}\right)^{1/2} \frac{\sigma^2}{(1-\alpha)} + (NT)^{-1/2}\frac{\sigma^2}{(1-\alpha)^2}\sum_{t=1}^{T}\frac{(1-\alpha^t)}{t}. \tag{A85}$$

Also, the second term in (A84) is $o_p(1)$ since it has zero mean and, using Lemma C2, its variance satisfies

$$\begin{aligned}\operatorname{var}\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}'(I_N - M_t)v_t\right) &= \frac{1}{NT}\sum_{t=1}^{T-1}\operatorname{var}[w_{t-1}'(I_N - M_t)v_t] \\ &= \frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E[w_{t-1}'(I_N - M_t)w_{t-1}] \\ &= \frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E[\mu_t^{*\prime}(I_N - M_t)\mu_t^*] \\ &= \frac{1}{T}O(\log T) \to 0.\end{aligned} \tag{A86}$$

Therefore, using (A12):

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*\prime} M_t v_t^* - \mu_{NT}^{\dagger} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' v_t + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4}{(1-\alpha^2)}\right). \tag{A87}$$

Moreover, in view of (26), by Cramer's theorem,

$$\left(\frac{x^{*\prime} M x^*}{NT}\right)^{-1}[(NT)^{-1/2} x^{*\prime} M v^* - \mu_{NT}^{\dagger}] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2) \tag{A88}$$

or

$$\sqrt{NT}(\hat{\alpha}_{GMM} - \alpha) - \left(\frac{x^{*\prime} M x^*}{NT}\right)^{-1}\mu_{NT}^{\dagger} \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2). \tag{A89}$$

The result follows from noticing that, since $\mu_{NT}^{\dagger} = O(\sqrt{T/N})$,

$$\left(\frac{x^{*\prime} M x^*}{NT}\right)^{-1}\mu_{NT}^{\dagger} = \frac{(1-\alpha^2)}{\sigma^2}\mu_{NT}^{\dagger} + o_p(1) = -\left(\frac{T}{N}\right)^{1/2}(1+\alpha) + O\left(\frac{\log T}{\sqrt{NT}}\right) + o_p(1). \tag{A90}$$

A.3. *LIML*

Lemma 3

PROOF OF (31): Using the results in Lemmae 1 and 2, simple algebra reveals that

$$\frac{1}{NT}(W^{*\prime} W^*) \xrightarrow{p} \frac{\sigma^2}{(1-\alpha^2)}\begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}, \tag{A91}$$

$$\frac{1}{NT}(W^{*\prime} M W^*) \xrightarrow{p} \frac{\sigma^2}{(1-\alpha^2)}\begin{pmatrix} \alpha^2 + \frac{c}{2}(1-\alpha^2) & \alpha \\ \alpha & 1 \end{pmatrix}. \tag{A92}$$

Since $\hat{\ell} = \min$ eigenvalue $[W^{*'}MW^{*}(W^{*'}W^{*})]^{-1}$, due to the continuity of the min eigenvalue function, $\hat{\ell}$ converges in probability to the smallest root of the equation

$$\text{(A93)} \qquad \det\left[\begin{pmatrix} \alpha^2 + \frac{c}{2}(1-\alpha^2) & \alpha \\ \alpha & 1 \end{pmatrix} - \ell \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}\right] = 0$$

or equivalently

$$\text{(A94)} \qquad (1-\alpha^2)(1-\ell)\left(\frac{c}{2} - \ell\right) = 0.$$

Thus, the roots are 1 and $(c/2)$, with the latter being the smallest provided $c \leq 2$.

Theorem 3

PROOF OF (33): From Lemmae 1, 2, and 3

$$\text{(A95)} \qquad (NT)^{-1}(x^{*'}Mv^{*} - \hat{\ell}x^{*'}v^{*}) \xrightarrow{p} 0$$

and

$$\text{(A96)} \qquad (NT)^{-1}(x^{*'}Mx^{*} - \hat{\ell}x^{*'}x^{*}) \xrightarrow{p} \left(1 - \frac{c}{2}\right)\frac{\sigma^2}{1-\alpha^2}$$

from which consistency of $\hat{\alpha}_{LIML}$ follows.

PROOF OF (34): Turning to asymptotic normality, in view of Lemma 3

$$\text{(A97)} \qquad (NT)^{-1/2}(x^{*'}Mv^{*} - \hat{\ell}x^{*'}v^{*}) - (\mu_{NT}^{\dagger} - \hat{\ell}\mu_{NT}) \\ = [(NT)^{-1/2}x^{*'}Mv^{*} - \mu_{NT}^{\dagger}] - \frac{c}{2}[(NT)^{-1/2}x^{*'}v^{*} - \mu_{NT}] + o_p(1).$$

Moreover, in view of (A11) and (A87) the expression above equals

$$\text{(A98)} \qquad \left(1 - \frac{c}{2}\right)(NT)^{-1/2}\sum_{t=1}^{T-1} w_{t-1}' v_t + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, \left(1-\frac{c}{2}\right)^2 \frac{\sigma^4}{(1-\alpha^2)}\right).$$

Now, by Cramer's theorem:

$$\text{(A99)} \qquad \left(\frac{x^{*'}Mx^{*} - \hat{\ell}x^{*'}x^{*}}{NT}\right)^{-1}\left[(NT)^{-1/2}(x^{*'}Mv^{*} - \hat{\ell}x^{*'}v^{*}) - (\mu_{NT}^{\dagger} - \hat{\ell}\mu_{NT})\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2)$$

or

$$\text{(A100)} \qquad \sqrt{NT}(\hat{\alpha}_{LIML} - \alpha) - \left(\frac{x^{*'}Mx^{*} - \hat{\ell}x^{*'}x^{*}}{NT}\right)^{-1}(\mu_{NT}^{\dagger} - \hat{\ell}\mu_{NT}) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2).$$

For $0 < c \leq 2$, the result follows from noticing that

$$\text{(A101)} \qquad \left(\frac{x^{*'}Mx^{*} - \hat{\ell}x^{*'}x^{*}}{NT}\right)^{-1}(\mu_{NT}^{\dagger} - \hat{\ell}\mu_{NT}) = \left[\left(1 - \frac{T}{2N}\right)\frac{\sigma^2}{1-\alpha^2}\right]^{-1}\left(\mu_{NT}^{\dagger} - \frac{T}{2N}\mu_{NT}\right) + o_p(1) \\ = (NT)^{1/2}\frac{(1+\alpha)}{(2N-T)} + O\left(\frac{2N}{(2N-T)}\frac{\log T}{\sqrt{NT}}\right) + o_p(1).$$

For $c = 0$, we have $\mu_{NT}^{\dagger} = o(1)$, $\hat{\ell} = o_p(1)$, and $\mu_{NT} = O[(N/T)^{1/2}]$. Nevertheless, it is still the case that $\hat{\ell}\mu_{NT} = o_p(1)$, which ensures that the asymptotic bias vanishes when $c = 0$. We prove the latter assertion by showing that when $c = 0$

$$\text{(A102)} \qquad \left(\frac{N}{T}\right)^{1/2}\hat{\ell} \xrightarrow{p} 0.$$

Since $\hat{\ell}$ is the minimum of the criterion given in (11), we have

$$\hat{\ell} \leq \frac{v^{*'}Mv^{*}}{v^{*'}v^{*}}. \tag{A103}$$

From the proof of (27) in Lemma 2 it is easy to see that the result

$$\left(\frac{N}{T}\right)^{1/2}\left(\frac{v^{*'}Mv^{*}}{NT}\right) \xrightarrow{p} \sigma^2 \frac{c^{1/2}}{2} \tag{A104}$$

also holds for $c = 0$. Moreover, since from Lemma 1[11]

$$\frac{v^{*'}v^{*}}{NT} \xrightarrow{p} \sigma^2 \tag{A105}$$

with $c = 0$, we have

$$\left(\frac{N}{T}\right)^{1/2}\left(\frac{v^{*'}Mv^{*}}{v^{*'}v^{*}}\right) \xrightarrow{p} 0, \tag{A106}$$

which given the inequality above implies that $(N/T)^{1/2}\hat{\ell} = o_p(1)$.

A.4. *Crude GMM*

Theorem 4

PROOF OF (36): The CIV estimation error is given by

$$\hat{\alpha}_{CIV} - \alpha = \left(\sum_{t=1}^{T-1} \Delta x_{t+1}' M_t \Delta x_{t+1}\right)^{-1} \sum_{t=1}^{T-1} \Delta x_{t+1}' M_t \Delta v_{t+1} \tag{A107}$$

where $\Delta x_{t+1} = \Delta w_t$, and we use

$$\Delta x_{t+1} = -(1-\alpha) w_{t-1} + v_t. \tag{A108}$$

Thus,

$$\begin{aligned} \frac{1}{NT}\sum_{t=1}^{T-1} \Delta x_{t+1}' M_t \Delta v_{t+1} &= -\frac{1}{NT}\sum_{t=1}^{T-1} v_t' M_t v_t + \frac{1}{NT}\sum_{t=1}^{T-1} v_t' M_t v_{t+1} \\ &\quad + (1-\alpha)\frac{1}{NT}\sum_{t=1}^{T-1} w_{t-1}' M_t v_t - (1-\alpha)\frac{1}{NT}\sum_{t=1}^{T-1} w_{t-1}' M_t v_{t+1}. \end{aligned} \tag{A109}$$

Note that the only term in the right-hand side that has nonzero mean is the first one since $E(v_t' M_t v_t) = \sigma^2 t$. From a similar derivation to (A80) and (A82) it turns out that

$$\frac{1}{NT}\sum_{t=1}^{T-1} v_t' M_t v_t \xrightarrow{p} \sigma^2 \frac{c}{2}. \tag{A110}$$

The variance of the second term is seen to tend to zero in a way similar to the first one. Finally, the third and fourth terms are shown to converge to zero in probability as a straightforward application of Lemma C2.

[11] The result (A105) follows immediately from Lemma 1 and the fact that $y^{*\prime}y^{*}/(NT)$ converges to the same probability limit as $x^{*\prime}x^{*}/(NT)$:

$$\frac{v^{*\prime}v^{*}}{NT} = \frac{y^{*\prime}y^{*}}{NT} - \alpha^2\left(\frac{x^{*\prime}x^{*}}{NT}\right) - 2\alpha\left(\frac{x^{*\prime}v^{*}}{NT}\right) \xrightarrow{p} \sigma^2.$$

PROOF OF (37): Next, using the results in the previous proof and Lemma C2:

$$\text{(A111)}\quad \frac{1}{NT}\sum_{t=1}^{T-1}\Delta x'_{t+1}M_t\Delta x_{t+1} = (1-\alpha)^2\frac{1}{NT}\sum_{t=1}^{T-1}w'_{t-1}M_t w_{t-1} + \frac{1}{NT}\sum_{t=1}^{T-1}v'_t M_t v_t - (1-\alpha)\frac{2}{NT}\sum_{t=1}^{T-1}w'_{t-1}M_t v_t \xrightarrow{p} (1-\alpha)^2\frac{\sigma^2}{1-\alpha^2} + \sigma^2\frac{c}{2}.$$

PROOF OF (38): The result follows immediately from the previous two.

A.5. *Random Effects Maximum Likelihood*

EXPRESSION FOR LOG DENSITY (40): The model can be written as

$$\text{(A112)}\quad By_i = \alpha y_{i0}d_i + u_i$$

where B is a $T \times T$ matrix given by

$$B = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ -\alpha & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -\alpha & 1 \end{pmatrix}$$

and $y_i = (y_{i1}, \dots, y_{iT})'$, $d_i = (1, 0, \dots, 0)'$, $u_{it} = \eta_i + v_{it}$, and $u_i = (u_{i1}, \dots, u_{iT})'$.

The conditional density of y_i given y_{i0} can be written as

$$\text{(A113)}\quad f(y_i \mid y_{i0}) = f(u_i \mid y_{i0})\det(B)$$

but $\det(B) = 1$, since B is triangular. Moreover,

$$\text{(A114)}\quad f(u_i \mid y_{i0}) = f(\bar{u}_i, u_i^* \mid y_{i0})|\det(H)|$$

where $H = (\iota_T/T, A')'$ is the triangular transformation matrix that produces $Hu_i = (\bar{u}_i, u_i^{*\prime})'$. Therefore, also $T^{1/2}|\det(H)| = 1$.

From condition A1 $\text{var}(u_i) = \sigma^2(I_T + \lambda_\eta \iota_T \iota'_T)$ where $\lambda_\eta = \sigma^2_\eta/\sigma^2$. Hence,

$$\text{(A115)}\quad \text{var}(Hu_i) = \sigma^2\begin{pmatrix} \frac{1}{T} + \lambda_\eta & 0 \\ 0 & I_{T-1} \end{pmatrix}.$$

Thus, the Gaussian density factors as

$$\text{(A116)}\quad f(\bar{u}_i, u_i^* \mid y_{i0}) = f(\bar{u}_i \mid y_{i0})f(u_i^*)$$

since $E(\bar{u}_i u_i^* \mid y_{i0}) = 0$ and u_i^* is independent of y_{i0}. Then, the result follows from noting that $f(\bar{u}_i \mid y_{i0})$ and $f(u_i^*)$ are, respectively, the univariate and $(T-1)$-variate normal densities $\mathcal{N}(\varphi y_{i0}, \omega^2)$ and $\mathcal{N}(0, \sigma^2 I_{T-1})$.

The zero-mean property of the score $E[\partial \log f(y_{i1}, \dots, y_{iT} \mid y_{i0})/\partial(\alpha, \varphi, \sigma^2, \omega^2)] = 0$ can be written as the following "GLS type" orthogonality conditions:

$$\text{(A117)}\quad E[x_i^{*\prime}(y_i^* - \alpha x_i^*)] = -\frac{\sigma^2}{\omega^2}E[\bar{x}_i(\bar{y}_i - \alpha\bar{x}_i - \varphi y_{i0})],$$

$$\text{(A118)}\quad \frac{1}{\omega^2}E[y_{i0}(\bar{y}_i - \alpha\bar{x}_i - \varphi y_{i0})] = 0,$$

$$\text{(A119)}\quad E[(y_i^* - \alpha x_i^*)'(y_i^* - \alpha x_i^*) - \sigma^2] = 0,$$

$$\text{(A120)}\quad E[(\bar{y}_i - \alpha\bar{x}_i - \varphi y_{i0})^2 - \omega^2] = 0,$$

or equivalently,

(A121) $$E[\widetilde{X}_i'\widetilde{\Omega}^{-1}(\tilde{y}_i - \widetilde{X}_i\beta)] = 0,$$

(A122) $$E[(\tilde{y}_i - \widetilde{X}_i\beta)(\tilde{y}_i - \widetilde{X}_i\beta)'] = \widetilde{\Omega},$$

where

$$\tilde{y}_i = \begin{pmatrix} \bar{y}_i \\ y_i^* \end{pmatrix}, \quad \widetilde{X}_i = \begin{pmatrix} \bar{x}_i & y_{i0} \\ x_i^* & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} \alpha \\ \varphi \end{pmatrix}, \quad \widetilde{\Omega} = \begin{pmatrix} \omega^2 & 0 \\ 0 & \sigma^2 I_{T-1} \end{pmatrix}.$$

Note that under Assumption A2, (A117) multiplied by N corresponds to expression (16).

Theorem 5

CONSISTENCY OF THE RML: From (41) $\hat{\alpha}_{RML}$ is the minimizer of

(A123) $$\log\left[\frac{1}{NT}(y^* - ax^*)'(y^* - ax^*)\right] + \frac{1}{(T-1)}\log\left[\frac{1}{N}(\bar{y} - a\bar{x})'S_0(\bar{y} - a\bar{x})\right].$$

As $T \to \infty$ regardless of the asymptotic behavior of N, the second term in (A123) vanishes so that the limiting criterion is the same as the log limiting criterion for within-groups. Consistency of RML then follows from the consistency of WG as $T \to \infty$. However, unlike WG, RML is also consistent when T is fixed and $N \to \infty$ provided conditions (A117)–(A120) hold (including time series homoskedasticity).

ASYMPTOTIC NORMALITY OF THE RML: The first and second derivatives at $a = \alpha$ of the concentrated log likelihood:

(A124) $$L(a) = -N(T-1)\log[(y^* - ax^*)'(y^* - ax^*)] - N\log[(\bar{y} - a\bar{x})'S_0(\bar{y} - a\bar{x})]$$

are given by

(A125) $$\frac{\partial L(\alpha)}{\partial a} = \left(\frac{v^{*\prime}v^*}{N(T-1)}\right)^{-1}(x^{*\prime}v^*) + \left(\frac{\bar{u}'S_0\bar{u}}{N}\right)^{-1}(\bar{x}'S_0\bar{u}),$$

(A126) $$\frac{1}{NT}\frac{\partial^2 L(\alpha)}{\partial a^2} = -\left(\frac{v^{*\prime}v^*}{N(T-1)}\right)^{-1}\left(\frac{x^{*\prime}x^*}{NT}\right) + 2\left(\frac{v^{*\prime}v^*}{NT}\right)^{-2}\left(\frac{x^{*\prime}v^*}{NT}\right)^2\left(\frac{T-1}{T}\right)$$
$$- \frac{1}{T}\left(\frac{\bar{u}'S_0\bar{u}}{N}\right)^{-1}\left(\frac{\bar{x}'S_0\bar{x}}{N}\right) + \frac{2}{T}\left(\frac{\bar{u}'S_0\bar{u}}{N}\right)^{-2}\left(\frac{\bar{x}'S_0\bar{u}}{N}\right)^2.$$

HESSIAN: We show that as both N and T tend to infinity, regardless of the relative rate of increase:

(A127) $$\frac{1}{NT}\frac{\partial^2 L(\alpha)}{\partial a^2} \xrightarrow{p} -\frac{1}{(1-\alpha^2)}.$$

To verify (A127), first note that from Lemma 1 as $T \to \infty$, regardless of whether N is fixed or tends to infinity,

$$\frac{x^{*\prime}v^*}{NT} \xrightarrow{p} 0, \quad \frac{x^{*\prime}x^*}{NT} \xrightarrow{p} \frac{\sigma^2}{1-\alpha^2}, \quad \frac{v^{*\prime}v^*}{NT} \xrightarrow{p} \sigma^2.$$

Moreover, as both N and T tend to infinity,

(A128) $$\text{plim}\left(\frac{\bar{u}'S_0\bar{u}}{N}\right) = \text{plim}\left(\frac{\bar{u}'\bar{u}}{N}\right) - \left(\text{plim}\,\frac{y_0'y_0}{N}\right)^{-1}\left(\text{plim}\,\frac{\bar{u}'y_0}{N}\right)^2$$
$$= \sigma_\eta^2 - \gamma_{00}^{-1}\gamma_{\eta 0}^2.$$

This is so because $E(\bar{u}'\bar{u}/N) = E(\bar{u}_i^2) = \sigma_\eta^2 + (\sigma^2/T) \to \sigma_\eta^2$, and $\text{var}(\bar{u}'\bar{u}/N) = N^{-1}\text{var}(\bar{u}_i^2) \to 0$, since given finite fourth moments for η_i and v_{it} $\text{var}(\bar{u}_i^2) = O(1)$. Similarly, $E(\bar{u}'y_0/N) = E(\bar{u}_i y_{i0}) = E(\eta_i y_{i0}) = \gamma_{\eta 0}$, and $\text{var}(\bar{u}'y_0/N) = N^{-1}\text{var}(\bar{u}_i y_{i0}) \to 0$, since $\text{var}(\bar{u}_i y_{i0}) = O(1)$.

Using similar arguments we obtain

$$\text{(A129)} \qquad \text{plim}\left(\frac{\bar{x}'S_0\bar{x}}{N}\right) = \text{plim}\left(\frac{\bar{x}'\bar{x}}{N}\right) - \left(\text{plim}\,\frac{y_0'y_0}{N}\right)^{-1}\left(\text{plim}\,\frac{\bar{x}'y_0}{N}\right)^2 = \frac{\sigma_\eta^2}{(1-\alpha)^2} - \gamma_{00}^{-1}\left(\frac{\sigma_\eta^2}{(1-\alpha)^2}\right)^2$$

and

$$\text{(A130)} \qquad \text{plim}\left(\frac{\bar{x}'S_0\bar{u}}{N}\right) = \text{plim}\left(\frac{\bar{x}'\bar{u}}{N}\right) - \left(\text{plim}\,\frac{y_0'y_0}{N}\right)^{-1}\left(\text{plim}\,\frac{\bar{x}'y_0}{N}\right)\left(\text{plim}\,\frac{\bar{u}'y_0}{N}\right) = \frac{\sigma_\eta^2}{(1-\alpha)} - \gamma_{00}^{-1}\left(\frac{\sigma_\eta^2}{(1-\alpha)^2}\right)\gamma_{\eta 0}.$$

SCORE: Now the scaled score can be written as

$$\text{(A131)} \qquad (NT)^{-1/2}\frac{\partial L(\alpha)}{\partial a} = \frac{1}{\sigma^2}(NT)^{-1/2}[x^{*\prime}v^* - E(x^{*\prime}v^*)] + \Psi_{NT} + o_p(1)$$

where

$$\text{(A132)} \qquad \Psi_{NT} = \left(\frac{N}{T}\right)^{1/2}\left[\left(\frac{\bar{u}'S_0\bar{u}}{N}\right)^{-1}\left(\frac{\bar{x}'S_0\bar{u}}{N}\right) + \frac{1}{\sigma^2}E\left(\frac{x^{*\prime}v^*}{N}\right)\right].$$

Moreover, using (A117) we have

$$\text{(A133)} \qquad \Psi_{NT} = \left(\frac{N}{T}\right)^{1/2}\left\{\left(\frac{\bar{u}'S_0\bar{u}}{N}\right)^{-1}\left(\frac{\bar{x}'S_0\bar{u}}{N}\right) - \frac{1}{\omega^2}\left[E(\bar{x}_i\bar{u}_i) - \varphi E(\bar{x}_i y_{i0})\right]\right\}.$$

Note also that in view of (A128) and (A130), as N and T tend to infinity,

$$\left(\frac{\bar{u}'S_0\bar{u}}{N}\right) - \omega^2 \xrightarrow{p} 0, \qquad \left(\frac{\bar{x}'S_0\bar{u}}{N}\right) - [E(\bar{x}_i\bar{u}_i) - \varphi E(\bar{x}_i y_{i0})] \xrightarrow{p} 0.$$

Thus, if N and T tend to infinity, provided $0 \le \lim(N/T) < \infty$, $\Psi_{NT} = o_p(1)$ and from result (20) in Theorem 1

$$\text{(A134)} \qquad (NT)^{-1/2}\frac{\partial L(\alpha)}{\partial a} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{(1-\alpha^2)}\right).$$

Given (A127) and (A134), the asymptotic normality result

$$\sqrt{NT}(\hat{\alpha}_{RML} - \alpha) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2)$$

follows from Theorem 4.1.3 in Amemiya (1985).

TABLE A1

MEDIANS, INTERQUARTILE RANGES, AND MEDIAN ABSOLUTE ERRORS OF THE ESTIMATORS ($N = 100$)[a]

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML
								$T^o = 10$							
median	0.065	0.186	0.196	0.127	0.202	0.318	0.474	0.492	0.348	0.499	0.554	0.724	0.784	0.353	0.796
iqr	0.047	0.067	0.069	0.077	0.055	0.048	0.080	0.084	0.098	0.058	0.044	0.109	0.133	0.153	0.078
mae	0.135	0.033	0.034	0.073	0.027	0.182	0.040	0.041	0.152	0.029	0.246	0.078	0.067	0.447	0.039
								$T^o = 25$							
median	0.149	0.187	0.194	0.036	0.200	0.435	0.480	0.490	0.199	0.500	0.714	0.761	0.783	0.175	0.799
iqr	0.026	0.031	0.032	0.041	0.027	0.025	0.032	0.034	0.051	0.027	0.021	0.034	0.043	0.069	0.025
mae	0.051	0.018	0.017	0.164	0.014	0.065	0.021	0.018	0.301	0.014	0.086	0.039	0.024	0.625	0.012
								$T^o = 50$							
median	0.175	0.187	0.192	−0.080	0.199	0.468	0.483	0.490	0.050	0.499	0.760	0.774	0.784	0.058	0.799
iqr	0.019	0.020	0.021	0.027	0.019	0.017	0.019	0.021	0.029	0.018	0.014	0.017	0.022	0.037	0.015
mae	0.025	0.014	0.012	0.280	0.010	0.032	0.017	0.012	0.450	0.009	0.040	0.026	0.017	0.742	0.007

[a] $\sigma_\eta^2 = 0.2$, $\sigma^2 = 1$, 1000 replications, iqr is the 75th–25th interquartile range; mae denotes the median absolute error.

TABLE A2

MEDIANS, INTERQUARTILE RANGES, AND MEDIAN ABSOLUTE ERRORS OF THE ESTIMATORS ($N = 100$)[a]

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML
								$T^o = 10$							
median	0.065	0.182	0.194	0.115	0.201	0.318	0.465	0.489	0.312	0.499	0.554	0.680	0.767	0.257	0.796
iqr	0.047	0.074	0.077	0.084	0.055	0.048	0.091	0.098	0.109	0.058	0.044	0.130	0.205	0.168	0.077
mae	0.135	0.037	0.038	0.085	0.028	0.182	0.050	0.049	0.188	0.029	0.246	0.120	0.104	0.543	0.039
								$T^o = 25$							
median	0.149	0.186	0.193	0.026	0.200	0.435	0.479	0.490	0.178	0.500	0.714	0.754	0.778	0.142	0.799
iqr	0.026	0.031	0.033	0.042	0.027	0.025	0.033	0.036	0.052	0.027	0.021	0.039	0.051	0.071	0.025
mae	0.051	0.019	0.017	0.174	0.014	0.065	0.023	0.020	0.322	0.013	0.086	0.046	0.028	0.658	0.012
								$T^o = 50$							
median	0.175	0.187	0.192	−0.087	0.199	0.468	0.483	0.490	0.039	0.499	0.760	0.772	0.782	0.047	0.799
iqr	0.019	0.020	0.022	0.027	0.019	0.017	0.020	0.023	0.030	0.018	0.014	0.018	0.024	0.037	0.015
mae	0.025	0.014	0.012	0.287	0.010	0.032	0.017	0.013	0.461	0.009	0.040	0.028	0.018	0.753	0.007

[a] $\sigma_\eta^2 = 1$, $\sigma^2 = 1$, 1000 replications, iqr is the 75th–25th interquartile range; mae denotes the median absolute error.

TABLE A3

MEDIANS, INTERQUARTILE RANGES, AND MEDIAN ABSOLUTE ERRORS OF THE ESTIMATORS ($N = 50$)[a]

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML
							$T^o = 10$								
median	0.063	0.171	0.189	0.068	0.200	0.317	0.450	0.484	0.242	0.499	0.556	0.674	0.764	0.197	0.795
iqr	0.068	0.091	0.097	0.102	0.079	0.067	0.103	0.115	0.130	0.084	0.060	0.140	0.212	0.186	0.110
mae	0.136	0.049	0.047	0.132	0.039	0.183	0.060	0.058	0.258	0.042	0.244	0.129	0.108	0.602	0.055
							$T^o = 25$								
median	0.149	0.175	0.185	−0.082	0.200	0.436	0.463	0.478	0.041	0.501	0.714	0.735	0.760	0.042	0.800
iqr	0.039	0.044	0.049	0.052	0.041	0.038	0.044	0.051	0.065	0.040	0.029	0.048	0.078	0.078	0.034
mae	0.050	0.030	0.027	0.282	0.021	0.064	0.037	0.030	0.459	0.020	0.086	0.065	0.045	0.758	0.017
							$T^o = 50$								
median	0.176	0.176	0.178	−0.234	0.200	0.468	0.468	0.468	−0.114	0.500	0.760	0.756	0.748	−0.043	0.800
iqr	0.027	0.028	0.031	0.029	0.028	0.024	0.025	0.033	0.033	0.025	0.019	0.023	0.048	0.037	0.019
mae	0.025	0.024	0.023	0.435	0.014	0.031	0.032	0.032	0.614	0.012	0.040	0.044	0.052	0.843	0.010

[a] $\sigma_\eta^2 = 0.2$, $\sigma^2 = 1$, 1000 replications; iqr is the 75th–25th interquartile range; mae denotes the median absolute error.

TABLE A4

MEDIANS, INTERQUARTILE RANGES, AND MEDIAN ABSOLUTE ERRORS OF THE ESTIMATORS ($N = 50$)[a]

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML	WG	GMM	LIML	CIV	RML
								$T^o = 10$							
median	0.063	0.165	0.185	0.047	0.200	0.317	0.436	0.474	0.193	0.499	0.556	0.622	0.714	0.123	0.796
iqr	0.068	0.102	0.112	0.114	0.079	0.067	0.121	0.143	0.148	0.084	0.060	0.168	0.373	0.196	0.112
mae	0.136	0.055	0.055	0.153	0.040	0.183	0.074	0.074	0.307	0.041	0.244	0.178	0.194	0.676	0.056
								$T^o = 25$							
median	0.149	0.172	0.182	−0.095	0.200	0.436	0.460	0.474	0.020	0.501	0.714	0.727	0.745	0.021	0.800
iqr	0.039	0.045	0.051	0.054	0.041	0.038	0.045	0.056	0.065	0.040	0.029	0.051	0.103	0.078	0.034
mae	0.050	0.032	0.029	0.295	0.021	0.064	0.042	0.033	0.480	0.020	0.086	0.073	0.059	0.779	0.017
								$T^o = 50$							
median	0.176	0.176	0.176	−0.242	0.200	0.468	0.467	0.466	−0.124	0.500	0.760	0.753	0.737	−0.050	0.800
iqr	0.027	0.029	0.033	0.030	0.028	0.024	0.026	0.037	0.033	0.025	0.019	0.024	0.060	0.037	0.019
mae	0.025	0.025	0.026	0.442	0.014	0.031	0.033	0.034	0.624	0.013	0.040	0.047	0.062	0.850	0.010

[a] $\sigma_\eta^2 = 1$, $\sigma^2 = 1$, 1000 replications; iqr is the 75th–25th interquartile range; mae denotes the median absolute error.

REFERENCES

AHN, S., AND P. SCHMIDT (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27.

ALONSO-BORREGO, C., AND M. ARELLANO (1999): "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, 17, 36–49.

AMEMIYA, T. (1985): *Advanced Econometrics*. Oxford: Basil Blackwell.

ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series.* New York: John Wiley & Sons.

——— (1978): "Repeated Measurements on Autoregressive Processes," *Journal of the American Statistical Association*, 73, 371–378.

ANDERSON, T. W., AND C. HSIAO (1981): "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598–606.

ARELLANO, M., AND S. R. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

ARELLANO, M., AND O. BOVER (1995): "Another Look at the Instrumental-Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

BEKKER, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

BLUNDELL, R., AND S. BOND (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

BLUNDELL, R., AND R. SMITH (1991): "Initial Conditions and Efficient Estimation in Dynamic Panel Data Models," *Annales d'Economie et de Statistique*, 20/21, 109–123.

CANOVA, F., AND A. MARCET (1995): "The Poor Stay Poor: Non-Convergence across Countries and Regions," Unpublished Manuscript, Universitat Pompeu Fabra, Barcelona.

HAHN, J., AND G. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica*, 70, 1639–1657.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

HOLTZ-EAKIN, D., W. NEWEY, AND H. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–1395.

JUDSON, R. A., AND A. L. OWEN (1997): "Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists," Finance and Economics Discussion Series 1997-3, Federal Reserve Board, Washington.

KIVIET, J. F. (1995): "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53–78.

KOENKER, R., AND J. A. MACHADO (1999): "GMM Inference when the Number of Moment Conditions is Large," *Journal of Econometrics*, 93, 327–344.

KUNITOMO, N. (1980): "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Society*, 75, 693–700.

MADDALA, G. S. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39, 341–358.

MORIMUNE, K. (1983): "Approximate Distributions of k-Class Estimators when the Degree of Overidentifiability Is Large Compared with the Sample Size," *Econometrica*, 51, 821–841.

NERLOVE, M. (1967): "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," *Economic Studies Quarterly*, 18, 42–74.

——— (1971): "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, 39, 359–387.

NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

PESARAN, M. H., AND R. SMITH (1995): "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79–113.

SARGAN, J. D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.