Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large

Jinyong Hahn
Brown University

Guido Kuersteiner
MIT

July, 2001

Acknowledgment: Helpful comments by Joshua Angrist, Gary Chamberlain, Jerry Hausman, Dale Jorgenson, Tony Lancaster, Whitney Newey, a Co-Editor, anonymous referees, and participants of seminars at BC/BU, Harvard/MIT, Johns Hopkins University, Pennsylvania State University, Rice University, Texas A & M, University of Virginia, 1998 CEME Conference, and 1999 Econometric Society North American Winter Meeting are appreciated.

Abstract

We consider a dynamic panel AR(1) model with fixed effects when both n and T are large. Under the "T fixed n large" asymptotic approximation, the ordinary least squares (OLS) or Gaussian maximum likelihood estimator (MLE) is known to be inconsistent due to the well-known incidental parameter problem. We consider an alternative asymptotic approximation where n and T grow at the same rate. It is shown that, although OLS or the MLE is asymptotically biased, a relatively simple fix to OLS or the MLE results in an asymptotically unbiased estimator. Under the assumption of Gaussian innovations, the bias-corrected MLE is shown to be asymptotically efficient by a Hajék type convolution theorem.

1 Introduction

In this paper, we consider estimation of the autoregressive parameter θ_0 of a dynamic panel data model with fixed effects. The model has additive individual time invariant intercepts (fixed effects) along with a parameter common to every individual. The total number of parameters is therefore equal to the number of individuals (n) plus the dimension of the common parameter. When the number of individuals (n) is large relative to the time series dimension (T), the ordinary least squares (OLS) or Gaussian maximum likelihood estimator (MLE) would lead to inconsistent estimates of the common parameter of interest. This is the well-known incidental parameter problem.[1] Inconsistency of OLS under T fixed n large asymptotics led to a focus on instrumental variables estimation in the recent literature. Most instrumental variables estimators are at least partly based on the intuition that first differencing yields a model free of fixed effects.[2] Despite its appeal as a procedure which avoids the incidental parameter problem, the instrumental variables based procedure is problematic as a general principle to deal with potentially nonlinear panel models because of its inherent reliance on first differencing. Except for a small number of cases where conditioning on some sufficient statistic eliminates fixed effects, there does not seem to exist any general strategy for potentially nonlinear panel models.

In this paper, we develop a strategy that could potentially be extended to nonlinear models by considering an alternative asymptotic approximation where both n and T are large. We analyze properties of OLS under this approximation for quite general innovation distributions. It is shown that OLS is consistent and asymptotically normal, although it is not centered at the true value of the parameter. The noncentrality parameter under our alternative asymptotic approximation implicitly captures bias of order $O\left(T^{-1}\right)$, which can be viewed as an alternative form of the incidental parameter problem. We develop a bias-corrected estimator by examining the noncentrality parameter. Our strategy can be potentially replicated in nonlinear panel models, although analytic derivations for nonlinear models are expected to be much more involved than in linear dynamic panel models. We can in principle iterate our strategy to eliminate biases of order $O\left(T^{-2}\right)$ or $O\left(T^{-3}\right)$, although we do not pursue such a route here.

Having removed the asymptotic bias, we raise an efficiency question for the case of Gaussian innovations where OLS is equivalent to the maximum likelihood estimator. Is the bias-corrected MLE asymptotically efficient among the class of all reasonable estimators? In order to assess efficiency, we derive a Hajék-type convolution theorem, and show that the asymptotic distribution of the bias-corrected MLE is equal to the minimal distribution in the convolution theorem.

Our alternative asymptotic approximation is expected to be of practical relevance if T is not too small compared to n as is the case for example in cross-country studies.[3] The properties of dynamic panel models are usually discussed under the implicit assumption that T is small and n is large relying on T fixed n large asymptotics. Such asymptotics seem quite natural when T is indeed very small compared

[1] See Neyman and Scott (1948) for a general discussion of the incidental parameter problem, and Nickell (1981) for its implication in the particular linear dynamic panel model of interest.

[2] For discussion of various instrumental variables estimators and moment restrictions, see Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1992), Chamberlain (1992), Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1995), and Hahn (1997).

[3] Inter-country comparison studies seem to be a reasonable application for our perspective. See Islam (1995) and/or Lee, Pesaran, and Smith (1998) for recent examples of inter-country comparison studies.

to n. In cases where T and n are of comparable size we expect our approximation to be more accurate.

2 Bias Corrected OLS (MLE) for Panel VARs with Fixed Effects

In this section, we consider estimation of the autoregressive parameter θ_0 in a dynamic panel model with fixed effects

$$y_{it}' = \alpha_i' + y_{it-1}'\theta_0' + \varepsilon_{it}', \quad i = 1, \ldots, n;\ t = 1, \ldots, T, \tag{1}$$

where y_{it} is an m-dimensional vector and ε_{it} is i.i.d. normal. We establish the asymptotic distribution of the OLS estimator (MLE) for θ_0 under the alternative asymptotics, and develop an estimator free of (asymptotic) bias. We go on to argue that the bias-corrected MLE is efficient using a Hajék type convolution theorem, and provide an intuitive explanation of efficiency by considering the limit of the Cramer-Rao lower bound. Finally, we point out that the asymptotic distribution of the bias-corrected MLE is robust to nonnormality by presenting an asymptotic analysis for a model where ε_{it} violates the normality assumption. We leave the efficiency analysis of models with nonnormal innovations for future research.

Model (1) may be understood as a parametric completion of the univariate dynamic panel AR(1) model with additional regressors. If we write $y_{it} = \left(Y_{it}, X_{it+1}'\right)'$, then the first component of the model (1) can be rewritten as

$$Y_{it} = c_i + \beta_0 \cdot Y_{it-1} + \gamma_0' X_{it} + e_{it}, \quad i = 1, \ldots, n;\ t = 1, \ldots, T \tag{2}$$

where c_i and $(\beta_0, \gamma_0')'$denote the first component of α_i and the first column of θ_0'. This implies that, under the special circumstances where X_{it} follows a first order VAR, we can regard model (1) as a completion of model (2). Under this interpretation, model (1) encompasses panel models with further regressors such as (2).

Even more generally, model (1) can be parametrized to be the reduced form of a dynamic simultaneous equation system in y_{it} allowing for higher order VAR dynamics as well as exogenous regressors. This requires imposing blockwise zero and identity restrictions on θ_0. It is well-known that MLE reduces to blockwise OLS as long the restrictions are block recursive. Even though we do not spell out the details of this interpretation of our model, this more general case could in principle be dealt with in our framework.

If we assume ε_{it} is i.i.d. over t and i, and has a zero mean multivariate normal distribution, then the MLE (fixed effects estimator/OLS) takes the form

$$\widehat{\theta}' = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \overline{y}_{i-}\right)\left(y_{it-1} - \overline{y}_{i-}\right)'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \overline{y}_{i-}\right)\left(y_{it} - \overline{y}_{i}\right)'\right),$$

where $\bar{y}_i \equiv \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{y}_{i-} \equiv \frac{1}{T}\sum_{t=1}^{T} y_{it-1}$. We examine properties of $\widehat{\theta}$ under potential nonnormality of ε_{it} under the alternative asymptotics. If the innovations ε_{it} are not normal then the resulting estimator $\widehat{\theta}$ is a pseudo-MLE, and does no longer possess the efficiency properties of the exact MLE. For this reason we impose the additional assumption of normality for our discussion of asymptotic efficiency later in this section. We impose the following conditions:

Condition 1 *(i)* ε_{it} *is* i.i.d. *across* i *and strictly stationary in* t *for each* i, $E[\varepsilon_{it}] = 0$ *for all* i *and* t, $E[\varepsilon_{it}\varepsilon_{is}'] = \Omega \cdot 1\{t = s\}$; *(ii)* $0 < \lim \frac{n}{T} \equiv \rho < \infty$; *(iii)* $\lim_{n\to\infty} \theta_0^n = 0$; *and (iv)* $\frac{1}{n}\sum_{i=1}^n |y_{i0}|^2 = O(1)$ *and* $\frac{1}{n}\sum_{i=1}^n |\alpha_i|^2 = O(1)$.

The innovations ε_{it} are uncorrelated but not independent over t. Their higher order dependence allows for conditional heteroskedasticity. In order to be able to establish central limit theorems for our estimators and to justify covariance matrix estimation we need to impose additional restrictions on the distribution of the innovations. The dependence is limited by a fourth order cumulant summability restriction slightly stronger than in Andrews (1991). These conditions could be related to more primitive mixing conditions on the underlying ε_{it} as shown in Andrews(1991). We define $u_{it}^* \equiv \sum_{j=0}^{\infty} \theta_0^j \varepsilon_{it-j}$.

Condition 2

$$\sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\mathrm{cum}_{j_1,\ldots,j_4}\left(u_{it_1}^*, \varepsilon_{it_2}, u_{it_3}^*, \varepsilon_{i0}\right)\right| < \infty \ \forall i \text{ and } j_1, \ldots, j_k \in \{1, \ldots, m\}.$$

In the same way as Andrews (1991), we define $z_{it} \equiv \left(I_m \otimes u_{it-1}^*\right)\varepsilon_{it}$ where I_m is the m-dimensional identity matrix and impose an additional eighth order moment restriction on ε_{it}, which takes the form of a fourth order cumulant summability condition on z_{it}.

Condition 3

$$\sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\mathrm{cum}_{j_1,\ldots,j_4}\left(z_{it_1}, z_{it_2}, z_{it_3}, z_{i0}\right)\right| < \infty \ \forall i \text{ and } j_1, \ldots, j_k \in \{1, \ldots, m\}.$$

Remark 1 *In Condition 1, our requirement that* $0 < \lim \frac{n}{T} \equiv \rho < \infty$ *corresponds to the choice of a particular set of asymptotic sequences. The choice of these sequences is guided by the desire to obtain asymptotic approximations that mimic certain moments of the finite sample distribution, in our case the mean, of the estimator. Bekker (1994, p.661) argues that the choice of a particular sequence can be justified by its ability to "generate acceptable approximations of known distributional properties of related statistics".*

In our case it seems most appropriate to investigate the properties of the score related to the dynamic panel model. After concentrating out the fixed effects, we are led to consider the normalized score process $S_{nT} = \frac{1}{\sqrt{nT}}\sum_{i=1}^n \sum_{t=1}^T \left(I_m \otimes \left(y_{it-1} - \overline{y}_{i-}\right)\right)\left(\varepsilon_{it} - \overline{\varepsilon}_i\right)$. *In the appendix, we show that*[4] $S_{nT} \xrightarrow{d} S$ *under the alternative asymptotics with* $n/T \to \rho$, *where* S *has a normal distribution with mean equal to* $-\sqrt{\rho}\left(I_m \otimes I_m - (I_m \otimes \theta_0)\right)^{-1} \mathrm{vec}(\Omega)$. *Clearly, under fixed* T *large* n *asymptotics the score process has an explosive mean leading to the inconsistency result. The exact finite sample bias for the score is given by* $E[S_{nT}] = -\sqrt{n/T}T^{-1}\sum_{t=1}^T \sum_{j=0}^T \left(I_m \otimes \theta_0^j\right)\mathrm{vec}(\Omega)$. *The term* $-\sqrt{\rho}T^{-1}\sum_{t=1}^T \sum_{j=0}^T \left(I_m \otimes \theta_0^j\right)\mathrm{vec}(\Omega)$ *converges to* $-\sqrt{\rho}\left(I_m \otimes I_m - (I_m \otimes \theta_0)\right)^{-1}\mathrm{vec}(\Omega) = E[S]$ *by the Toeplitz lemma*[5] *as* $T \to \infty$, *and is closer to* $E[S]$ *for small values of* θ_0. *In other words our asymptotic sequence preserves the mean of the score process in the limit. The form of the approximation error also may explain simulation findings indicating that the approximation improves for larger values of* T *and deteriorates with* θ *getting closer to the unit circle.*

[4] Lemma 6 in Appendix A.

[5] See Hall and Heyde, 1980, p. 31.

Our asymptotics may also be understood as an attempt to capture the bias of the score of order $O\left(T^{-1}\right)$. We show in the appendix that the score process is well approximated by a process, say S_{nT}^{*},[6] *such that $E\left[S_{nT}^{*}\right]=-\sqrt{\frac{n}{T}}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\left(I_m\otimes\theta_0^j\right)\operatorname{vec}(\Omega)$.*[7] *Because the term $\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\left(I_m\otimes\theta_0^j\right)\operatorname{vec}(\Omega)$ is of order $O(1)$, the approximate mean of the normalized score process can be elicited only by considering the alternative approximation where n and T grow to infinity at the same rate. The mean of the score process that our asymptotics captures may also be identified as the bias of the score up to $O\left(T^{-1}\right)$. Because the score $\frac{1}{\sqrt{nT}}S_{nT}$ is approximated by $\frac{1}{\sqrt{nT}}S_{nT}^{*}$, and because*

$$E\left[\frac{1}{\sqrt{nT}}S_{nT}^{*}\right]=-\frac{1}{T}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\left(I_m\otimes\theta_0^j\right)\operatorname{vec}(\Omega)=\frac{1}{T}\left(-\left(I_m\otimes I_m-\left(I_m\otimes\theta_0\right)\right)^{-1}\operatorname{vec}(\Omega)+o(1)\right),$$

we may understand $-\frac{1}{T}\left(I_m\otimes I_m-\left(I_m\otimes\theta_0\right)\right)^{-1}\operatorname{vec}(\Omega)$ as the bias of the score of order $O\left(T^{-1}\right)$.

Remark 2 *Condition 2 implies $\sum_{j=-\infty}^{\infty}\left|\operatorname{Cov}_{k_1,k_2}\left(z_{it},z_{it-j}\right)\right|<\infty$, because*

$$\begin{aligned}\operatorname{Cov}_{k_1,k_2}(z_{it},z_{it-j})&=\operatorname{cum}_{l_1,..,l_4}\left(u_{it-1}^{*},\varepsilon_{it},u_{it-1-j}^{*},\varepsilon_{it-j}\right)+\operatorname{Cov}_{l_1,l_3}\left(u_{it-1}^{*},u_{it-1-j}^{*}\right)\operatorname{Cov}_{l_2,l_4}\left(\varepsilon_{it},\varepsilon_{it-j}\right)\\&\quad+\operatorname{Cov}_{l_1,l_4}\left(u_{it-1}^{*},\varepsilon_{it-j}\right)\operatorname{Cov}_{l_3,l_2}\left(u_{it-1-j}^{*},\varepsilon_{it}\right),\end{aligned}$$

where $k_1=l_1m+l_2+1$ and $k_2=l_3m+l_4+1$ with $l_1,...,l_4\in\{0,1,...,m\}$. In this sense our Condition 2 is stronger than the first part of Assumption A in Andrews (1991). Condition 3 is identical to the second part of Assumption A in Andrews (1991).

Remark 3 *In the special case where ε_{it} is iid across i and t Conditions 2 and 3 are equivalent to $E\left[\left|\varepsilon_{it}^{(j)}\right|^8\right]<\infty$ for all j where $\varepsilon_{it}^{(j)}$ is the j-th element in ε_{it}. See Lemma 1 in Appendix A.*

We show below that the OLS estimator $\widehat{\theta}$ is consistent, but $\sqrt{nT}\operatorname{vec}\left(\widehat{\theta}'-\theta_0'\right)$ is not centered at zero:

Theorem 1 *Let y_{it} be generated by (1). Under Conditions 1, 2 and 3, we have*

$$\begin{aligned}&\sqrt{nT}\operatorname{vec}\left(\widehat{\theta}'-\theta_0'\right)\\&\quad\to\mathcal{N}\left(-\sqrt{\rho}(I_m\otimes\Upsilon)^{-1}\left(I_m\otimes I_m-\left(I_m\otimes\theta_0\right)\right)^{-1}\operatorname{vec}(\Omega),\left(I_m\otimes\Upsilon\right)^{-1}\left(\Omega\otimes\Upsilon+\mathcal{K}\right)\left(I_m\otimes\Upsilon\right)^{-1}\right),\end{aligned}$$

where $\Upsilon\equiv\Omega+\theta_0\Omega\theta_0'+\theta_0^2\Omega\left(\theta_0'\right)^2+\cdots$, $\mathcal{K}\equiv\sum_{t=-\infty}^{\infty}\mathcal{K}(t,0)$, $\mathcal{K}(t_1,t_2)\equiv E\left[\left(I_m\otimes u_{it_1-1}^{*}\right)\varepsilon_{it_1}\varepsilon_{it_2}'\left(I_m\otimes u_{it_2-1}^{*\prime}\right)\right]-E\left[\varepsilon_{it_1}\varepsilon_{it_2}'\right]\otimes E\left[u_{i0}^{*}u_{i0}^{*\prime}\right]$, and $u_{it}^{*}\equiv\sum_{j=0}^{\infty}\theta_0^j\varepsilon_{it-j}$. If in addition all the innovations ε_{it} are independent for all i and t then

$$\sqrt{nT}\operatorname{vec}\left(\widehat{\theta}'-\theta_0'\right)\to\mathcal{N}\left(-\sqrt{\rho}(I_m\otimes\Upsilon)^{-1}\left(I_m\otimes I_m-\left(I_m\otimes\theta_0\right)\right)^{-1}\operatorname{vec}(\Omega),\Omega\otimes\Upsilon^{-1}\right).$$

Proof. See Appendix A. ■

Under our alternative asymptotic sequence OLS is therefore consistent but has a limiting distribution that is not centered at zero. The non-centrality parameter results from correlation between the averaged error terms and the regressors y_{it-1}. Because averaging takes place for each individual the estimated

[6] The exact definition of S_{nT}^{*} is given in (11) in Appendix A.

[7] See Lemma 3.

sample means do not converge to constants fast enough to eliminate their effect on the limiting distribution. Under our asymptotics the convergence is however fast enough to eliminate the inconsistency problem found for fixed T large n asymptotic approximations.

When the innovations are not i.i.d. then the limiting distribution is affected by higher order moments reflecting the conditional heteroskedasticity in the data. The limiting covariance matrix $\Omega \otimes \Upsilon + \mathcal{K}$ can also be expressed as $\lim T^{-1} \sum_{t_1 t_2 = -T}^{T} E\left[\left(I_m \otimes u_{it_1-1}^{*}\right) \varepsilon_{it_1} \varepsilon_{it_2}' \left(I_m \otimes u_{it_2-1}^{*\prime}\right)\right]$. Standard tools for consistent and optimal estimation of $\Omega \otimes \Upsilon + \mathcal{K}$ were discussed in Andrews (1991). Under our conditions the results of Andrews are directly applicable.

Our theorem 1 predicts that $\operatorname{vec}\left(\widehat{\theta}' - \theta_0'\right)$ is approximately distributed as

$$\mathcal{N}\left(-\frac{1}{T}(I_m \otimes \Upsilon)^{-1}\left(I_m \otimes I_m - (I_m \otimes \theta_0)\right)^{-1} \operatorname{vec}(\Omega), \frac{1}{nT}(I_m \otimes \Upsilon)^{-1}(\Omega \otimes \Upsilon + \mathcal{K})(I_m \otimes \Upsilon)^{-1}\right).$$

Therefore, the noncentrality-parameter $-\sqrt{\rho}(I_m \otimes \Upsilon)^{-1}\left(I_m \otimes I_m - (I_m \otimes \theta_0)\right)^{-1} \operatorname{vec}(\Omega)$ can be viewed as a device to capture bias of order up to $O\left(T^{-1}\right)$.

Our bias-corrected OLS estimator is given by

$$\begin{aligned} \operatorname{vec}\left(\widehat{\widehat{\theta}}'\right) \equiv & \left[I_m \otimes \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \overline{y}_{i-}\right)\left(y_{it-1} - \overline{y}_{i-}\right)'\right)^{-1}\right] \\ & \cdot \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(I_m \otimes \left(y_{it-1} - \overline{y}_{i-}\right)\right)\left(y_{it} - \overline{y}_{i}\right)' + \frac{1}{T}\left(I_m \otimes I_m - \left(I_m \otimes \widehat{\theta}\right)\right)^{-1} \operatorname{vec}\left(\widehat{\Omega}\right)\right] \end{aligned} \tag{3}$$

where

$$\widehat{\Upsilon} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \overline{y}_{i-}\right)\left(y_{it-1} - \overline{y}_{i-}\right)', \quad \text{and} \quad \operatorname{vec}\left(\widehat{\Omega}\right) \equiv \left(I_m \otimes I_m - \left(\widehat{\theta} \otimes \widehat{\theta}\right)\right) \operatorname{vec}\left(\widehat{\Upsilon}\right). \tag{4}$$

We show below that the bias-corrected OLS estimator $\widehat{\widehat{\theta}}$ is consistent, and $\sqrt{nT}\operatorname{vec}\left(\widehat{\widehat{\theta}}' - \theta_0'\right)$ is centered at zero:

Theorem 2 *Let y_{it} be generated by (1). Then, under Conditions 1, 2 and 3, we have*

$$\sqrt{nT}\operatorname{vec}\left(\widehat{\widehat{\theta}}' - \theta_0'\right) \to \mathcal{N}\left(0, (I_m \otimes \Upsilon)^{-1}(\Omega \otimes \Upsilon + \mathcal{K})(I_m \otimes \Upsilon)^{-1}\right).$$

If in addition all the innovations ε_{it} are independent for all i and t then

$$\sqrt{nT}\operatorname{vec}\left(\widehat{\theta}' - \theta_0'\right) \to \mathcal{N}\left(0, \Omega \otimes \Upsilon^{-1}\right).$$

Proof. See Appendix B. ■

For the case of Gaussian innovations where OLS is equivalent to the MLE, we now show that the bias-corrected MLE is asymptotically efficient. We do so by showing that the asymptotic distribution of the bias-corrected MLE is 'minimal' in the sense of a Hajék type convolution theorem.[8] We show that the asymptotic distribution of any reasonable estimator can be written as a convolution of the 'minimal' normal distribution and some other arbitrary distribution. In this sense, the bias-corrected MLE can be understood to be asymptotically efficient.

[8] See Appendix C for the exact sense under which the asymptotic distribution of the bias corrected MLE is 'minimal'.

Condition 4 *(i)* $\varepsilon_{it} \sim \mathcal{N}(0,\Omega)$ i.i.d.*; (ii)* $0 < \lim \frac{n}{T} \equiv \rho < \infty$*; (iii)* $\lim_{n\to\infty} \theta_0^n = 0$*; and (iv)* $\frac{1}{n}\sum_{i=1}^{n} |y_{i0}|^2 = O(1)$ *and* $\frac{1}{n}\sum_{i=1}^{n} |\alpha_i|^2 = O(1)$.

In order to discuss efficiency we naturally have to guarantee that $\hat{\theta}$ is the exact MLE. For this reason we impose the additional requirement of normal innovations in condition (4).

Theorem 3 *Let y_{it} be generated by (1). Suppose that Condition (4) is satisfied. Then, the asymptotic distribution of any regular estimator of* $\mathrm{vec}(\theta_0)$ *cannot be more concentrated than* $\mathcal{N}\left(0,\Omega\otimes\Upsilon^{-1}\right)$.

Proof. See Appendix C. ■

It should be emphasized that Theorem 3 in itself does not say anything about the attainability of the bound $\Omega\otimes\Upsilon^{-1}$. The asymptotic variance bound it provides is a lower bound of the asymptotic variances of regular estimators. On the other hand, it is not clear whether such a bound is attainable. Comparison with Theorem 2 leads us to conclude that the bound is attained by the bias-corrected OLS estimator as long as the innovations ε_{it} are i.i.d. Gaussian.

Corollary 1 *Under Condition 4, the bias-corrected MLE* $\widehat{\widehat{\theta}}$ *is asymptotically efficient.*

3 Application to Univariate Dynamic Panel Models with Fixed Effects

In this section, we apply Theorems 1 and 2 in the previous section to the univariate stationary panel AR(1) model with fixed effects

$$y_{it} = \alpha_i + \theta_0 y_{it-1} + \varepsilon_{it}, \quad i = 1,\ldots,n;\ t = 1,\ldots,T. \tag{5}$$

We also consider estimation of fixed effects α_i in the univariate contexts. Finally, we examine how the result changes under the unit root. It turns out that the distribution of the MLE is quite sensitive to such a specification change. As such, we expect that our bias-corrected estimator will not be (approximately) unbiased under a unit root.

We first apply Theorems 1 and 2 to the univariate case. Obviously, Condition 4 would now read (i) $\varepsilon_{it} \sim \mathcal{N}(0,\Omega)$ *i.i.d.*; (ii) $0 < \lim \frac{n}{T} \equiv \rho < \infty$; (iii) $|\theta_0| < 1$; and (iv) $\frac{1}{n}\sum_{i=1}^{n} y_{i0}^2 = O(1)$ and $\frac{1}{n}\sum_{i=1}^{n} \alpha_i^2 = O(1)$. Note that the MLE (OLS) is given by

$$\widehat{\theta} = \frac{\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it}-\overline{y}_i\right)\cdot\left(y_{it-1}-\overline{y}_{i-}\right)}{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\overline{y}_{i-}\right)^2}.$$

Applying (3) and (4) to the univariate model, we obtain

$$\widehat{\widehat{\theta}} \equiv \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\overline{y}_{i-}\right)^2\right)^{-1}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\overline{y}_{i-}\right)\left(y_{it}-\overline{y}_i\right)+\frac{1}{T}\left(1-\widehat{\theta}\right)^{-1}\widehat{\Omega}\right]$$

where

$$\widehat{\Omega} = \left(1-\widehat{\theta}^2\right)\widehat{\Upsilon} = \left(1-\widehat{\theta}^2\right)\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\overline{y}_{i-}\right)^2\right).$$

Therefore, our bias-corrected estimator is given by

$$\widehat{\widehat{\theta}} = \widehat{\theta} + \frac{1}{T}\frac{1}{1-\widehat{\theta}}\left(1-\widehat{\theta}^2\right) = \widehat{\theta} + \frac{1}{T}\left(1+\widehat{\theta}\right) = \frac{T+1}{T}\widehat{\theta} + \frac{1}{T}. \tag{6}$$

Because $\Upsilon = \frac{\Omega}{1-\theta_0^2}$ in the univariate case, we can conclude from Theorem 2 that

$$\sqrt{nT}\left(\widehat{\widehat{\theta}} - \theta_0\right) \to \mathcal{N}\left(0, 1-\theta_0^2\right)$$

From Theorem 3, we can also conclude that $\widehat{\widehat{\theta}}$ is efficient under the alternative asymptotics where $n, T \to \infty$ at the same rate.

We now consider estimation of fixed effects α_i. Recently, Geweke and Keane (1996), Chamberlain and Hirano (1997), and Hirano (1998) examined predictive aspects of the dynamic panel model from a Bayesian perspective. From a Frequentist perspective, prediction requires estimation of individual specific intercept terms. We argue that intercept estimation is asymptotically unbiased to begin with, and is affected very little by bias-corrected estimation of θ_0. It follows that estimation of θ_0 can be separately analyzed for the purpose of prediction. Observe that the MLE of α_i is given by

$$a_i \equiv \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \widehat{\theta} y_{it-1}\right) = \alpha_i + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it} - \left(\widehat{\theta} - \theta_0\right)\frac{1}{T}\sum_{t=1}^{T} y_{it-1}. \tag{7}$$

so that

$$\begin{aligned}\sqrt{T}\left(a_i - \alpha_i\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} - \sqrt{T}\left(\widehat{\theta} - \theta_0\right)\cdot\frac{1}{T}\sum_{t=1}^{T} y_{it-1} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} - O_p\left(\frac{1}{\sqrt{n}}\right)\cdot O_p(1) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} + o_p(1).\end{aligned}$$

Because $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}$ converges in distribution to $N\left(0, \sigma_0^2\right)$ as $T \to \infty$, the MLE is asymptotically unbiased. Furthermore, we have

$$\sqrt{T}\left(\widehat{a}_i - \alpha_i\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} + o_p(1),$$

where $\widehat{a}_i$ denotes the estimator of α_i obtained by replacing the MLE $\widehat{\theta}$ in (7) by the bias-corrected estimator $\widehat{\widehat{\theta}}$. It follows that more efficient estimation of θ_0 does not affect the estimation of α_i.

We now consider the nonstationary case where $\theta_0 = 1$. We first consider a simple dynamic panel model with a unit root, where individual specific intercepts are all equal to zero but the econometrician does not know that. The econometrician therefore estimates fixed effects along with θ.

Theorem 4 *Suppose that (i)* $\varepsilon_{it} \sim \mathcal{N}\left(0, \sigma^2\right)$ i.i.d*; (ii)* $\alpha_i \equiv 0$*; (iii)* $\theta_0 = 1$*; and (iv)* $n, T \to \infty$.[9] *We then have*

$$\sqrt{nT^2}\left(\widehat{\theta} - \theta_0 + \frac{3}{T+1}\right) \to \mathcal{N}\left(0, \frac{51}{5}\right).$$

[9] No particular rate on the growth of n and T is imposed.

Proof. See Appendix D.1. ■

One obvious implication of Theorem 4 is that the bias correction for the stationary case is not expected to work under the unit root.

We now consider the case where individual specific intercepts are nonzero, and the econometrician estimates them along with θ.

Theorem 5 *Suppose that (i)* $\varepsilon_{it} \sim \mathcal{N}\left(0, \sigma^2\right)$ i.i.d*; (ii)* $\lim \frac{1}{n} \sum_{i=1}^{n} \alpha_i^2 > 0$*; (iii)* $\theta_0 = 1$*; and (iv)* $\lim \sqrt{\frac{n}{T}}$ *exists. We then have*

$$n^{1/2} T^{3/2} \left(\widehat{\theta} - \theta_0\right) \to \mathcal{N}\left(-\frac{6\sigma^2 \lim \sqrt{\frac{n}{T}}}{\lim \frac{1}{n} \sum_{i=1}^{n} \alpha_i^2}, \frac{12\sigma^2}{\lim \frac{1}{n} \sum_{i=1}^{n} \alpha_i^2}\right).$$

Proof. See Appendix D.2. ■

Although Theorem 5 shares the same feature as Theorem 1 as far as the asymptotic bias being proportional to $\lim \sqrt{\frac{n}{T}}$, it is quite clear that the bias correction for the stationary case does not work because the asymptotic bias under the unit root depends on $\lim \frac{1}{n} \sum_{i=1}^{n} \alpha_i^2$.

4 Monte Carlo

We conduct a small Monte Carlo experiment to evaluate the accuracy of our asymptotic approximations to the small sample distribution of the MLE and bias-corrected MLE. We generate samples from the model

$$y_{it} = \alpha_i + \theta_0 y_{it-1} + \varepsilon_{it}$$

where $y_{it} \in \mathbb{R}$, $\theta_0 \in \{0, .3, .6, .9\}$, $\alpha_i \sim \mathcal{N}(0, 1)$ independent across i, and $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ independent across i and t. We generate α_i and ε_{it} such that they are independent of each other. We chose $y_{i0} | \alpha_i \sim \mathcal{N}\left(\frac{\alpha_i}{1-\theta_0}, \frac{\mathrm{Var}(\varepsilon_{it})}{1-\theta_0^2}\right)$. The effective sample sizes we consider are $n = \{100, 200\}$ and $T \in \{5, 10, 20\}$.[10] For each sample of size n and T we compute the bias-corrected MLE $\widehat{\widehat{\theta}}$ based on formulation (6). We also compute the usual GMM estimator $\widehat{\theta}_{GMM}$ based on first differences

$$y_{it} - y_{it-1} = \theta_0 \left(y_{it-1} - y_{it-2}\right) + \varepsilon_{it}$$

using past levels $(y_{i0}, \ldots, y_{it-2})$ as instruments. In order to avoid the complexity of weight matrix estimation, we consider Arellano and Bover's (1995) modification.[11]

Finite sample properties of both estimators obtained by 5000 Monte Carlo runs are summarized in Table 1. We can see that both estimators have some bias problems. Unfortunately, our bias-corrected estimator does not completely remove the bias. This suggests that an even more careful small sample analysis based on higher order expansions of the distribution might be needed to account for the entire bias. On the other hand, the efficiency of $\widehat{\widehat{\theta}}$ measured by the root mean squared error (RMSE) often dominates that of the GMM estimator, suggesting that our crude higher order asymptotics and the related convolution theorem provide a reasonable prediction about the efficiency of the bias-corrected MLE.

[10] More extensive Monte Carlo results are available from authors upon request.

[11] Arellano and Bover (1995) proposed to use the moment restrictions obtained by Helmert's transformation of the same set of information. Strictly speaking, therefore, their estimator is not based on first differences with past level instruments.

5 Summary

In this paper, we considered a dynamic panel model with fixed effects where n and T are of the same order of magnitude. We developed a method to remove the asymptotic bias of OLS, and, under the assumption of Gaussian innovations, showed that the bias-corrected MLE is asymptotically efficient in the sense that its asymptotic variance equals that of the limit of the Cramer-Rao lower bound. Our simulation results compare our efficient bias-corrected MLE to more conventional GMM estimators. It turns out that our estimator has comparable bias properties and often dominates the GMM estimator in terms of mean squared error loss for the sample sizes that we think our procedure is most appropriate for.

Our theoretical result may be related to Kiviet's (1995) result. He derived an expression of approximate bias of the MLE, which is slightly different from ours. His expression for the approximate bias is a nonlinear function of the unknown parameter values including θ_0. He showed by simulation that the *infeasible* bias-corrected MLE, based on knowledge of θ_0, has much more desirable finite sample properties than various instrumental variable type estimators. Because his bias correction depends on the unknown parameter value θ_0, feasible implementation appears to require a preliminary estimator of θ_0. He considered instrumental variable type estimators as preliminary estimators in his simulation study, but he failed to produce any asymptotic theory for the corresponding estimator. Our bias-corrected estimator, which does not require a preliminary estimator of θ_0, may be understood as an implementable version of Kiviet's estimator.

It should be emphasized that some of the results in Sections 3 are independently found by Alvarez and Arellano (1998). They derived basically the same result (and more) for the MLE and other IV estimators under the assumption that (i) the initial observation has a stationary distribution, and (ii) the fixed effects are normally distributed with zero mean. Although our result is derived under slightly more general assumptions in that we do not impose such conditions, this difference should be regarded as mere technicality. The more fundamental difference is that they were concerned with the comparison of various estimators for dynamic panel data models whereas we are concerned with bias correction and efficiency. Phillips and Moon (1999) recently considered a panel model where both T and n are large. They considered asymptotic properties of OLS estimators for a panel cointegrating relation when both T and n go to infinity. This paper differs from theirs with respect to the assumption that $0 < \lim n/T < \infty$ whereas they assume $\lim n/T = 0$ as $n, T \to \infty$. It was shown in Section 3 that the asymptotic bias of the MLE (OLS) is proportional to $\sqrt{n/T}$. Phillips and Moon (1999) showed that the OLS estimator is consistent and asymptotically normal with zero mean. Although their setup is different from ours in the sense that their regressor is assumed to be nonstationary, it is plausible that their asymptotic unbiasedness of OLS critically hinges on the assumption that $\lim n/T = 0$.

One advantage of using bias-corrected MLE as the guiding principle for constructing estimators is that it can more naturally be extended to nonlinear models. Except for a few well-known examples for which ad-hoc solutions are available, we do not have any general method to deal with incidental parameter problems for general panel models with fixed effects. We expect MLE to be consistent and asymptotically normal even for nonlinear models under the alternative asymptotics where both n and T grow to infinity at the same rate. We also expect that the limiting distribution is not centered at zero. By estimating the non-centrality parameter and subtracting from the MLE, we can develop a bias corrected estimator, which is expected to dominate the MLE. GMM estimators on the other hand ultimately rely on transformations such as first differencing or similar averaging techniques to remove the individual fixed effects. Such transformations are inherently linear in nature and therefore not suited for generalizations to a nonlinear context.

Our technique can in principle be generalized to remove bias of higher order than T^{-1} by repeating the alternative asymptotic approximation scheme for an appropriately rescaled version of the bias-corrected estimator. We are planning to pursue this avenue in future research.

Our bias-corrected MLE is not expected to be asymptotically unbiased under a unit root. We leave development of a bias-corrected estimator robust to nonstationarity to future research.

Appendix

A Proof of Theorem 1

Theorem 1 is established by combining Lemmas 6 and 7 below. Note that

$$y_{it} = \theta_0^t y_{i0} + (I_m - \theta_0)^{-1} \left(1 - \theta_0^t\right) \alpha_i + \theta_0^{t-1} \varepsilon_{i1} + \theta_0^{t-2} \varepsilon_{i2} + \cdots + \varepsilon_{it}. \tag{8}$$

In the stationary case where $\lim_n \theta_0^n = 0$, we work with the stationary approximation to y_{it} which is given by

$$u_{it}^* \equiv \sum_{j=0}^{\infty} \theta_0^j \varepsilon_{it-j}, \quad t \geq 1 \tag{9}$$

$$y_{it}^* \equiv (I_m - \theta_0)^{-1} \alpha_i + u_{it}^*, \quad t \geq 0. \tag{10}$$

The vectorized representation of the OLS estimator for θ_0' is given by

$$\operatorname{vec}\left(\widehat{\theta}' - \theta_0'\right) = \left[I_m \otimes \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \overline{y}_{i-}\right) \left(y_{it-1} - \overline{y}_{i-}\right)' \right)^{-1} \right] \left[\sum_{i=1}^{n} \sum_{t=1}^{T} \left(I_m \otimes \left(y_{it-1} - \overline{y}_{i-}\right)\right) \left(\varepsilon_{it} - \overline{\varepsilon}_i\right) \right]$$

where $\bar{\varepsilon}_i \equiv \frac{1}{T} \sum_t^T \varepsilon_{it}$.

Lemma 1 *Let ε_{it} be iid across i and t and $E\left[\left|\varepsilon_{it}^{(j)}\right|^8\right] < \infty$. Assume Conditions 4 (ii) - (iv) hold. Then Conditions 2 and 3 hold.*

Proof. See Hahn and Kuersteiner (2001). ■

Lemma 2 *Let y_{it} be generated by (1). Also, let $S_{nT}^* \equiv \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \left(I_m \otimes u_{it}^*\right) \left(\varepsilon_{it} - \overline{\varepsilon}_i\right)$. Then, under Conditions 1, 2 and 3*

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(I_m \otimes \left(y_{it-1} - \overline{y}_{i-}\right)\right) \left(\varepsilon_{it} - \overline{\varepsilon}_i\right) = S_{nT}^* + o_p(1) \tag{11}$$

Proof. See Hahn and Kuersteiner (2001). ■

Lemma 3 *Let y_{it} be generated by (1). Under Conditions 1, 2 and 3,*

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(I_m \otimes u_{it}^*\right) \overline{\varepsilon}_i = \sqrt{\rho} \left(I_m \otimes I_m - \left(I_m \otimes \theta_0\right)\right)^{-1} \operatorname{vec}(\Omega) + o_p(1).$$

Proof. See Hahn and Kuersteiner (2001). ■

Lemma 4 *Assume ε_t is a sequence of independent, identically distributed random vectors with $E[\varepsilon_t] = 0$ for all t. Then $\operatorname{cum}_{j_1,\ldots j_k}(\varepsilon_{t_1}, \ldots, \varepsilon_{t_k}) = 0$ unless $t_1 = t_2 = \cdots = t_k$. In this case we define $\operatorname{cum}(j_1, \ldots, j_k) \equiv \operatorname{cum}_{j_1,\ldots j_k}(\varepsilon_t, \ldots, \varepsilon_t)$.*

Lemma 5 *Let Conditions 1, 2 and 3 be satisfied. Then*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(I_m\otimes u_{it-1}^*\right)\varepsilon_{it}\to\mathcal{N}\left(0,\Omega\otimes\Upsilon+\mathcal{K}\right)$$

where $\mathcal{K}=\sum_{t=-\infty}^{\infty}\mathcal{K}\left(t,0\right)$ *and* $\mathcal{K}\left(t_1,t_2\right)\equiv E\left[\left(I_m\otimes u_{it_1-1}^*\right)\varepsilon_{it_1}\varepsilon_{it_2}'\left(I_m\otimes u_{it_2-1}^{*\prime}\right)\right]-E\left[\varepsilon_{it_1}\varepsilon_{it_2}'\right]\otimes E\left[u_{i0}^*u_{i0}^{*\prime}\right]$. *If in addition all the innovations* ε_{it} *are independent for all* i *and* t *then*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(I_m\otimes u_{it-1}^*\right)\varepsilon_{it}\to\mathcal{N}\left(0,\Omega\otimes\Upsilon\right).$$

Proof. We need to check the generalized Lindeberg Feller condition for joint asymptotic normality as in Theorem 2 of Phillips and Moon (1999). A sufficient condition for the theorem to hold is that for all $\ell\in\mathbb{R}^{m^2}$ such that $\ell'\ell=1$ it follows $E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\ell'(I_m\otimes u_{it-1}^*)\varepsilon_{it}\right)^4\right]<\infty$ uniformly in i and T. Letting $\mathbf{z}_{it}=\ell'\left(I_m\otimes u_{it-1}^*\right)\varepsilon_{it}$ and noting that $E\left[\mathbf{z}_{it}\right]=0$ we show in Hahn and Kuersteiner (2001) that $\frac{1}{T^2}\sum_{t_1,\dots t_4=1}^{T}E\left[\mathbf{z}_{it_1}\mathbf{z}_{it_2}\mathbf{z}_{it_3}\mathbf{z}_{it_4}\right]<\infty$. Now consider

$$\begin{aligned}
&E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(I_m\otimes u_{it-1}^*\right)\varepsilon_{it}\right)^2\right]\\
&=\frac{1}{T}\sum_{t,s=1}^{T}E\left[\left(I_m\otimes u_{it-1}^*\right)\varepsilon_{it}\varepsilon_{is}'\left(I_m\otimes u_{is-1}^{*\prime}\right)\right]\\
&=\frac{1}{T}\sum_{t,s=1}^{T}\operatorname{vec}\left(E\left[u_{it-1}^*\varepsilon_{is}'\right]\right)\operatorname{vec}\left(E\left[u_{is-1}^*\varepsilon_{it}'\right]\right)+\frac{1}{T}\sum_{t,s=1}^{T}E\left[\varepsilon_{it}\varepsilon_{is}'\right]\otimes E\left[u_{it-1}^*u_{is-1}^{*\prime}\right]+\frac{1}{T}\sum_{t,s=1}^{T}\mathcal{K}\left(t,s\right)\\
&=\Omega\otimes\Upsilon+\mathcal{K}+o(1).
\end{aligned}$$

where $\mathcal{K}=\sum_{t_1=-\infty}^{\infty}\mathcal{K}(t_1,0)$. Note that $\operatorname{vec}\left(E\left[u_{it-1}^*\varepsilon_{is}'\right]\right)\operatorname{vec}\left(E\left[u_{is-1}^*\varepsilon_{it}'\right]\right)=0$ for all t and s and that $\frac{1}{T}\sum_{t,s=1}^{T}E\left[\varepsilon_{it}\varepsilon_{is}'\right]\otimes E\left[u_{it-1}^*u_{is-1}^{*\prime}\right]=\frac{1}{T}\sum_{t=1}^{T}E\left[\varepsilon_{it}\varepsilon_{it}'\right]\otimes E\left[u_{it-1}^*u_{it-1}^{*\prime}\right]=\Omega\otimes\Upsilon$ by strict stationarity. The last line of the display follows by Cesàro summability and stationarity. The second part of the theorem follows from Lemma 4, which implies that $\mathcal{K}\left(t_1,t_2\right)=0$ for all t_1 and t_2. ■

Lemma 6 *Let* y_{it} *be generated by (1). Under Conditions 1, 2 and 3,*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(I_m\otimes\left(y_{it-1}-\overline{y}_{i-}\right)\right)\left(\varepsilon_{it}-\overline{\varepsilon}_i\right)\to\mathcal{N}\left(-\sqrt{\rho}\left(I_m\otimes I_m-\left(I_m\otimes\theta_0\right)\right)^{-1}\operatorname{vec}\left(\Omega\right),\Omega\otimes\Upsilon+\mathcal{K}\right).$$

Proof. The result follows from Lemmas 2, 3, and 5. ■

Lemma 7 *Let* y_{it} *be generated by (1). Under Conditions 1, 2 and 3*

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\overline{y}_{i-}\right)\left(y_{it-1}-\overline{y}_{i-}\right)'=E\left[\left(y_{it-1}^*-Ey_{it-1}^*\right)\left(y_{it-1}^*-Ey_{it-1}^*\right)'\right]+o_p\left(1\right)=\Upsilon+o_p\left(1\right).$$

Proof. See Hahn and Kuersteiner (2001). ■

B Proof of Theorem 2

Because $\operatorname{vec}(\Upsilon) = (I_m - (\theta_0 \otimes \theta_0))^{-1} \operatorname{vec}(\Omega)$, and $\widehat{\theta} = \theta_0 + o_p(1)$, we have

$$\sqrt{\frac{n}{T}} \left(I_m \otimes I_m - \left(I_m \otimes \widehat{\theta} \right) \right)^{-1} \operatorname{vec}\left(\widehat{\Omega} \right) = \sqrt{\rho} \left(I_m \otimes I_m - (I_m \otimes \theta_0) \right)^{-1} \operatorname{vec}(\Omega) + o_p(1). \qquad (12)$$

Combining with Lemma 6, we obtain

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(I_m \otimes \left(y_{it-1} - \overline{y}_{i-} \right) \right) \left(\varepsilon_{it} - \overline{\varepsilon}_i \right) + \sqrt{\frac{n}{T}} \left(I_m \otimes I_m - \left(I_m \otimes \widehat{\theta} \right) \right)^{-1} \operatorname{vec}\left(\widehat{\Omega} \right) \xrightarrow{d} \mathcal{N}\left(0, \left(\Omega \otimes \Upsilon + \mathcal{K} \right) \right).$$

The conclusion follows by using Lemma 7.

C Proof of Theorem 3

For the discussion and derivation of the asymptotic variance bound, we adopt the same framework as in van der Vaart and Wellner (1996, p. 412). For this purpose, we discuss some of their notation. Let H be a linear subspace H of a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$, and let $P_{n,h}$ denote a probability measure on a measurable space $(\mathcal{X}_n, \mathcal{A}_n)$. Consider estimating a parameter $\kappa_n(h)$ based on an observation with law $P_{n,h}$. Now, let $\{\Delta_h : h \in H\}$ be the "iso-Gaussian process" with zero mean and covariance function $E[\Delta_{h_1} \Delta_{h_2}] = \langle h_1, h_2 \rangle$. We say that the sequence $(\mathcal{X}_n, \mathcal{A}_n, P_{n,h})$ is *asymptotically normal* if

$$\log \frac{dP_{n,h}}{dP_{n,0}} = \Delta_{n,h} - \frac{1}{2} \|h\|^2$$

for stochastic processes $\{\Delta_{n,h} : h \in H\}$ such that $\Delta_{n,h}$ converges weakly to Δ_h marginally under $P_{n,0}$. Now, consider the sequence of parameters $\kappa_n(h)$ belonging to a Banach space B, which is *regular* in the sense that $r_n(\kappa_n(h) - \kappa_n(0)) \to \dot{\kappa}(h)$ for every $h \in H$ for a continuous, linear map $\dot{\kappa} : H \to B$ and certain linear maps $r_n : B \to B$. A sequence of estimators τ_n is defined to be *regular* if $r_n(\tau_n - \kappa_n(h))$ converges weakly to the same measure L, say, regardless of h. The bound of the asymptotic variance of a regular estimator can be found from the following theorem based on the modification of van der Vaart and Wellner (1996, Theorem 3.11.2):

Theorem 6 *Assume that $(P_{n,h} : h \in H)$ is asymptotically normal. Also, suppose that (i) $h = (\delta, \Xi)$, $h_0 = (0,0)$; (ii) $\kappa_n(h) \equiv \xi_n \equiv \xi_0 + \frac{1}{r_n}\delta$ for some $\xi_0 \in \mathbb{R}$; and (iii) $\Delta_h \equiv \Delta_1 \cdot \delta + \Delta_2(\Xi)$. Further suppose that, with respect to the norm $\|\cdot\|$, (iv) the mapping $\dot{\kappa} : (\delta, \Xi) \to \delta$ is continuous; and (v) H is complete. Then, for every regular sequence of estimators $\{\tau_n\}$, we have*

$$r_n(\tau_n - \kappa_n(h)) \to \mathcal{N}\left(0, \mathrm{E}\left[\widetilde{\Delta}_1{}^2 \right]^{-1} \right) \oplus W$$

for some W, where $\oplus$ denotes convolution, and $\widetilde{\Delta}_1$ is the residual in the projection of Δ_1 on $\{\Delta_2(\Xi) : (\delta, \Xi) \in H\}$.[12]

[12] This theorem originally appeared in Hahn (1998), but is reproduced here for convenience.

We show that our setup is covered by the preceding theorem. Ignoring irrelevant constants, the joint likelihood of the model (1) is given by

$$\mathcal{L} = \frac{nT}{2} \log \det (\Psi) - \frac{1}{2} \sum_{i=1}^{n} \sum_{t=1}^{T} \operatorname{trace} (\Psi Z_{it} (\alpha_i, \theta)), \tag{13}$$

where $Z_{it} (\alpha_i, \theta) \equiv \left(y_{it} - \alpha_i - \theta' y_{it-1}\right) \left(y_{it} - \alpha_i - \theta' y_{it-1}\right)'$, and $\Psi = \Omega^{-1}$. We will localize the parameter. Let $\boldsymbol{\alpha}$ denote the sequence $(\alpha_1, \alpha_2, \ldots)$. We will attach subscript 0 to denote the 'truth'. We will localize the parameter around the 'truth', so that $\theta \equiv \theta(n) \equiv \theta_0 + \frac{1}{\sqrt{nT}} \widetilde{\theta}$, $\Psi \equiv \Psi(n) \equiv \Psi_0 + \frac{1}{\sqrt{nT}} \widetilde{\Psi}$, and $\boldsymbol{\alpha} \equiv \boldsymbol{\alpha}(n) \equiv \boldsymbol{\alpha}_0 + \frac{1}{\sqrt{nT}} \widetilde{\boldsymbol{\alpha}}$. Let $h \equiv \left(\widetilde{\boldsymbol{\alpha}}, \widetilde{\theta}, \widetilde{\Psi}\right)$, and let H denote a set of all possible values of $\left(\widetilde{\boldsymbol{\alpha}}, \widetilde{\theta}, \widetilde{\Psi}\right)$. Let $P_{n,h}$ denote the joint probability under parameters characterized by h. It can be shown that the sequence $(P_{n,h} : h \in H)$ is *asymptotically normal*:

Lemma 8 *Under $P_{n,0}$, we have*

$$\log \frac{dP_{n,h}}{dP_{n,0}} = \Delta_{n,h} - \frac{1}{2} \left\|\Delta_{n,h}\right\|^2 + o_p(1),$$

where

$$\begin{aligned} \Delta_{n,h} &\equiv \Delta_n \left(\widetilde{\alpha}, \widetilde{\theta}, \widetilde{\Psi}\right) \\ &= -\frac{1}{2\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \operatorname{trace} \left(\widetilde{\Psi} \left(\varepsilon_{it} \varepsilon_{it}' - \Omega\right)\right) + \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\widetilde{\boldsymbol{\alpha}}_i + \widetilde{\theta} y_{it-1}^*\right)' \Psi \varepsilon_{it} \end{aligned}$$

converges weakly (under $P_{n,0}$) to $\Delta_h \sim \mathcal{N}\left(0, \|h\|^2\right)$. Here, $\|h\|^2 = \langle h, h \rangle$, and

$$\begin{aligned} \left\langle \left(\widetilde{\alpha}_1, \widetilde{\theta}_1, \widetilde{\Psi}_1\right), \left(\widetilde{\alpha}_2, \widetilde{\theta}_2, \widetilde{\Psi}_2\right) \right\rangle &\equiv \frac{1}{2} \operatorname{vec}\left(\widetilde{\Psi}_2\right)' \operatorname{vec}(\Omega) \cdot \operatorname{vec}(\Omega)' \operatorname{vec}\left(\widetilde{\Psi}_1\right) + \operatorname{vec}\left(\widetilde{\theta}_2'\right)' (\Psi \otimes \Upsilon) \operatorname{vec}\left(\widetilde{\theta}_1'\right) \\ &+ \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \widetilde{\boldsymbol{\alpha}}_{1i}' \Psi \widetilde{\boldsymbol{\alpha}}_{2i} + \left(\operatorname{vec}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha_i \widetilde{\boldsymbol{\alpha}}_{1i}'\right)\right)' \left(\Psi \otimes \left(I_m - \theta'\right)^{-1}\right) \operatorname{vec}\left(\widetilde{\theta}_2'\right) \\ &+ \left(\operatorname{vec}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha_i \widetilde{\boldsymbol{\alpha}}_{2i}'\right)\right)' \left(\Psi \otimes \left(I_m - \theta'\right)^{-1}\right) \operatorname{vec}\left(\widetilde{\theta}_1'\right) \\ &+ \left(\operatorname{vec}\left(\widetilde{\theta}_2'\right)\right)' \left(\Psi \otimes (I_m - \theta)^{-1} \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha_i \alpha_i'\right) \left(I_m - \theta'\right)^{-1}\right) \operatorname{vec}\left(\widetilde{\theta}_1'\right). \end{aligned} \tag{14}$$

Proof. See Hahn and Kuersteiner (2001). ■

Now, note that we may write $\Delta_n \left(\widetilde{\alpha}, \widetilde{\theta}, \widetilde{\Psi}\right) = \operatorname{vec}\left(\widetilde{\theta}'\right)' \Delta_{1n} + \Delta_{2n}\left(\widetilde{\alpha}, \widetilde{\Psi}\right)$, where

$$\begin{aligned} \Delta_{1n} &\equiv \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\Psi \otimes y_{it-1}^*\right) \varepsilon_{it}, \\ \Delta_{2n}\left(\widetilde{\alpha}, \widetilde{\Psi}\right) &\equiv -\frac{1}{2\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \operatorname{trace}\left(\widetilde{\Psi}\left(\varepsilon_{it}\varepsilon_{it}' - \Omega\right)\right) + \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \widetilde{\boldsymbol{\alpha}}_i' \Psi \varepsilon_{it}. \end{aligned}$$

Theorem 6 implies that the 'minimal' asymptotic distribution is $\mathcal{N}\left(0, \left(E\left[\widetilde{\Delta}_1 \widetilde{\Delta}_1'\right]\right)^{-1}\right)$, where $\widetilde{\Delta}_1$ is the residual in the projection of Δ_1 on the linear space spanned by $\left\{\Delta_2\left(\widetilde{\alpha}, \widetilde{\Psi}\right)\right\}$. Here, Δ_1 and $\Delta_2\left(\widetilde{\alpha}, \widetilde{\Psi}\right)$

denote the 'limits' of Δ_{1n} and $\Delta_{2n}\left(\widetilde{\alpha},\widetilde{\Psi}\right)$. Lemma 9 below establishes that $\left\langle\widetilde{\Delta}_1,\widetilde{\Delta}_1'\right\rangle=\Psi\otimes\Upsilon$. Therefore, the minimum variance of estimation of $\operatorname{vec}(\theta_0)$ is given by the inverse of $\Psi\otimes\Upsilon$, or $\Omega\otimes\Upsilon^{-1}$.

Lemma 9 $\left\langle\widetilde{\Delta}_1,\widetilde{\Delta}_1'\right\rangle=\Psi\otimes\Upsilon$.

Proof. We first establish that

$$\widetilde{\Delta}_1\equiv\left[\begin{array}{c} e_1' \\ \vdots \\ e_{m^2}' \end{array}\right]\Delta_1-\left[\begin{array}{c} \Delta_2\left(D_1\left(I_m-\theta_0\right)^{-1}\alpha,0\right) \\ \vdots \\ \Delta_2\left(D_{m^2}\left(I_m-\theta_0\right)^{-1}\alpha,0\right) \end{array}\right],$$

where, e_j' denotes the jth row of I_{m^2}, and D_j is an $m\times m$ matrix such that $\operatorname{vec}\left(D_j'\right)=e_j$. We minimize the norm of $e_j'\Delta_1-\Delta_2\left(\widetilde{\alpha},\widetilde{\Psi}\right)$ for each j. From (14), we obtain

$$\begin{aligned}\left\|e_j'\Delta_1-\Delta_2\left(\widetilde{\alpha},\widetilde{\Psi}\right)\right\|^2&=e_j'\left(\Psi\otimes\left(I_m-\theta_0\right)^{-1}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\alpha_i\alpha_i'\right)\left(I_m-\theta_0'\right)^{-1}\right)e_j\\&\quad-2\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\widetilde{\boldsymbol{\alpha}}_i'\Psi D_j\left(I_m-\theta_0\right)^{-1}\alpha_i+\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\widetilde{\boldsymbol{\alpha}}_i'\Psi\widetilde{\boldsymbol{\alpha}}_i\\&\quad+e_j'\left(\Psi\otimes\Upsilon\right)e_j+\frac{1}{2}\left(\operatorname{trace}\left(\Omega\widetilde{\Psi}\right)\right)^2.\end{aligned}$$

Therefore, the minimum of $\left\|e_j'\Delta_1-\Delta_2\left(\widetilde{\alpha},\widetilde{\Psi}\right)\right\|^2$ is attained with $\widetilde{\Psi}=0$, and $\widetilde{\boldsymbol{\alpha}}_i'=D_j\left(I_m-\theta_0\right)^{-1}\alpha_i$. Observe that the (j,k)-element of $\left\langle\widetilde{\Delta}_1,\widetilde{\Delta}_1'\right\rangle$ is equal to

$$\left\langle e_j'\Delta_1-\Delta_2\left(D_j\left(I_m-\theta_0\right)^{-1}\alpha_i,0\right),e_k'\Delta_1-\Delta_2\left(D_k\left(I_m-\theta_0\right)^{-1}\alpha_i,0\right)\right\rangle.$$

After some tedious algebra, we can show that it is equal to $e_j'\left(\Psi\otimes\Upsilon\right)e_k$. In other words, $\left\langle\widetilde{\Delta}_1,\widetilde{\Delta}_1'\right\rangle=\Psi\otimes\Upsilon$. ■

D Proofs of Theorems 4 and 5

Ignore the i subscript whenever obvious. Let $H_T=I_T-\frac{1}{T}\ell_T\ell_T'$, $y=(y_1,\ldots,y_T)'$, $y_-=(y_0,\ldots,y_{T-1})'$, and $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_T)'$. We can write $\sum_{t=1}^T\left(\varepsilon_t-\overline{\varepsilon}\right)\left(y_{t-1}-\overline{y}_-\right)=\varepsilon'H_Ty_-$, and $\sum_{t=1}^T\left(y_{t-1}-\overline{y}_-\right)^2=y_-'H_Ty_-$. Here, ℓ_T denotes T-dimensional column vector consisting of ones. Note that

$$y_-=\left(\begin{array}{c}1\\1\\1\\\vdots\\1\end{array}\right)y_0+\left(\begin{array}{c}0\\1\\2\\\vdots\\T-1\end{array}\right)\alpha+\left[\begin{array}{ccccc}0&0&0&\cdots&0\\1&0&0&\cdots&0\\1&1&0&\cdots&0\\\vdots&\vdots&\vdots&\cdots&\cdots\\1&1&1&1&0\end{array}\right]\varepsilon\equiv\xi_{1T}y_0+\xi_{2T}\alpha+A_T\varepsilon.$$

Let $D_T \equiv H_T A_T$. We have $H_T \xi_{1T} = 0$, and hence, it follows that

$$
\begin{aligned}
H_T y_- &= \left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)\alpha + D_T\varepsilon, \\
\varepsilon' H_T y_- &= \varepsilon'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)\alpha + \varepsilon' D_T\varepsilon, \\
y_-' H_T y_- &= \left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)\alpha^2 + 2\alpha\varepsilon'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right) + \varepsilon' D_T' D_T\varepsilon
\end{aligned}
$$

In the special case where $\alpha = 0$, we have $\varepsilon' H_T y_- = \varepsilon' D_T \varepsilon, \qquad y_-' H_T y_- = \varepsilon' D_T' D_T \varepsilon.$

Lemma 10

$$
\begin{aligned}
&\operatorname{trace}\left(D_T' D_T\right) = \frac{1}{6}T^2 - \frac{1}{6}, \qquad \operatorname{trace}\left[\left(D_T' D_T\right)^2\right] = \frac{1}{90}T^4 + \frac{1}{36}T^2 - \frac{7}{180}, \\
&\operatorname{trace}\left(D_T' D_T^2\right) = -\frac{1}{12}T^2 + \frac{1}{12}, \\
&\operatorname{trace}\left(D_T + D_T'\right) = -T + 1, \qquad \operatorname{trace}\left(\left(D_T + D_T'\right)^2\right) = \frac{1}{6}T^2 + T - \frac{7}{6}, \\
&\operatorname{trace}\left(\left(D_T + D_T'\right)^3\right) = -\frac{1}{4}T^2 - T + \frac{5}{4}, \qquad \operatorname{trace}\left(\left(D_T + D_T'\right)^4\right) = \frac{1}{72}T^4 + \frac{11}{36}T^2 + T - \frac{95}{72}.
\end{aligned}
$$

Lemma 11 *As $n, T \to \infty$ it follows that $\frac{1}{nT^2}\sum_i \varepsilon' D_T' D_T \varepsilon = \frac{\sigma^2}{6} + o_p(1)$.*

Proof. Examining the cumulant generating function, we can see that the first two cumulants of $\varepsilon' D_T' D_T \varepsilon$ are equal to $\sigma^2 \operatorname{trace}(D_T' D_T)$, and $2\sigma^4 \operatorname{trace}\left[\left(D_T' D_T\right)^2\right]$. Using Lemma 10, we obtain the desired conclusion. ∎

Lemma 12 *As $n, T \to \infty$ it follows that $\frac{1}{\sqrt{nT}}\sum_i \varepsilon'\left(D_T + \frac{3}{T+1}D_T' D_T\right)\varepsilon \to \mathcal{N}\left(0, \frac{17\sigma^4}{60}\right)$.*

Proof. We first note that the fourth central moment of $\varepsilon' D_T \varepsilon + \frac{3}{T+1}\varepsilon' D_T' D_T \varepsilon$ can be bounded by eight times the sum of the fourth central moments of $\varepsilon' D_T \varepsilon$ and $\frac{3}{T+1}\varepsilon' D_T' D_T \varepsilon$, from which we can conclude that the fourth central moment of $\varepsilon' D_T \varepsilon + \frac{3}{T+1}\varepsilon' D_T' D_T \varepsilon$ is of order T^4. Examining the cumulant generating function, we can see that the first two cumulants of $\varepsilon'\left(D_T + \frac{3}{T+1}D_T' D_T\right)\varepsilon \equiv \varepsilon' G_T \varepsilon$ are given by $\frac{\sigma^2}{2}\operatorname{trace}\left(G_T + G_T'\right)$, and $\frac{\sigma^4}{2}\operatorname{trace}\left[\left(G_T + G_T'\right)^2\right]$. Using Lemma 10, we can show that

$$
\operatorname{trace}\left(G_T + G_T'\right) = 0, \qquad \operatorname{trace}\left[\left(G_T + G_T'\right)^2\right] = \frac{1}{30}\frac{17T^3 - 37T^2 + 37T - 17}{T+1}.
$$

Therefore, we have

$$
E\left[\varepsilon'\left(D_T + \frac{3}{T+1}D_T' D_T\right)\varepsilon\right] = 0, \qquad \operatorname{Var}\left(\varepsilon'\left(D_T + \frac{3}{T+1}D_T' D_T\right)\varepsilon\right) = \frac{\sigma^4}{60}\frac{17T^3 - 37T^2 + 37T - 17}{T+1}.
$$

Because the fourth central moment is of order T^4, and the variance is of order T^2, the Lyapounov condition is satisfied. Therefore, we have

$$
\frac{1}{\sqrt{n\frac{\sigma^4}{60}\frac{17T^3 - 37T^2 + 37T - 17}{T+1}}}\sum_i \varepsilon'\left(D_T + \frac{3}{T+1}D_T' D_T\right)\varepsilon \to \mathcal{N}(0, 1),
$$

from which the conclusion follows. ∎

Lemma 13 *Suppose that* $\lim \frac{1}{n}\sum_{i=1}^{n}\alpha_i^2 > 0$. *We then have*

$$\frac{1}{n^{1/2}T^{3/2}}\sum_i\left(\varepsilon' H_T y_- - \frac{\sigma^2}{2}(-T+1)\right) \to \mathcal{N}\left(0, \frac{\sigma^2}{12}\lim\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2\right)$$

Proof. Because $\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right) = -\frac{1}{12}T + \frac{1}{12}T^3$, we have

$$\sum_{i=1}^{n}\varepsilon_i'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)\alpha_i \sim \mathcal{N}\left(0, \sigma^2\left(-\frac{1}{12}T + \frac{1}{12}T^3\right)\sum_{i=1}^{n}\alpha_i^2\right) = O_p\left(n^{1/2}T^{3/2}\right). \tag{15}$$

It therefore suffices to prove that

$$\sum_i\left(\varepsilon' D_T\varepsilon - \frac{\sigma^2}{2}(-T+1)\right) = o_p\left(n^{1/2}T^{3/2}\right). \tag{16}$$

Examining the cumulant generating function, we can show that the first, and second cumulants of $\varepsilon' D_T\varepsilon$ are equal to $\frac{\sigma^2}{2}\operatorname{trace}\left[D_T + D_T'\right]$, and $\frac{\sigma^4}{2}\operatorname{trace}\left[(D_T + D_T')^2\right]$. Using Lemma 10 along with the well-known relation between cumulants and central moments, we obtain $E\left[\left(\varepsilon' D_T\varepsilon - \frac{\sigma^2}{2}(-T+1)\right)^2\right] = O\left(T^2\right)$, from which (16) follows. ■

Lemma 14 *Suppose that* $\lim \frac{1}{n}\sum_{i=1}^{n}\alpha_i^2 > 0$. *We then have* $\operatorname{plim}\frac{1}{nT^3}\sum_i y_-' H_T y_- = \frac{1}{12}\lim\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2$.

Proof. Note that

$$y_-' H_T y_- = \left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)\alpha^2 + 2\alpha\varepsilon'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right) + \varepsilon' D_T' D_T\varepsilon \tag{17}$$

Because $\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right) = -\frac{1}{12}T + \frac{1}{12}T^3$, we have

$$\sum_i\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)'\left(\xi_{2T} - \frac{T-1}{2}\ell_T\right)\alpha_i^2 = \left(-\frac{1}{12}T + \frac{1}{12}T^3\right)\sum_{i=1}^{n}\alpha_i^2 = O\left(nT^3\right).$$

From (15) and Lemma 11, we can see that the first term on the right hand side of (17) dominates the second and third terms. The conclusion follows. ■

D.1 Proof of Theorem 4

The conclusion follows from combining

$$\sqrt{n}T\left(\widehat{\theta} - \theta_0 + \frac{3}{T+1}\right) = \frac{\frac{1}{\sqrt{n}T}\sum_i\varepsilon'\left(D_T + \frac{3}{T+1}D_T' D_T\right)\varepsilon}{\frac{1}{nT^2}\sum_i\varepsilon' D_T' D_T\varepsilon}$$

with Lemmas 11 and 12.

D.2 Proof of Theorem 5

The conclusion follows from combining

$$n^{1/2}T^{3/2}\left(\widehat{\theta} - \theta_0\right) = \frac{\frac{1}{n^{1/2}T^{3/2}}\sum_i\left(\varepsilon' H_T y_- - \frac{\sigma^2}{2}(-T+1)\right) + \frac{1}{n^{1/2}T^{3/2}}\frac{\sigma^2}{2}n(-T+1)}{\frac{1}{nT^3}\sum_i y_-' H_T y_-}$$

with Lemmas 13 and 14.

References

[1] Ahn, S., and P. Schmidt (1995), "Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics* 68, pp. 5 - 28.

[2] Alvarez, J., and M. Arellano (1998), "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators", *unpublished manuscript.*

[3] Andrews, D. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica* 59, pp. 817-858.

[4] Arellano, M., and S. Bond (1992), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies* 58, pp. 277 - 297.

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

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

[7] Billingsley, P. (1995), *Probability and Measure*, Wiley, New York.

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

[9] Brillinger, D. (1981), *Time Series: Data Analysis and Theory*, Holden-Day, San Francisco.

[10] Chamberlain, G. (1998), "Econometrics and Decision Theory", *unpublished manuscript.*

[11] Chamberlain, G., and K. Hirano (1997), "Predictive Distributions Based on Longitudinal Earnings Data", unpublished manuscript.

[12] Geweke, J., and M. Keane (1996), "An Empirical Analysis of the Male Income Dynamics in the PSID", unpublished manuscript.

[13] Hahn, J. (1997), "Efficient Estimation of Panel Data Models with Sequential Moment Restrictions", *Journal of Econometrics* 79, pp. 1 - 21.

[14] Hahn, J. (1998), "Optimal Inference with Many Instruments", *unpublished manuscript.*

[15] Hahn, J. and G. Kuersteiner (2001), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large", *MIT workingpaper No. 01-17.*

[16] Hall, P. and C.C. Heyde (1980), *Martingale Limit Theory and Its Application*, Academic Press, San Diego.

[17] Hirano, K. (1998), "Semiparametric Bayesian Models for Dynamic Earnings Data", unpublished manuscript.

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

[19] Islam, N. (1995), "Growth Empirics: A Panel Data Approach", *Quarterly Journal of Economics* 110, pp. 1127 - 1170.

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

[21] Lee, K., H. Pesaran, and R. Smith (1998), "Growth Empirics: A Panel Data Approach - A Comment", *Quarterly Journal of Economics* 113, pp. 319 - 323.

[22] Magnus, J., and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*,Wiley, New York.

[23] Neyman, J., and E. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations", *Econometrica* 16, pp. 1 -31.

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

[25] Phillips, P., and H. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica* 67, pp. 1057-1111.

[26] van der Vaart, A., and J. Wellner (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York.

Table 1: Performance of Bias Corrected Maximum Likelihood Estimator

Case	T	N	θ_0	Bias $\widehat{\theta}_{GMM}$	Bias $\widehat{\widehat{\theta}}$	RMSE $\widehat{\theta}_{GMM}$	RMSE $\widehat{\widehat{\theta}}$	Test of Normality $\widehat{\theta}_{GMM}$	Test of Normality $\widehat{\widehat{\theta}}$
(1)	5	100	0	-0.011	-0.039	0.074	0.065	25.030	6.878
(2)	5	100	0.3	-0.027	-0.069	0.099	0.089	18.978	0.887
(3)	5	100	0.6	-0.074	-0.115	0.160	0.129	33.449	0.717
(4)	5	100	0.9	-0.452	-0.178	0.552	0.187	180.519	1.858
(5)	5	200	0	-0.006	-0.041	0.053	0.055	6.970	0.978
(6)	5	200	0.3	-0.014	-0.071	0.070	0.081	2.813	0.400
(7)	5	200	0.6	-0.038	-0.116	0.111	0.124	0.720	1.463
(8)	5	200	0.9	-0.337	-0.178	0.443	0.183	210.255	2.818
(9)	10	100	0	-0.011	-0.010	0.044	0.036	7.296	1.344
(10)	10	100	0.3	-0.021	-0.019	0.053	0.040	0.210	0.917
(11)	10	100	0.6	-0.045	-0.038	0.075	0.051	6.313	8.251
(12)	10	100	0.9	-0.218	-0.079	0.248	0.085	248.829	19.933
(13)	10	200	0	-0.006	-0.011	0.031	0.027	7.119	4.677
(14)	10	200	0.3	-0.011	-0.019	0.038	0.032	0.964	1.239
(15)	10	200	0.6	-0.025	-0.037	0.051	0.045	2.086	0.014
(16)	10	200	0.9	-0.152	-0.079	0.181	0.082	159.734	4.548
(17)	20	100	0	-0.011	-0.003	0.029	0.024	2.901	0.823
(18)	20	100	0.3	-0.017	-0.005	0.033	0.024	0.423	0.512
(19)	20	100	0.6	-0.029	-0.011	0.042	0.024	2.687	5.555
(20)	20	100	0.9	-0.100	-0.032	0.109	0.037	127.077	27.399
(21)	20	200	0	-0.006	-0.003	0.020	0.017	0.925	1.590
(22)	20	200	0.3	-0.009	-0.005	0.022	0.017	4.420	1.873
(23)	20	200	0.6	-0.016	-0.010	0.027	0.018	10.647	0.895
(24)	20	200	0.9	-0.065	-0.031	0.074	0.034	117.679	8.088

Simulations are based on 5000 replications. The fixed effects α_i and the innovations ϵ_{it} are assumed to have independent standard normal distributions. The initial observations y_{i0} are assumed to be generated by the stationary distribution $\mathcal{N}\left(\frac{\alpha_i}{1-\theta_0}, \frac{1}{1-\theta_0^2}\right)$. The last two columns report the statistic "sample size $\times$[skewness$^2/6 + (\text{kurtosis} - 3)^2/24$]". The statistic is asymptotically $\chi^2(2)$ under the null of normality. The upper and lower 2.5 percentile of $\chi^2(2)$ are 7.38 and .05. Normality of $\widehat{\theta}_{GMM}$ is rejected for cases 1, 2, 3, 4, 8, 12, 16, 20, 23, and 24. Normality of $\widehat{\widehat{\theta}}$ is rejected for cases 11, 12, 20, and 24.