BIAS IN DYNAMIC PANEL ESTIMATION WITH FIXED EFFECTS, INCIDENTAL TRENDS AND CROSS SECTION DEPENDENCE

By

Peter C.B. Phillips and Donggyu Sul

September 2003

Revised June 2004

COWLES FOUNDATION DISCUSSION PAPER NO. 1438

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence

Peter C.B. Phillips
Cowles Foundation, Yale University
University of Auckland & University of York

Donggyu Sul
Department of Economics
University of Auckland

June 8, 2004

Abstract

Explicit asymptotic bias formulae are given for dynamic panel regression estimators as the cross section sample size $N \to \infty$. The results extend earlier work by Nickell (1981) and later authors in several directions that are relevant for practical work, including models with unit roots, deterministic trends, predetermined and exogenous regressors, and errors that may be cross sectionally dependent. The asymptotic bias is found to be so large when incidental linear trends are fitted and the time series sample size is small that it changes the sign of the autoregressive coefficient. Another finding of interest is that, when there is cross section error dependence, the probability limit of the dynamic panel regression estimator is a random variable rather than a constant, which helps to explain the substantial variability observed in dynamic panel estimates when there is cross section dependence even in situations where N is very large. Some proposals for bias correction are suggested and finite sample performance is analyzed in simulations.

Keywords: Autoregression, Bias, Bias correction, Cross section dependence, Dynamic factors, Dynamic panel estimation, Incidental trends, Panel unit root.

JEL Classification Numbers: C33 Panel Data

First Draft: August, 2000
Completed Version: April, 2003
Revision: May, 2004

1 Introduction

In an influential paper, Nickell (1981) showed that in dynamic panel regressions the well known finite sample autoregressive bias (Orcutt, 1948; Kendall, 1954) in time series models persists asymptotically in large panels as the cross section sample size dimension $N \to \infty$. Nickell gave analytic formulae for this bias and found that its magnitude was considerable in many cases relevant to applied research. In consequence, bias reduction procedures have been proposed for practical implementation with a variety of dynamic panel estimators (e.g. Kiviet, 1995; Hahn and Kuersteiner, 2000). The literature is reviewed in Arrelano and Honoré (2000), Baltagi (2001) and Arrelano (2003).

The present paper extends this work in several directions that are relevant for empirical applications. The cases studied here include dynamic panel models with a unit root, deterministic linear trends, exogenous regressors, and errors that may be cross sectionally dependent. Many, and sometimes all, of these elements appear in applied work with dynamic panels. The main contribution of the paper is to provide new bias/inconsistency formulae for dynamic panel regressions in these cases, focusing on pooled least squares regression estimates. It is, of course, well known that instrumental variable and GMM procedures provide consistent estimates of dynamic coefficients in cases where pooled least squares is inconsistent (see Baltagi, 2001, Hsaio, 2003, and Arrelano, 2003, for recent overviews). However, these procedures are also known to suffer bias (Hahn, Hausman and Kuersteiner, 2001)) and, more significantly, weak instrumentation problems (Kruiniger, 2000; Hahn et al., 2001) when the dynamic coefficient is close to unity, as it often is in practical work. They can therefore be an unsatisfactory alternative in such cases, even when the time series sample size T is large, because of high variance (Phillips and Sul, 2003) and slow convergence (Moon and Phillips, 2004) problems. Hahn et al. (2001) have suggested a long difference estimator to alleviate some of these difficulties, but that estimator is not investigated here.

Two results of particular interest in the present paper are the size of the bias in models where incidental trends are extracted and the impact of cross section error dependence on the bias. In the first case, analytic formulae reveal that the inconsistency as the cross section sample size $N \to \infty$ can be huge when the time series sample size (T) is small and incidental trends are extracted in panel regression. For instance, our results show that when $T < 8$, the inconsistency in the estimate of a panel unit root is large enough to change the sign of the coefficient from positive to negative. Simulations confirm that this enormous asymptotic bias also manifests in finite (N) samples.

A second result of interest is the impact of heterogeneity and cross section error dependence on the bias. While mild heterogeneity has no asymptotic effect, cross section dependence has a major impact on the inconsistency of dynamic panel regression. Under cross section dependence, it is shown that the probability limit of the dynamic panel regression estimator is a random variable rather than a constant (as it is in the cross section independent case). The randomness of this limit as $N \to \infty$ helps to explain the substantial variability of dynamic panel estimates that is known to occur under cross section dependence even when N is very large (e.g., Phillips and Sul, 2003).

The remainder of the paper is organised as follows. Section 2 describes the panel models that are studied in the paper. Section 3 provides bias formulae for various cases under cross section independence and relates these to the existing literature. Section 4 considers the impact of cross section dependence on dynamic panel regression bias, looking at both stationary and unit root panels. Section 5 considers some bias reduction methods for both the cross section independent and dependent cases, and reports the results of some simulations. Section 6 concludes. The appendix contains derivations of the main results (Section 7) and a glossary of notation (Section 8).

2 Models

The panel regression models considered here fall into the following categories:

M1: (Fixed Effects) $\begin{cases} y_{it} = a_i + \rho y_{it-1} + \varepsilon_{it} & \rho \in (-1,1) \\ y_{it} = a_i + y^0_{it},\ y^0_{it} = \rho y^0_{it-1} + \varepsilon_{it} & \rho = 1 \end{cases}$

M2: (Incidental Linear Trends) $\begin{cases} y_{it} = a_i + b_i t + \rho y_{it-1} + \varepsilon_{it} & \rho \in (-1,1) \\ y_{it} = a_i + b_i t + y^0_{it},\ y^0_{it} = \rho y^0_{it-1} + \varepsilon_{it} & \rho = 1 \end{cases}$

M3: (Exogenous Regressors) $\tilde{y}_{it} = \rho \tilde{y}_{it-1} + \tilde{Z}'_{it}\beta + \tilde{\varepsilon}_{it}, \quad \rho \in (-1,1].$

In each case, the index i ($i = 1, ..., N$) stands for the i'th cross sectional unit and t ($t = 1, .., T$) indexes time series observations. The variables Z_{it} are exogenous. The affix notation on $\tilde{w}_t$ signifies that the series w_t has been detrended or demeaned and this will be clear from the context. Models M1 and M2 allow for both stationary ($|\rho| < 1$) and nonstationary ($\rho = 1$) cases. In M3, we allow for unit root and stationary y_{it} but do not consider here cases where Z_{it} may have nonstationary elements (i.e., the possibly cointegrated regression case). In the unit root cases, the initialization of y^0_{it} is taken to be $y^0_{i0} = O_p(1)$ and uncorrelated with $\{\varepsilon_{it}\}_{t\geq 1}$.

The cases of cross section independence and cross section dependence for the panel regression errors will be considered separately in Sections 3 and 4. We take first the case where the errors ε_{it} in the above models are independent across i. The following section derives explicit formulae for the asymptotic bias of the least squares estimates of ρ and β in that case, giving the inconsistency $\text{plim}_{N\to\infty}(\hat{\rho} - \rho)$ for each model where $\hat{\rho}$ is the panel least squares estimate of ρ. Section 4 studies the inconsistency of these estimates when there is cross section dependence.

3 Models with Cross Section Independence

This section includes three subsections, one for each model, and deals separately with the stationary and panel unit root cases. Before proceeding, one important difference in autoregressive bias between the time series AR(1) and panel AR(1) should be mentioned: there is negligible bias when the fixed effect is known (or zero) in the panel AR(1) model for large N. It is well known that the bias in an autogression with known mean arises from the asymmetry of the distribution of the least squares estimator $\hat{\rho}$ and is a finite sample (T) phenomenon. A similar phenomenon occurs in panel autoregressions with finite T and finite N when the mean is known (or, equivalently, set to zero). However, in panel autogressions with a known mean, the averaging across section eventually removes the asymmetry of the distribution as $N \to \infty$. Hence, for large N the distribution of $\hat{\rho}$ is close to symmetric about ρ and bias is negligible. Only when N is small is the bias important in the known fixed effect case.

On the other hand, when the fixed effect is estimated or when there are incidental trends to be removed, autoregressive bias can be large and it persists even as $N \to \infty$. As Orcutt (1948) pointed out, the removal of a mean or trend from the data in an autoregression produces an additional source of bias arising from the correlation of the error and the lagged dependent variable. In a panel model with incidental fixed effects and/or trends, this additional source of bias is not diminished as $N \to \infty$, as is well understood from Neyman and Scott (1948) and Nickell (1981). Interestingly, that inconsistency persists even as $T \to \infty$ when $\rho = 1 + c/T$ and the parameter being estimated is local to unity (Moon and Phillips, 1999, 2000 & 2004).

3.1 Fixed Effects Model M1

We first consider the stationary case where $\rho_i = \rho$, $|\rho| < 1$, under cross section error independence for ε_{it} and where the initial conditions are in the infinite past. The following explicit error condition is convenient.

Assumption A1: (error condition) *The ε_{it} have zero mean, finite $2+2\nu$ moments for some $\nu > 0$, are independent over i and t with $E(\varepsilon_{it}^2) = \sigma_i^2$ for all t, and $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \sigma_i^2 = \sigma^2$.*

Nickell (1981) assumed $iid(0, \sigma^2)$ errors ε_{it} but this is easily relaxed to allow for mild heterogeneity under regularity conditions of the type given in **A1**. The bias for the pooled least squares estimate of ρ in large cross section (N) asymptotics follows in the same way as Nickell (1981) and turns out to have the same form when there are heterogeneous errors. The calculations are straightforward and are not repeated here.

To illustrate, for the fixed effects model M1 the pooled least squares estimate of ρ has the form

$$\hat{\rho} = \rho + \frac{\sum_{t=1}^T \sum_{i=1}^N \tilde{y}_{it-1}\tilde{\varepsilon}_{it}}{\sum_{t=1}^T \sum_{i=1}^N \tilde{y}_{it-1}^2} = \rho + \frac{A_{NT}}{B_{NT}} = \rho + \frac{\frac{1}{N}A_{NT}}{\frac{1}{N}B_{NT}}. \tag{1}$$

Calculations analogous to those in Nickell (1981), but using the Markov strong law

$$\frac{1}{N}\sum_{i=1}^N \left(\varepsilon_{it}^2 - \sigma_i^2\right) \to_{a.s.} 0, \quad \frac{1}{N}\sum_{i=1}^N \varepsilon_{it}^2 \to_{a.s.} \sigma^2 \tag{2}$$

to accommodate cross section heterogeneity in ε_{it}, show that the limits of the numerator and denominator in (1) as $N \to \infty$ with T fixed have the same form as those in Nickell's case, viz.,

$$\text{plim}_{N\to\infty} \frac{1}{N}A_{NT} = -\frac{\sigma^2}{T}\frac{1}{1-\rho}\left[T - \frac{1-\rho^T}{1-\rho}\right] := -\sigma^2 A(\rho, T), \tag{3}$$

and

$$\text{plim}_{N\to\infty} \frac{1}{N}B_{NT} = \sigma^2 \frac{T-1}{1-\rho^2}\left\{1 - \frac{1}{T-1}\frac{2\rho}{1-\rho}\left[1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\right\} := \sigma^2 B(\rho, T). \tag{4}$$

Combining (3) and (4) we have the following simple extension of Nickell's (1981) bias result.

Proposition 1 *(Fixed Effects with $|\rho| < 1$) For model M1 with $|\rho| < 1$ and under Assumption A1, the inconsistency of the pooled least squares estimate of ρ as $N \to \infty$ is given by*

$$plim_{N\to\infty}(\hat{\rho} - \rho) = -\frac{1+\rho}{T-1}\left[1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\left\{1 - \frac{1}{T-1}\frac{2\rho}{1-\rho}\left[1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\right\}^{-1} \tag{5}$$

$$= G(\rho, T). \tag{6}$$

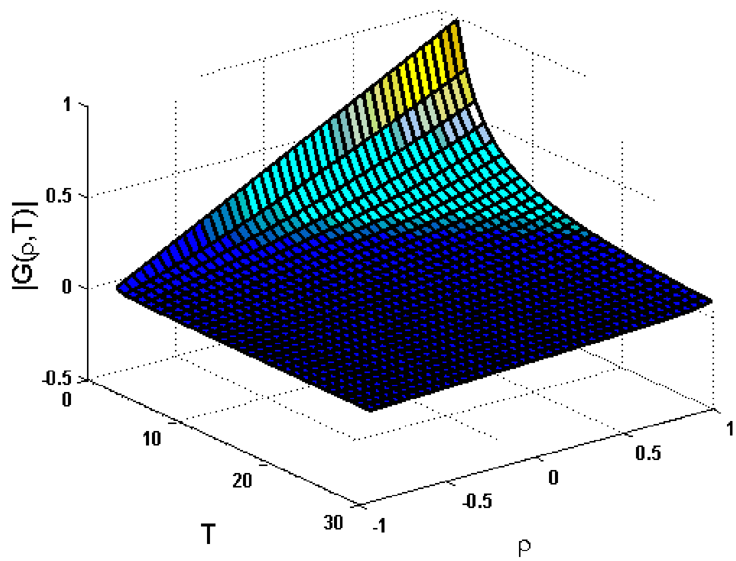

Figure 1: Asymptotic ($N \to \infty$) Bias Function $|G(\rho, T)| = -G(\rho, T)$ for Model M1.

For large T, the inconsistency has the expansion

$$G(\rho, T) = -\frac{1+\rho}{T-1}\left[1 + O(T^{-1})\right]. \tag{7}$$

Formula (5) is the same as that given by Nickell (1981) for the case of homogeneous errors[1]. Applying the third derivative version of l'Hôpital's rule directly to $G(\rho, T)$ with respect to ρ we obtain the limit behavior for the unit root case, viz., $\lim_{\rho \to 1} G(\rho, T) = -\frac{3}{T+1}$, and the inconsistency of the pooled least squares estimate for $\rho = 1$ follows

$$\text{plim}_{N\to\infty}(\hat{\rho} - 1) = -\frac{3}{T+1}, \tag{8}$$

a result that can be confirmed by more tedious direct calculation for the case $\rho = 1$.

Fig. 1 graphs the modulus of the inconsistency, $|G(\rho, T)| = -G(\rho, T)$, against ρ and T. As is clear from the figure, the magnitude of the asymptotic bias increases with ρ, and of course decreases as T increases.

[1] For $T = 3$, there is a typographical error in Nickell (1981), the correct formula being

$$\text{plim}_{N\to\infty}(\hat{\rho} - \rho) = -\frac{(1+\rho)(2+\rho)}{2(\rho+3)} \qquad T = 3.$$

3.2 Incidental Linear Trend Model M2

In this case there are heterogenous linear trends and constants as fixed effects. The pooled least squares estimate of ρ has the form $\hat{\rho} = C^y_{NT}/D_{NT}$, where

$$C^y_{NT} = \sum_{i=1}^{N}\left[\sum_{t=1}^{T}(y_{it}-y_{i\cdot})(y_{it-1}-y_{i\cdot-1}) - \frac{\sum_{t=1}^{T}[(t-\bar{t})(y_{it}-y_{i\cdot})]\sum_{t=1}^{T}[(t-\bar{t})(y_{it-1}-y_{i\cdot-1})]}{\sum_{t=1}^{T}(t-\bar{t})^2}\right],$$

and

$$D_{NT} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it-1}-y_{i\cdot-1} - \frac{\sum_{t=1}^{T}[(t-\bar{t})(y_{it-1}-y_{i\cdot-1})]}{\sum_{t=1}^{T}(t-\bar{t})^2}(t-\bar{t})\right]^2.$$

Setting $C_{NT} = C^y_{NT} - \rho D_{NT}$, the inconsistency as $N \to \infty$ with T fixed is

$$\text{p}\lim_{N\to\infty}(\hat{\rho}-\rho) = \frac{\text{plim}_{N\to\infty}\frac{1}{N}C_{NT}}{\text{plim}_{N\to\infty}\frac{1}{N}D_{NT}},$$

whose exact form and asymptotic (large T) representation are given in the following result.

Proposition 2 *(Linear Trend Fixed Effects with $|\rho| < 1$) As $N \to \infty$, for model M2 under Assumption A1, the inconsistency of the pooled least squares estimate for $\rho < 1$ is given by*

$$\begin{aligned} plim_{N\to\infty}(\hat{\rho}-\rho) &= -2\frac{1+\rho}{T-2}\left[1-\frac{1}{T-1}\frac{2}{1-\rho}C_1\right]\left[1-\frac{1}{T-2}\frac{4\rho}{1-\rho}D_1\right]^{-1} & (9)\\ &= H(\rho,T), & (10)\end{aligned}$$

where

$$\begin{aligned} C_1 &= 1-\frac{1}{T+1}\left(1+\frac{1-\rho^3}{(1-\rho)^3}\frac{1}{T}\right)+\left(\frac{1}{2}+\frac{1}{T+1}\left[\frac{1+2\rho}{1-\rho}+\frac{1-\rho^3}{(1-\rho)^3}\frac{1}{T}\right]\right)\rho^T, & (11)\\ D_1 &= 1-\frac{1}{T+1}\frac{2}{1-\rho}\left\{1+\frac{1}{T-1}\left[1-\frac{1-\rho^3}{T(1-\rho)^3}\left(1-\rho^T\right)+\left(\frac{3\rho}{1-\rho}+\frac{T+3}{2}\right)\rho^T\right]\right\}. & (12)\end{aligned}$$

For large T, the inconsistency has the following expansion

$$H(\rho,T) = -2\frac{1+\rho}{T-2}[1+O\left(T^{-1}\right)]. \tag{13}$$

Later calculations will extend these formulae to the case where the errors are cross section dependent. It is then useful to have explicit formulae for the numerator and denominator limits in the ratio (10) in order to highlight the differences between the two cases. These are as follows:

$$\begin{aligned} \text{plim}_{N\to\infty}\frac{1}{N}C_{NT} &= -\frac{\sigma^2}{T-1}\frac{2}{1-\rho}\left[(T-1)-\frac{2}{1-\rho}C_1\right] := -\sigma^2 C(\rho,T), & (14)\\ \text{plim}_{N\to\infty}\frac{1}{N}D_{NT} &= \sigma^2\frac{T-2}{1-\rho^2}\left[1-\frac{1}{T-2}\frac{4\rho}{1-\rho}D_1\right] := \sigma^2 D(\rho,T). & (15)\end{aligned}$$

From the expansions (13) and (7) for $H(\rho,T)$ and $G(\rho,T)$, it is apparent that the bias in the case of incidental trends is approximately twice that of the simple fixed effects model M1. For small T, the

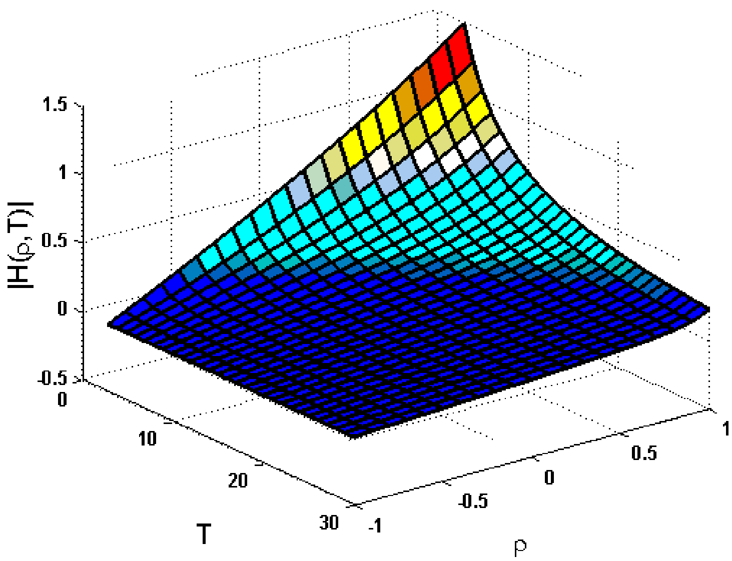

Figure 2: Asymptotic ($N \to \infty$) Bias Function $|H(\rho, T)| = -H(\rho, T)$ for Model M2.

magnitude of the bias in the trend model M2 is slightly larger than twice that of the fixed effects model M1. By direct calculation, the exact bias formula for some cases of small T are

$$H(\rho, T) = \begin{cases} -\frac{1}{2}\frac{\rho^2-3\rho-4}{\rho-3} & \text{for } T=3 \\ -\frac{1}{2}\frac{\rho^3-6\rho-5}{\rho^2-5} & \text{for } T=4 \\ -\frac{1}{2}\frac{2\rho^4+2\rho^3-5\rho^2-17\rho-12}{2\rho^3+2\rho^2-3\rho-15} & \text{for } T=5 \end{cases}. \tag{16}$$

and the bias differential (M2 - 2 $\times$ M1) is

$$|H(\rho, T)| - 2\,|G(\rho, T)| =$$

$$2G(\rho, T) - H(\rho, T) = \left\{ \begin{array}{ll} \frac{1}{2}\rho\frac{1-\rho^2}{3-\rho^2} > 0 & \text{for } T=3 \\ \frac{1}{2}\rho\frac{1-\rho^4-3\rho^3+\rho}{(\rho^2-5)(\rho^2+3\rho+6)} > 0 & \text{for } T=4 \\ \frac{1}{2}\rho\frac{2\rho^6+8\rho^5+13\rho^4-3\rho^2+12\rho+8}{(15-2\rho^3-2\rho^2+3\rho)(\rho^3+3\rho^2+6\rho+10)} > 0 & \text{for } T=5 \end{array} \right\} \text{ for } 0 \le \rho < 1$$

Fig. 2 graphs the modulus of the inconsistency, $|H(\rho, T)| = -H(\rho, T)$, against ρ and T. As is apparent from the figure, the inconsistency increases sharply in magnitude as ρ increases and as T decreases.

Applying the fifth derivative version of l'Hôpital's rule directly to $H(\rho, T)$ with respect to ρ we obtain the limit behavior for the unit root case, viz., $\lim_{\rho \to 1} H(\rho, T) = -7.5/(T+2)$. Thus, when y_{it} is a panel unit root process, the inconsistency for the pooled OLS estimator under model M2 is given by

$$\text{plim}_{N\to\infty}(\hat{\rho} - 1) = -\frac{7.5}{T+2}, \tag{17}$$

a result that was obtained by direct calculation in Harris and Tzavalis (1999). Comparing (17) with (8), we see that when $\rho = 1$ the bias for model M2 is more than twice that in model M1 for all $T > 3$. Table 1 shows corroborating results obtained by simulation.

Perhaps the most striking feature of the autoregressive bias in model M2 is that when T is small, the pooled least squares estimate of ρ is often negative even when the true autoregressive coefficient ρ is (near) unity. To illustrate the dramatic nature of these bias effects we show the results of detrending on a short time series panel. Fig. (3) shows a sample plot of data generated by the true panel relation between y_{it} and y_{it-1} for which $a_i = b_i = 0$ in M2 and with $\rho = 0.9$ and $T = 4$. This sample plot shows a clear positive relationship between y_{it} and y_{it-1} (the fitted $\hat{\rho} = 0.907$). After detrending the data by removing incidental trends, the sample plot of the new data is shown in Fig. 4, where the relationship between y_{it} and y_{it-1} is now seen to be clearly negative (the fitted $\hat{\rho} = -0.529$). The autoregressive bias in this case is so large that it distorts the correlation into the opposite direction: strongly positive autocorrelation ($\rho = 0.9$) becomes strong negative autocorrelation ($\bar{\rho} = \text{plim}_{N\to\infty}\hat{\rho} = 0.9 - 1.402 = -0.502$) in the detrended sample data. The reason for this distortion is clear. When T is small and there is positive autoregressive behavior in the panel y_{it}, incidental trend extraction (for each i) can have such a powerful effect on the configuration of the data that the detrended observations $\tilde{y}_{it}$ behave as if they were actually negatively autocorrelated.

Table 1: Asymptotic Bias in the Estimated Autoregressive Coefficient in the Linear Trend Model M2

	Absolute Bias: Model(Simulation)					
T	ρ =0.1	ρ =0.3	ρ =0.5	ρ =0.7	ρ =0.9	ρ =1.0
3	0.740(0.739)	0.891(0.890)	1.050(1.049)	1.220(1.219)	1.402(1.402)	1.500(1.499)
4	0.561(0.562)	0.690(0.690)	0.829(0.830)	0.982(0.982)	1.154(1.154)	1.250(1.250)
5	0.450(0.450)	0.558(0.558)	0.679(0.678)	0.816(0.815)	0.977(0.977)	1.071(1.071)
6	0.375(0.375)	0.466(0.466)	0.571(0.571)	0.694(0.694)	0.845(0.846)	0.938(0.938)
7	0.321(0.321)	0.399(0.398)	0.490(0.490)	0.601(0.600)	0.743(0.742)	0.833(0.833)
8	0.280(0.280)	0.348(0.348)	0.428(0.428)	0.528(0.528)	0.661(0.661)	0.750(0.750)
9	0.249(0.249)	0.308(0.308)	0.379(0.379)	0.470(0.470)	0.595(0.595)	0.682(0.682)

Note: $N = 5,000$, errors are drawn as *iid* $N(0, 1)$, the number of replications = 500, T = sample size used in the regression, $T + 1$ = the total number of observations of the dependent variable.

3.3 Exogenous Regressor Model M3

In many panel model applications, such as the original study by Balestra and Nerlove (1966) on the demand for natural gas, exogenous variables are included in addition to lagged dependent regressors in the specification. Another example that is important in ongoing practical work is the panel analysis of growth convergence, where specific covariates contributing to economic growth are included as well as dynamic effects. The effect of the presence of such variables can be analyzed in the context of models like M3.

Stacking cross section data first and then time series observations, model M3 can be written as

$$\tilde{y}_t = \rho\tilde{y}_{t-1} + \tilde{Z}_t'\beta + \tilde{\varepsilon}_t, \text{ and } \tilde{y} = \rho\tilde{y}_{-1} + \tilde{Z}\beta + \tilde{\varepsilon}, \text{ say,} \tag{18}$$

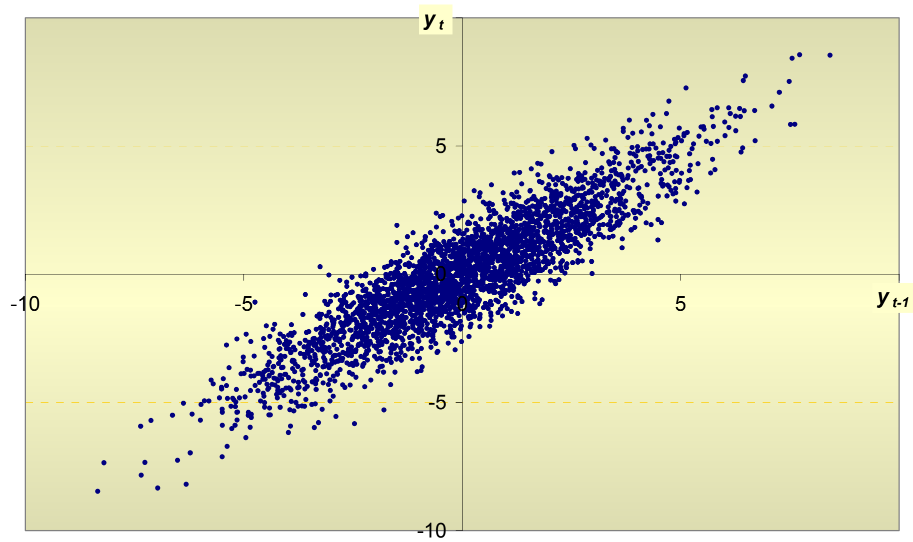

Figure 3: Sample Data before Detrending ($T = 4, N = 1,000, \rho = 0.9, \hat{\rho} = 0.90$)

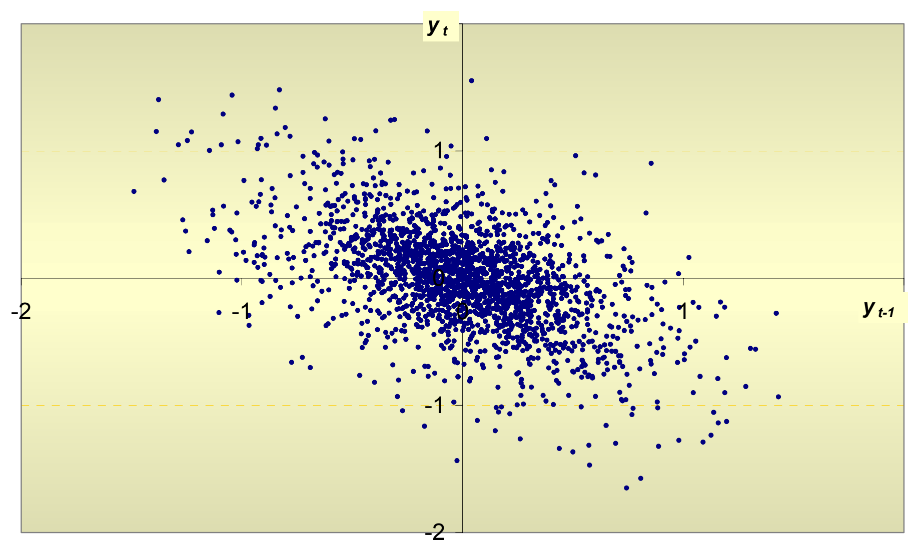

Figure 4: Sample Data after Detrending ($T = 4$, $N = 1,000$; $\rho = 0.9$, $\bar{\rho} = \text{plim}_{N\to\infty}\hat{\rho} = -0.502$, $\hat{\rho} = -0.53$).

where the tilde affix on $\tilde{w}$ signifies that the series w has been demeaned or detrended. Setting $Q_{\tilde{Z}} = I - \tilde{Z}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'$, we have

$$\text{plim}_{N\to\infty}\left(\hat{\rho}-\rho\right) = \left\{\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{y}_{-1}\right\}^{-1}\left\{\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{\varepsilon}\right\}, \tag{19}$$

and

$$\text{plim}_{N\to\infty}\left(\hat{\beta}-\beta\right) = -\left\{\text{plim}_{N\to\infty}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\left(\tilde{Z}'\tilde{y}_{-1}\right)\right\}\text{plim}_{N\to\infty}\left(\hat{\rho}-\rho\right). \tag{20}$$

Calculations similar to those in the preceeding section then lead to the following result on the inconsistency of these estimates.

Proposition 3 *(Exogenous Variables, Fixed and Trend Effects) As $N \to \infty$, for model M3 under Assumption A1 and with $|\rho| < 1$, the inconsistency of the pooled least squares estimate of ρ is given in the fixed effects case by*

$$plim_{N\to\infty}\left(\hat{\rho}-\rho\right) = -\frac{\sigma^2 A(\rho,T)}{\sigma^2 B(\rho,T) + \beta'\left[plim_{N\to\infty}\frac{1}{N}\tilde{Z}_{\rho,-1}'Q_{\tilde{Z}}\tilde{Z}_{\rho,-1}\right]\beta}, \tag{21}$$

and in the incidental trends case by

$$plim_{N\to\infty}\left(\hat{\rho}-\rho\right) = -\frac{\sigma^2 C(\rho,T)}{\sigma^2 D(\rho,T) + \beta'\left[plim_{N\to\infty}\frac{1}{N}\tilde{Z}_{\rho,-1}'Q_{\tilde{Z}}\tilde{Z}_{\rho,-1}\right]\beta}, \tag{22}$$

where $\tilde{Z}_{\rho,-1} = \left(\tilde{Z}_{\rho,0}', ..., \tilde{Z}_{\rho,T-1}'\right)'$ with $\tilde{Z}_{\rho,t} = \left(\tilde{Z}_{\rho,t}^1, ..., \tilde{Z}_{\rho,t}^N\right)'$ and $\tilde{Z}_{\rho,t}^i = \sum_{j=0}^{\infty}\rho^j\tilde{Z}_{it-j}$. The inconsistency of the pooled estimate of β is

$$plim_{N\to\infty}\left(\hat{\beta}-\beta\right) = -\left\{plim_{N\to\infty}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'\tilde{Z}_{\rho,-1}\beta\right\}plim_{N\to\infty}\left(\hat{\rho}-\rho\right). \tag{23}$$

These formulae continue to apply in the unit root case $\rho = 1$ upon replacement of $A(\rho,T)$, $B(\rho,T)$, $C(\rho,T)$, and $D(\rho,T)$ with $A(T)$, $B(T)$, $C(T)$, and $D(T)$, respectively, which are defined in (54) and (57), and $\tilde{Z}_{\rho,-1}$ by $\tilde{Z}_{1,-1} = \left(\tilde{Z}_{1,0}', ..., \tilde{Z}_{1,T-1}'\right)'$ where $\tilde{Z}_{1,t} = \left(\tilde{Z}_{1,t}^1, ..., \tilde{Z}_t^N\right)'$ and $\tilde{Z}_t^i = \sum_{j=0}^{t}\tilde{Z}_{it-j}$.

Note that when $\beta = 0$, the inconsistency (21) and (22) is the same as in the case of models M1 and M2 with no exogenous variables. When $\beta \neq 0$, the inconsistency is clearly smaller in absolute value than when there are no exogenous variables. Note that this is the opposite conclusion to that reached in Nickell (1981, p.1424). Nickell argued that the denominator in (19) is smaller than it is in the case of no exogenous variables because of the effect of the projection operator $Q_{\tilde{Z}}$ which reduces the magnitude of the sum of squares in the sense that $\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{y}_{-1} \leq \tilde{y}_{-1}'\tilde{y}_{-1}$. While this is certainly correct, the argument neglects the fact that when exogenous variables are present in the model they also affect the variability of the data $\tilde{y}_t$. In particular, when $|\rho| < 1$ we have

$$\tilde{y}_{it} = \sum_{j=0}^{\infty}\rho^j\tilde{Z}_{it-j}\beta + \sum_{j=0}^{\infty}\rho^j\tilde{\varepsilon}_{it} := \tilde{Z}_{\rho it}\beta + \tilde{y}_{it}^0, \text{ say} \tag{24}$$

and, using the stacked notation $\tilde{y} = \tilde{Z}_\rho\beta + \tilde{y}^0$ and its lagged variant, we find that

$$\begin{aligned}\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{y}_{-1} &= \beta'\left[\text{plim}_{N\to\infty}\frac{1}{N}\tilde{Z}_{\rho,-1}'Q_{\tilde{Z}}\tilde{Z}_{\rho,-1}\right]\beta + \text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}^{0\prime}\tilde{y}_{-1}^0 \\ &= \beta'\left[\text{plim}_{N\to\infty}\frac{1}{N}\tilde{Z}_{\rho,-1}'Q_{\tilde{Z}}\tilde{Z}_{\rho,-1}\right]\beta + \sigma^2 B(\rho,T).\end{aligned} \tag{25}$$

It is clear from (25) that we have the reverse inequality $\tilde{y}_{-1}' Q_{\tilde{Z}} \tilde{y}_{-1} \geq \tilde{y}_{-1}^{0\prime} \tilde{y}_{-1}^{0}$, the left side being the denominator for the case where exogenous variables are present in the model and the right side being the denominator for the case where there are no exogenous variables. Similar effects apply in the case of models with incidental trends. In short, the presence of exogenous variables reduces the extent of the inconsistency of $\hat{\rho}$ whenever these variables have a material effect on data variability, i.e. when $\beta \neq 0$.

An exception occurs in the case where the model has the following components form instead of (24):

$$\tilde{y}_{it} = \tilde{Z}_{it}\beta + \tilde{y}_{it}^{0}. \tag{26}$$

In this case, the fitted regression model M3 is replaced by

$$\tilde{y}_{it} = \rho \tilde{y}_{it-1} + \tilde{Z}_{it}\beta_1 + \tilde{Z}_{it-1}\beta_2 + \tilde{\varepsilon}_{it}, \text{ with } \beta_1 = \beta \text{ and } \beta_2 = \rho\beta. \tag{27}$$

and then $\tilde{y} = \rho \tilde{y}_{-1} + \tilde{Z}\gamma + \tilde{\varepsilon}$ with $\tilde{Z}$ comprising a stacked version of $(\tilde{Z}_{it}, \tilde{Z}_{it-1})$. It is apparent that instead of (25) we now have $\text{plim}_{N\to\infty} \frac{1}{N} \tilde{y}_{-1}' Q_{\tilde{Z}} \tilde{y}_{-1} = \sigma^2 B(\rho, T)$ and the Proposition continues to hold but without the second term in the denominator in (21) and (22). In this case, the inconsistency of $\hat{\rho}$ is unchanged by the presence of exogenous variables and the inconsistency of β is given by

$$\text{plim}_{N\to\infty} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -\beta \left\{ \text{plim}_{N\to\infty} (\hat{\rho} - \rho) \right\} \end{pmatrix} \tag{28}$$

in place of (23).

4 Models with Cross Section Dependence

Bai and Ng (2002), Forni, Hallin, Lippi and Reichlin (2000), Moon and Perron (2002), and Phillips and Sul (2003) provide some recent investigations of panel models with cross section dependence. In all these studies, the parametric form of dependence is based on a factor analytic structure. Broadly speaking, two types of factor models have been employed, the distinction resting on whether a dynamic structure is explicit or not. Forni, Lippi and Reichlin (1999), Moon and Perron (2002), and Phillips and Sul (2003) all use a factor structure where the dynamics are explicit in the system. The following model is a prototypical first order panel dynamic system

$$y_{it} = a_i + \rho_i y_{it-1} + u_{it}, \; u_{it} = \sum_{s=1}^{K} \delta_{is}\theta_{st} + \varepsilon_{it}, \tag{29}$$

where the errors u_{it} depend on K factors $\{\theta_{st} : s = 1, ..., K\}$ with factor loadings $\{\delta_{is} : s = 1, ..., K\}$, and ε_{it} is assumed to be $iid(0, \sigma_i^2)$. In this prototypical system, θ_{st} and ε_{it} are assumed to be independent of each other and each is assumed to be *iid*. Also, θ_{st} is taken to be cross sectionally independent of θ_{qt}.

The second type of model (e.g., Bai and Ng, 2002) uses a direct factor structure for the data of the form

$$y_{it} = \sum_{s=1}^{K} \lambda_{is} F_{st} + m_{it}. \tag{30}$$

In (30) there are again K factors and factor loadings $\{F_{st}, \lambda_{is} : s = 1, ..., K\}$, F_{st} may be correlated with F_{qt} and may have its own time series structure, and the residual m_{it} is assumed to be cross sectionally independent. When the dynamic factor model (29) has a homogeneous autoregressive coefficient ($\rho_i = \rho$), it can be viewed as a restricted version of the direct model (30) in which a common dynamic factor can be drawn from each of the individual factors and the error.

The impact of common factors on dynamic panel regression analysis can be illustrated in the simple case of a single factor with no fixed effects. Suppose $a_i = 0$ and $\rho_i = \rho$ in (29) for all i. Then, the data is generated according to $y_{it} = \rho y_{it-1} + \delta_i \theta_t + \varepsilon_{it}$, which we can write in a convenient components form as

$$y_{it} = y_{it}^0 + \delta_i z_t, \quad y_{it}^0 = \rho y_{it-1}^0 + \varepsilon_{it}, \quad z_t = \rho z_{t-1} + \theta_t. \tag{31}$$

Let $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 = m_\delta^2$ be finite. Then, straightforward calculations reveal that the probability limit of the pooled least squares estimate as $N \to \infty$ is

$$\text{plim}_{N\to\infty} (\hat{\rho} - \rho) = \frac{\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T y_{it-1} u_{it}}{\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T y_{it-1}^2} = \frac{m_\delta^2 \left(\sum_{t=1}^T z_{t-1} \theta_t \right)}{T \frac{\sigma^2}{1-\rho^2} + m_\delta^2 \sum_{t=1}^T z_{t-1}^2}. \tag{32}$$

Thus, even with no fixed effects, $\hat{\rho}$ is inconsistent and the inconsistency depends on the degree of cross section dependence and the variance ratio σ^2/m_δ^2. Importantly for fixed T, the bias is random and depends on the process z_t and factor θ_t. Obviously for large T and temporally independent common shocks $T^{-1} \sum_{t=1}^T z_{t-1} \theta_t = o_p(1)$, so that in this case the bias will be small.

While K is fixed and generally taken to be very small (typically $K = 1$ or 2) in most macro empirical studies, in microeconometric work it is often reasonable to think of the number of factors that influence behavior as being potentially large and possible infinite. For instance, in studies of earnings there are many observable factors in panel data sets such as the PSID and equally many unobservables. Also, there are often common factors for personal income data, such as region, family, male/female ratio, race composition, education and age composition, to mention just a few; and the number of these factors may increase as we collect more cross section observations. The number of factors may further vary across i and change over time.

Thus, we may, in principle at least, consider cases where $K \to \infty$ as $N \to \infty$ or where $K = \infty$, in which there are an infinite number of unobserved factors. In such cases, the component $\sum_{s=1}^K \delta_{is} \theta_{st}$ in (29) can be replaced by an infinite sum $\sum_{s=1}^\infty \delta_{is} \theta_{st}$, which may be interpreted as a spatial linear process and on whose coefficients δ_{is} some restrictions (and ordering) must be imposed to ensure convergence. Another approach is to normalize the coefficients δ_{is} by some function of the factor count index K and require the normalized coefficient δ_{isK} to be small enough in mean and variance as $K \to \infty$ to assure existence of suitable limits of the sample moments of the data. Some recent microeconometric work utilizing this approach is Altonji, Elder and Taber (2002). In their work, $\delta_{isK} = K^{-1/2} \delta_{is}$ and the δ_{is} and θ_s are taken to be covariance stationary and ergodic zero mean random variates over s for some given ordering and to satisfy a central limit theorem. If this approach were used above, (31) would be replaced by

$$y_{it} = y_{it}^0 + \sum_{s=1}^K \delta_{isK} z_{st}, \quad y_{it}^0 = \rho y_{it-1}^0 + \varepsilon_{it}, \quad z_{st} = \rho z_{st-1} + \theta_{st}. \tag{33}$$

Without going into details over regularity conditions, we can compare this case with result (32). By independence over i, we would have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \delta_{is} \delta_{ip} = E(\delta_{is} \delta_{ip}) = \mu(p - s), \quad \text{say,}$$

and by ergodicity

$$\lim_{K\to\infty}\frac{1}{K}\sum_{s=1}^{K-h} z_{st-1}\theta_{s+h,t} = E\left(z_{st-1}\theta_{s+h,t}\right) = \gamma_{z\theta}\left(h\right)$$

for each h. If $\xi_{st} = z_{st-1}\theta_{s+h,t} - \gamma_{z\theta}(h)$ and $K^{-1/2}\sum_{s=1}^{K-h}\xi_{st} = O_p(1)$, then, taking sequential limits as $N\to\infty$, followed by $K\to\infty$, we would have

$$\begin{aligned}
& \lim_{K\to\infty} \operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it-1}u_{it} \\
= & \lim_{K\to\infty}\frac{1}{K}\sum_{s,p=1}^{K}\mu\left(p-s\right)\sum_{t=1}^{T} z_{st-1}\theta_{pt} \\
= & \sum_{t=1}^{T}\lim_{K\to\infty}\frac{1}{K}\sum_{h=-K+1}^{K-1}\mu\left(h\right)\sum_{s=1-hI\{h<0\}}^{K-hI\{h>0\}} z_{st-1}\theta_{s+h,t} \\
= & \sum_{t=1}^{T}\lim_{K\to\infty}\sum_{h=-K+1}^{K-1}\mu\left(h\right)\frac{1}{K}\sum_{s=1-hI\{h<0\}}^{K-hI\{h>0\}}\left(\gamma_{z\theta}\left(h\right)+\xi_{st}\right) \\
= & \sum_{t=1}^{T}\lim_{K\to\infty}\left\{\frac{1}{K}\sum_{h=-K+1}^{K-1}\mu\left(h\right)\gamma_{z\theta}\left(h\right)+O_p\left(\frac{1}{\sqrt{K}}\right)\right\} \\
= & \ 0,
\end{aligned} \tag{34}$$

provided $\sum_{h=-\infty}^{\infty}\mu(h)\gamma_{z\theta}(h)$ is finite. Under this set-up, the dynamic panel estimation bias is zero in contrast to (32). Of course, this type of argument depends on the appropriateness of the weak dependence conditions, which in turn depends on the existence of some spatial ordering of the factors. Therefore, the circumstances under which (34) is more appropriate than (32) are complex and involve many other considerations that will not be pursued here.

In contrast, aggregate data may reasonably be thought of as having relatively fewer common factors because in the aggregation process, the effect of the micro common factors is averaged out. Moreover, with aggregate data, N is often considered to be fixed, as in the number of countries in cross country studies, whereas T continues to increase.

The analysis that follows is based on dynamic panel models of the type (29), where the time series structure is built explicitly into the system behavior of y_{it}. This facilitates comparisons with the cross section independent case of Nickell (1981) and corresponds with many models used in the empirical literature such as the original study by Balestra and Nerlove (1966). We consider first the case where there are no exogenous variables.

4.1 Fixed Effects

As in (29), the model extends M1 to accommodate cross section dependent errors as follows.

Model M1-CSD: (Fixed Effects) $$\begin{cases} y_{it} = a_i + \rho y_{it-1} + u_{it}, & \rho\in(-1,1) \\ y_{it} = a_i^0 + y_{it}^0,\ y_{it}^0 = \rho y_{it-1}^0 + u_{it}, & \rho = 1 \end{cases}$$

We deal first with the stationary case. In the unit root case, the initialization y_{i0}^0 is taken to be $O_p(1)$.

Assumption A2: (Cross Section Dependence) *The u_{it} have the factor component structure*

$$u_{it} = \sum_{s=1}^{K} \delta_{si}\theta_{st} + \varepsilon_{it} = \delta_i'\theta_t + \varepsilon_{it}, \tag{35}$$

*where the ε_{it} satisfy **A1,** the factors θ_t are $iid(0, \Sigma_\theta)$ over t and the factor loadings δ_i are nonrandom parameters satisfying $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\delta_i\delta_i' = M_\delta$. When $K = 1$, we set $\Sigma_\theta = \sigma_\theta^2$ and $M_\delta = m_\delta^2$.*

Under A2, we can develop an asymptotic theory for the pooled least squares estimate, $\hat{\rho}$, of the common dynamic coefficient ρ. It is convenient to use a sequential asymptotic argument with $N \to \infty$ followed by $T \to \infty$. This approach produces a result for the bias or inconsistency of $\hat{\rho}$ as $N \to \infty$ and the expression can conveniently be written in an asymptotic format that is valid as $T \to \infty$. This extends the earlier asymptotic expansion results (7) and (13) to the case of cross section dependence. The main result follows.

Proposition 4 *(Fixed Effects with $|\rho| < 1$) In model **M1-CSD** with errors u_{it} having the factor structure (35) and satisfying assumption **A2,** the pooled least squares estimate $\hat{\rho}$ is inconsistent as $N \to \infty$ and*

$$plim_{N\to\infty}(\hat{\rho} - \rho) = -\left[\sigma^2 A(\rho, T) + \psi_{AT}\right]\left[\sigma^2 B(\rho, T) + \psi_{BT}\right]^{-1}, \tag{36}$$

where $A(\rho, T)$ and $B(\rho, T)$ are defined in (3) and (4),

$$\psi_{AT} = -trace\left\{\sum_{t=1}^{T}\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)\left(\theta_t - \bar{\theta}\right)' M_\delta\right\}, \quad \bar{Z}_{\theta,-1} = T^{-1}\sum_{t=1}^{T} Z_{\theta t} \tag{37}$$

$$\psi_{BT} = trace\left\{\sum_{t=1}^{T}\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)' M_\delta\right\}, \tag{38}$$

and $Z_{\theta t} = \sum_{j=0}^{\infty}\rho^j\theta_{t-j}$. In the single factor ($K = 1$) case, the inconsistency (36) has the following asymptotic representation as $T \to \infty$

$$\begin{aligned} plim_{N\to\infty}(\hat{\rho} - \rho) \quad &= \quad -\frac{1+\rho}{T} - \frac{2\rho}{T}\frac{\sigma_\theta^2 m_\delta^2}{\sigma^2 + \sigma_\theta^2 m_\delta^2} \\ &\quad + \frac{\sigma_\theta^2 m_\delta^2}{\sigma^2 + \sigma_\theta^2 m_\delta^2}\left(g_{\theta T} - Eg_{\theta T}\right) + o_p\left(T^{-1}\right), \end{aligned} \tag{39}$$

where

$$g_{\theta T} = \frac{\sum_{t=1}^{T}\left(z_{\theta t-1} - \bar{z}_{\theta -1}\right)\left(\theta_t - \bar{\theta}\right)}{\sum_{t=1}^{T}\left(z_{\theta t-1} - \bar{z}_{\theta -1}\right)^2}$$

is the centred least squares estimate of the slope coefficient in a regression of $z_{\theta t}$ on $z_{\theta t-1}$ and a constant, and where $E\left(g_{\theta T}\right) = -\frac{1+3\rho}{T} + o\left(T^{-1}\right)$.

Remark 1 It is apparent from the form of (36) and (39) that the inconsistency of the panel estimate $\hat{\rho}$ as $N \to \infty$ is random, as distinct from the nonrandom expression that we normally get for bias or inconsistency, such as that given by (7) in the cross section independent case. Note, of course, that when the factor loadings $\delta_{si} = 0$ for all i and s, we have $M_\delta = 0$ and then (36) reduces to $G(\rho, T) = -A(\rho, T)/B(\rho, T)$, and the second term on the right side of (39) is zero. So, in this case, the results reduce to those that apply in the cross section independent case, viz. (5) and (7). When $\delta_{si} \neq 0$ and $M_\delta \neq 0$, then the components ψ_{AT} and ψ_{BT} in (36) are non zero random variables with positive variance. Likewise, the third term of (39) is nonzero. So the immediate contribution of cross section dependence is to introduce variability into the inconsistency of $\hat{\rho}$ and additional bias.

Remark 2 In the single factor model ($K = 1$), the inconsistency expression (39) involves the regression coefficient error $g_{\theta T}$ of $z_{\theta t}$, and (39) can be written as

$$\text{plim}_{N\to\infty}(\hat{\rho} - \rho) = -\frac{1+\rho}{T} - \frac{\sigma_\theta^2 m_\delta^2}{\sigma^2 + \sigma_\theta^2 m_\delta^2}\left[\frac{2\rho}{T} + (Eg_{\theta T} - g_{\theta T})\right] + o_p\left(T^{-1}\right).$$

The second term in this expansion of the inconsistency involves the factor $m_\delta^2\sigma_\theta^2/\left(\sigma^2 + m_\delta^2\sigma_\theta^2\right)$ which is less than unity and whose magnitude decreases as σ^2 increases. Hence, as the importance of the error component ε_{it} grows (i.e. as $\sigma^2 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sigma_i^2$ increases), then the relative importance of the random component in the inconsistency (arising from the presence of cross section dependence) diminishes.

Remark 3 Next consider the case where there is a large number of factors. To simplify, assume that the factors θ_{kt} are $iid(0,\sigma_\theta^2)$ over both k and t and with finite fourth moments, that $M_\delta = diag\left(m_1^2, m_2^2, ..., m_K^2\right)$ is diagonal, $\sup_k m_k^4 < \infty$, and that $K^{-1}\sum_{k=1}^K m_k^2 \to m^2 > 0$ as $K \to \infty$. is large. Then, setting $\xi_{kT} = \sum_{t=1}^T \left(Z_{\theta_k t-1} - \bar{Z}_{\theta_k,-1}\right)\theta_{kt}$ and noting that ξ_{kT} is iid over k with mean $E\left(\xi_{kT}\right) = \sigma_\theta^2 A\left(\rho, T\right)$ and finite variance, we find that

$$\begin{aligned} K^{-1}\psi_{AT} &= -K^{-1}\sum_{k=1}^K m_k^2\left\{\sum_{t=1}^T\left(Z_{\theta_k t-1} - \bar{Z}_{\theta_k,-1}\right)\left(\theta_{kt} - \bar{\theta}_k\right)\right\} = -K^{-1}\sum_{k=1}^K m_k^2\xi_{kT} \\ &= -K^{-1}\sum_{k=1}^K m_k^2 E\left(\xi_{kT}\right) - K^{-1}\sum_{k=1}^K m_k^2\left\{\xi_{kT} - E\left(\xi_{kT}\right)\right\} \\ &= m^2\sigma_\theta^2 A\left(\rho, T\right) + o_p\left(1\right), \quad \text{as } K \to \infty. \end{aligned}$$

In a similar way,

$$K^{-1}\psi_{BT} = K^{-1}\sum_{k=1}^K m_k^2\left\{\sum_{t=1}^T\left(Z_{\theta_k t-1} - \bar{Z}_{\theta_k,-1}\right)^2\right\} = m^2\sigma_\theta^2 B\left(\rho, T\right) + o_p\left(1\right), \quad \text{as } K \to \infty.$$

It follows that

$$\begin{aligned} \lim_{K\to\infty}\text{plim}_{N\to\infty}\left(\hat{\rho} - \rho\right) &= \lim_{K\to\infty}\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^C}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^C} \\ &= \lim_{K\to\infty}\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}} \\ &= \frac{A\left(\rho, T\right)}{B\left(\rho, T\right)} = G\left(\rho, T\right). \end{aligned} \tag{40}$$

Thus, when there are a large number of independent factors, the dynamic panel bias of the cross section dependent case becomes less random and as $K \to \infty$ it converges to the bias of the cross section independent case. Fig. 5 illustrates this effect by showing the bias distribution for various values of K, against that of the cross section independent case. This result appears to be relevant for micro panel data situations where large numbers of independent factors are involved.

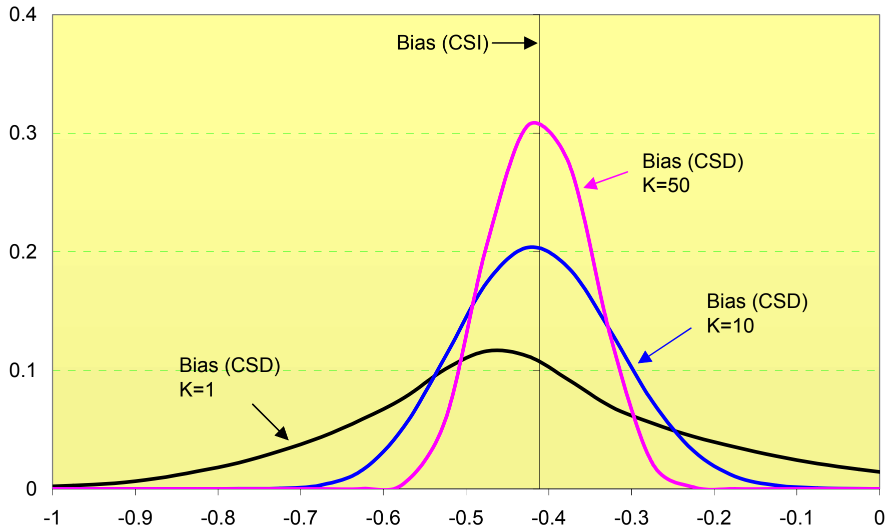

Figure 5: Random Bias under cross section dependence: $T = 5$, $\rho = 0.5$, $\delta_{is} - iidN\,(0, 1)$

Remark 4 In the unit root case $(\rho = 1)$, the same limit theory applies. In particular, (36) holds and

$$\text{plim}_{N\to\infty}(\hat{\rho} - 1) = -\left[\sigma^2 A(1,T) + \psi_{AT}\right]\left[\sigma^2 B(1,T) + \psi_{BT}\right]^{-1},$$

with $A(1,T) = (T-1)/2$ and $B(1,T) = (T-1)(T+1)/6$. When $K = 1$, we then get the expansion

$$\text{plim}_{N\to\infty}(\hat{\rho} - 1) = -\frac{3}{T} - \frac{\sigma_\theta^2 m_\delta^2}{\sigma^2 + \sigma_\theta^2 m_\delta^2}\left\{\frac{3}{T} + g_{\theta T}\right\} + o_p\left(T^{-1}\right)$$

in place of (39).

4.2 Incidental Trends

We take M2 and allow for errors u_{it} that satisfy Assumption A2:

Model M2-CSD (Incidental Trends) $\begin{cases} y_{it} = a_i + b_i t + \rho y_{it-1} + u_{it} & \rho \in (-1, 1) \\ y_{it} = a_i + b_i t + y_{it}^0, \; y_{it}^0 = \rho y_{it-1}^0 + u_{it} & \rho = 1 \end{cases}$

It will be convenient to define the following notation to represent the residual from linear detrending the variable w_t :

$$\begin{aligned} w_t^\tau &= w_t - \left\{\frac{2\,(2T+1)}{T\,(T-1)}\left(\sum_{t=1}^T w_t\right) - \frac{6}{T\,(T-1)}\sum_{t=1}^T t w_t\right\} \\ &\quad - \left\{-\frac{6}{T\,(T-1)}\sum_{t=1}^T w_t + \frac{12}{T\,(T^2-1)}\sum_{t=1}^T t w_t\right\} t \\ &= w_t - a_T^w - b_T^w t, \end{aligned}$$

where

$$a_T^w = \frac{2(2T+1)}{T(T-1)}\left(\sum_{t=1}^{T} w_t\right) - \frac{6}{T(T-1)}\sum_{t=1}^{T} tw_t, \quad b_T^w = \frac{12}{T(T^2-1)}\sum_{t=1}^{T} tw_t - \frac{6}{T(T-1)}\sum_{t=1}^{T} w_t.$$

Derivations similar to those of proposition 4 provide the following analogue of (36) and (39).

Proposition 5 *(Incidental Trends with $|\rho| < 1$) In model **M2-CSD** with errors u_{it} having the factor structure (35) and satisfying assumption **A2,** the pooled least squares estimate $\hat{\rho}$ is inconsistent as $N \to \infty$ and*

$$plim_{N\to\infty}(\hat{\rho}-\rho) = -\left[\sigma^2 C(\rho,T) + \psi_{CT}\right]\left[\sigma^2 D(\rho,T) + \psi_{DT}\right]^{-1}, \tag{41}$$

where $C(\rho,T)$ and $D(\rho,T)$ are defined in (14) and (15),

$$\psi_{CT} = -trace\left\{\sum_{t=1}^{T} Z_{\theta t-1}^{\tau}\theta_t^{\tau\prime} M_\delta\right\}, \tag{42}$$

$$\psi_{DT} = trace\left\{\sum_{t=1}^{T} Z_{\theta t-1}^{\tau} Z_{\theta t-1}^{\tau\prime}\right\}, \tag{43}$$

and where $Z_{\theta t} = \sum_{j=0}^{\infty}\rho^j\theta_{t-j}$ and $\tilde{Z}_{\theta t}^{\tau} = Z_{\theta t} - a_T^{Z_\theta} - b_T^{Z_\theta}t$ is detrended $Z_{\theta t}$ so is θ_t^{τ}. In the single factor ($K=1$) case, the inconsistency (41) has the following asymptotic representation as $T\to\infty$

$$plim_{N\to\infty}(\hat{\rho}-\rho) = -2\frac{1+\rho}{T} - \frac{\sigma_\theta^2 m_\delta^2}{\sigma^2+\sigma_\theta^2 m_\delta^2}\left[\frac{2\rho}{T} + (Eh_{\theta T} - h_{\theta T})\right] + o_p\left(T^{-1}\right), \tag{44}$$

where $h_{\theta T} = \psi_{CT}/\psi_{DT} = \sum_{t=1}^{T} z_{\theta t-1}^{\tau}\theta_t^{\tau} / \sum_{t=1}^{T}\left(z_{\theta t-1}^{\tau}\right)^2$ is the centred least squares estimate of the slope coefficient in a regression of $z_{\theta t}^{\tau}$ on $z_{\theta t-1}^{\tau}$, and where $E\left(h_{\theta T}\right) = -2\frac{1+2\rho}{T} + o\left(T^{-1}\right)$.

The unit root case for model M2-CSD is handled in a similar way. As in the M1-CSD.model, direct calculation is needed because it is no longer possible to extract the unit root case by taking the limit as $\rho\to 1$, in view of the randomness of the limit functions (42) and (44). The inconsistency of $\hat{\rho}$ for the case of unit root is given by

$$\begin{aligned}\text{plim}_{N\to\infty}\left(\hat{\rho}-1\right) &= \frac{\text{plim}_{N\to\infty}\frac{1}{N}C_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}D_{NT}^{C}} \\ &= -\frac{7.5}{T} - \frac{\mu_\delta^2\sigma_\theta^2}{\sigma^2+\mu_\delta^2\sigma_\theta^2}\left[\frac{3}{T} + h_{\theta T}\right] + o_p\left(T^{-1}\right)\end{aligned}$$

5 Bias Reduction and Simulations

5.1 Cross Section Independence

Under cross section independence, bias correction is straightforward especially when N is moderately large, regardless of the value of T. First, consider the bias correction strategy when there are no exogeneous variables. An unbiased estimator can be obtained through inversion of the mean function, i.e.,

$$\hat{\rho}_{\text{MUE}} = m^{-1}\left(\hat{\rho}\right).$$

where m^{-1} is the inverse of the function G for the fixed effects case and H for the case of a model with incidental trends. This estimator can be obtained by direct numerical calculation and can be called a "mean unbiased estimator". Simulations indicate the function m is one-to-one. End corrections can be implemented at unity, so that in effect

$$\hat{\rho}_{\text{MUE}} = 1 \text{ if } \begin{cases} \hat{\rho} \geq 1 - 3/T & \text{fixed effects case} \\ \hat{\rho} \geq 1 - 7.5/T & \text{linear trend case} \end{cases}$$

When there are exogenous regressors, bias correction is still fairly straightforward. To fix ideas, consider the case of only two exogenous regressors which affect y_{it} in levels and in quasi-differences as in .

$$y_{it} = a_i + \rho y_{it-1} + \gamma_1 w_{it} + \gamma_2 w_{it-1} + \beta z_{it} + \varepsilon_{it}, \quad \gamma_2 = -\gamma_1 \rho.$$

Here w_{it} may be regarded as affecting y_{it} in levels (i.e. after removing the autoregressive transformation) while z_{it} affects y_{it} in the quasi-difference form $y_{it} - \rho y_{it-1}$. As discussed earlier (c.f. (28)), the estimate $\hat{\gamma}_1$ does not suffer from asymptotic bias, while the biases of $\hat{\beta}$ and $\hat{\rho}$ depend on the true values of β and ρ. To separate the bias of $\hat{\rho}$ from β, we run a regression of y_{it} on $\{y_{it-1}, w_{it}, w_{it-1}\}$ with fixed effects, i.e.,

$$y_{it} = \hat{b}_i + \hat{\rho}_{-z} y_{it-1} + \hat{\gamma}_1 w_{it} + \hat{\gamma}_2 w_{it-1} + \hat{v}_{it},$$

The bias of the estimator $\hat{\rho}_{-z}$ is given by the functions G and H for fixed effects and for linear trends, respectively. Since $\text{plim}_{N\to\infty} \left(\hat{\rho}_{-z} - \rho \right) = m(\rho, T)$, asymptotically mean unbiased estimators can be defined as

$$\hat{\rho}_{\text{MUE}} = m^{-1}\left(\hat{\rho}_{-z} \right), \quad \hat{\gamma}_{2,\text{MUE}} = \hat{\gamma}_2 + \hat{\gamma}_1 \left(\hat{\rho}_{-z} - \hat{\rho}_{\text{MUE}} \right),$$

using (28). A bias corrected estimator of β can be obtained by running the following regression

$$y_{it} - \hat{\rho}_{\text{MUE}} y_{it-1} - \hat{\gamma}_1 w_{it} - \hat{\gamma}_{2,\text{MUE}} w_{it-1} = b_i + \beta z_{it} + \varepsilon_{it}$$

The panel least squares estimator in this regression is asymptotically mean unbiased since the asymptotic bias of $\hat{\rho}$ and $\hat{\gamma}_2$ has been removed.

5.2 Cross Section Dependence

We distinguish two general types of panel data. For micro panel data such as the PSID, the number of factors as well as the number of cross sectional units will often be large while the number of time periods is small. As shown earlier, when the factors are independent and the number of factors K is large, the randomness in the bias arising from cross section dependence is attenuated and the bias is similar to that which applies under cross section independence. In such cases, common time effects or time dummies is usually recommended and this helps to reduce the efficiciency loss arising from cross section dependence (Phillips and Sul, 2003).

In contrast, for aggregated panels like regional income and consumption data, the time dimension may be reasonably long but there may only be one or two common factors. As we have seen, in such cases the bias is random and depends on the unknown common factors, and pooled OLS has high variability as well as bias. The practical issue is to reduce bias and variability in estimation. One approach is to construct a feasible generalized least squares (FGLS) estimator, which can be accomplished either by using the iterative method of moments procedure in Phillips and Sul (2003) or by using the sample covariance matrix of the residuals $\hat{u}_{it} = \tilde{y}_{it} - \hat{\rho}^{\lambda}_{\text{MUE}}\tilde{y}_{it-1}$ where '~' stands for demeaned or detrended y_{it} and $\hat{\rho}^{\lambda}_{\text{MUE}}$ is defined below.

The properties of FGLS depend on the first stage estimator and if this estimator is inconsistent (like panel OLS), then so is FGLS. The mean unbiased estimator (based on the bias formula that applies under cross section independence) is also inconsistent under cross section dependence. Its bias for the case of fixed effects and a single common factor has asymptotic expansion given by

$$\text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{MUE}} - \rho\right) = -\left[\frac{2\rho}{T} + \left(Eg_{\theta T} - g_{\theta T}\right)\right]\frac{m_{\delta}^2\sigma_{\theta}^2}{\sigma^2 + m_{\delta}^2\sigma_{\theta}^2} + o_p\left(T^{-1}\right), \tag{45}$$

which is small for large T. The use of common time effects or time dummies in the regression can be shown to reduce this bias. That is, if the regression model is augmented as

$$y_{it} = a_i + \lambda_t + \rho y_{it-1} + u_{it},$$

and estimated by pooled OLS with a mean correction based on the cross section independent case (giving the estimate $\hat{\rho}^{\lambda}_{\text{MUE}}$), then the asymptotic bias of $\hat{\rho}^{\lambda}_{\text{MUE}}$ has the following expansion

$$\text{plim}_{N\to\infty}\left(\hat{\rho}^{\lambda}_{\text{MUE}} - \rho\right) = -\left[\frac{2\rho}{T} + \left(Eg_{\theta T} - g_{\theta T}\right)\right]\frac{\left(m_{\delta}^2 - \bar{\delta}^2\right)\sigma_{\theta}^2}{\sigma^2 + \left(m_{\delta}^2 - \bar{\delta}^2\right)\sigma_{\theta}^2} + o_p\left(T^{-1}\right),$$

where $\bar{\delta} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\delta_i$. Since

$$\frac{m_{\delta}^2\sigma_{\theta}^2}{\sigma^2 + m_{\delta}^2\sigma_{\theta}^2} - \frac{\left(m_{\delta}^2 - \bar{\delta}^2\right)\sigma_{\theta}^2}{\sigma^2 + \left(m_{\delta}^2 - \bar{\delta}^2\right)\sigma_{\theta}^2} = \frac{\bar{\delta}^2\sigma^2\sigma_{\theta}^2}{\left(\sigma^2 + m_{\delta}^2\sigma_{\theta}^2\right)\left(\sigma^2 + \left(m_{\delta}^2 - \bar{\delta}^2\right)\sigma_{\theta}^2\right)} \geq 0,$$

with equality holding when $\bar{\delta} = 0$, the mean corrected estimator with common time effects reduces bias and variation.

An alternative option is to attempt to eliminate the factor loading coefficients δ_i in the regression. One approach that has recently been considered in the literature is to project out the factor θ_t by including cross sectional averages of y_{it} and y_{it-1} in the regression (Pesaran, 2002). This can be accomplished by rewriting the model M1 in the following augmented regression form

$$\begin{aligned} y_{it} &= a_i^{+} + \rho y_{it-1} + c_{1i}\left(\frac{1}{N}\sum_{i=1}^{N} y_{it}\right) + c_{2i}\left(\frac{1}{N}\sum_{i=1}^{N} y_{it-1}\right) + \varepsilon_{it} \\ c_{1i} &= \frac{\delta_i}{\bar{\delta}}, \quad c_{2i} = -\rho\frac{\delta_i}{\bar{\delta}}, \quad a_i^{+} = a_i - \frac{\delta_i}{\bar{\delta}}\left(\bar{a} + \bar{\varepsilon}_{.t}\right) \end{aligned} \tag{46}$$

Multiple factors can be treated in a similar way. Let the cross section observations be classified into groups $\{A_k : k = 1, ..., K\}$ with counts $N_k = \#\{i \in A_k\}$ in each group and suppose $N_k/N \to r_k \neq 0$ for all k as $N \to \infty$. Further, let $\bar{\delta}_{A_k} = N_k^{-1}\sum_{i\in A_k}\delta_i$,define $D_K = \left[\bar{\delta}_{A_1}, ..., \bar{\delta}_{A_K}\right]$ and assume D_K is of full rank K. Set

$$\begin{aligned} \bar{y}_{A_k t} &= N_k^{-1}\sum_{i\in A_k} y_{it}, \quad \bar{y}_{Kt} = \left(\bar{y}_{A_1 t}, ..., \bar{y}_{A_K t}\right)', \\ \bar{a}_{A_k} &= N_k^{-1}\sum_{i\in A_k} a_i, \quad \bar{a}_K = \left(\bar{a}_{A_1}, ..., \bar{a}_{A_K}\right)', \\ \bar{\varepsilon}_{A_1 t} &= N_k^{-1}\sum_{i\in A_k} \varepsilon_{it}, \quad \bar{\varepsilon}_{Kt} = \left(\bar{\varepsilon}_{A_1 t}, ..., \bar{\varepsilon}_{A_K t}\right)'. \end{aligned}$$

Then,

$$\bar{y}_{A_k t} = \bar{a}_{A_k} + \rho \bar{y}_{A_k t-1} + \bar{\delta}'_{A_k} \theta_t + \bar{\varepsilon}_{A_k t},$$

and

$$\theta_t = D_K^{-1} \left(\bar{y}_{Kt} - \bar{a}_K - \rho \bar{y}_{Kt-1} - \bar{\varepsilon}_{Kt} \right)$$

In this case, the augmented regression has the form

$$y_{it} = a_i^+ + \rho y_{it-1} + \delta_i' D_K^{-1} \left(\bar{y}_{Kt} - \rho \bar{y}_{Kt-1} \right) + \varepsilon_{it}, \tag{47}$$

with

$$a_i^+ = a_i - \delta_i' D_K^{-1} \bar{a}_K - \delta_i' D_K^{-1} \bar{\varepsilon}_{Kt} = a_i - D_K^{-1} \bar{a}_K + o_p(1)$$

as $N \to \infty$. Again, (47) may be estimated in restricted or unrestricted form and the panel estimate of ρ may be adjusted for bias just as in the cross section independent case.[2]

5.3 Monte Carlo Studies

We consider two data generating processes (DGP)s. The first DGP is for the case of exogenous variables and is given by

$$y_{it} = \rho y_{it} + \beta z_{it} + \varepsilon_{it},$$

fitting both fixed effects and incidental trends. We consider various values of ρ but report the case of $\rho = 0.9$, which is representative, to save the space[3]. We set $\beta = 1$, and generate ε_{it} as *iid* $N(0,1)$. Table 2 reports the finite sample performance of pooled least squares and mean unbiased estimators as described in subsection 5.1. The results in columns B and D of the Table show that the bias of $\hat{\rho}_{\mathrm{MUE}}$ and $\hat{\beta}_{\mathrm{MUE}}$ is small in both cases and these estimates provide a clear improvement over panel least squares.

The second DGP covers the case of cross section dependence given by

$$y_{it} = \rho y_{it-1} + \delta_i \theta_t + \varepsilon_{it}$$

We set $\delta_i \equiv U[1,4]$, $\varepsilon_{it} \equiv iid\ N(0,1)$ and $\theta_t \equiv iid\ N(0,1)$. We consider six estimators: the least squares dummy variable (LSDV) estimator $\hat{\rho}$ (A); LSDV with common time effects $\hat{\rho}^{\lambda}$ (B); panel feasible generalized mean unbiased estimator (FGMUE) based on the residual covariance matrix calculated from $\hat{\rho}$ (C); panel FGMUE based on the residual covariance matrix calculated from $\hat{\rho}_{\mathrm{MUE}}$ (D); panel FGMUE

[2] Pesaran (2002) calls the regression in (46) a 'common correlated regression (CCR)'. Unfortunately, the bias of the CCR estimator cannot be reduced in a simple way by utilizing a mean bias function. To see this, define $\bar{y}_t = \left(\tilde{y}'_{\cdot t}, \tilde{y}'_{\cdot t-1} \right)'$, $M_y = y_t \left(\mathbf{y}_t' \mathbf{y}_t \right)^{-1} y_t'$, and $Q_y = I - M_y$ where $\tilde{y}_{\cdot t} = N^{-1} \sum_{i=1}^N \left(y_{it} - T^{-1} \sum_{t=1}^T y_{it} \right)$ and $y_{\cdot t-1} = N^{-1} \sum_{i=1}^N \left(y_{it-1} - T^{-1} \sum_{t=1}^T y_{it-1} \right)$. The asymptotic bias of the common correlated estimator $\hat{\rho}_{CCR}$ in (46) is given by $\text{plim}_{N\to\infty} \left(\hat{\rho}_{CCR} - \rho \right) = \left\{ \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \tilde{y}_i' \mathbf{Q}_y \tilde{y}_i \right\}^{-1} \left\{ \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \tilde{y}_i' \mathbf{Q}_y \tilde{\varepsilon}_i \right\}$. Note that the numerator term becomes $-\sigma_\varepsilon^2 A(\rho, T)$, which is the same as in the case of exogenous regressors. However, the denominator term contains an additional term. In particular, $\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \tilde{y}_i' Q_y \tilde{y}_i = \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \tilde{y}_i' \tilde{y}_i - \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \tilde{y}_i' M_y \tilde{y}_i = \sigma_\varepsilon^2 B(\rho, T) + \psi_{BT} - \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \tilde{y}_i' M_y \tilde{y}_i \neq \sigma_\varepsilon^2 B(\rho, T)$, where ψ_{BT} was defined in (38). The numerator term in (**??**) is the same as that without cross section dependence. This is because the ψ_{AT} term vanishes by virtue of the inclusion of cross sectional averages of $\bar{y}_t$ and $\bar{y}_{t-1}$ in (46). At the same time, the inclusion of $\bar{y}_t$ and $\bar{y}_{t-1}$ means that the denominator includes additional terms, thereby making bias correction more difficult.

[3] Full Excel formated tables are available requested upon authors.

based on the residual covariance matrix calculated from $\hat{\rho}^{\lambda}_{\mathrm{MUE}}$ (E); and the mean unbiased estimator after eliminating the factor loading coefficients through Pesaran's correlated common method (F). The residual covariance matrices for (C), (D) and (E) are estimated using iterative method of moments (Phillips and Sul, 2003). We set $T = 25, 50, 100, 200$ and $N = 10, 25, 50, 100$, which covers the most typical data dimensions in empirical studies with macro panel data.

Table 3 shows the results for the fixed effects and incidental trend cases, respectively. The mean unbiased estimator (E) shows the best performance both in terms of absolute bias and mean square error ratio. Meanwhile, the mean unbiased estimator based on Pesaran's estimator (F) is better than LSDV with common time effects but is inferior in comparison to other FGLS estimators.

6 Conclusion

The results of the present paper focus on dynamic bias in pooled panel regression, showing that the problem is particularly serious when trends are extracted and is pervasive in a range of cases that are relevant in applications. When cross section error dependence is present, problems of bias are confounded with increases in dispersion, which manifests itself even in the limit theory as $N \to \infty$ through a random probability limit.

The specific nature of the panel can play an important role in the bias and the possibility of bias correction. For micro panels, it is natural to assume that there are a number of common factors in the panel. In this case, the biases in pooled panel regressions can be corrected by utilizing mean unbiased functions in a straightforward way. In dynamic panel regressions with such micro panels, the bias correction methods differ depending on the way exogenous variables figure in the model. The original empirical study of the demand for natural gas by Balestra and Nerlove (1966) illustrates this point. Balestra and Nerlove fitted the following panel regression equation to estimate the demand for natual gas.

$$G_{it} = \alpha_i + \rho G_{it-1} + \beta p_{it} + \gamma_1 \Delta M_{it} + \gamma_2 M_{it-1} + \gamma_3 \Delta Y_{it} + \gamma_4 Y_{it-1} + u_{it}$$

where G_{it}, p_{it}, M_{it}, and Y_{it} represent quantity demanded for gas, the relative price of gas, population and per capita income at time t and for the i'th unit, respectively. This model fits the framework of model M3. The authors modelled the exogenous variables in such a way that population and per capita income affected G_{it} in levels but the relative price of gas affected G_{it} in first differences. As a result, the reported LSDV estimates of β are biased but those of γ_1 and γ_3 are unbiased.

For macro panel data, modelling cross section dependence is important. As a second illustration, we consider the study by Frankel and Rose (1996) who used a panel of 45 annual observations over 150 countries to examine the half life of deviations from purchasing power parity (PPP) by running the following panel regression equation[4]

$$q_{it} = a_i + \rho q_{it-1} + u_{it}, \tag{48}$$

where q_{it} is the logarithm of the real exchange rate. From the point estimate $\hat{\rho} = 0.88$, they calculated the half-life of the PPP deviation to be $\ln(0.5)/\ln(0.88) = 5.4$ years. As discussed, such estimates are biased and can be very inefficient in the presence of cross section dependence. To illustrate the empirical

[4] See Frankel and Rose (1996, table 3 p. 219). Similar results to those reported were obtained in an equation with time-specific intercepts.

effects of taking bias and cross section dependence into account in estimation, we reestimated the half-life of the PPP deviation from the same model (48) using an updated data set[5] involving 51 annual observations from 21 OECD countries. Table 4 displays the estimation results for all the estimates discussed earlier in the paper. The LSDV point estimate gives a half-life for PPP deviations of 3.4 years, whereas feasible generalized least squares estimates that adjust for bias and make allowance for potential cross section dependence in long run PPP deviations are more than twice as great. These empirical findings confirm that adjustments for dynamic panel bias and allowance for cross section dependence can have a major impact on estimates of key parameters like the half-life of PPP deviations.

Table 4: Estimation of Half Life of the PPP Deviation

	(A)	(B)	(C)	(D)	(E)	(F)
Coefficient Estimates	0.817	0.858	0.913	0.917	0.919	0.857
Half-Life Estimates	3.419	4.536	7.615	8.000	8.206	4.492

Legend: (A) = LSDV; (B) = LSDV with common time effect; (C) = FGMUE based on residual variance of LSDV; (D) = FGMUE based on residual variance of MUE with fixed effects; (E) = FGMUE based on residual variance of MUE with common time effects; (F) = MUE with Pesaran's correlated common estimator.

[5] Data for 21 countries over the period 1948-1998 was taken from the International Financial Statistics. The series involved annual price indices for each country and real exchange rates calculated from the individual national price indices and the end of the period spot exchange rates. The US dollar was chosen as the numeraire currency.

References

[1] Altonji, J., T. E. Elder and C. R. Taber (2002). “Selection on observed and unobserved variables: assessing the effectiveness of Catholic Schools” mimeographed, Yale University.

[2] Arrelano, M. (2003). *Panel Data Econometrics.* Oxford: Oxford University Press.

[3] Arrelano, M. and B. Honoré (2000) Panel data models: Some recent developments. in J. Heckman and E. Leamer, eds., *Handbook of Econometrics.* Amsterdam: North-Holland

[4] Bai, J. and S. Ng (2002). “Determining the number of factors in approximate factor models,” *Econometrica*, 191-221

[5] Balestra, P. and M. Nerlove (1966). “Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas,” *Econometrica*, 585-612.

[6] Baltagi, B. (2001). *Econometric Analysis of Panel Data*. New York: Wiley.

[7] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), “The generalized dynamic factor model: Identification and estimation,” *Review of Economics and Statistics*, 540–554.

[8] Frankel, J. and A.K. Rose (1996). “A panel projection on purchasing power parity: Mean reversion within and between countries,” *Journal of International Economics*, 209–224.

[9] Hahn, J., J. Hausman and G. Kuersteiner (2001), “Bias corrected instrumnetal variables estimation for dynamic panel models with fixed effects,” MIT Working Paper

[10] Hahn, J. and G. Kuersteiner (2002). “Asymptotically unbiased inference for a dynamic panel model with fixed effects,” *Econometrica*, 1639-1657.

[11] Harris, R.D.F. and E. Tzavalis (1999). “Inference for unit roots in dynamic panels where the time dimension is fixed,” *Journal of Econometrics* 201–226.

[12] Hsiao, C. (2003). *Analysis of Panel Data*. 2’nd Edition. Cambridge: Cambridge University Press.

[13] Kendall, M. G. (1954). “Note on the bias in the estimation of autocorrelation,” *Biometrika*, 403-404.

[14] Kiviet, J. F. (1995). “On bias, inconsistency and efficiency of some estimators in dynamic panel data models”. *Journal of Econometrics*, 68, 53-78.

[15] Kruiniger, H. (2000). “GMM estimation of dynamic panel data models with persistent data,” mimeo, Queen Mary, Unversity of London.

[16] Moon, H.R. and B. Perron, “Testing for a unit root in panels with dynamic factors,” mimeo, University of Southern California.

[17] Moon, H.R. and P.C.B. Phillips (1999) : Maximum likelihood estimation in panels with incidental trends, *Oxford Bulletin of Economics and Statistics, 61, 771–748.*

[18] Moon, H.R. and P.C.B. Phillips (2000) : Estimation of autoregressive roots near unity using panel data, *Econometric Theory*, 16, 927–997.

[19] Moon, H.R. and P.C.B. Phillips (2004) : GMM estimation of autoregressive roots near unity with panel data, *Econometrica*, 72, 467-522.

[20] Neyman, J and E. Scott (1948): Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1–32.

[21] Nickell, S. (1981): "Biases in dynamic models with fixed effects", *Econometrica*, 49, 1417–1426.

[22] Orcutt, G. H. (1948). "A study of the autoregressive nature of the times series used for Tinbergen's model of the economic system of the United States," *Journal of the Royal Statistical Society*, Series B, 1-45.

[23] Orcutt, G. H. and H. S. Winokur (1969). "First order autoregression: inference, estimation and prediction". *Econometrica*, 37, 1-14.

[24] Pesaran, H. (2002) "Estimation and inference in large heterogenous panels with cross section dependence" University of Cambridge DAE Working Paper No.0305.

[25] Phillips, P.C.B. and V. Solo (1992). "Asymptotics for linear processes," *Annals of Statistics*, 971–1001.

[26] Phillips, P.C.B. and D. Sul (2001). "Edgeworth expansions in the first order autoregressive model with trend" mimeo, University of Auckland.

[27] Phillips, P.C.B. and D. Sul (2003). "Dynamic panel estimation and homogeneity testing under cross section dependence," *The Econometrics Journal,* 6, 217-259.

7 Appendix

7.1 Proofs of Propositions

Proof of Proposition 2 Write the model in components form as $y_{it} = \alpha_i + \beta_i t + x_{it}$, where $x_{it} = \rho x_{it-1} + u_{it}$ for $t = 1, ..., T$. Then the panel least squares estimate of ρ is $\hat{\rho} = C_{NT}^x / D_{NT}^x$, where

$$C_{NT}^x = \sum_{i=1}^N \left[\sum_{t=1}^T (x_{it} - x_{i\cdot})(x_{it-1} - x_{i\cdot-1}) - \frac{\sum_{t=1}^T [(t-\bar{t})(x_{it} - x_{i\cdot})] \sum_{t=1}^T [(t-\bar{t})(x_{it-1} - x_{i\cdot-1})]}{\sum_{t=1}^T (t-\bar{t})^2} \right],$$

$$D_{NT}^x = \sum_{i=1}^N \left[\sum_{t=1}^T (x_{it-1} - x_{i\cdot-1})^2 - \frac{\left[\sum_{t=1}^T (t-\bar{t})(x_{it-1} - x_{i\cdot-1})\right]^2}{\sum_{t=1}^T (t-\bar{t})^2} \right],$$

using the sum notation $w_{i\cdot} = T^{-1} \sum_{t=1}^T w_{it}$, $w_{i\cdot-1} = T^{-1} \sum_{t=1}^T w_{it-1}$. Expanding the cross product moments in these expressions and standardizing by N^{-1}, probability limits are taken as $N \to \infty$ with T fixed. A typical term is evaluated in the following manner using a law of large numbers for heterogeneous sequences. First note that

$$\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N x_{it} x_{is} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N E[x_{it} x_{is}] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \frac{\rho^{|t-s|}}{1-\rho^2} = \sigma^2 \frac{\rho^{|t-s|}}{1-\rho^2}.$$

Then we have

$$\begin{aligned} \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T x_{it} \left(\sum_{s=1}^T s x_{is} \right) &= \sum_{t,s=1}^T s \left(\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N x_{it} x_{is} \right) = \frac{\sigma^2}{1-\rho^2} \sum_{t,s=1}^T s \rho^{|t-s|} \\ &= E\left\{ \sum_{t=1}^T x_t \sum_{s=1}^T s x_s \right\}, \end{aligned}$$

thereby writing the limit as a moment of a homogeneous (across i) process x_t which follows the stationary autoregression $x_t = \rho x_{t-1} + \varepsilon_t$ where ε_t is $iid\left(0, \sigma^2\right)$.

Let $C_{NT} = C_{NT}^x - \rho D_{NT}^x$. Using this approach, we find after some lengthy but routine derivations using the lemmas in Section 7.2 that the inconsistency as $N \to \infty$ with T fixed has the form

$$\text{p}\lim_{N\to\infty} (\hat{\rho} - \rho) = \frac{\text{plim}_{N\to\infty} \frac{1}{N} C_{NT}}{\text{plim}_{N\to\infty} \frac{1}{N} D_{NT}} = -\frac{C(\rho, T)}{D(\rho, T)}, \tag{49}$$

where

$$C(\rho, T) = -\frac{1}{T-1} \frac{2}{1-\rho} \left[(T-1) - \frac{2}{1-\rho} C_1 \right], \tag{50}$$

$$D(\rho, T) = \frac{T-2}{1-\rho^2} \left[1 - \frac{1}{T-2} \frac{4\rho}{1-\rho} D_1 \right], \tag{51}$$

with

$$\begin{aligned} C_1 &= 1 - \frac{1}{T+1} \left(1 + \frac{1-\rho^3}{(1-\rho)^3} \frac{1}{T} \right) + \left(\frac{1}{2} + \frac{1}{T+1} \left[\frac{1+2\rho}{1-\rho} + \frac{1-\rho^3}{(1-\rho)^3} \frac{1}{T} \right] \right) \rho^T, \\ D_1 &= 1 - \frac{1}{T+1} \frac{2}{1-\rho} \left\{ 1 + \frac{1}{T-1} \left[1 - \frac{1-\rho^3}{T(1-\rho)^3} \left(1-\rho^T\right) + \left(\frac{3\rho}{1-\rho} + \frac{T+3}{2} \right) \rho^T \right] \right\}. \end{aligned}$$

Upon further algebraic reduction the rational function limit (49) has the following explicit form in terms of constituent polynomials in ρ and T :

$$H(\rho,T)=-\frac{C(\rho,T)}{D(\rho,T)}=-2\rho\frac{a_1T^3+a_2T^2+a_3T+a_4}{b_1T^4+b_2T^3+b_3T^2+b_4T+b_5}, \tag{52}$$

where

$$\begin{aligned}
a_0&=-(1+\rho)(1-\rho)^3, & a_1&=-(1-\rho)a_0,\\
a_2&=a_0(2+\rho^T), & a_3&=-a_1-3\rho^T(1-\rho^2)^2,\\
a_4&=2(1+\rho)(1-\rho^3)(1-\rho^T), & b_1&=\rho(1-\rho)^4,\\
b_2&=2\rho a_0, & b_3&=(\rho-1)^2(12\rho^2-\rho(\rho+1)^2+4\rho^{T+2}),\\
b_4&=(1-\rho^2)((1-\rho)^2 2\rho+12\rho^{2+T}) \quad \text{and} & b_5&=8\rho^2(\rho^2+\rho+1)(\rho^T-1).
\end{aligned}$$

Adjusting (50)and (51) for dominant terms yields the following approximant:

$$\text{plim}_{N\to\infty}(\hat{\rho}-\rho)=-2\frac{1+\rho}{T-2}+O(T^{-2}).$$

For the first few values of T, the exact limit formulae work out as follows:

$$\text{plim}_{N\to\infty}(\hat{\rho}-\rho)=\begin{cases}-\frac{1}{2}\frac{\rho^2-3\rho-4}{\rho-3} & \text{for } T=3\\ -\frac{1}{2}\frac{\rho^3-6\rho-5}{\rho^2-5} & \text{for } T=4\\ -\frac{1}{2}\frac{2\rho^4+2\rho^3-5\rho^2-17\rho-12}{2\rho^3+2\rho^2-3\rho-15} & \text{for } T=5\end{cases}$$

The approximate formula, $-2(1+\rho)/(T-2)$ is usually smaller (in absolute value) than the exact formula when ρ is larger than (around) 0.7.

Proof of Proposition 3 From (19), $\text{plim}_{N\to\infty}(\hat{\rho}-\rho)=\left\{\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{y}_{-1}\right\}^{-1}\left\{\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{\varepsilon}\right\}$, and by virtue of exogeneity

$$\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{\varepsilon} &= \text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'\tilde{\varepsilon}-\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'\tilde{Z}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'\tilde{\varepsilon}\\
&= \text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'\tilde{\varepsilon}=-\sigma^2A(\rho,T),
\end{aligned}$$

as given in (3). Next, when $|\rho|<1$, we have

$$\tilde{y}_{it}=\sum_{j=0}^{\infty}\rho^j\tilde{Z}_{it-j}\beta+\sum_{j=0}^{\infty}\rho^j\tilde{\varepsilon}_{it}:=\tilde{Z}_{\rho it}\beta+\tilde{y}_{it}^0,$$

and, using the stacked notation $\tilde{y}=\tilde{Z}_\rho\beta+\tilde{y}^0$ and its lagged variant, we have as in (25)

$$\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}'Q_{\tilde{Z}}\tilde{y}_{-1} &= \beta'\left[\text{plim}_{N\to\infty}\frac{1}{N}\tilde{Z}_{\rho,-1}'Q_{\tilde{Z}}\tilde{Z}_{\rho,-1}\right]\beta+\text{plim}_{N\to\infty}\frac{1}{N}\tilde{y}_{-1}^{0\prime}\tilde{y}_{-1}^0\\
&= \beta'\left[\text{plim}_{N\to\infty}\frac{1}{N}\tilde{Z}_{\rho,-1}'Q_{\tilde{Z}}\tilde{Z}_{\rho,-1}\right]\beta+\sigma^2B(\rho,T),
\end{aligned}$$

where $B(\rho,T)$ is given in (4). It follows that

$$\text{plim}_{N\to\infty}(\hat{\rho}-\rho)=-\frac{\sigma^2A(\rho,T)}{\sigma^2B(\rho,T)+\beta'\left[\text{plim}_{N\to\infty}\frac{1}{N}\tilde{Z}_{\rho,-1}'Q_{\tilde{Z}}\tilde{Z}_{\rho,-1}\right]\beta}, \tag{53}$$

as given in (21). Results (22) and (23) follow in a similar way.

When $\rho = 1$, we have

$$\lim_{\rho\to 1} A(\rho, T) = A(T) = \frac{(T-1)}{2}, \quad \lim_{\rho\to 1} B(\rho, T) = B(T) = \frac{(T-1)(T+1)}{6}, \tag{54}$$

so that (53) becomes

$$\text{plim}_{N\to\infty} (\hat{\rho} - \rho) = -\frac{\sigma^2 A(T)}{\sigma^2 B(T) + \beta' \left[\text{plim}_{N\to\infty} \frac{1}{N} \tilde{Z}'_{1,-1} Q_{\tilde{Z}} \tilde{Z}_{1,-1}\right] \beta}, \tag{55}$$

in which $\tilde{Z}_{1,-1} = \left(\tilde{Z}'_{1,0}, ..., \tilde{Z}'_{1,T-1}\right)'$ with $\tilde{Z}_{1,t} = \left(\tilde{Z}^1_{1,t}, ..., \tilde{Z}^N_t\right)'$ and $\tilde{Z}^i_t = \sum_{j=0}^t \tilde{Z}_{it-j}$. The corresponding result in the incidental trends case is

$$\text{plim}_{N\to\infty} (\hat{\rho} - \rho) = -\frac{\sigma^2 C(T)}{\sigma^2 D(T) + \beta' \left[\text{plim}_{N\to\infty} \frac{1}{N} \tilde{Z}'_{1,-1} Q_{\tilde{Z}} \tilde{Z}_{1,-1}\right] \beta}, \tag{56}$$

where

$$\lim_{\rho\to 1} C(\rho, T) = C(T) = \frac{1}{2}(T-2), \quad \lim_{\rho\to 1} D(\rho, T) = D(T) = \frac{1}{15}\left(T^2 - 4\right), \tag{57}$$

as in (**??**) and (**??**). Formula (23) for the inconsistency of $\hat{\beta}$ continues to apply in the unit root case upon appropriate substitution of result (55) or (56).

Proof of Proposition 4 It is convenient here to use sequential asymptotics with $N \to \infty$ followed by $T \to \infty$. Write the panel least squares estimate under cross sectional dependence as

$$\hat{\rho} - \rho = \frac{A^C_{NT}}{B^C_{NT}}. \tag{58}$$

In the one factor ($K = 1$) case, the model is given by

$$y_{it} = \alpha_i + x_{it}, \; x_{it} = \rho x_{it-1} + u_{it}, \quad u_{it} = \delta_i \theta_t + \varepsilon_{it}, \tag{59}$$

and then

$$x_{it} = \delta_i \sum_{j=0}^{\infty} \rho^j \theta_{t-j} + \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j} := \delta_i z_{\theta t} + x^{\epsilon}_{it}, \text{ say.} \tag{60}$$

Since $y_{it} - \frac{1}{T}\sum y_{it} = x_{it} - \frac{1}{T}\sum x_{it}$, we have

$$\begin{aligned}
& \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it-1} - y_{i\cdot -1})(u_{it} - u_{i\cdot}) = \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it-1} - x_{i\cdot -1})(u_{it} - u_{i\cdot}) \\
= \; & \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} x_{it-1} u_{it} - \frac{1}{T} \sum_{t=1}^{T} x_{it-1} \sum_{t=1}^{T} u_{it}\right] \\
= \; & \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} \left(\delta_i z_{\theta t-1} + x^{\epsilon}_{it-1}\right)(\delta_i \theta_t + \varepsilon_{it}) - \frac{1}{T} \sum_{t=1}^{T} \left(\delta_i z_{\theta t-1} + x^{\epsilon}_{it-1}\right) \sum_{t=1}^{T} (\delta_i \theta_t + \varepsilon_{it})\right] \\
= \; & -\sigma^2 A(\rho, T) + \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i^2 \left[\sum_{t=1}^{T} z_{\theta t-1} \theta_t - \frac{1}{T} \sum_{t=1}^{T} z_{\theta t-1} \sum_{s=1}^{T} \theta_s\right],
\end{aligned} \tag{61}$$

where

$$\sigma^2 A(\rho, T) = \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \sum_{t=1}^{T} (\sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1}) \sum_{s=1}^{T} \varepsilon_{is}\right] = \frac{\sigma^2}{T} \frac{1}{1-\rho} \left[T - \frac{1-\rho^T}{1-\rho}\right],$$

as in (3). Using the fact that $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 = m_\delta^2$, (61) becomes

$$\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C} &= -\sigma^2 A(\rho,T) - m_\delta^2 \frac{1}{T}\sum_{t=1}^{T} z_{\theta t-1}\sum_{s=1}^{T}\theta_s + m_\delta^2\sum_{t=1}^{T} z_{\theta t-1}\theta_t \qquad (62)\\
&= -\sigma^2 A(\rho,T) + m_\delta^2\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)\left(\theta_t-\bar{\theta}\right). \qquad (63)
\end{aligned}$$

Dealing with the denominator in a similar fashion, we get

$$\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^{C} = \sigma^2 B(\rho,T) + m_\delta^2\left[\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2\right].$$

Note that

$$\begin{aligned}
\sigma^2 B(\rho,T) &= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\Big(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1}\Big)^2 - \frac{1}{T}\left(\sum_{t=1}^{T}\Big(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1}\Big)\right)^2\right]\\
&= \sigma^2\frac{T-1}{1-\rho^2}\left\{1-\frac{1}{T-1}\frac{2\rho}{1-\rho}\left[1-\frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\right\}. \qquad (64)
\end{aligned}$$

Combining the two results gives

$$\begin{aligned}
\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^{C}} &= \frac{-\sigma^2 A(\rho,T) + m_\delta^2\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)\left(\theta_t-\bar{\theta}\right)}{\sigma^2 B(\rho,T) + m_\delta^2\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2}\\
&= -\frac{\sigma^2 A(\rho,T)\left[\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2\right]^{-1} - m_\delta^2 g_{\theta T}}{\sigma^2 B(\rho,T)\left[\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2\right]^{-1} + m_\delta^2},
\end{aligned}$$

where

$$g_{\theta T} = \frac{\sum_{t=1}^{T} z_{\theta t-1}\theta_t - \frac{1}{T}\sum_{t=1}^{T} z_{\theta t-1}\sum_{s=1}^{T}\theta_s}{\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2} = \frac{\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)\left(\theta_t-\bar{\theta}\right)}{\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2},$$

which is the centred serial correlation coefficient of $z_{\theta t}$, viz., the centred least squares estimate of the slope coefficient in a regression of $z_{\theta t}$ on $z_{\theta t-1}$ and a constant. The density of $g_{\theta T}$ is studied in Phillips(1977) and Tanaka(1983). Its unconditional mean has a large T expansion given by

$$E\left(g_{\theta T}\right) = -\frac{1+3\rho}{T} + o\left(T^{-1}\right).$$

Letting $T\to\infty$ we have

$$\frac{1}{T}\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2 \to_p E\left(z_{\theta t}^2\right) = \frac{\sigma_\theta^2}{1-\rho^2},$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2 = \frac{\sigma_\theta^2}{1-\rho^2} + O_p\left(T^{-1/2}\right).$$

Hence,

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\left[\sum_{t=1}^{T}\left(z_{\theta t-1}-\bar{z}_{\theta-1}\right)^2\right] = T\left[m_\delta^2\frac{\sigma_\theta^2}{1-\rho^2} + O_p\left(T^{-1/2}\right)\right], \text{ as } T\to\infty. \qquad (65)$$

Taking limits as $N \to \infty$ followed by an expansion as $T \to \infty$, we have

$$\begin{aligned}
\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^{C}} &= -\frac{\sigma^2 A(\rho,T)\,T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}+o_p(1)\right]^{-1} - m_\delta^2 g_{\theta T}}{\sigma^2 B(\rho,T)\,T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}+o_p(1)\right]^{-1} + m_\delta^2} \\
&= -\frac{\sigma^2\frac{1}{1-\rho}T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}\right]^{-1} - m_\delta^2 g_{\theta T}}{\sigma^2\frac{1}{1-\rho^2}\left[\frac{\sigma_\theta^2}{1-\rho^2}\right]^{-1} + m_\delta^2} + o_p\left(T^{-1}\right) \\
&= -\frac{\frac{1+\rho}{T}\sigma^2 - \sigma_\theta^2 m_\delta^2 g_{\theta T}}{\sigma^2+\sigma_\theta^2 m_\delta^2} + o_p\left(T^{-1}\right) \\
&= -\frac{\frac{1+\rho}{T}\left(\sigma^2+\sigma_\theta^2 m_\delta^2\right) - \frac{1+\rho}{T}\sigma_\theta^2 m_\delta^2 - \sigma_\theta^2 m_\delta^2 g_{\theta T}}{\sigma^2+\sigma_\theta^2 m_\delta^2} + o_p\left(T^{-1}\right) \\
&= -\frac{\frac{1+\rho}{T}\left(\sigma^2+\sigma_\theta^2 m_\delta^2\right) + \frac{2\rho}{T}\sigma_\theta^2 m_\delta^2 - \sigma_\theta^2 m_\delta^2\left(g_{\theta T} - Eg_{\theta T}\right)}{\sigma^2+\sigma_\theta^2 m_\delta^2} + o_p\left(T^{-1}\right) \\
&= -\frac{1+\rho}{T} - \frac{2\rho}{T}\frac{\sigma_\theta^2 m_\delta^2}{\sigma^2+\sigma_\theta^2 m_\delta^2} + \frac{\sigma_\theta^2 m_\delta^2}{\sigma^2+\sigma_\theta^2 m_\delta^2}\left(g_{\theta T} - Eg_{\theta T}\right) + o_p\left(T^{-1}\right).
\end{aligned}$$

In the multi-factor case, we have $u_{it} = \delta_i'\theta_t + \varepsilon_{it}$ in (59) where θ_t is $iid\ (0, \Sigma_\theta)$ and Σ_θ is $K \times K$. Then, $Z_{\theta t} = \sum_{j=0}^{\infty}\rho^j\theta_{t-j}$, and $z_{i\theta t} := \delta_i' Z_{\theta t} = \sum_{j=0}^{\infty}\rho^j\delta_i'\theta_{t-j}$ is first order autoregressive and satisfies $z_{i\theta t} = \rho z_{i\theta t-1} + \theta_{it}$ where $\theta_{it} = \delta_i'\theta_t$ is $iid\ \left(0, \delta_i'\Sigma_\theta\delta_i\right)$. Then, in place of (60), we have

$$y_{it} = \alpha_i + x_{it}, \text{ with } x_{it} = z_{i\theta t} + x_{it}^{\epsilon}.$$

Proceeding as above, we obtain

$$\begin{aligned}
& \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i\cdot-1})(u_{it} - u_{i\cdot}) \\
= \ & \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(z_{i\theta t-1} + x_{it-1}^{\epsilon}\right)(\theta_{it}+\varepsilon_{it}) - \frac{1}{T}\sum_{t=1}^{T}\left(z_{i\theta t-1} + x_{it-1}^{\epsilon}\right)\sum_{t=1}^{T}(\theta_{it}+\varepsilon_{it})\right] \\
= \ & -\sigma^2 A(\rho,T) + \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(z_{i\theta t-1} - \bar{z}_{i\theta,-1}\right)\left(\theta_{it} - \bar{\theta}_i\right)\right] \\
= \ & -\sigma^2 A(\rho,T) + \text{trace}\left\{\sum_{t=1}^{T}\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)\left(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\right)' M_\delta\right\},
\end{aligned}$$

where $M_\delta = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i\delta_i'$ and $\boldsymbol{\theta}_t = (\theta_1, ..., \theta_k)$. In a similar manner, we find the following limit for the denominator

$$\text{plim}_{N\to\infty}\ \frac{1}{N}B_{NT}^{C} = \sigma^2 B(\rho,T) + \text{trace}\left\{\sum_{t=1}^{T}\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)' M_\delta\right\}.$$

It follows that

$$\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^{C}} = -\frac{\sigma^2 A(\rho,T) - \text{trace}\left\{\sum_{t=1}^{T}\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)\left(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\right)' M_\delta\right\}}{\sigma^2 B(\rho,T) + \text{trace}\left\{\sum_{t=1}^{T}\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)\left(Z_{\theta t-1} - \bar{Z}_{\theta,-1}\right)' M_\delta\right\}},$$

which gives the stated result.

Proof of Proposition 5 Define

$$\begin{aligned} x_{it-1}^{\tau} &= (x_{it-1} - x_{i\cdot-1})^2 - \frac{(t-\bar{t})(x_{it-1} - x_{i\cdot-1})}{\sum_{t=1}^{T}(t-\bar{t})^2}, \\ u_{it}^{\tau} &= (u_{it} - u_{i\cdot})^2 - \frac{(t-\bar{t})(u_{it} - u_{i\cdot})}{\sum_{t=1}^{T}(t-\bar{t})^2}. \end{aligned}$$

Then we have

$$C_{NT}^{C} = \sum_{i=1}^{N} x_{it-1}^{\tau} u_{it}^{\tau},$$

and

$$D_{NT}^{C} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it-1}^{\tau}\right)^2.$$

We derive an explicit form for the inconsistency

$$\text{plim}_{N\to\infty}(\hat{\rho}-\rho) = \text{plim}_{N\to\infty}\frac{\frac{1}{N}C_{NT}^{C}}{\frac{1}{N}D_{NT}^{C}}. \tag{66}$$

The data are generated by the model

$$y_{it} = a_i + b_i t + \rho y_{it-1} + u_{it}, \quad \rho \in (-1,1)$$

which has the alternate form

$$y_{it} = a_i^0 + b_i^0 t + x_{it}, \quad x_{it} = \rho x_{it-1} + u_{it} = \sum_{j=0}^{\infty}\rho^j u_{it-j}.$$

Linear detrending the variable x_{it} leads to the residual quantity

$$\begin{aligned} x_{it}^{\tau} &= x_{it} - \left\{\frac{2(2T+1)}{T(T-1)}\left(\sum_{t=1}^{T}x_{it}\right) - \frac{6}{T(T-1)}\sum_{t=1}^{T}tx_{it}\right\} \\ &\quad - \left\{\frac{12}{T(T^2-1)}\sum_{t=1}^{T}tx_{it} - \frac{6}{T(T-1)}\sum_{t=1}^{T}x_{it}\right\}t \\ &= x_{it} - g_T^{x_i} - h_T^{x_i}t, \end{aligned}$$

where

$$g_T^{x_i} = \frac{2(2T+1)}{T(T-1)}\left(\sum_{t=1}^{T}x_{it}\right) - \frac{6}{T(T-1)}\sum_{t=1}^{T}tx_{it}, \quad h_T^{x_i} = \frac{12}{T(T^2-1)}\sum_{t=1}^{T}tx_{it} - \frac{6}{T(T-1)}\sum_{t=1}^{T}x_{it}, \tag{67}$$

As in eq. (60), the detrended series when $K=1$ can be decomposed as

$$x_{it}^{\tau} = \delta_i z_{\theta t}^{\tau} + x_{it}^{\varepsilon\tau}$$

From the proof of Proposition 2 we have

$$\sigma^2 D(\rho,T) = \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(x_{it-1}^{\varepsilon\tau}\right)^2\right] = \sigma^2\frac{T-2}{1-\rho^2}\left[1-\frac{1}{T-2}\frac{4\rho}{1-\rho}D_1\right], \tag{68}$$

where D_1 is defined in (12). Then

$$\begin{aligned} \text{plim}_{N\to\infty}\ \frac{1}{N}D_{NT}^{c} &= \sigma^2 D(\rho,T) + \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\left[\sum_{t=1}^{T}\left(z_{\theta t-1}^{\tau}\right)^2\right] \\ &= \sigma^2 D(\rho,T) + m_\delta^2\sum_{t=1}^{T}\left(z_{\theta t-1}^{\tau}\right)^2, \end{aligned} \tag{69}$$

Letting $T\to\infty$ we have

$$\frac{1}{T}\sum_{t=1}^{T}\left(z_{\theta t-1}^{\tau}\right)^2 \to_p E\left(z_{\theta t-1}^2\right) = \frac{\sigma_\theta^2}{1-\rho^2},$$

and then

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\left[\sum_{t=1}^{T}\left(z_{\theta t-1}^{\tau}\right)^2\right] = T\left[m_\delta^2\frac{\sigma_\theta^2}{1-\rho^2}+o_p(1)\right],\ \text{as } T\to\infty. \tag{70}$$

Combining (70) and (68) with (69) yields

$$\begin{aligned} \text{plim}_{N\to\infty}\ \frac{1}{N}D_{NT}^{C} &= \sigma^2 D(\rho,T) + \frac{T}{1-\rho^2}m_\delta^2\sigma_\theta^2\left[1+o_p(1)\right],\ \text{as } T\to\infty. \\ &= \frac{T}{1-\rho^2}\left\{\sigma^2+m_\delta^2\sigma_\theta^2+o_p(1)\right\},\quad \text{as } T\to\infty \end{aligned} \tag{71}$$

Turning to the numerator of (66), we have

$$\begin{aligned} \text{plim}_{N\to\infty}\ \frac{1}{N}C_{NT}^{C} &= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it-1}^{\tau}u_{it}^{\tau} \\ &= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it-1}^{\varepsilon\tau}\varepsilon_{it}^{\tau} - \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\left[\sum_{t=1}^{T}z_{\theta t-1}^{\tau}\theta_t^{\tau}\right], \end{aligned} \tag{72}$$

where, from the proof of Proposition 2, we have

$$\begin{aligned} \sigma^2 C(\rho,T) &= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it-1}^{\varepsilon\tau}\varepsilon_{it}^{\tau} \\ &= \frac{\sigma^2}{T-1}\frac{2}{1-\rho}\left[(T-1)-\frac{2}{1-\rho}C_1\right], \end{aligned}$$

and C_1 is defined in (11). Since $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 = m_\delta^2$, (72) becomes

$$\text{plim}_{N\to\infty}\ \frac{1}{N}C_{NT}^{C} = -\sigma^2 C(\rho,T) - m_\delta^2 k_{\theta T}. \tag{73}$$

where $k_{\theta T} = \sum_{t=1}^{T} z_{\theta t-1}^{\tau}\theta_t^{\tau}$. Then, using (71) and (73), we have

$$\begin{aligned} \text{plim}_{N\to\infty}(\hat{\rho}-\rho) &= \frac{\text{plim}_{N\to\infty}C_{NT}^{C}}{\text{plim}_{N\to\infty}D_{NT}^{C}} \\ &= -\frac{\sigma^2 C(\rho,T)+m_\delta^2\sum_{t=1}^{T}z_{\theta t-1}^{\tau}\theta_t^{\tau}}{\sigma^2 D(\rho,T)+m_\delta^2\sum_{t=1}^{T}\left(z_{\theta t-1}^{\tau}\right)^2}, \end{aligned} \tag{74}$$

and the single factor ($K=1$) version of (42) follows. Extension to the multiple factor case follows in a straightforward way.

Phillips and Sul (2001) provide an asymptotic expansion the fitted autoregressive coefficient in an autoregression with trend. From this work we have

$$Eh_{\theta T} = E\frac{\sum_{t=1}^{T} z_{\theta t-1}^{\tau}\theta_t^{\tau}}{\sum_{t=1}^{T}\left(z_{\theta t-1}^{\tau}\right)^2} = -\frac{2+4\rho}{T} + O\left(T^{-2}\right),$$

and then, expanding the probability limit (74) as $T \to \infty$, we find

$$\begin{aligned}
\frac{\operatorname{plim}_{N\to\infty}\frac{1}{N}C_{NT}^{C}}{\operatorname{plim}_{N\to\infty}\frac{1}{N}D_{NT}^{C}} &= -\frac{\sigma^2 C(\rho,T)\,T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}+o_p(1)\right]^{-1}+m_\delta^2 h_{\theta T}}{\sigma^2 D(\rho,T)\,T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}+o_p(1)\right]^{-1}+m_\delta^2} \\
&= -\frac{\sigma^2\frac{2}{1-\rho}T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}\right]^{-1}+m_\delta^2 h_{\theta T}}{\sigma^2\frac{1}{1-\rho^2}\left[\frac{\sigma_\theta^2}{1-\rho^2}\right]^{-1}+m_\delta^2} + o_p\left(T^{-1}\right) \\
&= -\frac{2\frac{1+\rho}{T}\left(1+\frac{\sigma_\theta^2}{\sigma^2}m_\delta^2\right)-2\frac{1+\rho}{T}\frac{\sigma_\theta^2}{\sigma^2}m_\delta^2+\frac{\sigma_\theta^2}{\sigma^2}m_\delta^2 h_{\theta T}}{1+\frac{\sigma_\theta^2}{\sigma^2}m_\delta^2} + o_p\left(T^{-1}\right) \\
&= -2\frac{1+\rho}{T} - \frac{2\rho}{T}\frac{\sigma_\theta^2 m_\delta^2}{\sigma^2+\sigma_\theta^2 m_\delta^2} - \frac{\sigma_\theta^2 m_\delta^2}{\sigma^2+\sigma_\theta^2 m_\delta^2}\left(Eh_{\theta T}-h_{\theta T}\right) + o_p\left(T^{-1}\right),
\end{aligned}$$

as given in (44).

7.2 Additional Lemmas

The following two lemmas, whose proofs are straightforward and omitted, are used in calculating various results involving trend regression in the paper. They provide moment formulae for various sample moments of the (homogeneous) autoregression

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \rho \in (-1,1], \text{ with } \varepsilon_t \sim iid\left(0,\sigma^2\right), \tag{75}$$

in the stationary case ($|\rho| < 1$), where $\sigma_x^2 = \frac{\sigma^2}{1-\rho^2}$, and the unit root case ($\rho = 1$), where the initialization at $t = 0$ is $x_0 = 0$. The lemmas provide basic formulae from which reduced results can be obtained by further calculation or by the use of algebraic manipulation software such as Maple. The latter formulae are lengthy and are not repeated here.

Lemma 1 (Stationary x_t):

(a) $E\left(\sum_{t=1}^{T-1} t x_t\right)^2 = \sigma_x^2\left\{\sum_{t=1}^{T-1} t\sum_{j=t}^{T-1} j\rho^{j-t} + \sum_{t=2}^{T-1} t\sum_{j=1}^{t-1} j\rho^{t-j}\right\}$

(b) $E\left(\sum_{t=1}^{T-1} t x_t\right)\left(\sum_{t=1}^{T-1} x_t\right) = \sigma_x^2\left\{\sum_{t=1}^{T-1} t\sum_{j=t}^{T-1} \rho^{j-t} + \sum_{t=2}^{T-1} t\sum_{j=1}^{t-1} \rho^{t-j}\right\}$

(c) $E\left(\sum_{t=1}^{T-1} t x_t\right)\left(\sum_{t=2}^{T} t x_t\right) = \sigma_x^2\left\{\sum_{t=1}^{T-1} t\sum_{j=t}^{T-1} j\rho^{j-t+1} + \sum_{t=2}^{T-1} t\sum_{j=1}^{t-1} j\rho^{t-j-1}\right\}$

(d) $E\left(\sum_{t=1}^{T-1} t x_t\right)\left(\sum_{t=2}^{T} x_t\right) = \sigma_x^2\left\{\sum_{t=1}^{T-1} t\sum_{j=t}^{T-1} \rho^{j-t+1} + \sum_{t=2}^{T-1} t\sum_{j=1}^{t-1} \rho^{t-j-1}\right\}$

(e) $E\left(\sum_{t=1}^{T-1} t x_{t+1}\right)\left(\sum_{t=1}^{T-1} x_t\right) = \sigma_x^2\left\{\sum_{t=1}^{T-1} t\sum_{j=1}^{t} \rho^{j} + \sum_{t=1}^{T-2} t\sum_{j=1}^{T-t-1} \rho^{j-1}\right\}$

(f) $E\sum_{t=2}^{T} x_t \sum_{t=2}^{T} x_{t-1} = \sigma_x^2 \sum_{t=1}^{T-1}\sum_{j=t}^{T-1}\rho^{j-t+1} + \sum_{t=2}^{T-1}\sum_{j=1}^{t-1}\rho^{t-j-1}$

(g) $E\sum_{t=2}^{T} x_t x_{t-1} = (T-1)\rho\sigma_x^2$

(h) $E\left(\sum_{t=1}^{T-1} x_t\right)^2 = \sigma_x^2\left(T-1+\frac{2\rho}{1-\rho}\sum_{k=1}^{T-2}(1-\rho^k)\right)$

Lemma 2 (Unit Root x_t):

(a) $E\left(\sum_{t=1}^{T-1} t x_t\right)^2 = \sigma^2\left\{\sum_{t=1}^{T-1}\left(\frac{1}{6}t^2 - \frac{1}{6}t^4 + \frac{1}{2}T^2t^2 - \frac{1}{2}Tt^2\right)\right\}$

(b) $E\left(\sum_{t=1}^{T-1} t x_t\right)\left(\sum_{t=1}^{T-1} x_t\right) = \sigma^2\left\{\sum_{t=1}^{T-1}\left(-\frac{1}{2}t^3 - \frac{1}{2}t^2 + Tt^2\right)\right\}$

(c) $E\left(\sum_{t=1}^{T-1} t x_t\right)\left(\sum_{t=2}^{T} t x_t\right) = \sigma^2\left\{\sum_{t=1}^{T-1}\left(t\sum_{i=1}^{t-2} i(i+1) + t^2\sum_{i=t-1}^{T-1} i\right)\right\}$

(d) $E\left(\sum_{t=1}^{T-1} t x_t\right)\left(\sum_{t=2}^{T} x_t\right) = \sigma^2\left\{\sum_{t=1}^{T-1}\left(-\frac{1}{2}t^3 + \frac{1}{2}t^2 - t + Tt^2\right)\right\}$

(e) $E\left(\sum_{t=1}^{T-1} t x_{t+1}\right)\left(\sum_{t=1}^{T-1} x_t\right) = \sigma^2\left\{\sum_{t=1}^{T-1} t\sum_{j=1}^{t} j + \sum_{t=1}^{T-2} t(t+1)(T-t-1)\right\}$

(f) $E\sum_{t=2}^{T} x_t \sum_{t=2}^{T} x_{t-1} = \sigma^2\left\{\sum_{t=1}^{T-1}\sum_{j=1}^{t} j + \sum_{t=1}^{T-2}(t+1)(T-t-1)\right\}$

(g) $E\sum_{t=2}^{T} x_t x_{t-1} = \sigma^2\sum_{t=1}^{T-1} t$

(h) $E\left(\sum_{t=1}^{T-1} x_t\right)^2 = \sigma^2\left\{\sum_{t=1}^{T-1}\sum_{j=1}^{t} j + \sum_{t=1}^{T-2} t\sum_{j=t+1}^{T-1} 1\right\}$

8 Notation

$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
$\int_0^1 f$	$\int_0^1 f(r)dr$
$[\cdot]$	integer part
$:=$	definitional equality
CSD	Cross Section Dependent
CSI	Cross Section Independent
$w_{i\cdot}$	$T^{-1}\sum_{t=1}^{T} w_{it}$
$w_{i\cdot-1}$	$T^{-1}\sum_{t=1}^{T} w_{it-1}$
w_t^τ	$w_t - g_T^w - h_T^w t$detrended $w_t$see (67)
a_T^w	$\frac{2(2T+1)}{T(T-1)}\left(\sum_{t=1}^{T} w_t\right) - \frac{6}{T(T-1)}\sum_{t=1}^{T} t w_t$
b_T^w	$\frac{12}{T(T^2-1)}\sum_{t=1}^{T} t w_t - \frac{6}{T(T-1)}\sum_{t=1}^{T} w_t$
$\rightarrow_d$	weak convergence
$\rightarrow_p$	convergence in probability, almost surely

Table 2: Finite Sample Performance of Mean Unbiased Estimator with an Exogenous Variable: ($\rho = 0.9, \beta = 1$)

$$y_{it} = a_i + b_i t + \rho y_{it-1} + \beta z_{it} + u_{it}$$

Sample	Absolute Bias $\times T$				MSE Ratio	
	(A)	(B)	(C)	(D)	A/B	C/D
			Fixed Effects			
T= 5,N=1000	1.293	0.000	0.585	0.001	0.009	0.021
T= 10,N= 500	1.277	0.001	0.496	0.004	0.019	0.092
T= 25,N= 200	1.195	0.007	0.312	0.006	0.056	0.585
T= 50,N= 100	1.107	0.008	0.180	0.012	0.150	0.942
T=100,N= 50	1.047	0.026	0.096	0.024	0.420	0.996
			Linear Trend			
T= 5,N=1000	3.237	0.010	1.623	0.006	0.022	0.020
T= 10,N= 500	3.087	0.078	1.435	0.032	0.037	0.042
T= 25,N= 200	2.771	0.016	1.027	0.002	0.043	0.145
T= 50,N= 100	2.482	0.011	0.664	0.015	0.068	0.560
T=100,N= 50	2.241	0.021	0.368	0.026	0.158	0.941

Legend: Errors are drawn as *iid* $N(0,1)$, the number of replications = 10,000
A= $\hat{\rho}$ (Pooled OLS), B= $\hat{\rho}_{\mathrm{MUE}}$ (Mean unbiased estimator),
C=$\hat{\beta}$ (Pooled OLS), D=$\hat{\beta}_{\mathrm{MUE}}$(Mean unbiased estimator)

Table 3: Finite Sample Performance of Various Feasible Generalized Mean Unbiased Estimator Under Cross Section Dependence ($\rho = 0.9$)

		Bias $\times T$						MSE Ratio$\times$10				
T	N	(A)	(B)	(C)	(D)	(E)	(F)	B/A	C/A	D/A	E/A	F/A
					Fixed Effects							
25	10	-3.65	-2.63	-0.65	-0.30	-0.08	-1.57	3.71	1.05	0.91	0.81	2.45
25	25	-3.52	-2.69	-0.53	-0.15	0.03	-1.44	4.34	0.58	0.46	0.39	1.80
25	50	-3.49	-2.63	-0.49	-0.10	0.08	-1.40	4.16	0.38	0.27	0.21	1.55
25	100	-3.53	-2.64	-0.48	-0.09	0.10	-1.38	4.03	0.29	0.19	0.12	1.34
50	10	-3.64	-2.78	-0.45	-0.27	-0.19	-1.41	4.54	0.86	0.83	0.79	1.89
50	25	-3.58	-2.43	-0.34	-0.16	-0.06	-1.27	2.88	0.38	0.35	0.32	1.14
50	50	-3.51	-2.53	-0.29	-0.10	-0.01	-1.22	3.55	0.22	0.20	0.17	0.93
50	100	-3.51	-2.51	-0.27	-0.08	0.01	-1.18	3.42	0.13	0.11	0.09	0.78
100	10	-3.43	-2.51	-0.33	-0.25	-0.21	-1.30	3.84	0.94	0.93	0.92	1.73
100	25	-3.39	-2.45	-0.20	-0.11	-0.07	-1.14	3.45	0.37	0.36	0.35	0.87
100	50	-3.47	-2.40	-0.16	-0.07	-0.03	-1.08	3.06	0.19	0.18	0.18	0.61
100	100	-3.41	-2.40	-0.15	-0.06	-0.02	-1.06	3.13	0.10	0.10	0.09	0.50
200	10	-3.42	-2.47	-0.24	-0.20	-0.18	-1.18	3.67	1.01	1.00	1.00	1.49
200	25	-3.40	-2.47	-0.12	-0.09	-0.07	-1.04	3.70	0.40	0.40	0.40	0.68
200	50	-3.48	-2.44	-0.11	-0.07	-0.05	-1.03	3.18	0.20	0.20	0.20	0.45
200	100	-3.33	-2.32	-0.08	-0.04	-0.02	-1.00	2.95	0.11	0.10	0.10	0.33
					Linear Trend							
25	10	-6.66	-5.60	-1.25	-0.35	-0.01	-2.15	5.76	1.14	0.89	0.80	1.95
25	25	-6.57	-5.68	-1.18	-0.09	0.25	-2.00	6.28	0.71	0.52	0.47	1.37
25	50	-6.58	-5.64	-1.15	0.01	0.37	-1.95	6.15	0.52	0.35	0.32	1.14
25	100	-6.58	-5.63	-1.17	0.00	0.38	-1.95	6.03	0.43	0.26	0.21	1.00
50	10	-6.43	-5.52	-0.70	-0.24	-0.11	-1.77	6.10	0.83	0.80	0.79	1.50
50	25	-6.40	-5.13	-0.62	-0.14	0.02	-1.63	4.71	0.37	0.33	0.32	0.86
50	50	-6.30	-5.24	-0.58	-0.10	0.04	-1.59	5.38	0.22	0.18	0.16	0.71
50	100	-6.32	-5.22	-0.56	-0.08	0.06	-1.56	5.23	0.14	0.10	0.08	0.59
100	10	-5.97	-4.95	-0.47	-0.28	-0.23	-1.52	5.22	0.73	0.71	0.71	1.28
100	25	-5.88	-4.85	-0.31	-0.11	-0.06	-1.33	4.95	0.30	0.29	0.28	0.65
100	50	-6.00	-4.83	-0.28	-0.08	-0.03	-1.28	4.63	0.15	0.14	0.14	0.45
100	100	-5.87	-4.78	-0.26	-0.06	0.00	-1.26	4.71	0.08	0.07	0.07	0.38
200	10	-5.67	-4.65	-0.30	-0.22	-0.19	-1.28	4.74	0.80	0.79	0.79	1.16
200	25	-5.63	-4.66	-0.18	-0.09	-0.07	-1.14	4.79	0.32	0.31	0.31	0.54
200	50	-5.72	-4.62	-0.17	-0.08	-0.06	-1.13	4.35	0.16	0.16	0.16	0.36
200	100	-5.57	-4.49	-0.13	-0.04	-0.01	-1.09	4.12	0.08	0.08	0.08	0.26

Legend: (A) = LSDV; (B) = LSDV with common time effect; (C) = FGMUE based on residual variance of LSDV; (D) = FGMUE based on residual variance of MUE with fixed effects; (E) = FGMUE based on residual variance of MUE with common time effects; (F) = MUE with Pesaran's correlated common estimator.