

# 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*

July 15, 2003

## Abstract

Explicit asymptotic bias formulae are given for dynamic panel regression estimators as the cross section sample size  $N \rightarrow \infty$ . The results extend earlier work by Nickell (1981) 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.

*Keywords:* Autoregression, Bias, 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

## 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  $N \rightarrow \infty$ . 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) and Baltagi (2001)

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, and Hsiao, 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, 2002) and slow convergence (Moon and Phillips, 2002) 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  $N \rightarrow \infty$  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  $N \rightarrow \infty$  helps to explain the substantial variability that is observed in dynamic panel estimates under cross section dependence even in situations where  $N$  is very large (e.g., Phillips and Sul, 2002).

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. Both sections report some simulation findings on the adequacy of the asymptotics. Section 5 concludes and offers some thoughts on bias correction possibilities. The appendix contains derivations of the main results (Section 6) and a glossary of notation (Section 7).

## 2 Models

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

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

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

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

In each case, the index  $i$  ( $i = 1, \dots, N$ ) stands for the  $i$ 'th cross sectional unit and  $t$  ( $t = 1, \dots, T$ ) 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  $y_{it}^0$  is taken to be  $y_{i0}^0 = O_p(1)$  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  $\varepsilon_{it}$  in the above models are independent across  $i$ . The following section derives explicit formulae for the asymptotic bias of the least squares estimates of  $\rho$  and  $\beta$  in that case, giving the inconsistency  $\text{plim}_{N \rightarrow \infty}(\hat{\rho} - \rho)$  for each model where  $\hat{\rho}$  is the panel least squares estimate of  $\rho$ . 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  $N \rightarrow \infty$ . Hence, for large  $N$  the distribution of  $\hat{\rho}$  is close to symmetric about  $\rho$  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  $N \rightarrow \infty$ . 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  $N \rightarrow \infty$ , as is well understood from Neyman and Scott (1948) and Nickell (1981). Interestingly, that inconsistency persists even as  $T \rightarrow \infty$  when  $\rho = 1 + c/T$  and the parameter being estimated is local to unity (Moon and Phillips, 1999, 2000 & 2003).

#### 3.1 Fixed Effects Model M1

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

convenient.

**Assumption A1: (error condition)** *The  $\varepsilon_{it}$  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  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2 = \sigma^2$ .*

Nickell (1981) assumed  $iid(0, \sigma^2)$  errors  $\varepsilon_{it}$  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  $\rho$  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  $\rho$  has the form

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

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

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

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

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

and

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

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  $\rho$  as  $N \rightarrow \infty$  is given by*

$$\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = -\frac{1+\rho}{T-1} \left[ 1 - \frac{1}{T} \frac{1-\rho^T}{1-\rho} \right] \left\{ 1 - \frac{1}{T-1} \frac{2\rho}{1-\rho} \left[ 1 - \frac{1}{T} \frac{1-\rho^T}{1-\rho} \right] \right\}^{-1} \quad (5)$$

$$= G(\rho, T). \quad (6)$$

For large  $T$ , the inconsistency has the expansion

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

Formula (5) is the same as that given by Nickell (1981) for the case of homogeneous errors<sup>1</sup>. Applying the third derivative version of l'Hôpital's rule directly to  $G(\rho, T)$  with respect to  $\rho$  we obtain the limit

<sup>1</sup>For  $T = 3$ , there is a typographical error in Nickell (1981), the correct formula being

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

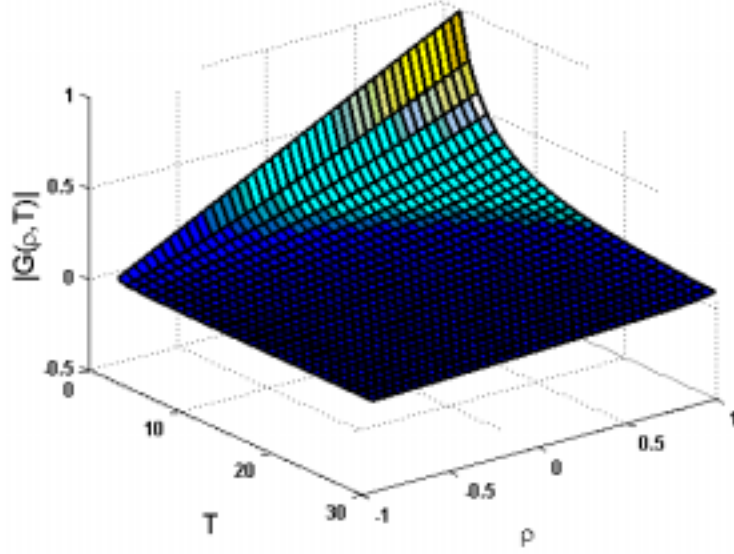


Figure 1: Asymptotic ( $N \rightarrow \infty$ ) Bias Function  $|G(\rho, T)| = -G(\rho, T)$  for Model M1.

behavior for the unit root case, viz.,  $\lim_{\rho \rightarrow 1} G(\rho, T) = -\frac{3}{T+1}$ , and the inconsistency of the pooled least squares estimate for  $\rho = 1$  follows

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

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  $\rho$  and  $T$ . As is clear from the figure, the magnitude of the asymptotic bias increases with  $\rho$ , and of course decreases as  $T$  increases.

### 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  $\rho$  has the form  $\hat{\rho} = C_{NT}^y / D_{NT}$ , where

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

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  $C_{NT} = C_{NT}^y - \rho D_{NT}$ , the inconsistency as  $N \rightarrow \infty$  with  $T$  fixed is

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

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

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

$$\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = -2 \frac{1+\rho}{T-2} \left[ 1 - \frac{1}{T-1} \frac{2}{1-\rho} C_1 \right] \left[ 1 - \frac{1}{T-2} \frac{4\rho}{1-\rho} D_1 \right]^{-1} \quad (9)$$

$$= H(\rho, T), \quad (10)$$

where

$$C_1 = 1 - \frac{1}{T+1} \left( 1 + \frac{1-\rho^3}{(1-\rho)^3} \frac{1}{T} \right) + \left( \frac{1}{2} + \frac{1}{T+1} \left[ \frac{1+2\rho}{1-\rho} + \frac{1-\rho^3}{(1-\rho)^3} \frac{1}{T} \right] \right) \rho^T, \quad (11)$$

$$D_1 = 1 - \frac{1}{T+1} \frac{2}{1-\rho} \left\{ 1 + \frac{1}{T-1} \left[ 1 - \frac{1-\rho^3}{T(1-\rho)^3} (1-\rho^T) + \left( \frac{3\rho}{1-\rho} + \frac{T+3}{2} \right) \rho^T \right] \right\}. \quad (12)$$

For large  $T$ , the inconsistency has the following expansion

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

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:

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} C_{NT} = -\frac{\sigma^2}{T-1} \frac{2}{1-\rho} \left[ (T-1) - \frac{2}{1-\rho} C_1 \right] := -\sigma^2 C(\rho, T), \quad (14)$$

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT} = \sigma^2 \frac{T-2}{1-\rho^2} \left[ 1 - \frac{1}{T-2} \frac{4\rho}{1-\rho} D_1 \right] := \sigma^2 D(\rho, T). \quad (15)$$

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 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

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

and the bias differential (M2 - 2 × M1) is

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

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

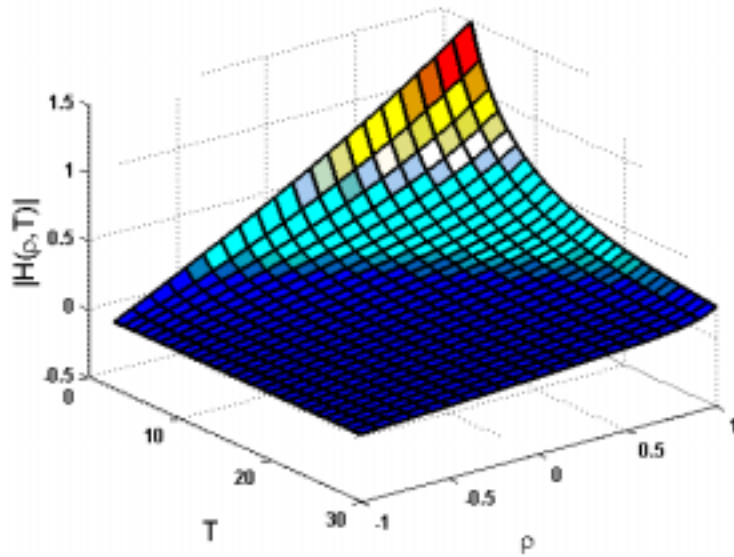


Figure 2: Asymptotic ( $N \rightarrow \infty$ ) Bias Function  $|H(\rho, T)| = -H(\rho, T)$  for Model M2.

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

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

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  $\rho$  is often negative even when the true autoregressive coefficient  $\rho$  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 ( $\bar{\rho} = \text{plim}_{N \rightarrow \infty} \hat{\rho} = 0.9 - 1.402 = -0.502$ ) 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.

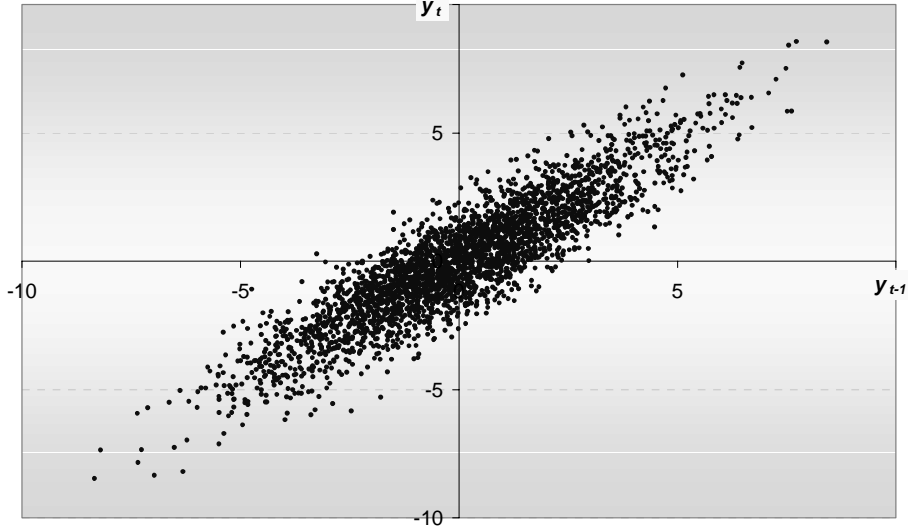


Figure 3: Sample Data before Detrending ( $T = 4, N = 1,000, \rho = 0.9, \hat{\rho} = 0.90$ )

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

Absolute Bias: Model(Simulation)						
$T$	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 1.0$
3	0.740(0.741)	0.891(0.896)	1.050(1.057)	1.220(1.223)	1.402(1.403)	1.500(1.500)
4	0.561(0.562)	0.690(0.695)	0.829(0.837)	0.982(0.987)	1.154(1.155)	1.250(1.250)
5	0.450(0.450)	0.558(0.562)	0.679(0.687)	0.816(0.822)	0.977(0.978)	1.071(1.072)
6	0.375(0.375)	0.466(0.469)	0.571(0.578)	0.694(0.701)	0.845(0.846)	0.938(0.938)
7	0.321(0.321)	0.400(0.402)	0.490(0.497)	0.601(0.609)	0.743(0.744)	0.833(0.833)
8	0.281(0.281)	0.348(0.350)	0.428(0.435)	0.528(0.537)	0.661(0.663)	0.750(0.750)
9	0.249(0.249)	0.308(0.310)	0.380(0.385)	0.470(0.478)	0.595(0.597)	0.682(0.682)

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

Figs. 5 and 6 show the actual biases (computed by simulation), the exact biases from (5) and (9), and the approximate biases from (7) and (13) with  $N = 1,000$  and  $T = 5$ , respectively. The exact formulae are seen to work very well for both cases, while the asymptotic approximations are satisfactory for small  $\rho$  but deteriorate in quality as  $\rho$  increases, particularly as it approaches unity.

### 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



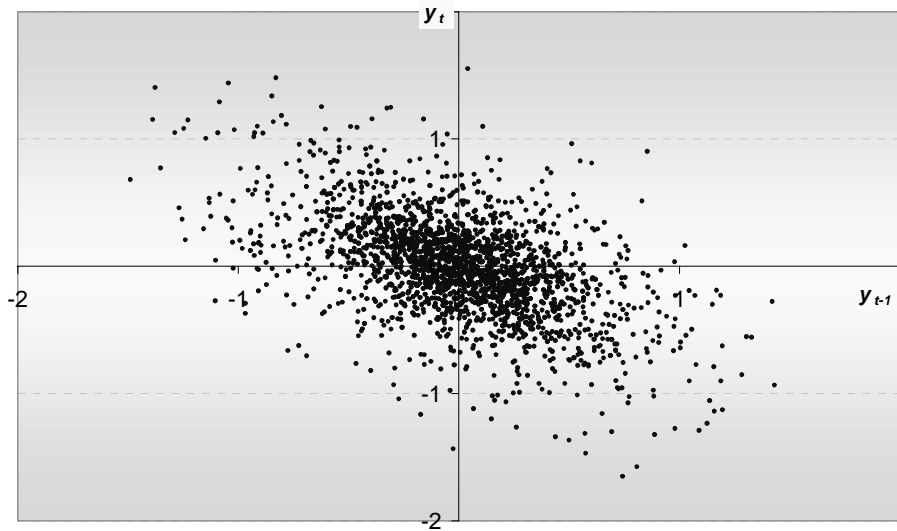


Figure 4: Sample Data after Detrending ( $T = 4$ ,  $N = 1,000$ ;  $\rho = 0.9$ ,  $\bar{\rho} = \text{plim}_{N \rightarrow \infty} \hat{\rho} = -0.502$ ,  $\hat{\rho} = -0.53$ ).

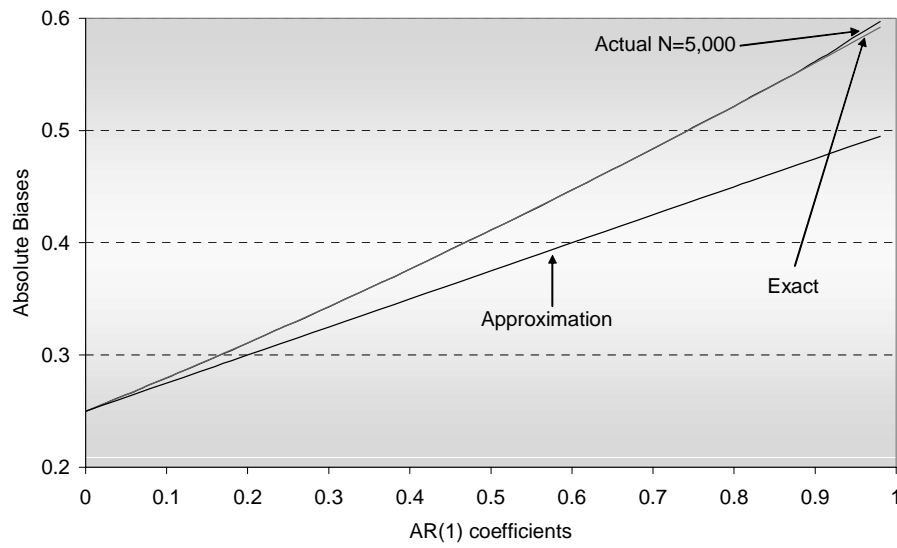


Figure 5: Absolute Biases for Fixed Effects Case M1 ( $T = 5$ )

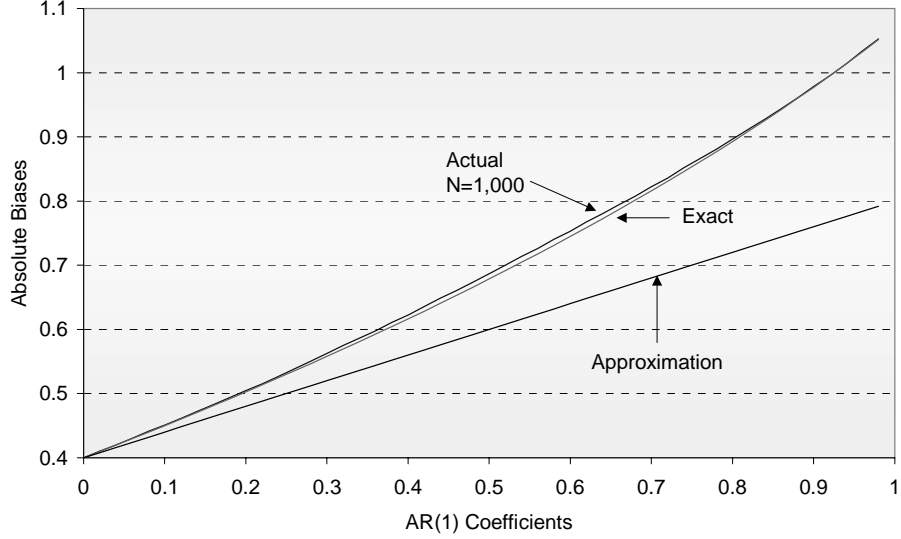


Figure 6: Absolute Biases for Linear Trend Case M2 ( $T = 5$ ).

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

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

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

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

and

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

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

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

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

and in the incidental trends case by

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

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

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

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 (50) and (53), and  $\tilde{Z}_{\rho, -1}$  by  $\tilde{Z}_{1, -1} = \left( \tilde{Z}'_{1, 0}, \dots, \tilde{Z}'_{1, T-1} \right)'$  where  $\tilde{Z}_{1, t} = \left( \tilde{Z}_{1, t}^1, \dots, \tilde{Z}_{1, t}^N \right)'$  and  $\tilde{Z}_{1, t}^i = \sum_{j=0}^t \tilde{Z}_{it-j}$ .

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

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

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

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

It is clear from (25) that we have the reverse inequality  $\tilde{y}'_{-1} Q_{\tilde{Z}} \tilde{y}_{-1} \geq \tilde{y}_{-1}^{0'} \tilde{y}_{-1}^0$ , 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):

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

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

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

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  $plim_{N \rightarrow \infty} \frac{1}{N} \tilde{y}'_{-1} Q_{\tilde{Z}} \tilde{y}_{-1} = \sigma^2 B(\rho, T)$  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  $\beta$  is given by

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

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 (2002) 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 Sul (2002) all use a factor structure where the dynamics are explicit in the system. The following model is a prototypical first order panel dynamic system

$$y_{it} = a_i + \rho_i y_{it-1} + u_{it}, \quad u_{it} = \sum_{s=1}^K \delta_{is} \theta_{st} + \varepsilon_{it}, \quad (28)$$

where the errors  $u_{it}$  depend on  $K$  factors  $\{\theta_{st} : s = 1, \dots, K\}$  with factor loadings  $\{\delta_{is} : s = 1, \dots, K\}$ , and  $\varepsilon_{it}$  is assumed to be  $iid(0, \sigma_i^2)$ . In this prototypical system,  $\theta_{st}$  and  $\varepsilon_{it}$  are assumed to be independent of each other and each is assumed to be  $iid$ . Also,  $\theta_{st}$  is taken to be cross sectionally independent of  $\theta_{qt}$ . While  $K$  is fixed and generally taken to be small (typically  $K = 1$  or  $2$ ) in practical work, we may, in principle at least, consider cases where  $K \rightarrow \infty$  as  $T, N \rightarrow \infty$ . The number of factors may also be varied across  $i$ .

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

$$y_{it} = \sum_{s=1}^K \lambda_{is} F_{st} + m_{it}. \quad (29)$$

In (29) there are again  $K$  factors and factor loadings  $\{F_{st}, \lambda_{is} : s = 1, \dots, K\}$ ,  $F_{st}$  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 (28) has a homogeneous autoregressive coefficient ( $\rho_i = \rho$ ), it can be viewed as a restricted version of the direct model (29) in which a common dynamic factor can be drawn from each of the individual factors and the error.

The analysis that follows is based on dynamic panel models of the type (28), 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 (28), the model extends M1 to accommodate cross section dependent errors as follows.

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

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*

$$u_{it} = \sum_{s=1}^K \delta_{si} \theta_{st} + \varepsilon_{it}, \quad (30)$$

where the  $\varepsilon_{it}$  satisfy **A1**, the factors  $\theta_{st}$  ( $s = 1, \dots, K$ ) are  $iid(0, \sigma_{s\theta}^2)$  over  $t$  and the factor loadings  $\delta_{si}$  are nonrandom parameters satisfying  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_{si}^2 = \mu_{\delta s}^2$ .

Under A2, we can develop an asymptotic theory for the pooled least squares estimate,  $\hat{\rho}$ , of the common dynamic coefficient  $\rho$ . It is convenient to use a sequential asymptotic argument with  $N \rightarrow \infty$  followed by  $T \rightarrow \infty$ . This approach produces a result for the bias or inconsistency of  $\hat{\rho}$  as  $N \rightarrow \infty$  and the expression can conveniently be written in an asymptotic format that is valid as  $T \rightarrow \infty$ . 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 (30) and satisfying assumption A2, the pooled least squares estimate  $\hat{\rho}$  is inconsistent as  $N \rightarrow \infty$  and*

$$plim_{N \rightarrow \infty} (\hat{\rho} - \rho) = - [\sigma^2 A(\rho, T) + \psi_{AT}] [\sigma^2 B(\rho, T) + \psi_{BT}]^{-1}, \quad (31)$$

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

$$\psi_{AT} = \sum_{s=1}^K \mu_{\delta s}^2 \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{\theta st} \frac{1}{\sqrt{T}} \sum_{p=1}^T \theta_{sp} \right], \quad (32)$$

$$\psi_{BT} = \sum_{s=1}^K \mu_{\delta s}^2 \sum_{t=1}^T (Z_{\theta st} - \bar{Z}_{\theta})^2, \quad (33)$$

and  $Z_{\theta st} = \sum_{j=0}^{\infty} \rho^j \theta_{st-j-1}$ . The inconsistency (31) has the following asymptotic representation as  $T \rightarrow \infty$

$$plim_{N \rightarrow \infty} (\hat{\rho} - \rho) = -\frac{1+\rho}{T} - \frac{1+\rho}{T} \frac{\sum_{s=1}^K \mu_{\delta s}^2 \sigma_{\theta s}^2 (\eta_{\theta s}^2 - 1)}{\sigma^2 + \sum_{s=1}^K \mu_{\delta s}^2 \sigma_{\theta s}^2} + o_{a.s.} \left( \frac{1}{T} \right) \quad (34)$$

where  $\eta_{\theta s}^2$  is  $iid \chi_1^2$  over  $s = 1, \dots, K$ .

**Remark 1** It is apparent from the form of (31) and (34) that the inconsistency of the panel estimate  $\hat{\rho}$  as  $N \rightarrow \infty$  is random, as distinct from the usual 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  $\mu_{\delta s}^2 = 0$  and then (31) reduces to  $G(\rho, T) = -A(\rho, T)/B(\rho, T)$ , and the second term on the right side of (34) 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  $\mu_{\delta s}^2 \neq 0$ , then the components  $\psi_{AT}$  and  $\psi_{BT}$  in (31) are non zero random variables with positive variance. Likewise, the second term of (34) is nonzero. So the immediate contribution of cross section dependence is to introduce variability into the inconsistency of  $\hat{\rho}$ .

**Remark 2** Expression (34) gives the inconsistency to  $O(T^{-1})$  and is the analogue of (7) for the cross section dependent case. The difference between (34) and (7) is the second term of (34), which involves the random  $\chi_1^2$  elements  $\{\eta_{\theta_s}^2 : s = 1, \dots, K\}$ . Since  $E\{\eta_{\theta_s}^2 - 1\} = 0$ , the mean of this additional term in (34) is zero. So the mean inconsistency is the same as the inconsistency in the cross section independent case to order  $O(T^{-1})$ . However, the variation that is induced by this second term is substantial even for moderate time series sample sizes such as  $T = 20$ , as is apparent from Fig. 7 below.

**Remark 3** In the single factor model ( $K = 1$ ) the inconsistency expression has the simple form

$$\text{plim}_{N \rightarrow \infty}(\hat{\rho} - \rho) = -\frac{1 + \rho}{T} - \frac{1 + \rho}{T} \frac{\mu_\delta^2 \sigma_\theta^2}{\sigma^2 + \mu_\delta^2 \sigma_\theta^2} [\eta_\theta^2 - 1] + o_{a.s.} \left( \frac{1}{T} \right).$$

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

**Remark 4** For the infinite factor case, define

$$u_{it} = \sum_{s=1}^{\infty} \delta_{is} \theta_{st} + \varepsilon_{it}, \quad \text{with} \quad \sum_{s=1}^{\infty} s |\delta_{is}| < \infty,$$

let  $\delta_i^2 = \sum_{s=1}^{\infty} \delta_{is}^2$ , and assume that  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 = \mu_\delta^2$ , uniformly in  $s$ , and  $\sum_{s=1}^{\infty} \mu_\delta^2 \sigma_{\theta_s}^2 < \infty$ . Then

$$\text{plim}_{N \rightarrow \infty}(\hat{\rho}_{pols} - \rho) = -\frac{1 + \rho}{T} - \frac{1 + \rho}{T} \left[ \frac{\sum_{s=1}^{\infty} \mu_\delta^2 \sigma_{\theta_s}^2 (\eta_{\theta_s}^2 - 1)}{\sigma^2 + \sum_{s=1}^{\infty} \mu_\delta^2 \sigma_{\theta_s}^2} \right] + o_{a.s.} \left( \frac{1}{T} \right). \quad (35)$$

In practical work, of course, only finite values of  $K$  are used or estimated by selection criteria (c.f. Bai and Ng, 2002).

The adequacy of the limit distribution theory for the inconsistency given in Proposition 4 can be assessed by simulation. The pooled least squares estimate  $\hat{\rho}$  was computed for the model M1-CSD with  $\rho = 0.5$ ,  $a_i = 0$ ,  $N = 5,000$  and for time series sample sizes  $T = 5, 10, 20$ , each with 20,000 replications. The dashed curves (designated ‘sim’) in Fig. 7 show kernel density estimates of the distribution of  $\hat{\rho}$  for each value of  $T$ . The solid curves (designated ‘asy’) show the distribution of  $\hat{\rho}$  in the limit as  $N \rightarrow \infty$  that is given by (31) in Proposition 4 for each value of  $T$ . The figure also shows the point value inconsistency (5) that applies in the CSI case for the same parameter value  $\rho = 0.5$ . We notice that the CSI inconsistency is a good approximation to the mean value of the random inconsistency in the cross section dependent (CSD) case, as asymptotic theory predicts. The magnitude of the inconsistency is seen to reduce as  $T$  increases, but is still large and has substantial variance even for  $T = 20$ .

We now turn to the unit root case for model M1-CSD. In the CSI model M1, continuity of the limit function (5) enabled the unit root case to be extracted by a simple limiting operation as  $\rho \rightarrow 1$ . This procedure is not possible in the CSD unit root case because the probability limit (31) is random and involves the stochastic elements (32) and (33) whose definitions depend on  $\rho$ , the underlying model and the initialization. Instead, deriving the limit directly we find the following.

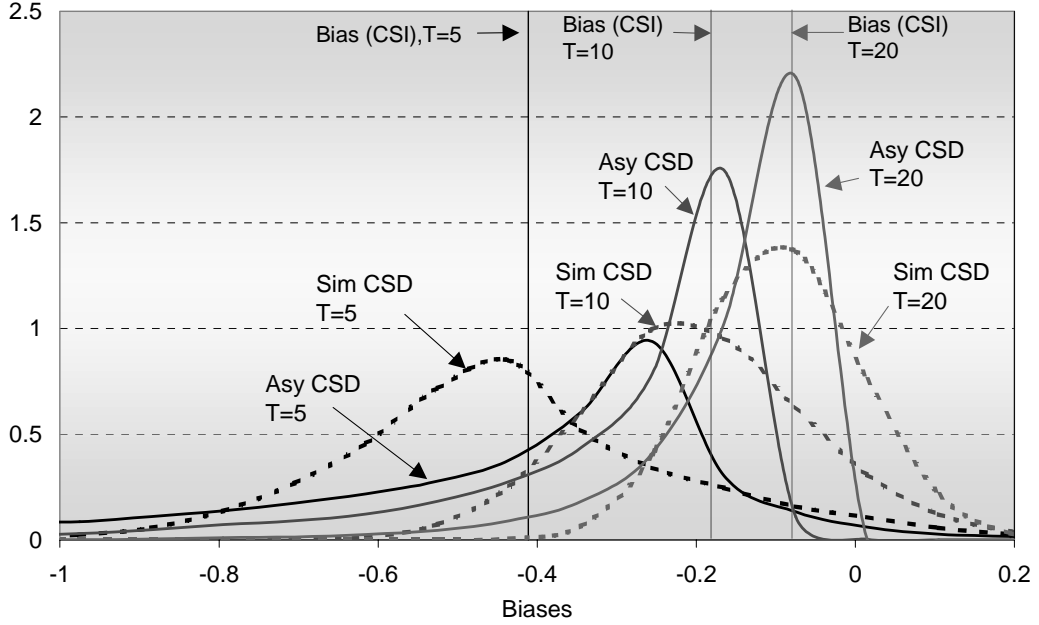


Figure 7: Simulated (Sim) and Asymptotic (Asy) Distributions of Inconsistency of  $\hat{\rho}$  based on  $N = 5,000$  and  $\rho = 0.5$ . (Legend: CSD – cross section dependent case; CSI – cross section independent case).

**Proposition 5** (*Fixed Effects with  $\rho = 1$* ) In model **M1-CSD** with errors  $u_{it}$  having the factor structure (30) and satisfying assumption **A2**, the pooled least squares estimate  $\hat{\rho}$  is inconsistent as  $N \rightarrow \infty$  and

$$plim_{N \rightarrow \infty}(\hat{\rho} - \rho) = -[\sigma^2 A(T) + \phi_{AT}] [\sigma^2 B(T) + \phi_{BT}]^{-1} \quad (36)$$

where

$$A(T) = \frac{1}{2}(T-1), \quad B(T) = \frac{(T-1)(T+1)}{6}$$

$$\phi_{AT} = \sum_{s=1}^K \mu_{\delta_s}^2 \left( \frac{1}{T} \sum_{p=1}^T S_{sp-1}^\theta \sum_{t=1}^T \theta_{st} \right), \quad \phi_{BT} = \sum_{s=1}^K \mu_{\delta_s}^2 \sum_{t=1}^T (S_{st-1}^\theta - S_{s,-1}^\theta)^2$$

and  $S_{st}^\theta = \sum_{p=1}^t \theta_{sp}$ . The inconsistency (36) has the following asymptotic representation as  $T \rightarrow \infty$

$$plim_{N \rightarrow \infty}(\hat{\rho} - \rho) = -\frac{3}{T+1} - \frac{1}{T+1} \frac{6 \sum_{s=1}^K \mu_{\delta_s}^2 \sigma_{\theta_s}^2 \left\{ \int_0^1 W_s(r) dr W_s(1) - 3 \int_0^1 \underline{W}_s(r)^2 dr \right\}}{\sigma^2 + 6 \sum_{s=1}^K \mu_{\delta_s}^2 \sigma_{\theta_s}^2 \int_0^1 \underline{W}_s(r)^2 dr} + o_{a.s.} \left( \frac{1}{T} \right) \quad (37)$$

where  $\{W_s : s = 1, \dots, K\}$  are independent standard Brownian motions and  $\underline{W}_s(r) = W_s - \int_0^1 W_s(r)$  is demeaned Brownian motion.

**Remark 5** When there is no cross section dependence  $\mu_{\delta_s}^2 = 0$  for all  $s$ , and then (36) reduces to

$$plim_{N \rightarrow \infty}(\hat{\rho} - \rho) = -\frac{A(T)}{B(T)} = -\frac{3}{T+1},$$

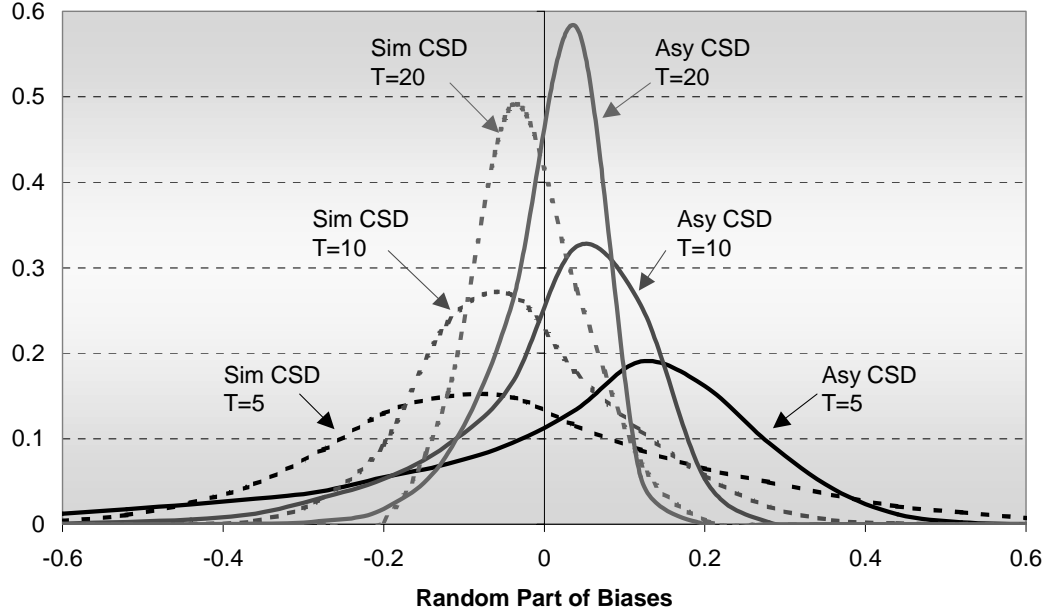


Figure 8: Simulated (Sim) and Asymptotic (Asy) Distributions of Random Parts of Inconsistency of  $\hat{\rho}$  based on  $N = 5,000$  and  $\rho = 1$ . (Legend: CSD – cross section dependent case; CSI – cross section independent case).

as given earlier in (8)

**Remark 6** As in the stationary case, the inconsistency (36) is random and depends on the common time effect processes  $\theta_{st}$ . For large  $T$ , (37) shows that the random second term in the probability limit depends on the Brownian motions to which partial sums of these common time effect processes converge. One major difference between (37) and the corresponding result (34) in the stationary case is that the random second term of (37) has a random denominator. Whereas the random second term in (34) has mean zero, this is not true of the random second term of (37). Only the numerator of the second term of (37) has mean zero as a simple calculation shows

$$E \left\{ \int_0^1 W_s(r) dr W_s(1) - 3 \int_0^1 \underline{W}_s(r)^2 dr \right\} = \frac{1}{2} - 3 \times \frac{1}{6} = 0.$$

The distribution of the term

$$\frac{1}{T+1} \frac{6 \sum_{s=1}^K \mu_{\delta_s}^2 \sigma_{\theta_s}^2 \left\{ \int_0^1 W_s(r) dr W_s(1) - 3 \int_0^1 \underline{W}_s(r)^2 dr \right\}}{\sigma^2 + 6 \sum_{s=1}^K \mu_{\delta_s}^2 \sigma_{\theta_s}^2 \int_0^1 \underline{W}_s(r)^2 dr}$$

in (37) is shown in Fig. 8. Again, the dispersion in the (random) inconsistency is substantial for small  $T$  and still appreciable for  $T = 20$ . As in the stationary case, the limit distribution apparently has a long left tail.



## 4.2 Incidental Trends

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

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

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

$$\begin{aligned} \tilde{w}_t &= w_t - \left\{ \frac{2(2T+1)}{T(T-1)} \left( \sum_{t=1}^T w_t \right) - \frac{6}{T(T-1)} \sum_{t=1}^T tw_t \right\} \\ &\quad - \left\{ -\frac{6}{T(T-1)} \sum_{t=1}^T w_t + \frac{12}{T(T^2-1)} \sum_{t=1}^T tw_t \right\} t \\ &= w_t - g_T^w - h_T^w t, \end{aligned}$$

where

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

Derivations similar to those of proposition 4 provide the following analogue of (31) and (34).

**Proposition 6** (*Incidental Trends with  $|\rho| < 1$* ) In model **M2-CSD** with errors  $u_{it}$  having the factor structure (30) and satisfying assumption **A2**, the pooled least squares estimate  $\hat{\rho}$  is inconsistent as  $N \rightarrow \infty$  and

$$plim_{N \rightarrow \infty} (\hat{\rho} - \rho) = - [\sigma^2 C(\rho, T) + \psi_{CT}] [\sigma^2 D(\rho, T) + \psi_{DT}]^{-1}, \quad (38)$$

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

$$\psi_{CT} = \sum_{s=1}^K \mu_{\delta_s}^2 \left\{ g_T^{\theta_s} \left( \sum_{t=1}^T Z_{\theta_s t} \right) + h_T^{\theta_s} \left( \sum_{t=1}^T t Z_{\theta_s t} \right) \right\}, \quad (39)$$

$$\psi_{DT} = \sum_{s=1}^K \mu_{\delta_s}^2 \sum_{t=1}^T \tilde{Z}_{\theta_s t}^2, \quad (40)$$

and where  $Z_{\theta_s t} = \sum_{j=0}^{\infty} \rho^j \theta_{st-j-1}$  and  $\tilde{Z}_{\theta_s t} = Z_{\theta_s t} - g_T^{Z_{\theta_s}} - h_T^{Z_{\theta_s}} t$  is detrended  $Z_{\theta_s t}$ . The inconsistency (38) has the following asymptotic representation as  $T \rightarrow \infty$

$$plim_{N \rightarrow \infty} (\hat{\rho} - \rho) = -2 \frac{1+\rho}{T} - \frac{1+\rho}{T} \frac{\sum_{s=1}^K \mu_{\delta_s}^2 \sigma_{s\theta}^2 (\eta_s^2 - 2)}{\sigma^2 + \sum_{s=1}^K \mu_{\delta_s}^2 \sigma_{s\theta}^2} + o_{a.s.} \left( \frac{1}{T} \right), \quad (41)$$

where  $\eta_s^2$  is iid  $\chi_2^2$  over  $s = 1, \dots, K$ .

As in the fixed effects case, the inconsistency is random and differs from the cross section independent (CSI) case by a term of  $O(T^{-1})$  involving a linear combination of independent centred  $\chi^2$  variates, each of which now have two degrees of freedom (reflecting the presence of both intercept and linear trend fixed effects in the panel regression). Again, since  $E\{\eta_s^2 - 2\} = 0$ , the mean inconsistency is the same as it is in the CSI case. Since the random component in (41) involves  $\chi_2^2$  variates, the variance of

$\sum_{s=1}^K \mu_{s\delta}^2 \sigma_{s\theta}^2 (\eta_s^2 - 2)$  is  $4 \sum_{s=1}^K \mu_{s\delta}^4 \sigma_{s\theta}^4$ , so that the random limit (41) will have greater dispersion than the limit (34) in the fixed effects case if

$$4 \sum_{s=1}^K \mu_{s\delta}^4 \sigma_{s\theta}^4 > \sum_{s=1}^K \mu_{s\delta}^2 \sigma_{s\theta}^2.$$

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  $\rho \rightarrow 1$ , in view of the randomness of the limit functions (39) and (41).

**Proposition 7** (*Incidental Trends with  $\rho = 1$* ) *In model M1-CSD with errors  $u_{it}$  having the factor structure (30) and satisfying assumption A2, the pooled least squares estimate  $\hat{\rho}$  is inconsistent as  $N \rightarrow \infty$  and*

$$plim_{N \rightarrow \infty}(\hat{\rho} - \rho) = -\frac{\sigma^2 C(T) + \phi_{CT}}{\sigma^2 D(T) + \phi_{DT}} \quad (42)$$

where

$$C(T) = \frac{1}{2}(T-2), \quad D(T) = \frac{(T-2)(T+2)}{15}$$

$$\phi_{CT} = \sum_{s=1}^K \mu_{s\delta}^2 \sum_{t=1}^T S_{t-1}^{\theta_s} (g_T^{\theta_s} + h_T^{\theta_s} t), \quad \phi_{DT} = \sum_{s=1}^K \mu_{s\delta}^2 \sum_{t=1}^T (\tilde{S}_{t-1}^{\theta_s})^2,$$

$S_{st}^\theta = \sum_{p=1}^t \theta_{sp}$  and  $\tilde{S}_{st}^\theta = S_{st}^\theta - g_T^{S_{\theta_s}} - h_T^{S_{\theta_s}} t$  is detrended  $S_{st}^\theta$ . The inconsistency (42) has the following asymptotic representation as  $T \rightarrow \infty$

$$plim_{N \rightarrow \infty}(\hat{\rho} - \rho) = -\frac{7.5}{T+2} - \frac{2}{T+2} \frac{\sum_{s=1}^K \mu_{s\delta}^2 \sigma_{\theta_s}^2 \left[ \int_0^1 (a_{\theta_s} + b_{\theta_s} r) dW_s(r) - 7.5 \int_0^1 \tilde{W}_s^2(r) dr \right]}{\sigma^2 + 15 \sum_{s=1}^K \mu_{s\delta}^2 \sigma_{\theta_s}^2 \int_0^1 \tilde{W}_s^2(r) dr} + o_{a.s.} \left( \frac{1}{T} \right) \quad (43)$$

where  $\{W_s : s = 1, \dots, K\}$  are independent standard Brownian motions and  $\tilde{W}_s(r) = W_s(r) - a_{\theta_s} - b_{\theta_s} r$  is detrended standard Brownian motion with coefficients

$$a_{\theta_s} = 4 \left\{ \int_0^1 W_s - \frac{3}{2} \int_0^1 r W_s \right\}, \quad b_{\theta_s} = 6 \left\{ 2 \int_0^1 r W_s - \int_0^1 W_s \right\} \quad (44)$$

Under cross section independence  $\mu_{\delta s}^2 = 0$  for all  $s$ , and (42) reduces to

$$plim_{N \rightarrow \infty}(\hat{\rho} - \rho) = -\frac{C(T)}{D(T)} = -\frac{7.5}{T+2},$$

corresponding to the earlier result (8). As in the fixed effects case (37), both the limit result (42) and the  $O(T^{-1})$  approximant (43) have random second terms involving random denominators. Here, the stochastic terms involve detrended Brownian motions, consonant with the detrending regression.

### 4.3 Exogenous Regressors

As in the CSI case, we start by using M3 and allow for errors  $u_{it}$  that satisfy Assumption A2:

**Model M3-CSD (Exogenous Variables)**  $\tilde{y}_{it} = \rho \tilde{y}_{it-1} + \tilde{Z}'_{it} \beta + \tilde{u}_{it} \quad \rho \in (-1, 1]$

Stacking observations, we have  $\tilde{y}_t = \rho\tilde{y}_{t-1} + \tilde{Z}'_t\beta + \tilde{u}_t$ , and  $\tilde{y} = \rho\tilde{y}_{-1} + \tilde{Z}\beta + \tilde{u}$ , say. Calculations similar to those in the preceding sections then lead to the following result on the inconsistency of these estimates. The notation is the same as that used earlier.

**Proposition 8** (*Exogenous Variables, Fixed and Trend Effects*) *As  $N \rightarrow \infty$ , for model M3 under Assumption A1 and with  $|\rho| < 1$ , the inconsistency of the pooled least squares estimate of  $\rho$  is given in the fixed effects case by*

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

and in the incidental trends case by

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

The inconsistency of the pooled estimate of  $\beta$  is

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

The formulae apply in the unit root case  $\rho = 1$  upon replacement of  $A(\rho, T)$ ,  $B(\rho, T)$ ,  $C(\rho, T)$ ,  $D(\rho, T)$  with  $A(T)$ ,  $B(T)$ ,  $C(T)$ ,  $D(T)$ , and  $\psi_{AT}$ ,  $\psi_{BT}$ ,  $\psi_{CT}$ ,  $\psi_{DT}$ , with  $\phi_{AT}$ ,  $\phi_{BT}$ ,  $\phi_{CT}$ ,  $\phi_{DT}$ , respectively, and  $\tilde{Z}_{\rho, -1}$  by  $\tilde{Z}_{1, -1}$ .

Again, the presence of exogenous variables generally reduces the extent of the inconsistency of  $\hat{\rho}$ , exceptions occurring when  $\beta = 0$  or when the model takes a components form like that of (26).

## 5 Conclusion and Remarks on Bias Correction

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  $N \rightarrow \infty$  through a random probability limit.

Against this background, bias correction methods or alternative approaches to estimation seem important considerations for practical work. In the case of cross section error independence, bias can be dealt with in a fairly straightforward manner by suitably designed correction mechanisms. For example, simple bias corrected estimates are produced by plugging in first stage estimates into bias formulae such as (6), (9), (21) or their asymptotic (in  $T$ ) approximants. Such methods have been suggested in earlier work on pooled least squares, instrumental variable and GMM estimators (see e.g., Kiviet, 1995; Hahn and Kuersteiner, 2002). It is also possible to use the formulae derived here to correct bias by inverting the mean function directly since in all the cases considered the mean function is monotonically increasing in  $\rho$ . Simulations we have conducted (not reported here) indicate that this method works very well and is generally much better than the use of plug-in corrections.

We illustrate the empirical effect of such bias correction in dynamic panel estimation by taking two applied studies of dynamic panels. First, in their study of the demand for natural gas, Balestra and Nerlove (1966) used annual panel data from 1950 to 1962 for 36 U.S. states and assuming cross section error independence, Balestra and Nerlove fitted the following panel regression equation (standard errors in parentheses)

$$G_{it} = \alpha_i + 0.68G_{it-1} - 0.203p_{it} - 0.014\Delta M_{it} + 0.033M_{it-1} + 0.013\Delta Y_{it} + 0.004Y_{it-1} + \text{error}$$

(0.063)      (0.053)      (0.022)      (0.005)      (0.008)      (0.01)

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 state, respectively. This model fits the framework of model M3. The coefficients of primary interest are  $\hat{\rho} = 0.68$ , which from the underlying theory equals  $1 - \hat{r}_g$ , where  $\hat{r}_g$  is the estimated depreciation rate of gas appliances, and  $\hat{\alpha}_1 = -0.203$ , which is the price elasticity of demand for gas consumption. While the information reported is not sufficient to implement the exact formulae given earlier for M3, we may compute an approximation to the M3 bias function (21) using  $G(\rho, T) = A(\rho, T)/B(\rho, T)$ , which is an upper bound to the bias and which is a valid approximation when  $B(\rho, T)$  dominates the denominator of (21). Inverting the mean function<sup>2</sup> of the pooled regression estimate gives a new estimate of 0.878 for  $\rho$  and the plug-in approximate mean unbiased estimate<sup>3</sup> gives the revised estimate  $0.82 = 0.68 + 1.68/12$ . Though they are only approximate in the present case, these corrections to  $\hat{\rho}$  appear substantial and they lead to very different implied depreciation rates.

As a second illustration, Frankel and Rose (1996) 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<sup>4</sup>

$$q_{it} = a_i + 0.88q_{it-1} + \text{error}$$

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. The plug-in approximate mean unbiased estimate gives an adjusted estimate of  $0.923 = 0.88 + 1.88/44$  which gives a half-life estimate of 8.6 years. Inverting the mean function gives the new estimate 0.934, for which the half life is 10.2 years, almost twice that of the original study. These adjustments indicate that PPP deviations are eroded much more slowly than the original estimates suggested.

Both the mean inversion and plug-in proposals for bias correction are designed for models with cross section independent errors. Under cross section dependence, as we have seen, the inconsistency is complicated by the presence of random elements. The random inconsistency of the autoregressive estimate cannot be so readily eliminated by mechanical bias correction procedures or by inversion. Indeed, when the inconsistency is random, issues of bias correction and variance reduction need to be addressed together. Earlier work by the authors (2002) showed that cross section error dependence can lead to very large increases in estimator variation. Given the prevalence of cross section dependence in

---

<sup>2</sup>Exact mean function tables have been computed for the constant and linear trend cases and are available at the web address: <http://yoda.eco.auckland.ac.nz/~dsul013/mf.htm>

<sup>3</sup>Here we use formula (7) and  $T = 13$ .

<sup>4</sup>See Frankel and Rose (1996, table 3 p. 219). The same results were also obtained in an equation with time-specific intercepts.

economic and financial panels, this matter is likely to be relevant in many practical studies and is an important issue to pursue in subsequent work.

## References

- [1] 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
- [2] Bai, J. and S. Ng (2002). “Determining the number of factors in approximate factor models,” *Econometrica*, 191-221
- [3] 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.
- [4] Baltagi, B. (2001). *Econometric Analysis of Panel Data*. New York: Wiley.
- [5] 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.
- [6] 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.
- [7] Hahn, J., J. Hausman and G. Kuersteiner (2001), “Bias corrected instrumental variables estimation for dynamic panel models with fixed effects,” MIT Working Paper
- [8] Hahn, J. and G. Kuersteiner (2002). “Asymptotically unbiased inference for a dynamic panel model with fixed effects,” *Econometrica*, 1639-1657.
- [9] 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.
- [10] Hsiao, C. (2003). *Analysis of Panel Data*. 2<sup>nd</sup> Edition. Cambridge: Cambridge University Press.
- [11] Kendall, M. G. (1954). “Note on the bias in the estimation of autocorrelation,” *Biometrika*, 403-404.
- [12] Kiviet, J. F. (1995). “On bias, inconsistency and efficiency of some estimators in dynamic panel data models”. *Journal of Econometrics*, 68, 53-78.
- [13] Kruiniger, H. (2000). “GMM estimation of dynamic panel data models with persistent data,” mimeo, Queen Mary, University of London.
- [14] Moon, H.R. and B. Perron, “Testing for a unit root in panels with dynamic factors,” mimeo, University of Southern California.
- [15] 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.
- [16] Moon, H.R. and P.C.B. Phillips (2000) : Estimation of autoregressive roots near unity using panel data, *Econometric Theory*, 16, 927–997.

- [17] Moon, H.R. and P.C.B. Phillips (2003) : GMM estimation of autoregressive roots near unity with panel data, Cowles Foundation Discussion Paper, No. 1390.
- [18] Neyman, J and E. Scott (1948): Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1–32.
- [19] Nickell, S. (1981): “Biases in dynamic models with fixed effects”, *Econometrica*, 49, 1417–1426.
- [20] 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.
- [21] Orcutt, G. H. and H. S. Winokur (1969). “First order autoregression: inference, estimation and prediction” . *Econometrica*, 37, 1-14.
- [22] Phillips, P.C.B. and V. Solo (1992). “Asymptotics for linear processes,” *Annals of Statistics*, 971–1001.
- [23] Phillips, P.C.B. and D. Sul (2002). “Dynamic panel estimation and homogeneity testing under cross section dependence,” *forthcoming The Econometrics Journal*.

## 6 Appendix

### 6.1 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  $t = 1, \dots, T$ . Then the panel least squares estimate of  $\rho$  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  $N \rightarrow \infty$  with  $T$  fixed. A typical term is evaluated in the following manner using a law of large numbers for heterogeneous sequences. First note that

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

Then we have

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

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  $\varepsilon_t$  is *iid*  $(0, \sigma^2)$ .

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 6.2 that the inconsistency as  $N \rightarrow \infty$  with  $T$  fixed has the form

$$\text{p lim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = \frac{\text{plim}_{N \rightarrow \infty} \frac{1}{N} C_{NT}}{\text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}} = -\frac{C(\rho, T)}{D(\rho, T)}, \quad (45)$$

where

$$C(\rho, T) = -\frac{1}{T-1} \frac{2}{1-\rho} \left[ (T-1) - \frac{2}{1-\rho} C_1 \right], \quad (46)$$

$$D(\rho, T) = \frac{T-2}{1-\rho^2} \left[ 1 - \frac{1}{T-2} \frac{4\rho}{1-\rho} D_1 \right], \quad (47)$$

with

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

Upon further algebraic reduction the rational function limit (45) has the following explicit form in terms of constituent polynomials in  $\rho$  and  $T$  :

$$H(\rho, T) = -\frac{C(\rho, T)}{D(\rho, T)} = -2\rho \frac{a_1 T^3 + a_2 T^2 + a_3 T + a_4}{b_1 T^4 + b_2 T^3 + b_3 T^2 + b_4 T + b_5}, \quad (48)$$

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 (46) and (47) for dominant terms yields the following approximant:

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

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

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

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

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

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

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)

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

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

$$\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = -\frac{\sigma^2 A(\rho, T)}{\sigma^2 B(\rho, T) + \beta' \left[ \text{plim}_{N \rightarrow \infty} \frac{1}{N} \tilde{Z}'_{\rho, -1} Q_{\tilde{Z}} \tilde{Z}_{\rho, -1} \right] \beta}, \quad (49)$$



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

When  $\rho = 1$ , we have

$$\lim_{\rho \rightarrow 1} A(\rho, T) = A(T) = \frac{(T-1)}{2}, \quad \lim_{\rho \rightarrow 1} B(\rho, T) = B(T) = \frac{(T-1)(T+1)}{6}, \quad (50)$$

so that (49) becomes

$$\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = -\frac{\sigma^2 A(T)}{\sigma^2 B(T) + \beta' \left[ \text{plim}_{N \rightarrow \infty} \frac{1}{N} \tilde{Z}'_{1,-1} Q_{\tilde{Z}} \tilde{Z}_{1,-1} \right] \beta}, \quad (51)$$

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

$$\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = -\frac{\sigma^2 C(T)}{\sigma^2 D(T) + \beta' \left[ \text{plim}_{N \rightarrow \infty} \frac{1}{N} \tilde{Z}'_{1,-1} Q_{\tilde{Z}} \tilde{Z}_{1,-1} \right] \beta}, \quad (52)$$

where

$$\lim_{\rho \rightarrow 1} C(\rho, T) = C(T) = \frac{1}{2}(T-2), \quad \lim_{\rho \rightarrow 1} D(\rho, T) = D(T) = \frac{1}{15}(T^2-4), \quad (53)$$

as in (93) and (90). Formula (23) for the inconsistency of  $\hat{\beta}$  continues to apply in the unit root case upon appropriate substitution of result (51) or (52).

**Proof of Proposition 4** It is convenient here to use sequential asymptotics with  $N \rightarrow \infty$  followed by  $T \rightarrow \infty$  and to employ an embedding argument. Write the panel least squares estimates under cross sectional dependence as

$$\hat{\rho} - \rho = \frac{A_{NT}^C}{B_{NT}^C}. \quad (54)$$

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

$$y_{it} = a_i + \rho y_{it-1} + u_{it}, \quad u_{it} = \delta_i \theta_t + \varepsilon_{it}.$$

Then, noting that

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T y_{it-1} u_{it} = \sum_{t=1}^T \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{it-1} u_{it} = \sum_{t=1}^T E(y_{it-1} u_{it}) = 0, \quad (55)$$

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i \sum_{t=1}^T \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \sum_{s=1}^T \varepsilon_{is} = \sum_{t=1}^T \sum_{s=1}^T \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i \varepsilon_{is} = 0, \quad (56)$$

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) \sum_{s=1}^T \theta_s = \sum_{t=1}^T \sum_{s=1}^T \theta_s \sum_{j=0}^{\infty} \rho^j \left[ \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{it-j-1} \right] = 0, \quad (57)$$

and we have

$$\begin{aligned}
\text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^C &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - y_{i,t-1})(u_{it} - u_{i,t-1}) = -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T y_{it-1}(u_{it} - u_{i,t-1}) \\
&= -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j u_{it-j-1} \right) \sum_{s=1}^T u_{is} \\
&= -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left( \left[ \frac{\delta_i^2}{T} \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) \sum_{s=1}^T \theta_s \right] + \left[ \frac{1}{T} \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) \sum_{s=1}^T \varepsilon_{is} \right] \right) \\
&\quad - \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{\delta_i}{T} \left( \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) \sum_{s=1}^T \varepsilon_{is} + \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) \sum_{s=1}^T \theta_s \right) \\
&= -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \frac{\delta_i^2}{T} \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) \sum_{s=1}^T \theta_s \right] - \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{T} \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) \sum_{s=1}^T \varepsilon_{is} \right] \\
&= -\sigma^2 A(\rho, T) - \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \frac{\delta_i^2}{T} \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) \sum_{s=1}^T \theta_s \right], \tag{58}
\end{aligned}$$

where

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

as in (3). Setting  $Z_{\theta t} = \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1}$  and using the fact that  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 = \mu_\delta^2$ , (58) becomes

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^C = -\sigma^2 A(\rho, T) - \mu_\delta^2 \frac{1}{T} \sum_{t=1}^T Z_{\theta t} \sum_{s=1}^T \theta_s. \tag{59}$$

By taking sequential limits as  $N \rightarrow \infty$  followed by  $T \rightarrow \infty$ , we have the further result

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \frac{\delta_i^2}{T} \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) \sum_{s=1}^T \theta_s \right] = \mu_\delta^2 \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{\theta t} \right) \left( \frac{1}{\sqrt{T}} \sum_{s=1}^T \theta_s \right) \rightarrow \mu_\delta^2 \xi_\theta \eta_\theta, \tag{60}$$

where, by standard central limit theory, as  $T \rightarrow \infty$  we have

$$\left[ \begin{array}{c} \frac{1}{\sqrt{T}} \sum_{s=1}^T \theta_s \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{\theta t} \end{array} \right] \rightarrow_d \left[ \begin{array}{c} \xi_\theta \\ \zeta_\theta \end{array} \right] = N \left( 0, \left[ \begin{array}{cc} \sigma_\theta^2 & \frac{\sigma_\theta^2}{1-\rho} \\ \frac{\sigma_\theta^2}{1-\rho} & \frac{\sigma_\theta^2}{(1-\rho)^2} \end{array} \right] \right). \tag{61}$$

Since  $\xi_\theta$  is perfectly correlated with  $\eta_\theta$ , we can write

$$\xi_\theta \zeta_\theta = \xi_\theta^2 \frac{1}{1-\rho} \quad a.s.$$

By suitable augmentation of the probability space and embedding arguments, we may now write the convergence in (61) as an almost sure convergence and then (60) can be written as

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \frac{\delta_i^2}{T} \sum_{t=1}^T Z_{\theta t} \sum_{s=1}^T \theta_s \right] = \mu_\delta^2 \xi_\theta^2 \frac{1}{1-\rho} + o_{a.s.}(1),$$

where the final term is  $o_{a.s.}(1)$  as  $T \rightarrow \infty$ . Hence, we have the asymptotic ( $N, T \rightarrow \infty$ ) approximation

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^C = -\frac{\sigma^2}{1-\rho} - \mu_\delta^2 \xi_\theta^2 \frac{1}{1-\rho} + o_{a.s.}(1), \text{ as } T \rightarrow \infty. \quad (62)$$

Note that the second term in this limit is a random quantity, while the leading term is  $-\lim_{T \rightarrow \infty} A(\rho, T)$ , the same expression as that occurring in the cross section independent case.

Dealing with the denominator in a similar fashion, we get

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^C &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j u_{it-j-1} \right)^2 - \frac{1}{T} \left( \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j u_{it-j-1} \right) \right)^2 \right] \\ &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right)^2 - \frac{1}{T} \left( \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) \right)^2 \right] \\ &\quad + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right)^2 - \frac{1}{T} \left( \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) \right)^2 \right]. \end{aligned}$$

Note that

$$\begin{aligned} \sigma^2 B(\rho, T) &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right)^2 - \frac{1}{T} \left( \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) \right)^2 \right] \\ &= \sigma^2 \frac{T-1}{1-\rho^2} \left\{ 1 - \frac{1}{T-1} \frac{2\rho}{1-\rho} \left[ 1 - \frac{1}{T} \frac{1-\rho^T}{1-\rho} \right] \right\}, \end{aligned} \quad (63)$$

as in (4). We then have

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^C &= \sigma^2 B(\rho, T) + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T (Z_{\theta t} - \bar{Z}_\theta)^2 \right] \\ &= \sigma^2 B(\rho, T) + \mu_\delta^2 \sum_{t=1}^T (Z_{\theta t} - \bar{Z}_\theta)^2. \end{aligned} \quad (64)$$

Letting  $T \rightarrow \infty$  we have

$$\frac{1}{T} \sum_{t=1}^T (Z_{\theta t} - \bar{Z}_\theta)^2 \xrightarrow{a.s.} E(Z_{\theta t}^2) = \frac{\sigma_\theta^2}{1-\rho^2},$$

and then

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T (Z_{\theta t} - \bar{Z}_\theta)^2 \right] = T \left[ \mu_\delta^2 \frac{\sigma_\theta^2}{1-\rho^2} + o_{a.s.}(1) \right], \text{ as } T \rightarrow \infty. \quad (65)$$

Combining (65) with (63) and (64) yields

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^C = \frac{T}{1-\rho^2} [\sigma^2 + \mu_\delta^2 \sigma_\theta^2 + o_{a.s.}(1)], \text{ as } T \rightarrow \infty. \quad (66)$$

The proposition now follows. First, we have from (59) and (64)

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) &= \frac{\text{plim}_{N \rightarrow \infty} A_{NT}^C}{\text{plim}_{N \rightarrow \infty} B_{NT}^C} \\ &= -\frac{\sigma^2 A(\rho, T) + \mu_\delta^2 \frac{1}{T} \sum_{t=1}^T Z_{\theta t} \sum_{s=1}^T \theta_s}{\sigma^2 B(\rho, T) + \mu_\delta^2 \sum_{t=1}^T (Z_{\theta t} - \bar{Z}_\theta)^2}, \end{aligned}$$

which gives result (31). Second, combining (62) and (66), we have

$$\begin{aligned} \text{plim}_{N \rightarrow \infty}(\hat{\rho} - \rho) &= -\frac{1}{T} \left[ \frac{\sigma^2}{1-\rho} + \mu_\delta^2 \xi_\theta^2 \frac{1}{1-\rho} + o_{a.s.}(1) \right] \left[ \frac{\sigma^2}{1-\rho^2} + \mu_\delta^2 \frac{\sigma_\theta^2}{1-\rho^2} + o_{a.s.}(1) \right]^{-1} \\ &= -\frac{1+\rho}{T} \left[ 1 + \frac{\mu_\delta^2}{\sigma^2} \sigma_\theta^2 - \frac{\mu_\delta^2}{\sigma^2} \sigma_\theta^2 + \frac{\mu_\delta^2}{\sigma^2} \xi_\theta^2 \right] \left[ 1 + \frac{\mu_\delta^2}{\sigma^2} \sigma_\theta^2 \right]^{-1} + o_{a.s.} \left( \frac{1}{T} \right) \\ &= -\frac{1+\rho}{T} - \frac{1+\rho}{T} \frac{\mu_\delta^2 \sigma_\theta^2}{\sigma^2 + \mu_\delta^2 \sigma_\theta^2} [\eta_\theta^2 - 1] + o_{a.s.} \left( \frac{1}{T} \right), \end{aligned}$$

where  $\eta_\theta^2 = \xi_\theta^2 / \sigma_\theta^2$  is  $\chi_1^2$ , giving (34).

The extension to the multi factor case ( $K > 1$ ) is straightforward and is omitted.

**Proof of Proposition 5** The pooled least squares estimate satisfies (54) and, since  $y_{it} = a_i + y_{it}^0$  and  $y_{it}^0 = \sum_{s=1}^t u_{is} + y_{i0}^0 = S_{it} + y_{i0}^0$ , we have

$$A_{NT}^C = \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - y_{i \cdot -1}) (u_{it} - u_{i \cdot}) = \sum_{i=1}^N \sum_{t=1}^T S_{it-1} (u_{it} - u_{i \cdot}).$$

As in the proof of Proposition 4,  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N S_{it-1} u_{it} = 0$ , so that

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^C = -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T S_{it-1} \sum_{t=1}^T u_{it}.$$

We proceed with the one factor ( $K = 1$ ) model, the multi-factor ( $K > 1$ ) case following in a straightforward way. Using the decomposition  $S_{it} = \delta_i \sum_{s=1}^t \theta_s + \sum_{s=1}^t \varepsilon_{is} := \delta_i S_{it}^\theta + S_{it}^\varepsilon$ , we can write

$$\begin{aligned} \sum_{i=1}^N \sum_{t=1}^T S_{it-1} \sum_{t=1}^T u_{it} &= \sum_{i=1}^N S_{i \cdot -1} \delta_i \sum_{t=1}^T \theta_t - \sum_{i=1}^N S_{i \cdot -1} \sum_{t=1}^T \varepsilon_{it} \\ &= \sum_{i=1}^N \delta_i^2 S_{i \cdot -1}^\theta \sum_{t=1}^T \theta_t + \sum_{i=1}^N \delta_i S_{i \cdot -1}^\varepsilon \sum_{t=1}^T \theta_t \\ &\quad - \sum_{i=1}^N \delta_i S_{i \cdot -1}^\theta \sum_{t=1}^T \varepsilon_{it} - \sum_{i=1}^N S_{i \cdot -1}^\varepsilon \sum_{t=1}^T \varepsilon_{it}, \end{aligned}$$

and so

$$\begin{aligned} \text{p lim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^C &= - \left( \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \right) \left( S_{\cdot -1}^\theta \sum_{t=1}^T \theta_t \right) - \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N S_{i \cdot -1}^\varepsilon \sum_{t=1}^T \varepsilon_{it} \\ &= -\mu_\delta^2 \left( \frac{1}{T} \sum_{p=1}^T S_{p-1}^\theta \sum_{t=1}^T \theta_t \right) - \text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^\varepsilon, \text{ say.} \end{aligned}$$

Since  $\varepsilon_{it}$  is independent over  $i$  and  $t$ , we have  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{it-j-1} \varepsilon_{it} = 0$  for all  $j \geq 0$ , and then

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^\varepsilon &= -\frac{1}{T} \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left( \sum_{j=0}^{t-1} \varepsilon_{it-j-1} \right) \sum_{t=1}^T \varepsilon_{it} \right] \\ &= -\frac{1}{T} \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \right) [1 + 2 + 3 + \cdots + T - 1] \\ &= -\frac{\sigma^2}{T} \sum_{i=1}^{T-1} t = -\frac{\sigma^2}{2} (T-1) := -\sigma^2 A(T), \text{ say.} \end{aligned} \quad (67)$$

It follows that

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^C = -\sigma^2 A(T) - \mu_\delta^2 \left( \frac{1}{T} \sum_{p=1}^T S_{p-1}^\theta \sum_{t=1}^T \theta_t \right) := -\sigma^2 A(T) - \phi_{AT} \quad (68)$$

When there is no cross section dependence  $\mu_\delta^2 = 0$ , and the first term of (68) is the corresponding result for the numerator limit in the CSI unit root case.

Dealing with the denominator of (54) in a similar way, we get  $S_{it} = \delta_i \sum_{s=1}^t \theta_s + \sum_{s=1}^t \varepsilon_{is} := \delta_i S_t^\theta + S_{it}^\varepsilon$ , and then

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^C &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - y_{i \cdot -1})^2 \\ &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T S_{it-1}^2 - \frac{1}{T} \left( \sum_{t=1}^T S_{it-1} \right)^2 \right] \\ &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T (S_{t-1}^\theta)^2 - \frac{1}{T} \left( \sum_{t=1}^T S_{t-1}^\theta \right)^2 \right] \\ &\quad + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T (S_{it-1}^\varepsilon)^2 - \frac{1}{T} \left( \sum_{t=1}^T S_{it-1}^\varepsilon \right)^2 \right] \\ &= \mu_\delta^2 \sum_{t=1}^T (S_{t-1}^\theta - S_{\cdot -1}^\theta)^2 + \text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^\varepsilon, \end{aligned} \quad (69)$$

since cross product terms of  $S_{it}^\varepsilon$  and  $\delta_i S_t^\theta$  have zero probability limit as  $N \rightarrow \infty$ , and where  $B_{NT}^\varepsilon = \sum_{i=1}^N \sum_{t=1}^T (S_{it-1}^\varepsilon - S_{i \cdot -1}^\varepsilon)^2$ . Decompose the second term of (69) as follows

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^\varepsilon &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \left( \sum_{j=0}^{t-1} \varepsilon_{it-j-1} \right)^2 - \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{T} \left( \sum_{t=1}^T \left( \sum_{j=0}^{t-1} \varepsilon_{it-j-1} \right) \right)^2 \\ &= B_{aT} - B_{bT}, \end{aligned}$$

where

$$\begin{aligned}
B_{aT} &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left( \sum_{j=0}^{t-1} \varepsilon_{it-j-1} \right)^2 \right] \\
&= \sum_{t=1}^T \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left( \sum_{j=0}^{t-1} \varepsilon_{it-j-1} \right)^2 \\
&= \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \right) \sum_{t=1}^T (t-1) = \sigma^2 \frac{T(T-1)}{2},
\end{aligned} \tag{70}$$

and

$$\begin{aligned}
B_{bT} &= \frac{1}{T} \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left( \sum_{t=1}^T S_{it-1}^\varepsilon \right)^2 \\
&= \frac{1}{T} \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left( \sum_{t=1}^T S_{it-1}^{\varepsilon^2} + 2 \sum_{t=1}^{T-1} S_{it-1}^\varepsilon S_{it}^\varepsilon + \cdots + 2 \sum_{t=1}^1 S_{it-1}^\varepsilon S_{iT-1}^\varepsilon \right) \\
&= \frac{\sigma^2}{T} \left[ \frac{T(T-1)}{2} + 2 \sum_{s=1}^{T-2} \sum_{r=s+1}^T r \right] = \frac{\sigma^2}{T} \left[ \frac{T(T-1)}{2} + 2 \sum_{s=1}^{T-2} \frac{s(s+1)}{2} \right] \\
&= \frac{\sigma^2 (T-1)(2T-1)}{6}.
\end{aligned} \tag{71}$$

Combining (70) and (71), we have

$$\begin{aligned}
\text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^\varepsilon &= \sigma^2 \frac{T(T-1)}{2} - \frac{\sigma^2 (T-1)(2T-1)}{6} \\
&= \sigma^2 \frac{(T-1)(T+1)}{6} = \sigma^2 B(T), \text{ say.}
\end{aligned}$$

Thus

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^C = B(T) + \mu_\delta^2 \sum_{t=1}^T (S_{t-1}^\theta - S_{\cdot-1}^\theta)^2 = \sigma^2 B(T) + \phi_{BT}. \tag{72}$$

The stated result (36) now follows by combining (68) and (72), giving

$$\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = \frac{\text{plim}_{N \rightarrow \infty} \frac{1}{N} A_{NT}^C}{\text{plim}_{N \rightarrow \infty} \frac{1}{N} B_{NT}^C} = - \frac{\sigma^2 A(T) + \phi_{AT}}{\sigma^2 B(T) + \phi_{BT}},$$

as required.

By standard weak convergence arguments (e.g., Phillips and Solo, 1992), we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T \cdot \rfloor} \theta_t \rightarrow_d \sigma_\theta W(\cdot) = BM(\sigma_\theta^2),$$

and then

$$T^{-2} \phi_{BT} = \mu_\delta^2 T^{-2} \sum_{t=1}^T (S_{t-1}^\theta - S_{\cdot-1}^\theta)^2 \rightarrow_d \mu_\delta^2 \sigma_\theta^2 \int_0^1 \underline{W}(r)^2 dr \tag{73}$$

where  $\underline{W}(r) = W(r) - \int_0^1 W$  is demeaned standard Brownian motion. Also,

$$T^{-1} \phi_{AT} = \mu_\delta^2 \frac{1}{T} \sum_{p=1}^T \frac{S_{p-1}^\theta}{\sqrt{T}} \sum_{t=1}^T \frac{\theta_t}{\sqrt{T}} \rightarrow_d \mu_\delta^2 \sigma_\theta^2 \int_0^1 W(r) dr W(1). \tag{74}$$

By suitable augmentation of the probability space and embedding arguments, we can write the weak convergence in (73) and (74) as almost sure convergence and then we may write

$$\begin{aligned}\frac{\sigma^2 A(T) + \phi_{AT}}{(T-1)(T+1)} &= \frac{1}{T+1} \left\{ \frac{\sigma^2}{2} + \frac{1}{T-1} \phi_{AT} \right\} \\ &= \frac{1}{T+1} \left\{ \frac{\sigma^2}{2} + \mu_\delta^2 \sigma_\theta^2 \int_0^1 W(r) dr W(1) + o_{a.s.}(1) \right\},\end{aligned}$$

and

$$\begin{aligned}\frac{\sigma^2 B(T) + \phi_{BT}}{(T-1)(T+1)} &= \frac{1}{6} \sigma^2 + \frac{\mu_\delta^2}{(T-1)(T+1)} \sum_{t=1}^T (S_{t-1}^\theta - S_{\cdot-1}^\theta)^2 \\ &= \frac{1}{6} \sigma^2 + \mu_\delta^2 \sigma_\theta^2 \int_0^1 \underline{W}(r)^2 dr + o_{a.s.}(1).\end{aligned}$$

It follows that

$$\begin{aligned}-\frac{A(T) + \phi_{AT}}{B(T) + \phi_{BT}} &= -\frac{1}{T+1} \frac{\frac{\sigma^2}{2} + \mu_\delta^2 \sigma_\theta^2 \int_0^1 W(r) dr W(1) + o_{a.s.}(1)}{\frac{1}{6} \sigma^2 + \mu_\delta^2 \sigma_\theta^2 \int_0^1 \underline{W}(r)^2 dr + o_{a.s.}(1)} \\ &= -\frac{1}{T+1} \frac{3\sigma^2 + 18\mu_\delta^2 \sigma_\theta^2 \int_0^1 \underline{W}(r)^2 dr + 6\mu_\delta^2 \sigma_\theta^2 \left[ \int_0^1 W(r) dr W(1) - 3 \int_0^1 \underline{W}(r)^2 dr \right]}{\sigma^2 + 6\mu_\delta^2 \sigma_\theta^2 \int_0^1 \underline{W}(r)^2 dr} + o_{a.s.}\left(\frac{1}{T}\right) \\ &= -\frac{3}{T+1} - \frac{1}{T+1} \frac{6\mu_\delta^2 \sigma_\theta^2 \left\{ \int_0^1 W(r) dr W(1) - 3 \int_0^1 \underline{W}(r)^2 dr \right\}}{\sigma^2 + 6\mu_\delta^2 \sigma_\theta^2 \int_0^1 \underline{W}(r)^2 dr} + o_{a.s.}\left(\frac{1}{T}\right),\end{aligned}\quad (75)$$

giving result (37).

Simple calculations show that  $E\left(\int_0^1 W(r) dr W(1)\right) = \frac{1}{2}$ , and

$$E\left\{\int_0^1 \underline{W}(r)^2 dr\right\} = E\left\{\int_0^1 W(r)^2 dr - \left(\int_0^1 W(r) dr\right)^2\right\} = \frac{1}{2} - \frac{1}{3} = \frac{1}{6},$$

so that

$$E\left\{\int_0^1 W(r) dr W(1) - 3 \int_0^1 \underline{W}(r)^2 dr\right\} = 0.$$

However, the mean of the second term in (75) is not zero.

**Proof of Proposition 6** Write

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

and

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

We derive an explicit form for the inconsistency

$$\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = \text{plim}_{N \rightarrow \infty} \frac{\frac{1}{N} C_{NT}^C}{\frac{1}{N} D_{NT}^C}. \quad (76)$$

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

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

where

$$g_T^{x_i} = \frac{2(2T+1)}{T(T-1)} \left( \sum_{t=1}^T x_{it} \right) - \frac{6}{T(T-1)} \sum_{t=1}^T t x_{it}, \quad h_T^{x_i} = \frac{12}{T(T^2-1)} \sum_{t=1}^T t x_{it} - \frac{6}{T(T-1)} \sum_{t=1}^T x_{it}, \quad (77)$$

and this notation will be used extensively below. When  $K = 1$ , we have  $u_{it} = \delta_i \theta_t + \varepsilon_{it}$  and then

$$x_{it} = \delta_i \sum_{j=0}^{\infty} \rho^j \theta_{t-j} + \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j} = \delta_i x_{it}^{\theta} + x_{it}^{\varepsilon}, \text{ say,}$$

from which we deduce that  $\tilde{x}_{it} = \delta_i \tilde{x}_{it}^{\theta} + \tilde{x}_{it}^{\varepsilon}$ . Using (55) – (57), proceeding as in the proof of Proposition 4, and working first with the denominator, we have

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^c &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \hat{y}_{it-1}^2 = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{x}_{it-1}^2 \\ &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (\delta_i \tilde{x}_{it-1}^{\theta} + \tilde{x}_{it-1}^{\varepsilon})^2 \\ &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \sum_{t=1}^T \left\{ (\tilde{x}_{it-1}^{\theta})^2 + (\tilde{x}_{it-1}^{\varepsilon})^2 \right\} \\ &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T (\tilde{x}_{it-1}^{\theta})^2 \right] + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T (\tilde{x}_{it-1}^{\varepsilon})^2 \right]. \end{aligned}$$

From the proof of Proposition 2 we have

$$\sigma^2 D(\rho, T) = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T (\tilde{x}_{it-1}^{\varepsilon})^2 \right] = \sigma^2 \frac{T-2}{1-\rho^2} \left[ 1 - \frac{1}{T-2} \frac{4\rho}{1-\rho} D_1 \right], \quad (78)$$

where  $D_1$  is defined in (12). We then have

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^c &= \sigma^2 D(\rho, T) + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T \tilde{Z}_{\theta t}^2 \right] \\ &= \sigma^2 D(\rho, T) + \mu_{\delta}^2 \sum_{t=1}^T \tilde{Z}_{\theta t}^2, \end{aligned} \quad (79)$$



where  $Z_{\theta t} = \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1}$  and  $\tilde{Z}_{\theta t} = Z_{\theta t} - g_T^{Z_{\theta}} - h_T^{Z_{\theta}} t$  is detrended  $Z_{\theta t}$ . Letting  $T \rightarrow \infty$  we have

$$\frac{1}{T} \sum_{t=1}^T \tilde{Z}_{\theta t}^2 \rightarrow_{a.s.} E(Z_{\theta t}^2) = \frac{\sigma_{\theta}^2}{1-\rho^2},$$

and then

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T \tilde{Z}_{\theta t}^2 \right] = T \left[ \mu_{\delta}^2 \frac{\sigma_{\theta}^2}{1-\rho^2} + o_{a.s.}(1) \right], \text{ as } T \rightarrow \infty. \quad (80)$$

Combining (80) and (78) with (79) yields

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^C &= \sigma^2 D(\rho, T) + \frac{T}{1-\rho^2} \mu_{\delta}^2 \sigma_{\theta}^2 [1 + o_{a.s.}(1)], \text{ as } T \rightarrow \infty. \\ &= \frac{T}{1-\rho^2} \{ \sigma^2 + \mu_{\delta}^2 \sigma_{\theta}^2 + o_{a.s.}(1) \}, \text{ as } T \rightarrow \infty \end{aligned} \quad (81)$$

Turning to the numerator of (76), we have

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} C_{NT}^C &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{it-1} \tilde{u}_{it} = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{x}_{it-1} \tilde{u}_{it} \\ &= -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T x_{it-1} \tilde{u}_{it} \\ &= -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left\{ \left( \sum_{j=0}^{\infty} \rho^j u_{it-j-1} \right) (g_T^{u_i} + h_T^{u_i} t) \right\} \right] \\ &= -\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \delta_i^2 \sum_{t=1}^T \left\{ \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) (g_T^{\theta} + h_T^{\theta} t) \right\} \right] \\ &\quad - \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left\{ \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) (g_T^{\varepsilon_i} + h_T^{\varepsilon_i} t) \right\} \right] \\ &= -\sigma^2 C(\rho, T) - \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \delta_i^2 \sum_{t=1}^T \left( \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1} \right) (g_T^{\theta} + h_T^{\theta} t) \right], \end{aligned} \quad (82)$$

where, from the proof of Proposition 2, we have

$$\begin{aligned} \sigma^2 C(\rho, T) &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T x_{it-1}^{\varepsilon} \tilde{\varepsilon}_{it} \\ &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left\{ \left( \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j-1} \right) (g_T^{\varepsilon_i} + h_T^{\varepsilon_i} t) \right\} \right] \\ &= \frac{\sigma^2}{T-1} \frac{2}{1-\rho} \left[ (T-1) - \frac{2}{1-\rho} C_1 \right], \end{aligned}$$

and  $C_1$  is defined in (11). Setting  $Z_{\theta t} = \sum_{j=0}^{\infty} \rho^j \theta_{t-j-1}$  as before and using the fact that  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 = \mu_{\delta}^2$ , (83) becomes

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} C_{NT}^C = -\sigma^2 C(\rho, T) - \mu_{\delta}^2 g_T^{\theta} \left( \sum_{t=1}^T Z_{\theta t} \right) - \mu_{\delta}^2 h_T^{\theta} \left( \sum_{t=1}^T t Z_{\theta t} \right). \quad (84)$$

From (81) and (84) we have

$$\begin{aligned} \text{plim}_{N \rightarrow \infty}(\hat{\rho} - \rho) &= \frac{\text{plim}_{N \rightarrow \infty} C_{NT}^C}{\text{plim}_{N \rightarrow \infty} D_{NT}^C} \\ &= -\frac{\sigma^2 C(\rho, T) + \mu_\delta^2 \left\{ g_T^\theta \left( \sum_{t=1}^T Z_{\theta t} \right) + h_T^\theta \left( \sum_{t=1}^T t Z_{\theta t} \right) \right\}}{\sigma^2 D(\rho, T) + \mu_\delta^2 \sum_{t=1}^T \tilde{Z}_{\theta t}^2}, \end{aligned}$$

and (39) of the Proposition follows.

Next we take sequential limits with  $N \rightarrow \infty$  followed by  $T \rightarrow \infty$ . By standard limit arguments as  $T \rightarrow \infty$  we have

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T \theta_t \\ \frac{1}{\sqrt{T^3}} \sum_{t=1}^T t \theta_t \end{bmatrix} \rightarrow_d \begin{bmatrix} B_\theta(1) \\ \int_0^1 r B_\theta(r) dr \end{bmatrix} := \begin{bmatrix} \xi_\theta \\ \nu_\theta \end{bmatrix} = N \left( 0, \begin{bmatrix} \sigma_\theta^2 & \frac{1}{2} \sigma_\theta^2 \\ \frac{1}{2} \sigma_\theta^2 & \frac{1}{3} \sigma_\theta^2 \end{bmatrix} \right), \quad (85)$$

where  $B_\theta$  is Brownian motion with variance  $\sigma_\theta^2$ , and

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{\theta t} \\ \frac{1}{\sqrt{T^3}} \sum_{t=1}^T t Z_{\theta t} \end{bmatrix} \rightarrow_d \frac{1}{1-\rho} \begin{bmatrix} B_\theta(1) \\ \int_0^1 r B_\theta(r) dr \end{bmatrix} = \frac{1}{1-\rho} \begin{bmatrix} \xi_\theta \\ \nu_\theta \end{bmatrix}. \quad (86)$$

Further

$$\begin{aligned} & g_T^\theta \left( \sum_{t=1}^T Z_{\theta t} \right) + h_T^\theta \left( \sum_{t=1}^T t Z_{\theta t} \right) \\ &= \left\{ \frac{2(2T+1)}{T(T-1)} \left( \sum_{t=1}^T \theta_t \right) - \frac{6}{T(T-1)} \sum_{t=1}^T t \theta_t \right\} \left( \sum_{t=1}^T Z_{\theta t} \right) \\ & \quad + \left\{ \frac{12}{T(T^2-1)} \sum_{t=1}^T t \theta_t - \frac{6}{T(T-1)} \sum_{t=1}^T \theta_t \right\} \left( \sum_{t=1}^T t Z_{\theta t} \right) \\ & \rightarrow_d \frac{1}{1-\rho} \{4\xi_\theta - 6\nu_\theta\} \xi_\theta + \frac{1}{1-\rho} \{12\nu_\theta - 6\xi_\theta\} \nu_\theta \\ &= \frac{1}{1-\rho} \{4\xi_\theta^2 - 12\xi_\theta \nu_\theta + 12\nu_\theta^2\} \\ &= \frac{1}{1-\rho} [\xi_\theta, \nu_\theta] \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \begin{bmatrix} \xi_\theta \\ \nu_\theta \end{bmatrix} \\ &= \frac{\sigma_\theta^2}{1-\rho} [\xi_\theta, \nu_\theta] \begin{bmatrix} \sigma_\theta^2 & \frac{1}{2} \sigma_\theta^2 \\ \frac{1}{2} \sigma_\theta^2 & \frac{\sigma_\theta^2}{3} \end{bmatrix}^{-1} \begin{bmatrix} \xi_\theta \\ \nu_\theta \end{bmatrix} \\ &= \frac{\sigma_\theta^2}{1-\rho} \eta^2, \end{aligned} \quad (87)$$

where  $\eta^2$  is a  $\chi_2^2$  variate.

It follows from (84) and (87), using the same embedding argument as in the proof of Proposition 4, that

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} C_{NT}^C &= -\sigma^2 C(\rho, T) - \frac{\mu_\delta^2 \sigma_\theta^2 \eta^2}{1-\rho} [1 + o_{a.s.}(1)], \text{ as } T \rightarrow \infty \\ &= -\frac{\sigma^2}{T-1} \frac{2}{1-\rho} \left[ (T-1) - \frac{2}{1-\rho} C_1 \right] - \frac{\mu_\delta^2 \sigma_\theta^2 \eta^2}{1-\rho} + o_{a.s.}(1) \end{aligned} \quad (88)$$

Combine (81) and (88), we get

$$\begin{aligned}
\text{plim}_{N \rightarrow \infty}(\hat{\rho} - \rho) &= \frac{-\sigma^2 C(\rho, T) - \frac{\mu_\delta^2 \sigma_\theta^2 \eta^2}{1-\rho} + o_{a.s.}(1)}{\frac{T}{1-\rho^2} \{\sigma^2 + \mu_\delta^2 \sigma_\theta^2 + o_{a.s.}(1)\}} \\
&= -\frac{\frac{2\sigma^2}{1-\rho} + \frac{\mu_\delta^2 \sigma_\theta^2 \eta^2}{1-\rho} + o_{a.s.}(1)}{\frac{T}{1-\rho^2} \{\sigma^2 + \mu_\delta^2 \sigma_\theta^2 + o_{a.s.}(1)\}} \\
&= -\frac{\frac{1}{1-\rho} \{2\sigma^2 + 2\mu_\delta^2 \sigma_\theta^2 + \mu_\delta^2 \sigma_\theta^2 (\eta^2 - 2) + o_{a.s.}(1)\}}{\frac{T}{1-\rho^2} \{\sigma^2 + \mu_\delta^2 \sigma_\theta^2 + o_{a.s.}(1)\}} \\
&= -2\frac{1+\rho}{T} - \frac{1+\rho}{T} \frac{\mu_\delta^2 \sigma_\theta^2 (\eta^2 - 2)}{\sigma^2 + \mu_\delta^2 \sigma_\theta^2} + o_{a.s.}\left(\frac{1}{T}\right),
\end{aligned}$$

which gives the required approximant (41) as  $T \rightarrow \infty$  when  $K = 1$ . The case  $K > 1$  follows in a straightforward manner.

**Proof of Proposition 7** The model is  $y_{it} = a_i + b_i t + x_{it}$ , where  $x_{it} = x_{it-1} + u_{it} = \sum_{s=1}^t u_{is} + x_{i0} = S_{it} + x_{i0}$  and  $u_{it} = \delta_i \theta_t + \varepsilon_{it}$  when  $K = 1$ . The panel least squares estimate is

$$\text{plim}_{N \rightarrow \infty}(\hat{\rho} - \rho) = \frac{\text{plim}_{N \rightarrow \infty} \frac{1}{N} C_{NT}^C}{\text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^C}. \quad (89)$$

As before,  $S_{it} = \delta_i \sum_{s=1}^t \theta_s + \sum_{s=1}^t \varepsilon_{is} := \delta_i S_t^\theta + S_t^\varepsilon$  and we have

$$\begin{aligned}
\text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^C &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{it-1}^2 = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{x}_{it-1}^2 \\
&= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \tilde{S}_{it-1}^2 \right] \\
&= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 \left[ \sum_{t=1}^T \left( \tilde{S}_{t-1}^\theta \right)^2 \right] + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \left( \tilde{S}_{it-1}^\varepsilon \right)^2 \right] \\
&= \mu_\delta^2 \sum_{t=1}^T \left( \tilde{S}_{t-1}^\theta \right)^2 + \text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^\varepsilon,
\end{aligned}$$

since cross product terms of  $S_{it}^\varepsilon$  and  $\delta_i S_t^\theta$  have zero probability limit as  $N \rightarrow \infty$ . As in the unit root CSI case, we get

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^\varepsilon = \sigma^2 D(T) = \frac{1}{15} \sigma^2 (T^2 - 4), \quad (90)$$

so that

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} D_{NT}^C = \frac{1}{15} \sigma^2 (T^2 - 4) + \mu_\delta^2 \sum_{t=1}^T \left( \tilde{S}_{t-1}^\theta \right)^2. \quad (91)$$

Next,

$$\begin{aligned}
\text{p lim}_{N \rightarrow \infty} \frac{1}{N} C_{NT}^C &= -\text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T S_{it-1} \tilde{u}_{it} \\
&= -\text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \{S_{it-1} (g_T^{u_i} + h_T^{u_i} t)\} \right] \\
&= -\text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \delta_i^2 \sum_{t=1}^T \{S_{t-1}^\theta (g_T^\theta + h_T^\theta t)\} \right] \\
&\quad -\text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \{S_{t-1}^{\varepsilon_i} (g_T^{\varepsilon_i} + h_T^{\varepsilon_i} t)\} \right] \\
&= -\sigma^2 C(T) - \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \delta_i^2 \sum_{t=1}^T S_{t-1}^\theta (g_T^\theta + h_T^\theta t) \right] \\
&= -\sigma^2 C(T) - \mu_\delta^2 \sum_{t=1}^T S_{t-1}^\theta (g_T^\theta + h_T^\theta t)
\end{aligned} \tag{92}$$

where

$$\begin{aligned}
\sigma^2 C(T) &= \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T S_{t-1}^{\varepsilon_i} \tilde{\varepsilon}_{it} = \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[ \sum_{t=1}^T \{S_{t-1}^{\varepsilon_i} (g_T^{\varepsilon_i} + h_T^{\varepsilon_i} t)\} \right] \\
&= \frac{1}{2} \sigma^2 (T-2).
\end{aligned} \tag{93}$$

Combining (93), (92) and (91) in (89) we get

$$\begin{aligned}
\text{p lim}_{N \rightarrow \infty} (\hat{\rho} - \rho) &= -\frac{\sigma^2 C(T) + \phi_{CT}}{\sigma^2 D(T) + \phi_{DT}} \\
&= -\frac{\frac{1}{2} \sigma^2 (T-2) + \mu_\delta^2 \sum_{t=1}^T S_{t-1}^\theta (g_T^\theta + h_T^\theta t)}{\frac{1}{15} \sigma^2 (T^2 - 4) + \mu_\delta^2 \sum_{t=1}^T (\tilde{S}_{t-1}^\theta)^2}
\end{aligned} \tag{94}$$

where

$$\phi_{CT} = \mu_\delta^2 \sum_{t=1}^T S_{t-1}^\theta (g_T^\theta + h_T^\theta t), \quad \phi_{DT} = \mu_\delta^2 \sum_{t=1}^T (\tilde{S}_{t-1}^\theta)^2,$$

as required for (42).

Next consider the asymptotic approximant to (94) as  $T \rightarrow \infty$ . Standard weak convergence arguments give

$$\frac{1}{T^2} \sum_{t=1}^T (\tilde{S}_{t-1}^\theta)^2 \rightarrow_d \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr, \tag{95}$$

where  $\tilde{W}_\theta(r) = W_\theta(r) - a_\theta - b_\theta r$  is detrended standard Brownian motion  $W_\theta$  with

$$a_\theta = 4 \left\{ \int_0^1 W_\theta - \frac{3}{2} \int_0^1 r W_\theta \right\}, \quad b_\theta = 6 \left\{ 2 \int_0^1 r W_\theta - \int_0^1 W_\theta \right\},$$

which leads to

$$T^{-2} \phi_{DT} = \mu_\delta^2 T^{-2} \sum_{t=1}^T (\tilde{S}_{t-1}^\theta)^2 \rightarrow_d \mu_\delta^2 \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr. \tag{96}$$

Now consider

$$\begin{aligned}
T^{-1}\phi_{CT} &= \mu_\delta^2 T^{-1} \sum_{t=1}^T S_{t-1}^\theta (g_T^\theta + h_T^\theta t) \\
&= \mu_\delta^2 \left\{ \frac{2(2T+1)}{T^2(T-1)} \left( \sum_{t=1}^T \theta_t \right) - \frac{6}{T^2(T-1)} \sum_{t=1}^T t\theta_t \right\} \sum_{t=1}^T S_{t-1}^\theta \\
&\quad + \mu_\delta^2 T^{-1} \left\{ \frac{12}{T(T^2-1)} \sum_{t=1}^T t\theta_t - \frac{6}{T(T-1)} \sum_{t=1}^T \theta_t \right\} \sum_{t=1}^T tS_{t-1}^\theta \\
&= \mu_\delta^2 \left\{ \left( \frac{4}{T^{3/2}} \sum_{t=1}^T S_{t-1}^\theta \right) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T \theta_t \right) - \left( \frac{6}{T^{3/2}} \sum_{t=1}^T S_{t-1}^\theta \right) \left( \frac{1}{T^{3/2}} \sum_{t=1}^T t\theta_t \right) \right\} \\
&\quad + \mu_\delta^2 \left\{ \left( \frac{12}{T^{3/2}} \sum_{t=1}^T tS_{t-1}^\theta \right) \left( \frac{1}{T^{3/2}} \sum_{t=1}^T t\theta_t \right) - \left( \frac{6}{T^{3/2}} \sum_{t=1}^T tS_{t-1}^\theta \right) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T \theta_t \right) \right\} + o_p(1) \\
&\rightarrow_d \mu_\delta^2 \sigma_\theta^2 \left\{ 4 \left( \int_0^1 W_\theta \right) W_\theta(1) - \left( 6 \int_0^1 W_\theta \right) \left( \int_0^1 r dW_\theta \right) \right\} \\
&\quad + \mu_\delta^2 \sigma_\theta^2 \left\{ 12 \left( \int_0^1 r W_\theta \right) \left( \int_0^1 r dW_\theta \right) - 6 \left( \int_0^1 r W_\theta \right) W_\theta(1) \right\} \\
&= \mu_\delta^2 \sigma_\theta^2 \left\{ 4W_\theta(1) \left[ \int_0^1 W_\theta - \frac{3}{2} \int_0^1 r W_\theta \right] \right\} + \mu_\delta^2 \left\{ 6 \int_0^1 r dW_\theta \left[ 2 \int_0^1 r W_\theta - \int_0^1 W_\theta \right] \right\} \\
&= \mu_\delta^2 \sigma_\theta^2 \left\{ a_\theta W_\theta(1) + b_\theta \int_0^1 r dW_\theta \right\} \\
&= \mu_\delta^2 \sigma_\theta^2 \left\{ \int_0^1 (a_\theta + b_\theta r) dW_\theta(r) \right\}. \tag{97}
\end{aligned}$$

By suitable augmentation of the probability space and embedding arguments, we can write the weak convergence in (96) and (97) as almost sure convergence and then we may write using (93)

$$\begin{aligned}
\frac{1}{(T-2)(T+2)} \{ \sigma^2 C(T) + \phi_{CT} \} &= \frac{1}{T+2} \left\{ \frac{\sigma^2}{2} + \frac{1}{T-2} \phi_{AT} \right\} \\
&= \frac{1}{T+2} \left\{ \frac{\sigma^2}{2} + \mu_\delta^2 \sigma_\theta^2 \left[ \int_0^1 (a_\theta + b_\theta r) dW_\theta(r) \right] + o_{a.s.}(1) \right\},
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{(T-2)(T+2)} \{ \sigma^2 D(T) + \phi_{DT} \} &= \frac{1}{15} \sigma^2 + \frac{\mu_\delta^2}{(T-2)(T+2)} \sum_{t=1}^T \left( \tilde{S}_{t-1}^\theta \right)^2 \\
&= \frac{1}{15} \sigma^2 + \mu_\delta^2 \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr + o_{a.s.}(1).
\end{aligned}$$

It follows that

$$\begin{aligned}
\frac{\sigma^2 C(T) + \phi_{CT}}{\sigma^2 D(T) + \phi_{DT}} &= -\frac{1}{T+2} \frac{\frac{\sigma^2}{2} + \mu_\delta^2 \sigma_\theta^2 \left[ \int_0^1 (a_\theta + b_\theta r) dW_\theta(r) \right] + o_{a.s.}(1)}{\frac{1}{15} \sigma^2 + \mu_\delta^2 \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr + o_{a.s.}(1)} \quad (1) \\
&= -\frac{7.5}{T+2} \frac{\sigma^2 + 2\mu_\delta^2 \sigma_\theta^2 \left[ \int_0^1 (a_\theta + b_\theta r) dW_\theta(r) \right] + o_{a.s.}\left(\frac{1}{T}\right)}{\sigma^2 + 15\mu_\delta^2 \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr} \\
&= -\frac{7.5}{T+2} \frac{\sigma^2 + 15\mu_\delta^2 \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr + 2\mu_\delta^2 \sigma_\theta^2 \left[ \int_0^1 (a_\theta + b_\theta r) dW_\theta(r) - 7.5 \int_0^1 \tilde{W}_\theta^2(r) dr \right] + o_{a.s.}\left(\frac{1}{T}\right)}{\sigma^2 + 15\mu_\delta^2 \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr} \\
&= -\frac{7.5}{T+2} - \frac{2}{T+2} \frac{\mu_\delta^2 \sigma_\theta^2 \left[ \int_0^1 (a_\theta + b_\theta r) dW_\theta(r) - 7.5 \int_0^1 \tilde{W}_\theta^2(r) dr \right] + o_{a.s.}\left(\frac{1}{T}\right)}{\sigma^2 + 15\mu_\delta^2 \sigma_\theta^2 \int_0^1 \tilde{W}_\theta^2(r) dr},
\end{aligned}$$

giving result (43).

**Proof of Proposition 8** We have  $\text{plim}_{N \rightarrow \infty} (\hat{\rho} - \rho) = \left\{ \text{plim}_{N \rightarrow \infty} \frac{1}{N} \tilde{y}'_{-1} Q_{\tilde{Z}} \tilde{y}_{-1} \right\}^{-1} \left\{ \text{plim}_{N \rightarrow \infty} \frac{1}{N} \tilde{y}'_{-1} Q_{\tilde{Z}} \tilde{u} \right\}$ , and by exogeneity

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

as in (59), where  $A(\rho, T)$  and  $\psi_{AT}$  are as in Proposition 4. When  $|\rho| < 1$ , we have

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

and then, just as in the proof of Proposition 3, we get

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

where  $B(\rho, T)$  and  $\psi_{BT}$  are given in Proposition 4. Result ( ) follows and the same argument gives result ( ) for the case of detrended data.

When  $\rho = 1$ , we get by straightforward combination of the arguments of Propositions 3 and 5

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

in which  $\tilde{Z}_{1, -1} = \left( \tilde{Z}'_{1,0}, \dots, \tilde{Z}'_{1, T-1} \right)'$  with  $\tilde{Z}_{1, t} = \left( \tilde{Z}_{1, t}^1, \dots, \tilde{Z}_{1, t}^N \right)'$  and  $\tilde{Z}_{1, t}^i = \sum_{j=0}^t \tilde{Z}_{1, t-j}^i$ . The result in the incidental trends case similarly follows from Propositions 3 and ??.

## 6.2 Additional Lemmas

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

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \rho \in (-1, 1], \quad \text{with } \varepsilon_t \sim iid(0, \sigma^2), \quad (98)$$

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)  $E \left( \sum_{t=1}^{T-1} tx_t \right)^2 = \sigma_x^2 \left\{ \sum_{t=1}^{T-1} t \sum_{j=t}^{T-1} j \rho^{j-t} + \sum_{t=2}^{T-1} t \sum_{j=1}^{t-1} j \rho^{t-j} \right\}$
- (b)  $E \left( \sum_{t=1}^{T-1} tx_t \right) \left( \sum_{t=1}^{T-1} x_t \right) = \sigma_x^2 \left\{ \sum_{t=1}^{T-1} t \sum_{j=t}^{T-1} \rho^{j-t} + \sum_{t=2}^{T-1} t \sum_{j=1}^{t-1} \rho^{t-j} \right\}$
- (c)  $E \left( \sum_{t=1}^{T-1} tx_t \right) \left( \sum_{t=2}^T tx_t \right) = \sigma_x^2 \left\{ \sum_{t=1}^{T-1} t \sum_{j=t}^{T-1} j \rho^{j-t+1} + \sum_{t=2}^{T-1} t \sum_{j=1}^{t-1} j \rho^{t-j-1} \right\}$
- (d)  $E \left( \sum_{t=1}^{T-1} tx_t \right) \left( \sum_{t=2}^T x_t \right) = \sigma_x^2 \left\{ \sum_{t=1}^{T-1} t \sum_{j=t}^{T-1} \rho^{j-t+1} + \sum_{t=2}^{T-1} t \sum_{j=1}^{t-1} \rho^{t-j-1} \right\}$
- (e)  $E \left( \sum_{t=1}^{T-1} tx_{t+1} \right) \left( \sum_{t=1}^{T-1} x_t \right) = \sigma_x^2 \left\{ \sum_{t=1}^{T-1} t \sum_{j=1}^t \rho^j + \sum_{t=1}^{T-2} t \sum_{j=1}^{T-t-1} \rho^{j-1} \right\}$
- (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)  $E \left( \sum_{t=1}^{T-1} tx_t \right)^2 = \sigma^2 \left\{ \sum_{t=1}^{T-1} \left( \frac{1}{6}t^2 - \frac{1}{6}t^4 + \frac{1}{2}T^2t^2 - \frac{1}{2}Tt^2 \right) \right\}$
- (b)  $E \left( \sum_{t=1}^{T-1} tx_t \right) \left( \sum_{t=1}^{T-1} x_t \right) = \sigma^2 \left\{ \sum_{t=1}^{T-1} \left( -\frac{1}{2}t^3 - \frac{1}{2}t^2 + Tt^2 \right) \right\}$
- (c)  $E \left( \sum_{t=1}^{T-1} tx_t \right) \left( \sum_{t=2}^T tx_t \right) = \sigma^2 \left\{ \sum_{t=1}^{T-1} \left( t \sum_{i=1}^{t-2} i(i+1) + t^2 \sum_{i=t-1}^{T-1} i \right) \right\}$
- (d)  $E \left( \sum_{t=1}^{T-1} tx_t \right) \left( \sum_{t=2}^T x_t \right) = \sigma^2 \left\{ \sum_{t=1}^{T-1} \left( -\frac{1}{2}t^3 + \frac{1}{2}t^2 - t + Tt^2 \right) \right\}$
- (e)  $E \left( \sum_{t=1}^{T-1} tx_{t+1} \right) \left( \sum_{t=1}^{T-1} x_t \right) = \sigma^2 \left\{ \sum_{t=1}^{T-1} t \sum_{j=1}^t j + \sum_{t=1}^{T-2} t(t+1)(T-t-1) \right\}$
- (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)  $E \left( \sum_{t=1}^{T-1} x_t \right)^2 = \sigma^2 \left\{ \sum_{t=1}^{T-1} \sum_{j=1}^t j + \sum_{t=1}^{T-2} t \sum_{j=t+1}^{T-1} 1 \right\}$

## 7 Notation

$o_{a.s.}(1)$	tends to zero almost surely
$O_{a.s.}(1)$	bounded almost surely
$\int_0^1 f$	$\int_0^1 f(r)dr$
$[\cdot]$	integer part
$:=$	definitional equality
CSD	Cross Section Dependent
CSI	Cross Section Independent
$w_i$	$T^{-1} \sum_{t=1}^T w_{it}$
$w_{i,-1}$	$T^{-1} \sum_{t=1}^T w_{it-1}$
$\tilde{w}_t$	$w_t - g_T^w - h_T^w t$ detrended $w_t$ ; see (77)
$g_T^w$	$\frac{2(2T+1)}{T(T-1)} \left( \sum_{t=1}^T w_t \right) - \frac{6}{T(T-1)} \sum_{t=1}^T tw_t$
$h_T^w$	$\frac{12}{T(T^2-1)} \sum_{t=1}^T tw_t - \frac{6}{T(T-1)} \sum_{t=1}^T w_t$
$\rightarrow_d$	weak convergence
$\rightarrow_p \rightarrow_{a.s.}$	convergence in probability, almost surely
$W(r)$	standard Brownian motion
$\underline{W}(r) = W - \int_0^1 W$	demeaned standard Brownian motion
$\tilde{W}(r) = W(r) - a - br$	detrended standard Brownian motion
$a$	$4 \left\{ \int_0^1 W - \frac{3}{2} \int_0^1 rW \right\}$
$b$	$6 \left\{ 2 \int_0^1 rW_s - \int_0^1 W_s \right\}$