The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators

Javier Alvarez
CEMFI
Manuel Arellano
CEMFI

Working Paper No. 9808
December 1998

We thank Gary Chamberlain, Jan Magnus, Enrique Sentana, Neil Shephard and seminar audiences at CEMFI, Institute for Fiscal Studies, 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. (E-mail addresses: alvarez@cemfi.es, arellano@cemfi.es).

CEMFI, Casado del Alisal 5, 28014 Madrid, Spain.
Tel: 34 91 4290551, fax: 34 91 4291056, www.cemfi.es.

Abstract

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 T, N and $(2N - T)$, 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 MLE with unrestricted initial conditions when both T and N tend to infinity.

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 (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^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 behaviour 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 case closeness to OLS (the WG estimator) becomes increasingly desirable since the "simultaneity bias" tends to zero as T tends to infinity. Nevertheless, WG, GMM and LIML turn out to exhibit a bias term in their asymptotic distributions, which are of orders T, N and $2N - T$, respectively. Provided $T < N$, the GMM bias is always smaller that the WG bias, and the LIML bias is smaller than the other two. When $T = N$ the three biases are all equal. Since the GMM and LIML estimators are only defined for $N \geq T - 1$, the asymptotics $T/N \to c$ is a relevant one to consider here. 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, although the standard formulae for fixed T estimated variances remain consistent (and often more reliable) estimates of the asymptotic variances as T tends to infinity. 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 maximum likelihood estimator (RML) which 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

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

where $\mid \alpha \mid < 1$ and v_{it} has zero mean given $\eta_i, y_{i0}, ..., y_{it-1}$. 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 $y_i = (y_{i1}, ..., y_{iT})'$, $x_i = (x_{i1}, ..., x_{iT})'$, ι_T is a $T \times 1$ vector of ones, and $v_i = (v_{i1}, ..., v_{iT})'$.

The within-groups estimator The within-groups or covariance estimator is given by

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

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 $Var(v_i) = \sigma^2 I_T$, the $(T-1) \times 1$ vector of errors in orthogonal deviations $v_i^* = Av_i$ also has $Var(v_i^*) = \sigma^2 I_{T-1}$.[1] Notice that

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

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

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

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

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

where $z_{it} = (x_{i1}, ..., x_{it})'$, 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 which can be written as:

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

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

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

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

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

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

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

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

$$\widehat{\alpha}_{GMM} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} Z_t (Z_t' Z_t)^{-1} Z_t' y_t^*}{\sum_{t=1}^{T-1} x_t^{*\prime} Z_t (Z_t' Z_t)^{-1} Z_t' x_t^*} \tag{9}$$

where x_t^* and y_t^* are the $N \times 1$ vectors whose i-th elements are x_{it}^* and y_{it}^*, respectively, and Z_t is the $N \times t$ matrix whose i-th row is z_{it}'. Notice that this GMM estimator is only defined for $N \geq T - 1$. Finally, $\widehat{\alpha}_{GMM}$ can also be written using the equations in first differences as opposed to orthogonal deviations (cf. Arellano and Bover, 1995). In such case:

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

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) analogue estimator solves the following problem:

$$\widehat{\alpha}_{LIML} = \arg\min_a \frac{(y^* - ax^*)' Z (Z'Z)^{-1} Z' (y^* - ax^*)}{(y^* - ax^*)'(y^* - ax^*)} \tag{11}$$

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 likelihood estimator; it is only a LIML analogue estimator in the sense of the instrumental-variable interpretation given by Sargan (1958) to the original LIML estimator.[2]

We can write down a simple explicit expression for $\widehat{\alpha}_{LIML}$ by noticing that the minimized criterion in (11) is the following minimum eigenvalue:

$$\widehat{\ell} = \min eigenvalue[W^{*\prime}Z(Z'Z)^{-1}Z'W^*(W^{*\prime}W^*)^{-1}] \tag{12}$$

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

Now the first order conditions for (11) are

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

from which we obtain:

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

[2] 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 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:

$$v_{it} \mid z_{it}, \eta_i \sim iid\ N(0, \sigma^2) \quad (t = 1, ..., T; \;\; i = 1, ..., N) \tag{A1}$$

where $z_{it} = (y_{i0}, ..., y_{it-1})'$.

$$y_{i0} \mid \eta_i \sim id\ N\left(\frac{\eta_i}{1-\alpha}, \frac{\sigma^2}{1-\alpha^2}\right) \quad (i = 1, ..., N) \tag{A2}$$

So that,

$$z_{iT+1} \mid \eta_i \sim id\ N\left(\frac{\eta_i}{1-\alpha}\iota_{T+1}, \frac{\sigma^2}{1-\alpha^2}V\right)$$

where ι_{T+1} is a $(T+1) \times 1$ vector of ones and V is the autoregressive matrix whose (t, s) element is given by $\alpha^{|t-s|}$. Finally, we assume

$$\eta_i \sim iid\ N(0, \sigma^2_\eta) \quad (i = 1, ..., N) \tag{A3}$$

Taken together, assumptions *A1*, *A2* and *A3* imply that the z_{iT+1} are *iid* normal random vectors with

$$E(z_{iT+1}) = \frac{E(\eta_i)}{(1-\alpha)}\iota_{T+1} = 0$$

$$Var(z_{iT+1}) = Var\left[E\left(z_{iT+1} \mid \eta_i\right)\right] + E\left[Var(z_{iT+1} \mid \eta_i)\right] = \frac{\sigma^2_\eta}{(1-\alpha)^2}\iota_{T+1}\iota'_{T+1} + \frac{\sigma^2}{(1-\alpha^2)}V.$$

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 or on the distribution of the unobserved heterogeneity.

3.1 The WG estimator

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

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

The results collected in the following Lemma are useful in establishing the asymptotic properties of the WG estimator.

Lemma 1 *Under assumptions* A1, A2 *and* A3*:*

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

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

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

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

Proof: See Appendix.

It is well known that $\widehat{\alpha}_{WG}$ is consistent as $T \to \infty$ regardless of the asymptotic behaviour 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 $p\lim(x^{*'}v^*/NT) = 0$. Together with (18), this implies that

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

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, A2, *and* A3 *hold. Then, as* $T \to \infty$, *regardless of whether* N *is fixed or tends to infinity:*

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

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

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

Proof:

Let us write

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

where w_{it} is the pure AR(1) process given by:

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

In view of (16) we have

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

Subtracting μ_{NT} from the expression above:

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

where

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

We now show that R_{NT} is $o_p(1)$ as $T \to \infty$. Clearly $E(R_{NT}) = 0$. Moreover, after some algebra we obtain

$$Var(R_{NT}) = \frac{\sigma^4}{(1-\alpha)^2}\left[\frac{2}{T} - \frac{2}{T^2}\frac{(1+2\alpha)(1-\alpha^T)}{(1-\alpha^2)} + \frac{1}{T^3}\left(\frac{1-\alpha^T}{1-\alpha}\right)^2\right]$$

so that $\lim_{T\to\infty} Var(R_{NT}) = 0$, which suffices to establish that R_{NT} is $o_p(1)$.

Finally, from a standard central limit theorem for autoregressive processes (cf. T.W. Anderson, 1971, ch. 5, Theorem 5.5.7, and T.W. Anderson, 1978) we have

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

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

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

which establishes the first result of the theorem.

Next, in view of (18), by Cramer's theorem we have

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

or

$$\sqrt{NT}(\widehat{\alpha}_{WG} - \alpha) - \left(\frac{x^{*\prime}x^*}{NT}\right)^{-1}\mu_{NT} \xrightarrow{d} N(0, 1-\alpha^2)$$

Using similar arguments as in the proof of (20) it can be shown that

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

where, in view of (A8),

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

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

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

Hence, by the delta method, provided $N/T^3 \to 0$

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

and therefore

$$\begin{aligned}\left(\frac{x^{*\prime}x^*}{NT}\right)^{-1}\mu_{NT} &= \frac{(1-\alpha^2)}{\sigma^2}\left[1+\frac{1}{T}\frac{(1+\alpha)}{(1-\alpha)}\right]\mu_{NT} + o_p(1) \\ &= -\left(\frac{N}{T}\right)^{1/2}(1+\alpha) - \left(\frac{N}{T^3}\right)^{1/2}\frac{(1+\alpha)(\alpha+\alpha^T)}{(1-\alpha)} + o_p(1)\end{aligned}$$

The second result of the theorem follows from noticing that when $N/T^3 \to 0$ the second term of the rhs in the expression above is also $o(1)$.

QED.

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:

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

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 the latter are well defined, namely, when $N \geq T-1$, and in datasets with $N \geq T-1$ 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 (1998) under slightly more general conditions. Hahn's paper has a different focus since he is concerned with the development of an efficient 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):

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

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

Lemma 2 *Under assumptions* A1, A2 *and* A3*:*

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

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

$$Var\left(\frac{x^{*\prime} M v^*}{(NT)^{1/2}} \right) = \frac{1}{T} \sigma^2 \sum_{t=1}^{T-1} E(x_{it}^* z_{it}')[E(z_{it} z_{it}')]^{-1} E(z_{it} x_{it}^*) + o(1) \to \frac{\sigma^4}{(1-\alpha^2)} \tag{25}$$

$$Cov\left(\frac{x^{*\prime} v^*}{(NT)^{1/2}}, \frac{x^{*\prime} M v^*}{(NT)^{1/2}} \right) \to \frac{\sigma^4}{(1-\alpha^2)} \tag{26}$$

$$\frac{1}{NT}(x^{*\prime} M x^*) \xrightarrow{p} \frac{\sigma^2}{1-\alpha^2} \tag{27}$$

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

$$\frac{1}{NT}(v^{*\prime} M v^*) \xrightarrow{p} \sigma^2 \frac{c}{2} \tag{28}$$

Proof: See Appendix.

The condition $(\log T)/N \to 0$ provides a limit on how slow N can tend to infinity relative to T. Since the GMM estimator is only defined for $N \geq T-1$ 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 $\widehat{\alpha}_{GMM}$ 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* $(\log T)/N \to 0, \widehat{\alpha}_{GMM}$ *is consistent for* α*:*

$$\widehat{\alpha}_{GMM} \xrightarrow{p} \alpha \tag{29}$$

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

$$\sqrt{NT}\left[\widehat{\alpha}_{GMM} - \left(\alpha - \frac{1}{N}(1+\alpha)\right)\right] \xrightarrow{d} N(0, 1-\alpha^2) \tag{30}$$

Proof: 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 (27) $(x^{*'}Mx^*)/NT$ is bounded in probability.

Turning to asymptotic normality, using (24) let us define

$$\mu_{NT}^{+} = E\left[(NT)^{-1/2}x^{*'}Mv^*\right] = -\left(\frac{T}{N}\right)^{1/2}\frac{\sigma^2}{(1-\alpha)} + (NT)^{-1/2}\frac{\sigma^2}{(1-\alpha)^2}\sum_{t=1}^{T}\frac{(1-\alpha^t)}{t}$$

We shall rely on the identity

$$(NT)^{-1/2}x^{*'}Mv^* - \mu_{NT}^{+} = (NT)^{-1/2}x^{*'}v^* - \mu_{NT} - R_{NT}^{+}$$

where

$$R_{NT}^{+} = (NT)^{-1/2}x^{*'}(I-M)v^* - (\mu_{NT} - \mu_{NT}^{+})$$

By construction, $E(R_{NT}^{+})=0$. Moreover, in view of Lemmae 1 and 2:

$$Var(R_{NT}^{+})=Var\left(\frac{x^{*\prime}v^{*}}{(NT)^{1/2}}\right)+Var\left(\frac{x^{*\prime}Mv^{*}}{(NT)^{1/2}}\right)-2Cov\left(\frac{x^{*\prime}v^{*}}{(NT)^{1/2}},\frac{x^{*\prime}Mv^{*}}{(NT)^{1/2}}\right)=o(1)$$

Therefore, $R_{NT}^{+}=o_p(1)$ and from result (20) in Theorem 1 we have

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

and in view of (27), by Cramer's theorem:

$$\left(\frac{x^{*\prime}Mx^{*}}{NT}\right)^{-1}\left[(NT)^{-1/2}x^{*\prime}Mv^{*}-\mu_{NT}^{+}\right]\xrightarrow{d}N(0,1-\alpha^2)$$

or

$$\sqrt{NT}(\widehat{\alpha}_{GMM}-\alpha)-\left(\frac{x^{*\prime}Mx^{*}}{NT}\right)^{-1}\mu_{NT}^{+}\xrightarrow{d}N(0,1-\alpha^2)$$

The result follows from noticing that since $\mu_{NT}^{+}=O(1)$

$$\left(\frac{x^{*\prime}Mx^{*}}{NT}\right)^{-1}\mu_{NT}^{+}\;=\;\frac{(1-\alpha^2)}{\sigma^2}\mu_{NT}^{+}+o_p(1)=-\left(\frac{T}{N}\right)^{1/2}(1+\alpha)+o_p(1)$$

QED.

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 $\widehat{\alpha}_{GMM}$ 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 $\widehat{\alpha}_{GMM}$ is that in our context as T tends to infinity the "simultaneity bias" tends to zero, and so closeness of $\widehat{\alpha}_{GMM}$ to $\widehat{\alpha}_{WG}$ for larger values of T becomes a desirable property of the GMM estimator.

The theorem also shows that as $T \to \infty$, $\widehat{\alpha}_{GMM}$ 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 are asymptotically equivalent and have 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. Therefore, provided $T < N$, the GMM bias will always be smaller than the WG bias, and when $T = N$ the two biases will coincide.

Finally, notice that in view of (25) the standard formulae for fixed T estimated variances of $\widehat{\alpha}_{GMM}$ remain consistent estimates of the asymptotic variances as $T \to \infty$. This is important because, unlike the limiting distribution, the exact distribution of $\widehat{\alpha}_{GMM}$ does depend on the variance of the individual effect. Therefore, for some parameter values there may be substantial differences between the fixed T and the large T approximations to the variance of the GMM estimator. This situation is in contrast with that for the WG estimator, whose exact distribution is invariant to σ_η^2.

For a more general class of problems, Koenker and Machado (1996) found that $q^3/N \to 0$ was a sufficient condition for the validity of conventional asymptotic inference about 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 conventional asymptotic inferences about $\widehat{\alpha}_{GMM}$ are valid. Since here $q = T(T-1)/2$, we have found a much tighter condition for the validity of standard fixed T inferences in the dynamic panel data context.

3.3 The LIML estimator

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

$$\widehat{\alpha}_{LIML} - \alpha = \frac{x^{*\prime} M v^* - \widehat{\ell}(x^{*\prime} v^*)}{x^{*\prime} M x^* - \widehat{\ell}(x^{*\prime} x^*)} \tag{31}$$

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

Lemma 3 *Under assumptions* A1, A2, *and* A3 *as both* N *and* T *tend to infinity, and* $T/N \to c, 0 \leq c \leq 2$

$$\widehat{\ell} \xrightarrow{p} \frac{c}{2} \tag{32}$$

Proof: Using the results in Lemmae 1 and 2, simple algebra reveals that

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

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

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

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

or equivalently

$$(1-\alpha^2)(1-\ell)(\frac{c}{2} - \ell) = 0$$

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

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

$$\frac{(y^* - dx^*)' M (y^* - dx^*)}{(y^* - dx^*)'(y^* - dx^*)} \xrightarrow{p} \frac{(d-\alpha)^2 + (1-\alpha^2)c/2}{(d-\alpha)^2 + (1-\alpha^2)} \tag{33}$$

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 only well defined for $(T-1)/N \leq 1$. The following theorem considers consistency and asymptotic normality of $\widehat{\alpha}_{LIML}$.

Theorem 3 *(Consistency and asymptotic normality of the LIML estimator). Let conditions* A1, A2 *and* A3 *hold. Then as both N and T tend to infinity, provided $T/N \to c, 0 \leq c \leq 2, \widehat{\alpha}_{LIML}$ is consistent for α:*

$$\widehat{\alpha}_{LIML} \xrightarrow{p} \alpha \tag{34}$$

Moreover,

$$\sqrt{NT}\left[\widehat{\alpha}_{LIML} - \left(\alpha - \frac{1}{(2N-T)}(1+\alpha)\right)\right] \xrightarrow{d} N(0, 1-\alpha^2) \tag{35}$$

Proof: From Lemmae 1, 2 and 3

$$(NT)^{-1}(x^{*'} M v^* - \widehat{\ell} x^{*'} v^*) \xrightarrow{p} 0$$

and

$$(NT)^{-1}(x^{*'} M x^* - \widehat{\ell} x^{*'} x^*) \xrightarrow{p} (1 - \frac{c}{2})\frac{\sigma^2}{1-\alpha^2}$$

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

Turning to asymptotic normality, in view of Lemma 3

$$\begin{aligned} & (NT)^{-1/2}(x^{*\prime}Mv^* - \widehat{\ell}x^{*\prime}v^*) - (\mu_{NT}^{+} - \widehat{\ell}\mu_{NT}) \\ = \quad & [(NT)^{-1/2}x^{*\prime}Mv^* - \mu_{NT}^{+}] - \frac{c}{2}[(NT)^{-1/2}x^{*\prime}v^* - \mu_{NT}] + o_p(1) \end{aligned}$$

Moreover, due to Theorem 2, the expression above satisfies

$$(1-\frac{c}{2})[(NT)^{-1/2}x^{*\prime}v^* - \mu_{NT}] + o_p(1) \xrightarrow{d} N\left(0, (1-\frac{c}{2})^2\frac{\sigma^4}{(1-\alpha^2)}\right)$$

Now, by Cramer's theorem:

$$\left(\frac{x^{*\prime}Mx^* - \widehat{\ell}x^{*\prime}x^*}{NT}\right)^{-1}[(NT)^{-1/2}(x^{*\prime}Mv^* - \widehat{\ell}x^{*\prime}v^*) - (\mu_{NT}^{+} - \widehat{\ell}\mu_{NT})]) \xrightarrow{d} N(0, 1-\alpha^2)$$

or

$$\sqrt{NT}(\widehat{\alpha}_{LIML} - \alpha) - \left(\frac{x^{*\prime}Mx^* - \widehat{\ell}x^{*\prime}x^*}{NT}\right)^{-1}(\mu_{NT}^{+} - \widehat{\ell}\mu_{NT}) \xrightarrow{d} N(0, 1-\alpha^2)$$

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

$$\begin{aligned} \left(\frac{x^{*\prime}Mx^* - \widehat{\ell}x^{*\prime}x^*}{NT}\right)^{-1}(\mu_{NT}^{+} - \widehat{\ell}\mu_{NT}) &= [(1-\frac{T}{2N})\frac{\sigma^2}{1-\alpha^2}]^{-1}(\mu_{NT}^{+} - \frac{T}{2N}\mu_{NT}) + o_p(1) \\ &= (NT)^{1/2}\frac{(1+\alpha)}{(2N-T)} + o_p(1) \end{aligned}$$

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

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

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

$$\widehat{\ell} \leq \frac{v^{*\prime} M v^{*}}{v^{*\prime} v^{*}}.$$

From the proof of (28) in Lemma 2 it is easy to see that the result

$$\left(\frac{N}{T}\right)^{1/2} \left(\frac{v^{*\prime} M v^{*}}{NT}\right) \xrightarrow{p} \sigma^2 \frac{c^{1/2}}{2}$$

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

$$\frac{v^{*\prime} v^{*}}{NT} \xrightarrow{p} \sigma^2$$

with $c = 0$, we have

$$\left(\frac{N}{T}\right)^{1/2} \left(\frac{v^{*\prime} M v^{*}}{v^{*\prime} v^{*}}\right) \xrightarrow{p} 0$$

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

QED

The theorem shows that like GMM, the LIML estimator is consistent despite $T \to \infty$ and $T/N \to c$. Also, $\widehat{\alpha}_{LIML}$ 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 bias is the smallest of the three, and when $T = N$ the three biases are equal.

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 case, however, the optimal weighting matrix becomes $[Z^{'}(I_N \nwarrow\!\!\searrow H)Z]^{-1}$ 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

$$\widehat{\alpha}_{CIV} = \frac{\Delta x^{'}Z(Z^{'}Z)^{-1}Z^{'}\Delta y}{\Delta x^{'}Z(Z^{'}Z)^{-1}Z^{'}\Delta x} \tag{36}$$

For fixed T as N tends to infinity, this estimator is asymptotically inefficient relative to $\widehat{\alpha}_{GMM}$, but it is still consistent and asymptotically normal, and as such it may be regarded as a computationally simpler alternative to $\widehat{\alpha}_{GMM}$ (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 "simultaneity bias" in first differences does not tend to zero as $T \to \infty$, there may be more fundamental differences between $\widehat{\alpha}_{CIV}$ and $\widehat{\alpha}_{GMM}$ 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 $T/N \to c$, $0 \leq c < \infty$*

$$\frac{1}{NT}(\Delta x^{'}M\Delta v) \xrightarrow{p} -\sigma^2\frac{c}{2} \tag{37}$$

$$\frac{1}{NT}(\Delta x^{'}M\Delta x) \xrightarrow{p} \sigma^2\left(\frac{c}{2} + \frac{1-\alpha}{1+\alpha}\right) \tag{38}$$

and

$$\widehat{\alpha}_{CIV} \xrightarrow{p} \alpha - \frac{(1+\alpha)}{2}\left(\frac{c}{2-(1+\alpha)(2-c)/2}\right) \tag{39}$$

Proof: See Appendix.

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 $\widehat{\alpha}_{CIV}$ 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 $\widehat{\alpha}_{GMM}$ and $\widehat{\alpha}_{CIV}$ there is only a second order difference in precision, whereas when $T/N \to c > 0$, $\widehat{\alpha}_{GMM}$ is still consistent but $\widehat{\alpha}_{CIV}$ is not.

3.5 The random effects ML estimator

In this section we discuss the random effects ML estimator $\widehat{\alpha}_{RML}$ based on assumptions *A1*, *A3* and

$$y_{i0} \mid \eta_i \sim id\ N(\delta\eta_i, \omega_{oo}^2) \quad (i = 1, ..., N) \qquad (A2')$$

Note that in *A2* we have $\delta = 1/(1-\alpha)$ and $\omega_{oo}^2 = \sigma^2/(1-\alpha^2)$, but here δ and ω_{oo} are free parameters. Thus $\widehat{\alpha}_{RML}$ 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.[3]

As shown in the Appendix, under *A1*, *A2'* and *A3* the log density of $(y_{i1}, ..., y_{iT})$ given y_{i0} can be written as

$$\ln f(y_{i1}, ..., y_{iT} \quad | \quad y_{i0}) = -\frac{(T-1)}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}(y_i^* - \alpha x_i^*)'(y_i^* - \alpha x_i^*) \\ -\frac{1}{2} \ln \omega^2 - \frac{1}{2\omega^2}(\overline{y}_i - \alpha \overline{x}_i - \varphi y_{i0})^2. \tag{40}$$

where $\overline{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$, $\overline{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$, and (φ, ω^2) are a reparameterization of (δ, σ_η^2) given by $\varphi = \delta \sigma_\eta^2 / Var(y_{i0})$, and $\omega^2 = \sigma_\eta^2 - \varphi^2 Var(y_{i0}) + \sigma^2/T$. Hence, by concentrating φ, ω^2, and σ^2 out of the log likelihood, the RML estimator can be expressed as

$$\widehat{\alpha}_{RML} = \arg\min_{a} \left\{ \ln\left[(y^* - ax^*)'(y^* - ax^*)\right] + \frac{1}{(T-1)} \ln\left[(\overline{y} - a\overline{x})' S_0 (\overline{y} - a\overline{x})\right] \right\} \tag{41}$$

where $S_0 = I_N - y_0 y_0' / (y_0' y_0)$, $y_0 = (y_{10}, ..., y_{N0})'$, $\overline{y} = (\overline{y}_1, ..., \overline{y}_N)'$, and $\overline{x} = (\overline{x}_1, ..., \overline{x}_N)'$.[4] Consistency and asymptotic normality of $\widehat{\alpha}_{RML}$ is considered in the following Theorem.

[3] We thank Gary Chamberlain for suggesting us to consider the RML estimator in this context.

[4] The estimator in (41) does not restrict σ_η^2 to be non-negative. Parameterizing the joint log likelihood in terms of $\delta, \omega_{oo,}^2 \alpha, \sigma^2$ and $\lambda = \sigma_\eta^2 / \sigma^2$, we may obtain ML estimates of α that enforce $\lambda \geq 0$, from a concentrated likelihood which is only a function of α and λ (see Appendix). In such case, a boundary solution at $\lambda = 0$ may occur. This problem was discussed by Maddala (1971).

Theorem 5 *(Consistency and asymptotic normality of the RML estimator) Let conditions* A1, A2 *and* A3 *hold. Then as both* N *and* T *tend to infinity,* $\widehat{\alpha}_{RML}$ *is consistent for* α*:*

$$\widehat{\alpha}_{RML} \xrightarrow{p} \alpha \tag{42}$$

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

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

Proof*: See Appendix.*

Obviously, the RML estimator is also consistent and asymptotically normal for fixed T as $N \to \infty$ under the stated conditions, but in such 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).[5]

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.

[5] 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.

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 1 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 2. For all cases we conducted 1000 replications from the model specified in sections 2 and 3 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 3 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 1 and 2, 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 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 loses 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 1 and 2), 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 $N = 50$, $T^0 = 50$, $\sigma_\eta^2 = 1$, 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 efficient when both N and T tend to infinity. They are asymptotically efficient in the sense of attaining the same asymptotic variance as the covariance estimator as $T \to \infty$. 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 T, N, and $(2N - T)$, 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 which 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 normality 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.

References

[1] Ahn, S. and P. Schmidt (1995), "Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics*, 68, 5-27.

[2] Alonso-Borrego, C. and M. Arellano (1999), "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data", *Journal of Business & Economic Statistics*, 17, January.

[3] Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell, Oxford.

[4] Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, John Wiley & Sons, New York.

[5] Anderson, T.W. (1978), "Repeated Measurements on Autoregressive Processes", *Journal of the American Statistical Association*, 73, 371-378.

[6] Anderson, T.W. and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606.

[7] 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.

[8] Arellano, M. and O. Bover (1995), "Another Look at the Instrumental-Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68, 29-51.

[9] Bekker, P.A. (1994), "Alternative Approximations to the Distributions of Instrumental Variable Estimators", *Econometrica*, 62, 657-681.

[10] Blundell, R. and S. Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, 87, 115-143.

[11] 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.

[12] Canova, F. and A. Marcet (1995), "The Poor Stay Poor: Non-Convergence across Countries and Regions", Unpublished manuscript, Universitat Pompeu Fabra, Barcelona.

[13] Hahn, J. (1998), "Asymptotically Unbiased Inference of Dynamic Panel Data Model with Fixed Effects When Both n and T are Large", Unpublished manuscript, University of Pennsylvania.

[14] 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.

[15] Holtz-Eakin, D., W. Newey and H. Rosen (1988), "Estimating Vector Autoregressions with Panel Data, *Econometrica*, 56, 1371-1395.

[16] 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.

[17] Kiviet, J.F. (1995), "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models", *Journal of Econometrics*, 68, 53-78.

[18] Koenker, R. and J.A. Machado (1996), "GMM Inference when the Number of Moment Conditions is Large", Unpublished manuscript, University of Illinois, Champaign.

[19] 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.

[20] Maddala, G. S. (1971), "The Use of Variance Components Models in Pooling Cross Section and Time Series Data", *Econometrica*, 39, 341-358.

[21] 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.

[22] 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.

[23] Nerlove, M. (1971), "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections", *Econometrica*, 39, 359-387.

[24] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417-1426.

[25] Pesaran, M.H. and R. Smith (1995), "Estimating Long-Run Relationships form Dynamic Heterogeneous Panels", *Journal of Econometrics*, 68, 79-113.

[26] Sargan, J.D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, 26, 393-415.

A Appendix

Lemma 1

Proof of (16): Firstly, note that

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

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 results follows.

Proof of (17): Due to cross-sectional independence

$$Var\left(\frac{x^{*'}v^*}{(NT)^{1/2}}\right) = \frac{1}{T}Var(x_i^{'}Q_T v_i). \tag{A3}$$

Moreover, due to joint normality of x_i and v_i

$$Var(x_i^{'}Q_T v_i) = tr[Q_T E(x_i x_i^{'})Q_T E(v_i v_i^{'})] + tr[Q_T E(x_i v_i^{'})Q_T E(x_i v_i^{'})], \tag{A4}$$

but since $E(v_i v_i^{'}) = \sigma^2 I_T$, the first term simplifies and we have

$$Var(x_i^{'}Q_T v_i) = \sigma^2 E(x_i^{'}Q_T x_i) + \sigma^4 tr(Q_T C_T Q_T C_T). \tag{A5}$$

As for the second term, it equals

$$tr(Q_T C_T Q_T C_T) = tr(C_T C_T) - \frac{2}{T}\iota_T^{'}(C_T C_T)\iota_T + \frac{1}{T^2}(\iota_T^{'}C_T\iota_T)^2 \tag{A6}$$

Noticing that $tr(C_T C_T) = 0$ and adding up the elements of $C_T C_T$, the second term turns out to be $o(T)$. Therefore, we are only left with the first term, which can be written as

$$E(x_i^{'}Q_T x_i) = E(w_i^{'}w_i) - \frac{1}{T}\iota_T^{'}E(w_i w_i^{'})\iota_T \tag{A7}$$

where w_i is a pure AR process with elements $w_{it} = y_{it} - \eta_i/(1-\alpha)$. Thus

$$E(x_i^{'}Q_T x_i) = T\frac{\sigma^2}{(1-\alpha^2)} - \frac{1}{T}\frac{\sigma^2}{(1-\alpha^2)}\iota_T^{'}V_T\iota_T. \tag{A8}$$

Since $\iota_T^{'}V_T\iota_T/T \to (1+\alpha)/(1-\alpha)$, it follows that

$$\frac{1}{T}Var(x_i^{'}Q_T v_i) = \frac{\sigma^2}{T}E(x_i^{'}Q_T x_i) + o(1) \to \frac{\sigma^4}{(1-\alpha^2)}. \tag{A9}$$

Proof of (18): Notice that we have already established that

$$E\left(\frac{x^{*'}x^*}{NT}\right) = \frac{1}{T}E(x_i^{'}Q_T x_i) \to \frac{\sigma^2}{(1-\alpha^2)} \tag{A10}$$

We now establish convergence in probability by proving that the variance of $x^{*'}x^*/(NT)$ tends to zero as $T \to \infty$.

We have that

$$Var\left(\frac{x^{*'}x^*}{NT}\right) = \frac{1}{NT^2}Var(x_i^{'}Q_T x_i) \tag{A11}$$

and due to normality of x_i

$$\begin{aligned} Var(x_i^{'}Q_T x_i) &= 2tr[Q_T E(w_i w_i^{'})Q_T E(w_i w_i^{'})] = \frac{2\sigma^4}{(1-\alpha^2)^2}tr(Q_T V_T Q_T V_T) \\ &= \frac{2\sigma^4}{(1-\alpha^2)^2}[tr(V_T V_T) - \frac{2}{T}\iota_T^{'}(V_T V_T)\iota_T + \frac{1}{T^2}(\iota_T^{'}V_T\iota_T)^2] \end{aligned} \tag{A12}$$

Direct evaluation shows that these terms are $o(T^2)$. For example, we have

$$tr(V_T V_T) = T\frac{(1+\alpha^2)}{(1-\alpha^2)} - \frac{2\alpha^2(1-\alpha^{2T})}{(1-\alpha^2)^2} \tag{A13}$$

Therefore, as $T \to \infty$ regardless of whether N is fixed or not

$$\frac{1}{T^2} Var(x_i^{'} Q_T x_i) \to 0 \tag{A14}$$

Lemma 2

Proof of (24): Letting $M_t = Z_t (Z_t^{'} Z_t)^{-1} Z_t^{'}$ we have

$$E(x^{*'} M v^*) = \sum_{t=1}^{T-1} E(x_t^{*'} M_t v_t^*) = \sum_{t=1}^{T-1} E\left\{ tr[M_t E_t(v_t^* x_t^{*'})] \right\} \tag{A15}$$

where $E_t(.)$ denotes an expectation conditional on Z_t. Since $E_t(v_t^*) = 0$, $E_t(v_t^* x_t^{*'})$ is the conditional covariance between v_t^* and x_t^*, which due to joint normality of v_t^*, x_t^*, and Z_t does not depend on Z_t. Therefore, $E_t(v_t^* x_t^{*'}) = E(v_t^* x_t^{*'})$. Moreover, by cross-sectional independence

$$E(v_t^* x_t^{*'}) = E(v_{it}^* x_{it}^*) I_N. \tag{A16}$$

Hence, using the fact that $tr(M_t) = t$, we have

$$E(x^{*'} M v^*) = \sum_{t=1}^{T-1} t E(v_{it}^* x_{it}^*) = \sum_{t=1}^{T-1} t\ a_t^{'} E(v_i x_i^{'}) a_t = tr\left[E(v_i x_i^{'}) \sum_{t=1}^{T-1} t\ a_t a_t^{'} \right] \tag{A17}$$

where a_t is the t-th row of the $(T-1) \times T$ orthogonal deviations operator A. By direct calculation it can be shown that

$$\sum_{t=1}^{T-1} t\ a_t a_t^{'} = \sum_{s=2}^{T} H_s Q_s H_s^{'} \tag{A18}$$

where H_s is a selection matrix of order $T \times s$ given by $H_s = (0 : I_s)^{'}$. Using this result we have

$$E(x^{*'} M v^*) = \sum_{t=2}^{T} E(x_i^{'} H_t Q_t H_t^{'} v_i) \tag{A19}$$

Notice that $H_t^{'}v_i = (v_{i(T-t+1)}, ..., v_{iT})^{'}$. Thus, using (16) and (A1) it turns out that

$$E(x_i^{'}H_tQ_tH_t^{'}v_i) = -\frac{\sigma^2}{(1-\alpha)}\left[1 - \frac{1}{t}\left(\frac{1-\alpha^t}{1-\alpha}\right)\right] \tag{A20}$$

Therefore,

$$E(x^{*'}Mv^*) = -\frac{\sigma^2}{(1-\alpha)}\sum_{t=2}^{T}\left[1 - \frac{1}{t}\left(\frac{1-\alpha^t}{1-\alpha}\right)\right] = -T\frac{\sigma^2}{(1-\alpha)}\left[1 - \frac{1}{T(1-\alpha)}\sum_{t=1}^{T}\frac{(1-\alpha^t)}{t}\right]. \tag{A21}$$

Proof of (25): We have

$$Var\left(\frac{x^{*'}Mv^*}{(NT)^{1/2}}\right) = \frac{1}{NT}\sum_{t=1}^{T-1}Var(x_t^{*'}M_tv_t^*) + \frac{1}{NT}\sum_{t\neq s}Cov(x_t^{*'}M_tv_t^*, x_s^*M_sv_s^*) \tag{A22}$$

We first consider a variance term. Given the variance decomposition

$$Var(x_t^*M_tv_t^*) = Var[E_t(x_t^{*'}M_tv_t^*)] + E[Var_t(x_t^{*'}M_tv_t^*)], \tag{A23}$$

from the proof of (24) $E_t(x_t^{*'}M_tv_t^*)$ does not depend on Z_t and therefore the first term on the rhs of (A23) vanishes.

Next, since conditional on Z_t, x_t^* and v_t^* are jointly normal, $E_t(v_t^*) = 0$, and M_t can be held constant given Z_t

$$Var_t(x_t^{*'}M_tv_t^*) = tr[M_tE_t(x_t^*x_t^{*'})M_tE_t(v_t^*v_t^{*'})] + tr[M_tE_t(x_t^*v_t^{*'})M_tE_t(x_t^*v_t^{*'})] \tag{A24}$$

From the proof of (24), $E_t(x_t^*v_t^{*'}) = E(x_{it}^*v_{it}^*)I_N$, and also $E_t(v_t^*v_t^{*'}) = \sigma^2 I_N$. Therefore

$$Var_t(x_t^{*'}M_tv_t^*) = \sigma^2 tr[M_tE_t(x_t^*x_t^{*'})] + [E(x_{it}^*v_{it}^*)]^2 t \tag{A25}$$

Let us now consider the linear projections

$$x_t^* = Z_t\pi_t + \varepsilon_t \qquad (t = 1, ..., T-1) \tag{A26}$$

Due to joint normality of x_t^* and Z_t

$$E_t(x_t^* x_t^{*'}) = Z_t \pi_t \pi_t^{'} Z_t^{'} + \sigma_{\varepsilon t}^2 I_N \tag{A27}$$

where

$$\sigma_{\varepsilon t}^2 = E(x_{it}^{*2}) - E(x_{it}^* z_{it}^{'})[E(z_{it} z_{it}^{'})]^{-1} E(z_{it} x_{it}^*) \tag{A28}$$

Therefore,

$$Var_t(x_t^{*'} M_t v_t^*) = \sigma^2 \pi_t^{'} (Z_t^{'} Z_t) \pi_t + \sigma^2 \sigma_{\varepsilon t}^2 t + [E(x_{it}^* v_{it}^*)]^2 t \tag{A29}$$

and

$$Var(x_t^{*'} M_t v_t^*) = N\sigma^2 E(x_{it}^* z_{it}^{'})[E(z_{it} z_{it}^{'})]^{-1} E(z_{it} x_{it}^*) + \sigma^2 \sigma_{\varepsilon t}^2 t + [E(x_{it}^* v_{it}^*)]^2 t. \tag{A30}$$

We turn to consider a covariance term. Assuming that $t > s$ and given the variance decomposition

$$\begin{aligned} Cov(x_t^{*'} M_t v_t^*, x_s^{'} M_s v_s^*) &= Cov[E_t(x_t^{*'} M_t v_t^*), E_t(x_s^{*'} M_s v_s^*)] + \\ &\quad E[Cov_t(x_t^{*'} M_t v_t^*, x_s^{*'} M_s v_s^*)] \end{aligned} \tag{A31}$$

As before, since $E_t(x_t^{*'} M_t v_t^*)$ does not depend on Z_t, the first term on the rhs vanishes.

Moreover, due to conditional normality and the fact that $E_t(v_t^*) = 0$

$$\begin{aligned} Cov_t(x_t^{*'} M_t v_t^*, x_s^{*'} M_s v_s^*) &= tr[M_t E_t(x_t^* x_s^{*'}) M_s E_t(v_s^* v_t^{*'})] + \\ &\quad tr[M_t E_t(x_t^* v_s^{*'}) M_s E_t(x_s^* v_t^{*'})] \end{aligned} \tag{A32}$$

Since $E_t(v_s^* v_t^{*'}) = E(v_{is}^* v_{it}^*) I_N = 0$, the first of the two terms vanishes. Moreover, $E_t(x_s^* v_t^{*'}) = E(x_{is}^* v_{it}^*) I_N$, so we obtain

$$\begin{aligned} E\left[Cov_t(x_t^{*'} M_t v_t^*, x_s^{*'} M_s v_s^*)\right] &= E(x_{is}^* v_{it}^*) E\left\{tr\left[M_t E_t(x_t^* v_s^{*'}) M_s\right]\right\} \quad \text{(A33)} \\ &= E(x_{is}^* v_{it}^*) E\left\{tr\left[E_t(Z_t \pi_t v_s^{*'} M_s)\right] + tr\left[E_t(\varepsilon_t v_s^{*'}) M_s M_t\right]\right\} \end{aligned}$$

Finally,

$$Cov(x_t^{*'} M_t v_t^*, x_s^{*'} M_s v_s^*) = sE(x_{is}^* v_{it}^*) E(x_{it}^* v_{is}^*) \tag{A34}$$

given that $tr(M_s M_t) = s$, $E_t(\varepsilon_t v_s^{'}) = E(\varepsilon_{it} v_{is}) I_N$, and $E_s(Z_t \pi_t v_s^{*'}) = E(z_{it}^{'} \pi_t v_{is}^*) I_N$. Substituting (A30) and (A34) into (A22) we obtain

$$Var\left(\frac{x^{*'} M v^*}{(NT)^{1/2}}\right) = \frac{\sigma^2}{T} \sum_{t=1}^{T-1} E(x_{it}^* z_{it}^{'}) [E(z_{it} z_{it}^{'})]^{-1} E(z_{it} x_{it}^*) + R_t^o \tag{A35}$$

where

$$R_t^o = \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} t\sigma_{\varepsilon t}^2 + \frac{1}{NT} \sum_{t=1}^{T-1} tr\left[K_t^{'} E(x_i^* v_i^{*'}) K_t K_t^{'} E(x_i^* v_i^{*'}) K_t\right] \tag{A36}$$

where K_t is a selection matrix of order $(T-1) \times t$ given by $K_t = (0 : I_t)^{'}$.

After some messy algebra we find that

$$\sum_{t=1}^{T-1} t\sigma_{\varepsilon t}^2 = O(T \log T) \tag{A37}$$

and therefore the first term of R_t^o is $o(1)$ provided $(\log T)/N \to 0$. The second term of R_t^o can be written as

$$\frac{1}{NT} \sum_{t=2}^{T} tr\left[(H_t Q_t H_t^{'}) E(x_i v_i^{'}) (H_t Q_t H_t^{'}) E(x_i v_i^{'})\right] \tag{A38}$$

taking into account that in fact $A^{'} K_t K_t^{'} A = H_{t+1} Q_{t+1} H_{t+1}^{'}$. Hence this term contains a sum of terms of the type given in (A6) above which also turns out to be $o(1)$. Therefore, $R_t^o = o(1)$, what establishes the first part of (25).

Concerning the leading term of (A35), after some algebra it can be seen to take the following expression:

$$\frac{\sigma^2}{T} \sum_{t=1}^{T-1} E(x_{it}^* z_{it}^{'}) \left[E(z_{it} z_{it}^{'})\right]^{-1} E(z_{it} x_{it}^*) = \frac{\sigma^4}{(1-\alpha^2)}$$

$$+\frac{1}{T}\frac{\sigma^4}{(1-\alpha^2)}\left[\left(\frac{\alpha}{1-\alpha}\right)^2\left(\sum_{t=2}^{T}\frac{(1-\alpha^{t-1})^2}{t(t-1)}\right)-\left(\sum_{t=1}^{T}\frac{1}{t}\right)-\frac{2\alpha}{(1-\alpha)}\left(\sum_{t=2}^{T}\frac{(1-\alpha^{t-1})}{t}\right)\right]$$
$$-\frac{1}{T}\frac{\sigma^4}{(1-\alpha^2)}\lambda\sum_{t=2}^{T}\left(\left[t-1+\left(\frac{\alpha}{1-\alpha}\right)^2\frac{(1-\alpha^{t-1})^2}{(t-1)}-\frac{2\alpha(1-\alpha^{t-2})}{1-\alpha}\right]\right.$$
$$\left.\frac{(1+\alpha)}{t(1-\alpha+\lambda[2\alpha+(T-t+1)(1-\alpha)])}\right) \quad \text{(A39)}$$

where $\lambda=\sigma_\eta^2/\sigma^2$. Since the last two terms are $o(1)$ the proof to the second part follows.

Proof of (26): We have

$$Cov\left(\frac{x^{*'}v^*}{(NT)^{1/2}},\frac{x^{*'}Mv^*}{(NT)^{1/2}}\right)=\frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}E(x_t^{*'}v_t^*x_s^{*'}M_sv_s^*)-\frac{1}{NT}E(x^{*'}v^*)E(x^{*'}Mv^*) \quad \text{(A40)}$$

Conditional on Z_s, due to conditional normality and the fact that $E_s(v_s^*)=0$ we have

$$E_s(x_t^{*'}v_t^*x_s^{*'}M_sv_s^*)=tr[E_s(x_t^*v_t^{*'})]tr[E_s(x_s^*v_s^{*'})M_s]$$
$$+tr[E_s(x_s^*x_t^{*'})E_s(v_t^*v_s^{*'})M_s]+tr[E_s(x_s^*v_t^{*'})E_s(x_t^*v_s^{*'})M_s] \quad \text{(A41)}$$

Note that the expected value of the first term is:

$$E\{tr[E_s(x_t^*v_t^{*'})]tr[E_s(x_s^*v_s^{*'})M_s]\}=sE(x_{is}^*v_{is}^*)E(x_{it}^*v_{it}^*) \quad \text{(A42)}$$

which will cancel with the last term in (A40).

Since for $t\neq s$ $\mathrm{E}_s(v_t^*v_s^{*'})=E(v_{it}^*v_{is}^*)I_N$, the second term vanishes except when $t=s$ in which case its expected value is given by

$$\sigma^2\sigma_{\varepsilon t}^2t+N\sigma^2E(x_{it}^*z_{it}^{'})[E(z_{it}z_{it}^{'})]^{-1}E(z_{it}x_{it}^*)$$

For the third term we obtain:

$$E\{tr[E_s(x_s^*v_t^{*'})E_s(x_t^*v_s^{*'})M_s]\}=\begin{cases} sE(x_{it}^*v_{is}^*)E(x_{is}^*v_{it}^*) & \text{if } t\geq s\\ sE(x_{it}^*v_{is}^*)E(v_{it}^*\varepsilon_{is})+NE(x_{it}^*v_{is}^*)E(\pi_s^{'}z_{is}v_{it}^*) & \text{if } t<s\end{cases} \quad \text{(A43)}$$

Therefore, collecting terms, the covariance is given by:

$$Cov\left(\frac{x^{*'}v^*}{(NT)^{1/2}}, \frac{x^{*'}Mv^*}{(NT)^{1/2}}\right) = \frac{\sigma^2}{T}\sum_{t=1}^{T-1} E(x_{it}^* z_{it}^{'})[E(z_{it}z_{it}^{'})]^{-1}E(z_{it}x_{it}^*)$$

$$+\frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\sigma_{\varepsilon_t}^2 t + \frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} sE(x_{it}^* v_{is}^*)E(x_{is}^* v_{it}^*) +$$
$$\frac{1}{NT}\sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}(N-s)E(x_{it}^* v_{is}^*)E(\pi_s^{'} z_{is} v_{it}^*) \tag{A44}$$

An expression for the leading term of (A44) is given in (A39), and it is seen to converge to $\sigma^4/(1-\alpha^2)$. Moreover, according to (A37), the second term is $o(1)$. Thus, it remains to show that the other two terms in (A44) tend to zero as N and T tend to infinity. We begin by considering the third term in (A44)

$$\frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} sE(x_{it}^* v_{is}^*)E(x_{is}^* v_{it}^*) = \frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} s\ a_t^{'}E(x_i v_i^{'})a_s a_s^{'}E(x_i v_i^{'})a_t$$

$$= \frac{1}{NT}tr\left[\left(\sum_{t=1}^{T-1} a_t a_t^{'}\right)E(x_i v_i^{'})\left(\sum_{s=1}^{T-1} s\ a_s a_s^{'}\right)E(x_i v_i^{'})\right]$$

$$= \frac{1}{NT}\sum_{s=2}^{T} tr[QE(x_i v_i^{'})H_s Q_s H_s^{'}E(x_i v_i^{'})], \tag{A45}$$

since $Q = \sum_{t=1}^{T-1} a_t a_t^{'}$ and, in view of (A18), $\sum_{s=1}^{T-1} s\ a_s a_s^{'} = \sum_{s=2}^{T} H_s Q_s H_s^{'}$.

Now letting $v_{i[s]} = H_s^{'}v_i$ and $x_{i[s]} = H_s^{'}x_i$, we have

$$\frac{1}{NT}\sum_{s=2}^{T} tr[QE(x_i v_i^{'})H_s Q_s H_s^{'}E(x_i v_i^{'})] = \frac{1}{NT}\sum_{s=2}^{T} tr[E(x_i v_{i[s]}^{'})E(x_{i[s]}v_i^{'})]$$

$$-\frac{1}{NT}\sum_{s=2}^{T}\frac{1}{T}\iota_T^{'}E(x_i v_{i[s]}^{'})E(x_{i[s]}v_i^{'})\iota_T - \frac{1}{NT}\sum_{s=2}^{T}\frac{1}{s}\iota_s^{'}E(x_{i[s]}v_i^{'})E(x_i v_{i[s]}^{'})\iota_s$$

$$+\frac{1}{NT}\sum_{s=2}^{T}\frac{1}{sT}[\iota_T^{'}E(x_i v_{i[s]}^{'})\iota_T][\iota_s^{'}E(x_{i[s]}v_i^{'})\iota_T] \tag{A46}$$

Notice that $tr[E(x_i v'_{i[s]})E(x_{i[s]} v'_i)] = 0$. Moreover, direct calculation of the remaining three terms reveals that they are $o(1)$.

We turn to consider the fourth term to the right of (A44). Taking into account that $E(\pi'_s z_{is} v^*_{it}) = 0$ for $s \leq t$, after some manipulations, we obtain:

$$\frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}(N-s)E(x^*_{it}v^*_{is})E(\pi'_s z_{is} v^*_{it}) = \frac{1}{NT}\sum_{s=1}^{T-1}(N-s)tr[QE(x_i v^*_{is})E(\pi'_s z_{is} v'_i)] =$$

$$\frac{1}{NT}\sum_{s=1}^{T-1}(N-s)E(\pi'_s z_{is} v'_i)E(x_i v^*_{is}) - \frac{1}{NT^2}\sum_{s=1}^{T-1}(N-s)\iota'_T E(x_i v^*_{is})E(\pi'_s z_{is} v'_i)\iota_T \tag{A47}$$

Direct evaluation shows that the first term to the right of (A47) is zero. On the other hand, after some algebra, the second term of (A47) can be seen to take the following expression :

$$\frac{1}{NT^2}\sum_{s=1}^{T-1}(N-s)\iota'_T E(x_i v^*_{is})E(\pi'_s z_{is} v'_i)\iota_T =$$

$$\frac{\sigma^4}{NT^2}\sum_{s=1}^{T-1}(N-s)\left(\frac{T-s}{T-s+1}\right)\left(1-\frac{\alpha(1-\alpha^{T-s})}{(1-\alpha)(T-s)}\right)$$

$$\left[\left(\frac{1-\alpha^{T-s}}{(1-\alpha)^2(T-s)}\right)-\left(\frac{\alpha^{T-s}}{1-\alpha}\right)\right]\left[\left(\frac{1-\alpha^{s-1}}{1-\alpha}\right)-\left(\frac{\lambda(s-1)}{(1-\alpha)+\lambda\left(2\alpha+s\left(1-\alpha\right)\right)}\right)\right] \tag{A48}$$

This expression involves products of terms which tend to zero as N and T tend to infinity from which the proof of (26) follows.

Proof of (27): We shall rely on the identity:

$$\frac{1}{NT}(x^{*'}Mx^*) = \frac{1}{NT}(x^{*'}x^*) - R^1_{NT} \tag{A49}$$

where

$$R_{NT}^1 = \frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*'} P_t x_t^* = \frac{1}{NT}\sum_{t=1}^{T-1} \varepsilon_t' P_t \varepsilon_t$$

and P_t denotes the matrix $(I - M_t)$. Notice that the second equality results from substituting the linear projections given by (A26).

We know from (18) that the probability limit of the first term on the rhs of (A49) is $\sigma^2/(1-\alpha^2)$. We now show that R_{NT}^1 is $o_p(1)$. Firstly, we have

$$E(R_{NT}^1) = \frac{1}{NT}\sum_{t=1}^{T-1} E\left\{tr[P_t E_t(\varepsilon_t \varepsilon_t')]\right\} \tag{A50}$$

Since $E_t(\varepsilon_t) = 0$, $E_t(\varepsilon_t \varepsilon_t^{'})$ is the conditional variance of ε_t, which due to the joint normality of ε_t and Z_t does not depend on Z_t. Moreover, by cross-sectional independence $E_t(\varepsilon_t \varepsilon_t^{'}) = \sigma_{\varepsilon t}^2 I_N$ and using the fact that $tr(P_t) = N - t$, we obtain

$$E(R_{NT}^1) = \frac{1}{T}\sum_{t=1}^{T-1} \sigma_{\varepsilon t}^2 - \frac{1}{NT}\sum_{t=1}^{T-1} \sigma_{\varepsilon t}^2 t \tag{A51}$$

We know from the proof of (25) that each term to the right of (A50) converges to zero as N and T tend to infinity provided that $(\log T)/N \to 0$. Next, we consider the variance of R_{NT}^1

$$Var(R_{NT}^1) = \frac{1}{N^2T^2}\sum_{t=1}^{T-1} Var(\varepsilon_t' P_t \varepsilon_t) + \frac{1}{N^2T^2}\sum_{t \neq s} Cov(\varepsilon_t' P_t \varepsilon_t, \varepsilon_s' P_s \varepsilon_s) \tag{A52}$$

We first consider a variance term. Using the variance decomposition

$$Var(\varepsilon_t' P_t \varepsilon_t) = Var[E_t(\varepsilon_t' P_t \varepsilon_t)] + E[Var_t(\varepsilon_t^{'} P_t \varepsilon_t)] \tag{A53}$$

since $E_t(\varepsilon_t' P_t \varepsilon_t)$ does not depend on Z_t, the first term on the rhs vanishes. Next, since conditional on Z_t, ε_t is normal, $E_t(\varepsilon_t) = 0$, and P_t can be held constant given Z_t

$$Var_t(\varepsilon_t^{'} P_t \varepsilon_t) = 2tr[P_t E_t(\varepsilon_t \varepsilon_t') P_t E_t(\varepsilon_t \varepsilon_t')] = 2\sigma_{\varepsilon t}^4 (N - t) \tag{A54}$$

Therefore,

$$\frac{1}{N^2T^2}\sum_{t=1}^{T-1} Var(\varepsilon_t^{'}P_t\varepsilon_t) = \frac{2}{N^2T^2}\sum_{t=1}^{T-1}\sigma_{\varepsilon t}^4(N-t) \tag{A55}$$

We turn to consider a covariance term. Assuming $t > s$ and given the variance decomposition

$$Cov(\varepsilon_t'P_t\varepsilon_t, \varepsilon_s'P_s\varepsilon_s) = Cov[E_t(\varepsilon_t'P_t\varepsilon_t), E_t(\varepsilon_s'P_s\varepsilon_s)] + E[Cov_t(\varepsilon_t'P_t\varepsilon_t, \varepsilon_s'P_s\varepsilon_s)], \tag{A56}$$

as before, since $E_t(\varepsilon_t'P_t\varepsilon_t)$ does not depend on Z_t, the first term on the rhs vanishes. Moreover, due to conditional normality and the fact that $E_t(\varepsilon_t) = 0$

$$Cov_t(\varepsilon_t'P_t\varepsilon_t, \varepsilon_s'P_s\varepsilon_s) = 2tr[P_tE_t(\varepsilon_t\varepsilon_s')P_sE_t(\varepsilon_t\varepsilon_s^{'})] \tag{A57}$$

Since $E_t(\varepsilon_t\varepsilon_s') = E(\varepsilon_{it}\varepsilon_{is})I_N$ and given that $tr(P_tP_s) = N - t$, we obtain

$$Cov_t(\varepsilon_t'P_t\varepsilon_t, \varepsilon_s'P_s\varepsilon_s) = 2E^2(\varepsilon_{it}\varepsilon_{is})(N-t), \text{ for } t > s \tag{A58}$$

Therefore,

$$\frac{1}{N^2T^2}\sum_{t\neq s} Cov(\varepsilon_t'P_t\varepsilon_t, \varepsilon_s'P_s\varepsilon_s) = \frac{4}{N^2T^2}\sum_{s=1}^{T-2}\sum_{t=s+1}^{T-1} E^2(\varepsilon_{it}\varepsilon_{is})(N-t) \tag{A59}$$

and

$$Var(R_{NT}^1) = \frac{2}{N^2T^2}\sum_{t=1}^{T-1}\sigma_{\varepsilon t}^4(N-t) + \frac{4}{N^2T^2}\sum_{s=1}^{T-2}\sum_{t=s+1}^{T-1} E^2(\varepsilon_{it}\varepsilon_{is})(N-t) \tag{A60}$$

After some tedious algebra it can be found that the two terms in (A60) are $O[N^2T^2/(N-T)]$ and $O\{N^2T^2/[(N-T)\log T]\}$, respectively, and therefore both converge to zero as N and T tend to infinity. Hence, R_{NT}^1 converges to zero in probability and the result follows.

Proof of (28): We have

$$E(v^{*'}Mv^*) = \sum_{t=1}^{T-1} E(v_t^{*'}M_tv_t^*) = \sum_{t=1}^{T-1} E\{tr[M_tE_t(v_t^*v_t^{*'})]\} \tag{A61}$$

Since $E_t(v_t^*) = 0$, $E_t(v_t^* v_t^{*'})$ is the conditional variance, which due to joint normality of v_t^* and Z_t does not depend on Z_t. Therefore, $E_t(v_t^* v_t^{*'}) = E(v_t^* v_t^{*'}) = \sigma^2 I_N$. Hence, using the fact that $tr(M_t) = t$, we obtain

$$E\left(\frac{v^{*'}Mv^*}{NT}\right) = \frac{\sigma^2}{NT}\sum_{t=1}^{T-1} t \tag{A62}$$

Notice that this term converges to $c/2$ as N and T tend to infinity and $T/N \rightarrow c$, where $0 \leq c < \infty$. Next, we establish convergence in probability by proving that the variance of $(v^{*'}Mv^*)/NT$ converges to zero. We have that

$$Var\left(\frac{v^{*'}Mv^*}{NT}\right) = \frac{1}{N^2T^2}\sum_{t=1}^{T-1} Var(v_t^{*'}M_t v_t^*) + \frac{1}{N^2T^2}\sum_{t\neq s} Cov(v_t^{*'}M_t v_t^*, v_s^{*'}M_s v_s^*) \tag{A63}$$

We first consider a variance term. Given the variance decomposition

$$Var(v_t^{*'}M_t v_t^*) = Var[E_t(v_t^{*'}M_t v_t^*)] + E[Var_t(v_t^{*'}M_t v_t^*)],$$

since $E_t(v_t^{*'}M_t v_t^*)$ does not depend on Z_t, the first term vanishes. Next, since conditional on Z_t, v_t^* is normal with zero mean, we have

$$Var_t(v_t^{*'}M_t v_t^*) = 2tr[M_t E_t(v_t^* v_t^{*'}) M_t E_t(v_t^* v_t^{*'})] = 2\sigma^4 t \tag{A64}$$

Therefore,

$$\frac{1}{N^2T^2}\sum_{t=1}^{T-1} Var(v_t^{*'}M_t v_t^*) = \frac{2\sigma^4}{N^2T^2}\sum_{t=1}^{T-1} t \tag{A65}$$

We turn to consider a covariance term. Assuming that $t > s$ and given the covariance decomposition

$$\begin{aligned} Cov(v_t^{*'}M_t v_t^*, v_s^{*'}M_s v_s^*) = {} & Cov[E_t(v_t^{*'}M_t v_t^*), E_t(v_s^{*'}M_s v_s^*)] + \\ & E[Cov_t(v_t^{*'}M_t v_t^*, v_s^{*'}M_s v_s^*)], \end{aligned} \tag{A66}$$

as before, since $E_t(v_t^{*'} M_t v_t^*)$ does not depend on Z_t, the first term on the rhs vanishes. Moreover, due to conditional normality and the fact that $E_t(v_t^*) = 0$

$$Cov(v_t^{*'} M_t v_t^*, v_s^{*'} M_s v_s^*) = 2tr[M_t E_t(v_t^* v_s^{*'}) M_s E_t(v_s^* v_t^{*\prime})] \tag{A67}$$

Since $E_t(v_t^* v_s^{*'}) = E(v_t^* v_s^{*'}) = 0$ for $t \neq s$, this term is equal to zero. Thus, we obtain

$$Var\left(\frac{v^{*'} M v^*}{NT}\right) = \frac{2\sigma^4}{N^2 T^2} \sum_{t=1}^{T-1} t \tag{A68}$$

This term converges to zero as N and T tend to infinity and $T/N \to c$, where $0 \leq c < \infty$, from which the result follows.

Theorem 4

Proof of (37): We have

$$E(\Delta x^{'} M \Delta v) = \sum_{t=1}^{T-1} E(\Delta x_{t+1}^{'} M_t \Delta v_{t+1}) = \sum_{t=1}^{T-1} E\left\{tr[M_t E_t(\Delta v_{t+1} \Delta x_{t+1}^{'})]\right\} \tag{A69}$$

where $\Delta x_{t+1} \equiv \Delta y_t$ and Δv_{t+1} are $(N \times 1)$ vectors whose $i-th$ elements are Δy_{it} and Δv_{it+1} respectively.

Since $E_t(\Delta v_{t+1}) = 0$, $E_t(\Delta v_{t+1} \Delta x_{t+1}^{'})$ is the conditional covariance between Δv_{t+1} and Δx_{t+1}, which due to joint normality of Δv_{t+1}, Δx_{t+1}, and Z_t does not depend on Z_t. Moreover, by cross-sectional independence and using the fact that $tr(M_t) = t$, we have

$$E\left(\frac{\Delta x^{'} M \Delta v}{NT}\right) = \frac{1}{NT} \sum_{t=1}^{T-1} E(\Delta v_{it+1} \Delta y_{it}) t = \frac{-\sigma^2}{NT} \sum_{t=1}^{T-1} t \tag{A70}$$

Therefore, this term converges to $-\sigma^2(c/2)$ as N and T tend to infinity, provided that $T/N \to c$, where $0 \leq c < \infty$. We now establish convergence in probability by proving that the variance of $(\Delta x^{'} M\Delta v)/NT$ tends to zero given our asymptotics. Thus, we have

$$Var\left(\frac{\Delta x^{'} M\Delta v}{NT}\right) = \frac{1}{N^2T^2}\sum_{t=1}^{T-1} Var(\Delta x^{'}_{t+1}M_t\Delta v_{t+1})+$$

$$\frac{1}{N^2T^2}\sum_{t\neq s} Cov(\Delta x^{'}_{t+1}M_t\Delta v_{t+1}, \Delta x^{'}_{s+1}M_s\Delta v_{s+1}) \qquad \text{(A71)}$$

We first consider a variance term. Given the variance decomposition

$$Var(\Delta x^{'}_{t+1}M_t\Delta v_{t+1}) = Var[E_t(\Delta x^{'}_{t+1}M_t\Delta v_{t+1})] + E[Var_t(\Delta x^{'}_{t+1}M_t\Delta v_{t+1})] \qquad \text{(A72)}$$

Since $E_t(\Delta x^{'}_{t+1}M_t\Delta v_{t+1})$ does not depend on Z_t, the first term on the rhs of (A72) vanishes. Moreover, since conditional on Z_t, Δx_{t+1} and Δv_{t+1} are jointly normal, $E_t(\Delta v_{t+1}) = 0$, and M_t can be held constant given Z_t

$$Var_t(\Delta x^{'}_{t+1}M_t\Delta v_{t+1}) = tr[M_tE_t(\Delta x_{t+1}\Delta v^{'}_{t+1})M_tE_t(\Delta x_{t+1}\Delta v^{'}_{t+1})]+$$

$$tr[M_tE_t(\Delta x_{t+1}\Delta x^{'}_{t+1})M_tE_t(\Delta v_{t+1}\Delta v^{'}_{t+1})] \qquad \text{(A73)}$$

By cross-sectional independence $E_t(\Delta x_{t+1}\Delta v^{'}_{t+1}) = -\sigma^2 I_N$ and $E_t(\Delta v_{t+1}\Delta v^{'}_{t+1}) = 2\sigma^2 I_N$, so that

$$Var_t(\Delta x^{'}_{t+1}M_t\Delta v_{t+1}) = \sigma^4 t + 2\sigma^2 tr[M_tE_t(\Delta x_{t+1}\Delta x^{'}_{t+1})] \qquad \text{(A74)}$$

Let us now consider the linear projections

$$\Delta x_{t+1} = Z_t\pi_{dt} + \xi_t \qquad (t = 1, ..., T-1) \qquad \text{(A75)}$$

Due to joint normality of Δx_{t+1} and Z_t

$$E_t(\Delta x_{t+1}\Delta x^{'}_{t+1}) = Z_t\pi_{dt}\pi^{'}_{dt}Z^{'}_t + \sigma^2_{\xi t}I_N \qquad \text{(A76)}$$

where

$$\sigma_{\xi t}^2 = E[(\Delta x_{it+1})^2] - E(\Delta x_{it+1} z_{it}^{'})[E(z_{it} z_{it}^{'})]^{-1} E(\Delta x_{it+1} z_{it}) \tag{A77}$$

Hence, by inserting (A76) into (A74), we have

$$Var_t(\Delta x_{t+1}^{'} M_t \Delta v_{t+1}) = \sigma^4 t + 2\sigma^2 \sigma_{\xi t}^2 t + 2\sigma^2 \pi_{dt}^{'} Z_t Z_t^{'} \pi_{dt} \tag{A78}$$

and

$$Var(\Delta x_{t+1}^{'} M_t \Delta v_{t+1}) = \sigma^4 t + 2\sigma^2 \sigma_{\xi t}^2 t + 2\sigma^2 N E(\Delta x_{it+1} z_{it}^{'})[E(z_{it} z_{it}^{'})]^{-1} E(\Delta x_{it+1} z_{it}) \tag{A79}$$

We turn to consider a covariance term. Assuming that $s > t$ and given the variance decomposition

$$Cov(\Delta x_{t+1}^{'} M_t \Delta v_{t+1}, \Delta x_{s+1}^{'} M_s \Delta v_{s+1}) = Cov[E_s(\Delta x_{t+1}^{'} M_t \Delta v_{t+1}), E_s(\Delta x_{s+1}^{'} M_s \Delta v_{s+1})] +$$

$$E[Cov_s(\Delta x_{t+1}^{'} M_t \Delta v_{t+1}, \Delta x_{s+1}^{'} M_s \Delta v_{s+1})], \tag{A80}$$

as before, since $E_s(\Delta x_{s+1}^{'} M_s \Delta v_{s+1})$ does not depend on Z_s, the first term on the rhs vanishes. Moreover, due to conditional normality and the fact that $E_s(\Delta v_{s+1}) = 0$

$$Cov_s(\Delta x_{t+1}^{'} M_t \Delta v_{t+1}, \Delta x_{s+1}^{'} M_s \Delta v_{s+1}) = tr[M_t E_s(\Delta x_{t+1} \Delta x_{s+1}^{'}) M_s E_s(\Delta v_{s+1} \Delta v_{t+1}^{'})] +$$

$$tr[M_t E_s(\Delta x_{t+1} \Delta v_{s+1}^{'}) M_s E_s(\Delta x_{s+1} \Delta v_{t+1}^{'})] \tag{A81}$$

Firstly, $E_s(\Delta v_{s+1} \Delta v_{t+1}^{'}) = E(\Delta v_{is+1} \Delta v_{it+1}) I_N$ by cross-sectional independence, and $E(\Delta v_{is+1} \Delta v_{it+1}) = -\sigma^2$ if $s = t+1$ and zero otherwise. Moreover, $E_s(\Delta x_{t+1} v_{s+1}^{'}) = E(\Delta y_{it} \Delta v_{is+1}) I_N = 0$. Therefore, the covariance terms are equal to zero unless $s = t+1$. In this case, by inserting the linear projections of Δx_{t+1} and Δx_{t+2} into (A81) and after some manipulations, we have

$$Cov(\Delta x'_{t+1} M_t \Delta v_{t+1}, \Delta x'_{t+2} M_{t+1} \Delta v_{t+1}) = -\sigma^2 [N\pi'_{dt} E(z_{it} z'_{it+1})\pi_{dt+1} + tE(\xi_{it} z'_{it+1}\pi_{dt+1})] \tag{A82}$$

By collecting the terms in (A79) and (A82) we obtain the following expression

$$Var\left(\frac{\Delta x' M \Delta v}{NT}\right) = \frac{\sigma^4}{N^2T^2}\sum_{t=1}^{T-1} t + \frac{2\sigma^2}{N^2T^2}\sum_{t=1}^{T-1}\sigma^2_{\xi t} t +$$

$$\frac{2\sigma^2}{NT^2}\sum_{t=1}^{T-1} E(\Delta x_{it+1} z'_{it})[E(z_{it} z'_{it})]^{-1} E(\Delta x_{it+1} z_{it}) - \frac{2\sigma^2}{NT^2}\sum_{t=1}^{T-2} E(\Delta x_{it+1} z'_{it+1}\pi_{dt+1})$$

$$-\frac{2\sigma^2}{N^2T^2}\sum_{t=1}^{T-2}(t-N)E(\xi_{it} z'_{it+1}\pi_{dt+1}) \tag{A83}$$

Clearly, the first term converges to zero as N and T tend to infinity. Moreover, it can be shown that

$$\frac{2\sigma^2}{N^2T^2}\sum_{t=1}^{T-1}\sigma^2_{\xi t} t = \frac{2\sigma^4}{N^2T^2}\sum_{t=1}^{T-1}\left\{t + \frac{\lambda(1-\alpha)t}{(1-\alpha)+\lambda[2\alpha + t(1-\alpha)]}\right\} \tag{A84}$$

$$\frac{2\sigma^2}{N^2T^2}\sum_{t=1}^{T-1} E(\Delta x_{it+1} z'_{it})[E(z_{it} z'_{it})]^{-1} E(\Delta x_{it+1} z_{it}) = \frac{2\sigma^4}{N^2T^2}\left(\frac{1-\alpha}{1+\alpha}\right)\frac{T-1}{N^2T^2} -$$

$$\frac{2\sigma^4}{N^2T^2}\sum_{t=1}^{T-1}\frac{\lambda(1-\alpha)}{1-\alpha+\lambda[2\alpha+t(1-\alpha)]} \tag{A85}$$

$$-\frac{2\sigma^2}{NT^2}\sum_{t=1}^{T-2} E(\Delta x_{it+1} z'_{it+1}\pi_{dt+1}) = \frac{2\sigma^4}{NT^2}\left(\frac{1-\alpha}{1+\alpha}\right)(T-2) \tag{A86}$$

Notice that the second, the third and the fourth terms on the rhs of (A83) are $o(1)$. Finally, the last term on the rhs of (A83) turns out to be

$$-\frac{2\sigma^2}{N^2T^2}\sum_{t=1}^{T-2}(t-N)E(\xi_{it} z'_{it+1}\pi_{dt+1}) = \frac{2\sigma^4(1-\alpha)}{N^2T^2}\left[\frac{(T-2)(T-1)}{2} - N\right] +$$

$$\frac{2\sigma^4(1-\alpha)^2}{N^2T^2}\sum_{t=1}^{T-2}\frac{\lambda(t-N)}{1-\alpha+\lambda[2\alpha+t(1-\alpha)]}-$$

$$\frac{2\sigma^4(1-\alpha)}{N^2T^2}\sum_{t=1}^{T-2}\frac{\lambda(t-N)}{1-\alpha+\lambda[2\alpha+(t+1)(1-\alpha)]}\left(\frac{1+\alpha-\alpha^2}{1+\alpha}\right)-$$

$$\frac{2\sigma^4(1-\alpha)^2}{N^2T^2}\sum_{t=1}^{T-2}\left\{\frac{\lambda(t-N)}{1-\alpha+\lambda[2\alpha+(t+1)(1-\alpha)]}\right\}\left\{\frac{\lambda}{1-\alpha+\lambda[2\alpha+t(1-\alpha)]}\right\} \quad \text{(A87)}$$

Hence, this term contains a sum of terms which are $o(1)$. Therefore, the variance converges to zero as N and T tend to infinity, from which the result follows.

Proof of (38): We shall rely on the identity

$$\frac{1}{NT}(\Delta x^{'}M\Delta x)=\frac{1}{NT}(\Delta x^{'}\Delta x)-R_{NT}^2 \quad \text{(A88)}$$

where

$$R_{NT}^2=\frac{1}{NT}\sum_{t=1}^{T-1}\Delta x_{t+1}^{\prime}P_t\Delta x_{t+1}=\frac{1}{NT}\sum_{t=1}^{T-1}\xi_t^{\prime}P_t\xi_t \quad \text{(A89)}$$

and P_t denotes the matrix $(I-M_t)$. Notice that the second equality follows from substituting the linear projections given by (A75). We now derive the probability limits of each term on the rhs of (A89). Firstly, we have

$$E\left(\sum_{t=1}^{T-1}\frac{\Delta x_{t+1}^{'}\Delta x_{t+1}}{NT}\right)=\frac{1}{T}E(\Delta x_i^{'}\Delta x_i) \quad \text{(A90)}$$

where Δx_i is the $(T-1)\times 1$ vector whose $t-th$ element is Δx_{it+1}. Therefore,

$$E\left(\sum_{t=1}^{T-1}\frac{\Delta x_{t+1}^{'}\Delta x_{t+1}}{NT}\right)=\frac{1}{T}\sum_{t=1}^{T-1}E[(\Delta x_{it+1})^2]=\frac{2\sigma^2(T-1)}{(1+\alpha)T} \quad \text{(A91)}$$

This term converges to $2\sigma^2/(1+\alpha)$ as N and T tend to infinity. Due to cross-sectional independence

$$Var\left(\sum_{t=1}^{T-1}\frac{\Delta x_{t+1}^{'}\Delta x_{t+1}}{NT}\right)=\frac{Var(\Delta x_i^{'}\Delta x_i)}{NT^2} \quad \text{(A92)}$$

Moreover, due to joint normality of Δx_i

$$Var(\Delta x_i^{'}\Delta x_i) = 2tr[E(\Delta x_i \Delta x_i^{'})E(\Delta x_i \Delta x_i^{'})] \tag{A93}$$

After some algebra, we find that

$$Var\left(\sum_{t=1}^{T-1} \frac{\Delta x_{t+1}^{'}\Delta x_{t+1}}{NT}\right) = \frac{1}{NT}\frac{4\sigma^2(3+\alpha)}{(1+\alpha)^3} - \frac{1}{NT^2}\frac{16\sigma^4}{(1+\alpha)^3} - \frac{1}{NT^2}\frac{4\sigma^4\alpha^2(1-\alpha^{2T-4})}{(1+\alpha)^4} \tag{A94}$$

Notice that the variance converges to zero as N and T tend to infinity. Therefore, the probabiliy limit of the first term on the rhs of (A87) is $2\sigma^2/(1+\alpha)$.

We now consider the second term on the rhs of (A88). Firstly, we have

$$E(R_{NT}^2) = \frac{1}{NT}\sum_{t=1}^{T-1} E\left\{tr[P_t E_t(\xi_t \xi_t')]\right\}, \tag{A95}$$

since $E_t(\xi_t) = 0$, $E_t(\xi_t\xi_t^{'})$ is the conditional variance, which due to joint normality of ξ_t and Z_t does not depend on Z_t. Hence, using the fact that $tr(P_t) = N - t$,

$$E(R_{NT}^2) = \frac{1}{NT}\sum_{t=1}^{T-1}\sigma_{\xi t}^2(N-t) = \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\left\{(N-t) + \frac{\lambda(1-\alpha)(N-t)}{(1-\alpha)+\lambda[2\alpha+t(1-\alpha)]}\right\} \tag{A96}$$

Therefore, this term converges to $\sigma^2(1-\frac{c}{2})$ as N and T tend to infinity and the ratio $T/N \to c$, where $0 \leq c < \infty$. On the other hand, we have

$$Var(R_{NT}^1) = Var\left(\sum_{t=1}^{T-1}\frac{\xi_t' P_t \xi_t}{NT}\right) = \frac{1}{N^2T^2}\sum_{t=1}^{T-1} Var(\xi_t^{'}P_t\xi_t) + \frac{1}{N^2T^2}\sum_{t\neq s} Cov(\xi_t^{'}P_t\xi_t, \xi_s^{'}P_s\xi_s) \tag{A97}$$

We first consider a variance term. Given the variance decomposition

$$Var(\xi_t^{'} P_t \xi_t) = Var[E_t(\xi_t^{'} P_t \xi_t)] + E[Var_t(\xi_t^{'} P_t \xi_t)] \tag{A98}$$

since $E_t(\xi_t^{'} P_t \xi_t)$ does not depend on Z_t, the first term on the rhs vanishes. Moreover, since conditional on Z_t, ξ_t is normal, $E_t(\xi_t) = 0$, and P_t can be held constant given Z_t

$$Var_t(\xi_t^{'} P_t \xi_t) = 2tr[P_t E_t(\xi_t \xi_t^{'}) P_t E_t(\xi_t \xi_t^{'})] = 2\sigma_{\xi t}^4 (N-t) \tag{A99}$$

Therefore,

$$\frac{1}{N^2T^2} \sum_{t=1}^{T-1} Var(\xi_t^{'} P_t \xi_t) = \frac{2}{N^2T^2} \sum_{t=1}^{T-1} \sigma_{\xi t}^4 (N-t) \tag{A100}$$

We turn to consider a covariance term. Assuming that $t > s$ and given the variance decomposition

$$Cov(\xi_t^{'} P_t \xi_t, \xi_s^{'} P_s \xi_s) = Cov[E_t(\xi_t^{'} P_t \xi_t), E_t(\xi_s^{'} P_s \xi_s)] + E[Cov_t(\xi_t^{'} P_t \xi_t, \xi_s^{'} P_s \xi_s)] \tag{A101}$$

as before, since $E_t(\xi_t^{'} P_t \xi_t)$ does not depend on Z_t, the first term on the rhs of (A101) vanishes. Moreover, due to conditional normality and the fact that $E_t(\xi_t) = 0$

$$Cov_t(\xi_t^{'} P_t \xi_t, \xi_s^{'} P_s \xi_s) = 2tr[P_t E_t(\xi_t \xi_s^{'}) P_s E_t(\xi_s \xi_t^{'})] = 0 \tag{A102}$$

To see this, note that $E_t(\xi_t \xi_s^{'}) = E(\xi_{it} \xi_{is}) I_N$ and $E(\xi_{it} \xi_{is})$ is zero for $t < s$ due to $E_t(\xi_t) = 0$ and the fact that ξ_s is a function of Z_t. Therefore,

$$Var(R_{NT}^2) = Var\left(\sum_{t=1}^{T-1} \frac{\xi_t^{\prime} P_t \xi_t}{NT}\right) = \frac{2}{N^2T^2} \sum_{t=1}^{T-1} \sigma_{\xi t}^4 (N-t) = \frac{2\sigma^4}{N^2T^2} \sum_{t=1}^{T-1} (N-t) +$$

$$\frac{2\sigma^4 (1-\alpha)^2}{N^2T^2} \sum_{t=1}^{T-1} \frac{\lambda^2 (N-t)}{\{(1-\alpha) + \lambda[2\alpha + t(1-\alpha)]\}^2} +$$

$$\frac{4\sigma^4(1-\alpha)}{N^2T^2}\sum_{t=1}^{T-1}\frac{(N-t)}{(1-\alpha)+\lambda[2\alpha+t(1-\alpha)]} \tag{A103}$$

Hence, this term contains a sum of terms which are $o(1)$ and R^2_{NT} converges in probability to $\sigma^2(1-\frac{c}{2})$. Given the probability limit of the leading term on the rhs of (A88) derived above, the result follows.

Proof of (39): The result follows immediately from (37) and (38).

Random effects maximum likelihood

Expression for log density (40): The model can be written as:

$$By_i = \alpha y_{i0} d_i + u_i \tag{A104}$$

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

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

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

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

$$f(y_i \mid y_{i0}) = f(u_i \mid y_{i0}) \det(B) \tag{A105}$$

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

$$f(u_i \mid y_{i0}) = f(\overline{u}_i, u_i^* \mid y_{i0}) \det(H) \tag{A106}$$

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

From conditions *A1* and *A3*, $u_i \sim N\left(0, \sigma^2(I_T + \lambda\iota_T\iota_T')\right)$ where $\lambda = \sigma^2_\eta/\sigma^2$. Hence,

$$Hu_i \sim N\left[0, \sigma^2 \begin{pmatrix} \frac{1}{T}+\lambda & 0 \\ 0 & I_{T-1} \end{pmatrix}\right]. \tag{A107}$$

From *A2'* and *A3*, $\eta_i \mid y_{i0}$ is also normally distributed with $E(\eta_i \mid y_{i0}) = \varphi y_{i0}$ and $Var(\eta_i \mid y_{i0}) = \sigma_\eta^2 - \varphi^2 Var(y_{i0})$, where $\varphi = \delta\sigma_\eta^2/Var(y_{i0})$, and $Var(y_{i0}) = (\delta^2\sigma_\eta^2 + \omega_{oo}^2)$. Then, the result in (40) follows from noting that we have:

$$Hu_i \mid y_{i0} \sim N\left[\begin{pmatrix} \varphi y_{i0} \\ 0 \end{pmatrix}, \begin{pmatrix} \omega^2 & 0 \\ 0 & \sigma^2 I_{T-1} \end{pmatrix}\right] \tag{A108}$$

That is, $E(\overline{u}_i \mid y_{i0}) = \varphi y_{i0}$, $E(u_i^* \mid y_{i0}) = 0$, and $\omega^2 = Var(\overline{u}_i \mid y_{i0}) = [\sigma_\eta^2 - \varphi^2 Var(y_{i0})] + \sigma^2/T$.

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

$$E\left(x_i^{*\prime}(y_i^* - \alpha x_i^*)\right) = -\frac{\sigma^2}{\omega^2} E\left(\overline{x}_i(\overline{y}_i - \alpha\overline{x}_i - \varphi y_{i0})\right) \tag{A109}$$

$$\frac{1}{\omega^2} E\left(y_{i0}(\overline{y}_i - \alpha\overline{x}_i - \varphi y_{i0})\right) = 0 \tag{A110}$$

$$E\left((y_i^* - \alpha x_i^*)'(y_i^* - \alpha x_i^*) - \sigma^2\right) = 0 \tag{A111}$$

$$E\left((\overline{y}_i - \alpha\overline{x}_i - \varphi y_{i0})^2 - \omega^2\right) = 0 \tag{A112}$$

Note that under assumption *A2*, (A109) multiplied by N corresponds to expression (16).

The concentrated joint likelihood as a function of α and λ: Under *A1*, *A2'* and *A3*, $z_{i(T+1)} \sim N(0, \Omega^*)$ where

$$\Omega^* = \begin{pmatrix} (\delta^2\sigma_\eta^2 + \omega_{oo}^2) & \delta\sigma_\eta^2 \iota_T' \\ \delta\sigma_\eta^2 \iota_T & \sigma^2(I_T + \lambda\iota_T\iota_T') \end{pmatrix}.$$

Thus, the joint log likelihood of $z_{i(T+1)} = (y_{i0}, y_{i1}, ..., y_{iT})'$ is given by:

$$L(\alpha, \lambda, \sigma^2, \delta, \omega_{oo}^2) = -\frac{N}{2}\log\ \det(\Omega^*) - \frac{1}{2}\sum_{i=1}^N z_{i(T+1)}' \Omega^{*-1} z_{i(T+1)}.$$

Concentrating σ^2, δ, and ω_{oo}^2 out of $L(\alpha,\lambda,\sigma^2,\delta,\omega_{oo}^2)$, we obtain

$$\begin{aligned} L(\alpha,\lambda) &= -\frac{NT}{2}\ln\left[\sum_{i=1}^{N}u_i'\left(I_T-\frac{\lambda}{(1+\lambda T)}\iota_T\iota_T'\right)u_i\right]-\frac{N}{2}\ln(1+\lambda T) \\ &\quad -\frac{N}{2}\ln(\overline{u}'S_0\overline{u})+\frac{N}{2}\ln(\overline{u}'\overline{u}). \end{aligned}$$

This criterion can be used to obtain ML estimates of α that enforce $\lambda \geq 0$. If $L(\alpha,\lambda)$ is concentrated further using the MLE of λ in the absence of the inequality constraint, we obtain the estimation criterion (41) which only depends on α.

Consistency of the RML: From (41) $\widehat{\alpha}_{RML}$ is the minimizer of

$$\ln\left[\frac{1}{NT}(y^*-ax^*)'(y^*-ax^*)\right]+\frac{1}{(T-1)}\ln\left[\frac{1}{N}(\overline{y}-a\overline{x})'S_0(\overline{y}-a\overline{x})\right] \quad \text{(A113)}$$

As $T \rightarrow \infty$ regardless of the asymptotic behaviour of N, the second term in (A113) 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 $T \rightarrow \infty$. However, unlike WG, RML is also consistent when T is fixed and $N \rightarrow \infty$ provided conditions (A109)-(A112) hold (including time series homoskedasticity).

Asymptotic Normality of the RML: The first and second derivatives at $a=\alpha$ of the concentrated log likelihood:

$$L(a)=-N(T-1)\ln\left[(y^*-ax^*)'(y^*-ax^*)\right]-N\ln\left[(\overline{y}-a\overline{x})'S_0(\overline{y}-a\overline{x})\right] \quad \text{(A114)}$$

are given by

$$\frac{\partial L(\alpha)}{\partial a}=\left(\frac{v^{*\prime}v^*}{N(T-1)}\right)^{-1}(x^{*\prime}v^*)+\left(\frac{\overline{u}'S_0\overline{u}}{N}\right)^{-1}(\overline{x}'S_0\overline{u}) \quad \text{(A115)}$$

$$\frac{1}{NT}\frac{\partial^2 L(\alpha)}{\partial a^2} = -\left(\frac{v^{*\prime}v^*}{N(T-1)}\right)^{-1}\left(\frac{x^{*\prime}x^*}{NT}\right)+2\left(\frac{v^{*\prime}v^*}{NT}\right)^{-2}\left(\frac{x^{*\prime}v^*}{NT}\right)^2\left(\frac{T-1}{T}\right)$$
$$-\frac{1}{T}\left(\frac{\overline{u}'S_0\overline{u}}{N}\right)^{-1}\left(\frac{\overline{x}'S_0\overline{x}}{N}\right)+\frac{2}{T}\left(\frac{\overline{u}'S_0\overline{u}}{N}\right)^{-2}\left(\frac{\overline{x}'S_0\overline{u}}{N}\right)^2. \quad (A116)$$

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

$$\frac{1}{NT}\frac{\partial^2 L(\alpha)}{\partial a^2} \xrightarrow{p} -\frac{1}{(1-\alpha^2)}. \quad (A117)$$

To verify (A117), first note that from Lemmae 1 and 3 as $T\to\infty$, regardless of whether N is fixed or tends to infinity:

$$\frac{1}{NT}(x^{*\prime}v^*) \xrightarrow{p} 0$$
$$\frac{1}{NT}(x^{*\prime}x^*) \xrightarrow{p} \frac{\sigma^2}{(1-\alpha^2)}$$
$$\frac{v^{*\prime}v^*}{NT} \xrightarrow{p} \sigma^2$$

Moreover, as both N and T tend to infinity:

$$p\lim\left(\frac{\overline{u}'S_0\overline{u}}{N}\right) = p\lim\left(\frac{\overline{u}'\overline{u}}{N}\right)-\left(p\lim\frac{y_0'y_0}{N}\right)^{-1}\left(p\lim\frac{\overline{u}'y_0}{N}\right)^2$$
$$= \sigma_\eta^2-\left(E(y_{i0}^2)\right)^{-1}\left(\delta\sigma_\eta^2\right)^2$$

This is so because $E(\overline{u}'\overline{u}/N)=E(\overline{u}_i^2)=\sigma_\eta^2+(\sigma^2/T)\to\sigma_\eta^2$, and $Var(\overline{u}'\overline{u}/N)= N^{-1}Var(\overline{u}_i^2)\to 0$, since due to normality $Var(\overline{u}_i^2)=2\left[E(\overline{u}_i^2)\right]^2$. Similarly, $E(\overline{u}'y_0/N)=E(\overline{u}_iy_{i0})=E(\eta_iy_{i0})=\delta\sigma_\eta^2$, and $Var(\overline{u}'y_0/N)=N^{-1}Var(\overline{u}_iy_{i0})\to 0$, since due to normality $Var(\overline{u}_iy_{i0})=E(\overline{u}_i^2)E(y_{i0}^2)+\left[E(\overline{u}_iy_{i0})\right]^2$.

Using similar arguments we obtain:

$$p\lim\left(\frac{\overline{x}'S_0\overline{x}}{N}\right) = p\lim\left(\frac{\overline{x}'\overline{x}}{N}\right)-\left(p\lim\frac{y_0'y_0}{N}\right)^{-1}\left(p\lim\frac{\overline{x}'y_0}{N}\right)^2$$
$$= \frac{\sigma_\eta^2}{(1-\alpha)^2}-\left(E(y_{i0}^2)\right)^{-1}\left(\frac{\sigma_\eta^2}{(1-\alpha)^2}\right)^2$$

and

$$
\begin{aligned}
p\lim\left(\frac{\overline{x}'S_0\overline{u}}{N}\right) &= p\lim\left(\frac{\overline{x}'\overline{u}}{N}\right) - \left(p\lim\frac{y_0'y_0}{N}\right)^{-1}\left(p\lim\frac{\overline{x}'y_0}{N}\right)\left(p\lim\frac{\overline{u}'y_0}{N}\right) \\
&= \frac{\sigma_\eta^2}{(1-\alpha)} - \left(E(y_{i0}^2)\right)^{-1}\left(\frac{\sigma_\eta^2}{(1-\alpha)^2}\right)\left(\delta\sigma_\eta^2\right).
\end{aligned}
$$

Score: Now the scaled score can be written as

$$
(NT)^{-1/2}\frac{\partial L(\alpha)}{\partial a} = \frac{1}{\sigma^2}(NT)^{-1/2}[x^{*\prime}v^* - E(x^{*\prime}v^*)] + \Upsilon_{NT} + o_p(1) \quad \text{(A118)}
$$

where

$$
\Upsilon_{NT} = \left(\frac{\overline{u}'S_0\overline{u}}{N}\right)^{-1}(NT)^{-1/2}(\overline{x}'S_0\overline{u}) + \frac{1}{\sigma^2}(NT)^{-1/2}E(x^{*\prime}v^*). \quad \text{(A119)}
$$

Moreover, from (A109) and the fact that $\varphi = E(\overline{u}'y_0)/E(y_0'y_0)$ we have

$$
\frac{1}{\sigma^2}E(x^{*\prime}v^*) = -\frac{1}{\omega^2}\left[E(\overline{x}'\overline{u}) - \varphi E(\overline{x}'y_0)\right] = -\frac{1}{\omega^2}\left[E(\overline{x}'\overline{u}) - E(\overline{x}'y_0)E(\overline{u}'y_0)/E(y_0'y_0)\right]. \quad \text{(A120)}
$$

Hence,

$$
\Upsilon_{NT} = \left(\frac{N}{T}\right)^{1/2}\left[\left(\frac{\overline{u}'S_0\overline{u}}{N}\right)^{-1}\left(\frac{\overline{x}'S_0\overline{u}}{N}\right) - \frac{1}{\omega^2}\left(\frac{E(\overline{x}'\overline{u}) - E(\overline{x}'y_0)E(\overline{u}'y_0)/E(y_0'y_0)}{N}\right)\right]. \quad \text{(A121)}
$$

Note also that $(\overline{u}'S_0\overline{u}/N) - \omega^2 \xrightarrow{p} 0$ as N and T tend to infinity. Thus, if N and T tend to infinity, provided $0 \leq \lim(N/T) < \infty$, $\Upsilon_{NT} = o_p(1)$ and from result (20) in Theorem 1

$$
(NT)^{-1/2}\frac{\partial L(\alpha)}{\partial a} \xrightarrow{d} N\left(0, \frac{1}{(1-\alpha^2)}\right) \quad \text{(A122)}
$$

Given (A117) and (A122), the asymptotic normality result

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

follows from Theorem 4.1.3 in Amemiya (1985).

Table 1

Medians, interquartile ranges, and median absolute errors of the estimators ($N = 100$)

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	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

$\sigma_\eta^2 = 0,\ \sigma^2 = 1$, 1000 replications, iqr is the 75th-25th interquantile range;
mae denotes the median absolute error.

Table 2

Medians, interquartile ranges, and median absolute errors of the estimators ($N = 50$)

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	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

$\sigma_\eta^2 = 0,\ \sigma^2 = 1$, 1000 replications, iqr is the 75th-25th interquantile range;
mae denotes the median absolute error.

Table 3

Asymptotic biases of the estimates

	$\alpha = 0.2$				$\alpha = 0.5$				$\alpha = 0.8$			
	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

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 $c = T/N$.

Table A1

Medians, interquartile ranges, and median absolute errors of the estimators ($N = 100$)

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	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

$\sigma_\eta^2 = 0.2$, $\sigma^2 = 1$, 1000 replications, iqr is the 75th-25th interquantile range;
mae denotes the median absolute error.

Table A2

Medians, interquartile ranges, and median absolute errors of the estimators ($N = 100$)

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	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

$\sigma_\eta^2 = 1,\ \sigma^2 = 1$, 1000 replications, iqr is the 75th-25th interquantile range;
mae denotes the median absolute error.

Table A3

Medians, interquartile ranges, and median absolute errors of the estimators ($N = 50$)

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	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

$\sigma_\eta^2 = 0.2$, $\sigma^2 = 1$, 1000 replications; iqr is the 75th-25th interquantile range;
mae denotes the median absolute error.

Table A4

Medians, interquartile ranges, and median absolute errors of the estimators ($N = 50$)

	$\alpha = 0.2$					$\alpha = 0.5$					$\alpha = 0.8$				
	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

$\sigma_\eta^2 = 1,\ \sigma^2 = 1$, 1000 replications; iqr is the 75th-25th interquantile range;
mae denotes the median absolute error.