Common Correlated Effects Estimation of Dynamic Panels with Cross-Sectional Dependence

Tom De Groote*[1] and Gerdie Everaert[1]

[1]*SHERPPA, Ghent University*

Preliminary, February 2010

Abstract

This paper studies estimation of dynamic panels with error cross-sectional dependence generated by unobserved common stochastic components using the common correlated effects pooled (CCEP) estimator suggested by Pesaran [Econometrica, 2006]. Pesaran shows that for a static model, the CCEP estimator is consistent for fixed T and $N \to \infty$. We now analyse the properties of this estimator in a dynamic setting, allowing for dynamics in both the model and in the common components. We first show that consistency of the CCEP estimator requires that next to N also T should tend to infinity in this setting. Second, the asymptotic bias, for fixed T, of the CCEP estimator is larger than that of the infeasible pooled OLS estimator, which assumes that the common components are known. Third, we propose a restricted CCEP estimator which is also consistent for both T and N tending to infinity but has an asymptotic bias, for fixed T, that is smaller than that of the unrestricted CCEP estimator and similar to that of the infeasible pooled OLS estimator. Monte Carlo experiments demonstrate that also the small sample properties of the restricted CCEP estimator are superior to that of the unrestricted CCEP estimator and not much worse than those of the infeasible pooled OLS estimator for T not too small.

JEL Classification: C13, C15, C23

Keywords: Cross-Sectional Dependence; Dynamic Panel; Common Correlated Effects

1 Introduction

In its influential paper, Nickell (1981) demonstrated that, when applied to a dynamic panel data model with individual effects, the within groups (WG) estimator is inconsistent for fixed T and $N \to \infty$. Nowadays various alternative estimators are available ranging from, among others, general method of moments (GMM) estimators (Arellano and Bond, 1991; Blundell and Bond,

*Corresponding author: Tweekerkenstraat 2, B-9000 Gent, Belgium. Email: tom.degroote@UGent.be. Website: http://www.sherppa.be.

1998), over analytical bias corrected within estimators (Kiviet, 1995; Bun and Carree, 2005) to a bootstrap based bias corrected within estimator (Everaert and Pozzi, 2007). However, new challenges arise when it comes to the estimation of dynamic panel data models. The recent panel data literature shifted its attention to the estimation of models with error cross-sectional dependence. A particular form that has become popular is a common factor error structure with a fixed number of unobserved common factors with individual-specific factor loadings (see e.g. Coakley et al., 2002; Phillips and Sul, 2003). For example, Serlenga and Shin (2007) estimate a gravity equation of bilateral trade among 15 European countries. Their model introduces common time-specific effects that capture determinants such as the business cycle or deal with globalization issues. This introduces error cross-sectional dependence of the factor-type. Their empirical results show that this approach provides more sensible results than the traditional approach using fixed time dummies.

The most obvious implication of error cross-sectional dependence is that standard panel data estimators such as WG are inefficient and estimated standard errors are biased and inconsistent. Phillips and Sul (2003) for instance show that if there is high cross-sectional correlation there may not be much to gain from pooling the data. However, cross-sectional dependence can also induce a bias and even result in inconsistent estimates. In general, inconsistency arises when the observed explanatory variables are correlated with the unobserved common factors (see e.g. Pesaran, 2006), i.e. this is an omitted variables bias which does not disappear as $T \to \infty$, $N \to \infty$ or both. In dynamic panel data models more specifically, Phillips and Sul (2007) show that the unobserved common factors introduce additional small sample bias and variability in the inconsistency of the WG estimator as $N \to \infty$ for fixed T even under the assumption of temporarily independent factors such that these are not correlated with the lagged dependent variable. This bias disappears as $T \to \infty$. Sarafidis and Robertson (2009) show that also standard dynamic panel data IV and GMM estimators (either in levels or first-differences) are inconsistent as $N \to \infty$ for fixed T as the moment conditions used by these estimators are invalid under error cross-sectional dependence.

In this paper we further analyse the impact of error cross-sectional dependence in the context of a linear dynamic panel data model. In order to keep the bias formulae as simple as possible, we do not include individual effects and consider a single unobserved factor.[1] We first extend the work of Phillips and Sul (2007) by deriving explicit bias formulae for the pooled OLS (POLS) estimator in the case of a temporally dependent factor. In line with Phillips and Sul, we find that under such circumstances the probability limit of the POLS estimator is a random variable rather than a constant for fixed T and $N \to \infty$. However, the POLS estimator is now also found to be inconsistent for $T \to \infty$ as the temporal dependence in the unobserved factor implies that it is correlated with the lagged dependent variable. As a benchmark, we also present asymptotic bias formulae for an infeasible POLS estimator, which includes the unobserved factor as an explanatory variable. This infeasible POLS estimator is inconsistent for fixed T and $N \to \infty$ but consistent

[1] Both restrictions will be relaxed in future versions of the paper.

for $T \to \infty$. Second, we extend the work of Pesaran (2006) by analysing the asymptotic behaviour of the common correlated effects pooled (CCEP) estimator in a dynamic panel data setting. The basic idea of CCEP estimation is to filter out the common components by orthogonalizing on the cross-section averages of both regressand and regressors such that as $N \to \infty$ the differential effects of the common components are eliminated. Contrary to the static model, the CCEP estimator is no longer consistent for $N \to \infty$ and fixed T in a dynamic panel data model. In fact, the inconsistency of the CCEP estimator is random for $N \to \infty$ and fixed T but disappears as $N, T \to \infty$ jointly. Furthermore, the asymptotic bias term for fixed T is bigger than that of the infeasible POLS estimator. Note that the standard CCEP estimator ignores the restrictions on the individual-specific factor loadings as implied by the derivation of the specification of the model augmented with cross-sectional averages. Imposing these restrictions, the restricted CCEP estimator is inconsistent for fixed T and $N \to \infty$, but its asymptotic bias is smaller than that of the unrestricted CCEP estimator and similar to that of the infeasible POLS estimator. Thus, the restricted CCEP is to be preferred over its unrestricted version.

We demonstrate the small sample properties of the POLS and CCEP estimators using Monte Carlo simulations. First, the naive POLS estimator is biased for temporally dependent factors. Second, the infeasible POLS is biased for small T, with the bias increasing in the degree of temporal dependence in the common factor. Third, both the unrestricted and the restricted CCEP have a higher bias than the infeasible POLS estimator for small values of T but the restricted CCEP clearly outperform the unrestricted CCEP and is not much worse than the infeasible POLS estimator for moderate values of T. Finally, the Monte Carlo simulation also illustrates the inefficiency of the naive POLS estimator and the substantial bias and inconsistency of its estimated standard errors.

The remainder of this paper is organized as follows: Section 2 sets out the basic model and its assumptions. Section 3 explores the asymptotic bias of, respectively a naive POLS, an infeasible POLS, the unrestricted and restricted CCEP in a dynamic context with error cross-sectional dependence. Section 4 contains the results of the Monte Carlo experiments. Section 5 concludes and outlines some future challenges.

2 Model and assumptions

We consider the following first-order autoregressive panel data model

$$y_{it} = \rho y_{i,t-1} + \nu_{it}, \qquad |\rho| \leq 1, \qquad (i = 1, \ldots, N; \quad t = 1, \ldots, T), \tag{1}$$

with y_{it} being the observation of the dependent variable of the ith cross-sectional unit at time t. For notational convenience we assume y_{i0} is observed. We further assume:

Assumption A1. *(Cross-section dependence) The error term* ν_{it} *has a single-factor structure*

$$\nu_{it} = \gamma_i F_t + \varepsilon_{it}, \tag{2}$$

$$F_t = \theta F_{t-1} + \mu_t, \qquad |\theta| \leq 1, \tag{3}$$

where F_t *is an individual-invariant time-specific unobserved effect with* $\mu_t \sim i.i.d.\left(0, \sigma_\mu^2\right)$*. The individual-specific factor loadings* γ_i *are nonrandom parameters satisfying* $\underset{N\to\infty}{plim} \frac{1}{N}\sum_{i=1}^{N}\gamma_i^2 = m_\gamma^2$*.* ε_{it} *satisfies A2.*

The restriction of a single-factor structure is for expositional purposes only.

Assumption A2. *(Error condition)* $\varepsilon_{it} \sim i.i.d.\left(0, \sigma_\varepsilon^2\right)$ *across* i *and* t *and is independent of* F_s *for all* i, t, s.

The model in equations (1)-(3) can be written in a convenient component form as

$$y_{it} = y_{it}^+ + \gamma_i F_t^+, \quad y_{it}^+ = \rho y_{i,t-1}^+ + \varepsilon_{it}, \quad F_t^+ = (1-\rho L)^{-1} F_t = (\rho + \theta) F_{t-1}^+ - \rho\theta F_{t-2}^+ + \mu_t. \tag{4}$$

For further discussion, stacking (1)-(2) for each i yields

$$y_i = \rho y_{i,-1} + \gamma_i F + \varepsilon_i, \tag{5}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$, $y_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$, $F = (F_1, \ldots, F_T)'$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$.

3 Estimators

In this section we analyse the properties of various estimators for the model in equations (1)-(3). We start with two 'extreme' approaches, i.e. the naive pooled OLS (POLSn) estimator which ignores cross-sectional dependence and the infeasible pooled OLS (POLSi) estimator which adds the unobserved common factor F_t as an explanatory variable to the model. Next, we analyse an unrestricted and a restricted version of the CCEP estimator suggested by Pesaran (2006) and compare their properties to those of the POLSn and POLSi estimators.

3.1 Naive POLS

We first analyse the asymptotic properties of the POLSn estimator, which is given by

$$\widehat{\rho}_{\text{POLSn}} = \frac{(1/NT)\sum_{i=1}^{N} y_{i,-1}' y_i}{(1/NT)\sum_{i=1}^{N} y_{i,-1}' y_{i,-1}} = \rho + \frac{(1/NT)\sum_{i=1}^{N} y_{i,-1}' (\gamma_i F + \varepsilon_i)}{(1/NT)\sum_{i=1}^{N} y_{i,-1}' y_{i,-1}}. \tag{6}$$

It is convenient to use sequential asymptotics with $N \to \infty$ followed by $T \to \infty$. All proofs are in the appendix.

Proposition 1. *In model (1) with errors having the factor structure (2)-(3) and satisfying A1-A2, the POLSn estimator is inconsistent as $N \to \infty$ and*

$$\underset{N\to\infty}{plim}\left(\widehat{\rho}_{POLSn}-\rho\right)=\frac{m_\gamma^2\frac{1}{T}\sum_{t=1}^{T}F_{t-1}^{+}F_t}{\frac{\sigma_\varepsilon^2}{1-\rho^2}+m_\gamma^2\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2}. \tag{7}$$

The inconsistency in (7) has the following asymptotic representation as $T \to \infty$

$$\underset{N\to\infty}{plim}\left(\widehat{\rho}_{POLSn}-\rho\right)=\frac{\theta\left(1-\rho^2\right)}{\left(1+\theta\rho\right)}\left(1+\frac{\left(1-\theta^2\right)\left(1-\theta\rho\right)}{\left(1+\theta\rho\right)}\frac{\sigma_\varepsilon^2}{m_\gamma^2\sigma_\mu^2}\right)^{-1}+o_p\left(1\right). \tag{8}$$

Equation (7) shows that $\widehat{\rho}_{\text{POLSn}}$ is inconsistent as $N \to \infty$ with the inconsistency depending on the persistence (θ) of the common factor and on the variance ratio $\sigma_\varepsilon^2/m_\gamma^2\sigma_\mu^2$. For a temporally independent factor, the asymptotic bias will be small. Importantly, the asymptotic bias is random for fixed T and depends on the particular values of F_t. Equation (8) shows that the POLSn estimator is inconsistent even for both N and $T \to \infty$ when $\theta \neq 0$. Essentially, this is an omitted variables bias as $\theta \neq 0$ implies $E\left(y_{i,t-1}F_t\right) \neq 0$ such that omitting F_t from the regression results in an inconsistent estimator for ρ.

3.2 Infeasible POLS

The POLSi estimator, including F_t as a regressor, is given by

$$\widehat{\rho}_{\text{POLSi}}=\frac{\left(1/NT\right)\sum_{i=1}^{N}y_{i,-1}'M_Fy_i}{\left(1/NT\right)\sum_{i=1}^{N}y_{i,-1}'M_Fy_{i,-1}}=\rho+\frac{\left(1/NT\right)\sum_{i=1}^{N}y_{i,-1}'M_F\varepsilon_i}{\left(1/NT\right)\sum_{i=1}^{N}y_{i,-1}'M_Fy_{i,-1}}, \tag{9}$$

where $M_F=I_T-F\left(F'F\right)^{-1}F'$.

Proposition 2. *In model (1) with errors having the factor structure (2)-(3) and satisfying A1-A2, the POLSi estimator is inconsistent as $N \to \infty$ and*

$$\underset{N\to\infty}{plim}\left(\widehat{\rho}_{POLSi}-\rho\right)=-\frac{\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}}{\frac{1}{1-\rho^2}\left(1-\frac{1}{T}\left(1+2\rho\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}\right)\right)+\frac{m_\gamma^2}{\sigma_\varepsilon^2}h_F}, \tag{10}$$

where $g_{F,t}=\sum_{s=t+1}^{T}\tau_{s,s-t}$ with $\tau_{s,s-t}$ being the $(s,s-t)$th element in $F\left(F'F\right)^{-1}F'$ and $h_F=\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2\left(1-\frac{\left(\frac{1}{T}\sum_{t=1}^{T}F_tF_{t-1}^{+}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}F_t^2\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2}\right)$.

The inconsistency in (10) has the following asymptotic representation as $T \to \infty$

$$\underset{N\to\infty}{plim}\left(\widehat{\rho}_{POLSi}-\rho\right)=-\frac{1}{T}\frac{\theta\left(1-\rho^2\right)}{1-\theta\rho}\left(1+\frac{m_\gamma^2}{\left(1-\theta\rho\right)^2}\frac{\sigma_\mu^2}{\sigma_\varepsilon^2}\right)^{-1}+o_p\left(\frac{1}{T}\right). \tag{11}$$

Thus, even with no fixed effects and observing the factor F_t, $\widehat{\rho}_{\text{POLSi}}$ is inconsistent for $N \to \infty$ and fixed T, with the asymptotic bias depending on the degree of inertia in F_t, being on average negative (positive) for positive (negative) values of θ. Importantly, the asymptotic bias is random as it depends on the particular realization of the process F_t. For large values of T or a temporally independent factor ($\theta = 0$), the bias will be small.

3.3 CCEP

The CCEP estimator suggested by Pesaran (2006) eliminates the unobserved common factors by including cross-section averages of the dependent and the explanatory variables. Taking cross-sectional averages of equation (1) gives

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} y_{it} &= \rho\frac{1}{N}\sum_{i=1}^{N} y_{i,t-1} + F_t\frac{1}{N}\sum_{i=1}^{N}\gamma_i + \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it},\\ \overline{y}_t &= \rho\overline{y}_{t-1} + \overline{\gamma}F_t + \overline{\varepsilon}_t, \end{aligned} \tag{12}$$

which can be solved for F_t as

$$F_t = \frac{1}{\overline{\gamma}}(\overline{y}_t - \rho\overline{y}_{t-1} - \overline{\varepsilon}_t). \tag{13}$$

Inserting this expression in equation (1) yields the following augmented form[2]

$$\begin{aligned} y_{it} &= \rho y_{i,t-1} + \frac{\gamma_i}{\overline{\gamma}}(\overline{y}_t - \rho\overline{y}_{t-1} - \overline{\varepsilon}_t) + \varepsilon_{it},\\ &= \rho y_{i,t-1} + \gamma_{1i}\overline{y}_t + \gamma_{2i}\overline{y}_{t,-1} + \widetilde{\varepsilon}_{it}, \end{aligned} \tag{14}$$

with $\gamma_{1i} = \gamma_i / \overline{\gamma}$, $\gamma_{2i} = \rho\gamma_{1i}$ and $\widetilde{\varepsilon}_{it} = \varepsilon_{it} - \gamma_{1i}\overline{\varepsilon}_t$.

3.3.1 Unrestricted CCEP

The unrestricted CCEP estimator, i.e. ignoring the restriction on γ_{2i}, for ρ in equation (14) is given by

$$\widehat{\rho}_{\text{CCEPu}} = \frac{(1/NT)\sum_{i=1}^{N} y'_{i,-1}M_G y_i}{(1/NT)\sum_{i=1}^{N} y'_{i,-1}M_G y_{i,-1}} = \rho + \frac{(1/NT)\sum_{i=1}^{N} y'_{i,-1}M_G\widetilde{\varepsilon}_i}{(1/NT)\sum_{i=1}^{N} y'_{i,-1}M_G y_{i,-1}}, \tag{15}$$

where $M_G = I_T - G(G'G)^{-1}G'$ and $G = (\overline{y}, \overline{y}_{-1})$.

Theorem 1. *In model (1) with errors having the factor structure (2)-(3) and satisfying A1-A2, the CCEPu estimator is inconsistent as $N \to \infty$ and*

$$\underset{N\to\infty}{plim}\,(\widehat{\rho}_{CCEPu} - \rho) = -\frac{\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}g^{+}_{F,t}}{\frac{1}{1-\rho^2}\left(1 - \frac{2}{T}\left(1 + \rho\sum_{t=1}^{T-1}\rho^{t-1}g^{+}_{F,t}\right)\right)}, \tag{16}$$

[2]See Phillips and Sul (2007) or an expression with multiple factors.

where $g^{+}_{F,t} = \sum_{s=t+1}^{T} \tau^{+}_{s,s-t}$ *with* $\tau^{+}_{s,s-t}$ *being the* $(s, s-t)\,th$ *element in* $H\left(H'H\right)^{-1}H'$ *and* $H = \left(F^{+}, F^{+}_{-1}\right)$.

The inconsistency (16) has the following asymptotic representation as $T \to \infty$

$$\underset{N\to\infty}{plim}\left(\widehat{\rho}_{CCEPu} - \rho\right) = -\frac{1}{T}\left(\frac{\theta(1-\rho^2)}{1-\theta\rho} + \rho\right) + o_p\left(\frac{1}{T}\right). \tag{17}$$

The implication of Theorem 1 is that the CCEPu estimator is consistent for $(N, T)_{seq} \to \infty$ but has a different asymptotic bias term compared to the POLSi estimator. More specifically

$$T\underset{N\to\infty}{\text{plim}}\left(\widehat{\rho}_{\text{CCEPu}} - \widehat{\rho}_{\text{POLSi}}\right) = -\rho - \left(1 + \frac{\left(1-\theta\rho\right)^2 \sigma^2_{\varepsilon}}{m^2_{\gamma}\sigma^2_{\mu}}\right)^{-1}\frac{\theta(1-\rho^2)}{1-\theta\rho} + o_p\left(1\right), \tag{18}$$

which implies that for the relevant case of both $\rho > 0$ and $\theta > 0$, the CCEPu estimator has a relatively bigger downward asymptotic bias compared to the POLSi estimator.

3.3.2 Restricted CCEP

The restricted CCEP estimator, i.e. taking into account the restriction on γ_{2i}, for ρ in equation (14) can be obtained by minimizing the objective function

$$S_{NT}\left(\rho, F\right) = \frac{1}{NT}\sum_{i=1}^{N}\left(y_i - \rho y_{i,-1}\right)' M_F \left(y_i - \rho y_{i,-1}\right). \tag{19}$$

Although F is not observed when estimating ρ and similarly, ρ is not observed when estimating F, we can replace the unobserved quantities by initial estimates and iterate until convergence. The continuously-updated estimator for (ρ, F) is defined as

$$\left(\widehat{\rho}_{\text{CCEPr}}, \widehat{F}\right) = \underset{\rho,F}{\text{argmin}}\, S_{NT}\left(\rho, F\right) \tag{20}$$

More specifically, $\left(\widehat{\rho}_{\text{CCEPr}}, \widehat{F}\right)$ is the solution to the following two equations

$$\widehat{\rho}_{\text{CCEPr}} = \frac{\left(1/NT\right)\sum_{i=1}^{N} y'_{i,-1} M_{\widehat{F}} y_i}{\left(1/NT\right)\sum_{i=1}^{N} y'_{i,-1} M_{\widehat{F}} y_{i,-1}} = \rho + \frac{\left(1/NT\right)\sum_{i=1}^{N} y'_{i,-1} M_{\widehat{F}} \widetilde{\varepsilon}_i}{\left(1/NT\right)\sum_{i=1}^{N} y'_{i,-1} M_{\widehat{F}} y_{i,-1}}, \tag{21}$$

$$\widehat{F} = \overline{y} - \widehat{\rho}_{\text{CCEPr}}\ \overline{y}_{-1} \tag{22}$$

where $M_{\widehat{F}} = I_T - \widehat{F}\left(\widehat{F}'\widehat{F}\right)^{-1}\widehat{F}'$.

Theorem 2. *In model (1) with errors having the factor structure (2)-(3) and satisfying A1-A2, the CCEPr estimator is consistent as* $(N, T)_{seq} \to \infty$,

$$\widehat{\rho}_{CCEPr} \underset{p}{\to} \rho, \tag{23}$$

with large T expansion given by

$$T \underset{N\to\infty}{plim} \left(\widehat{\rho}_{CCEPr} - \widehat{\rho}_{POLSi}\right) = o_p\left(1\right). \tag{24}$$

The theorem shows that like POLSi and CCEPu, the CCEPr estimator is consistent as $(N,T)_{seq} \to \infty$ but unlike the CCEPu estimator, its asymptotic bias as $N \to \infty$ is similar to that of the POLSi estimator.

4 Monte Carlo simulation

In this section we investigate the small sample properties of the POLSn, POLSi, CCEPu and CCEPr estimators under cross-sectional dependence. We are interested in the effects of (i) the extent of cross-sectional dependence, (ii) the degree of inertia in both the factor and dependend variable and (iii) the relative importance of the variance of the factor loadings and the idiosyncratic errors. In order to correctly interpret the results , comparability between the different experiments is necessary. To this extent, we take into consideration the remarks made by Kiviet (1995) on (i) the relative impact of the various error terms and (ii) the importance of the signal-to-noise ratio. Sarafidis et al. (2009) extend the Monte Carlo design of Kiviet (1995) towards an error structure containing common components. Our design slightly deviates from theirs since our model does not include individual effects, nor a vector of regressors x_{it}.

4.1 Experimental design

Data are generated according to (1)-(3) where ε_{it}, μ_t and γ_i are randomly drawn in each replication from $\sim i.i.d.N\left(0,\sigma_\varepsilon\right)$, $\sim i.i.d.N\left(0,\sigma_\mu\right)$ and $\sim i.i.d.U\left[\gamma_L,\gamma_U\right]$, respectively. $y_{i,-49}$ is set to zero and the first 50 observations are discarded.
We control for the relative impact of the common correlated effects $\gamma_i F_t$ and the idiosyncratic error terms ε_{it} on y_{it}. Using (A-2) and noting that $Var(\gamma_i F_t^+) = m_\gamma^2 \sigma_{F^+}^2$, the variance of y_{it} rewrites to

$$\begin{aligned} Var(y_{it}) &= Var(y_{it}^+) + Var(\gamma_i F_t^+), \\ &= \frac{\sigma_\varepsilon^2}{1-\rho^2} + \frac{(1+\theta\rho) m_\gamma^2 \sigma_\mu^2}{(1-\rho^2)(1-\theta^2)(1-\theta\rho)}, \\ &= (1+\omega^2)\frac{\sigma_\varepsilon^2}{1-\rho^2}, \end{aligned} \tag{25}$$

where $\omega^2 = \frac{m_\gamma^2\sigma_\mu^2/((1-\theta^2)(1-\theta\rho))}{\sigma_\varepsilon^2/(1+\theta\rho)}$. We therefore choose σ_μ^2 by setting

$$\sigma_\mu^2 = \frac{(1-\theta^2)(1-\theta\rho)}{1+\theta\rho}\frac{\omega^2\sigma_\varepsilon^2}{m_\gamma^2}, \tag{26}$$

where ω^2 controls for the degree of cross-sectional dependence. In this setting, changing the specifications of the distribution of γ_i only affects the amount of heterogeneity since an increase in m_γ^2 is offset by a decrease of σ_μ^2. Higher values of ω^2 correspond to a higher amount of error cross-sectional dependence.

Furthermore, we want to keep the signal to noise, σ_s^2, fixed over the various experiments. Using (25) and (26), the signal-to-noise ratio is given by

$$\begin{aligned}\sigma_s^2 &= Var(y_{it} - \nu_{it}), \\ &= \left(\rho^2 + \omega^2\left(1 - \frac{(1-\rho^2)(1-\theta\rho)}{1+\theta\rho}\right)\right)\frac{\sigma_\varepsilon^2}{1-\rho^2}.\end{aligned} \tag{27}$$

For a given ω^2, σ_s^2 and m_γ^2, the values for σ_ε^2 and σ_μ^2 follow from (27) and (26).

We conduct experiments for combinations of the following parameter values: $\rho \in \{0.4; 0.6; 0.8\}$, $\theta \in \{0; 0.4; 0.8\}$ and $\omega \in \{0.5; 1; 2\}$. Experiment 1 serves as a point of reference and has the following settings: $\rho = 0.6$, $\theta = 0.4$, $\omega = 1$, $\sigma_s^2 = 3$ and $\gamma \sim i.i.d.U(0.5, 2.1)$. The other six simulations verify the impact on the small sample properties of the estimators for different degrees of error cross section dependence and a higher persistence in both y_{it} and the common factors. The signal to noise ratio is set to 3. γ_i is drawn from $\sim i.i.d.U\,[0.5, 2.1]$. T and N respectively take the following values:{5;10;20;30;40;50} and {20;50}. All experiments are based on 5000 iterations.

4.2 Simulation Results

First we examine the performances of the two 'extreme' estimators, namely the naive and infeasable POLS. Experiment 1 indicates that the naive POLS is biased in a small sample. Furthermore, this bias persists as T or N increases. Proposition 1 reveals that the bias emerges from the persistence in the common factors. Simulation results for experiment 5, which generates the common factors by drawing them randomly from a normal distribution, establish that the bias of the naive POLS stems from the inertia residing within the factors. The naive POLS exhibits a small bias for experiment 5, whereas, for the other experiments, it is severly biased. Experiment 4, 6 and 7 illustrate that increasing the level of persistency in the factors or the degree of error cross-sectional dependence, aggravates the bias. Increasing the dynamics of y_{it} on the other hand, has the opposite effect (see experiment 2 and 3). In addition, doubling ω^2 reduces the efficiency of the naive POLS. The standard errors substantially underestimate the standard deviations in all experiments.

Table 1: Small sample properties of POLSn, POLSi, CCEPu and CCEPr

	POLSn				POLSi				CCEPu				CCEPr			
(T,N)	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 1: $\rho = 0.6$, $\theta = 0.4$, $\omega = 1$, $\sigma_s^2 = 3$ and $\gamma \sim i.i.d.U(0.5, 2.1)$																
(5,20)	0.061	0.074	0.142	0.155	-0.034	0.071	0.093	0.099	-0.230	0.117	0.224	0.321	-0.126	0.098	0.204	0.240
(10,20)	0.073	0.052	0.102	0.126	-0.017	0.046	0.050	0.053	-0.102	0.068	0.098	0.142	-0.046	0.064	0.095	0.106
(20,20)	0.085	0.036	0.076	0.115	-0.009	0.031	0.032	0.034	-0.050	0.044	0.051	0.071	-0.021	0.043	0.051	0.055
(30,20)	0.090	0.029	0.065	0.111	-0.007	0.025	0.025	0.026	-0.032	0.035	0.038	0.050	-0.013	0.034	0.038	0.040
(40,20)	0.093	0.025	0.056	0.109	-0.005	0.021	0.022	0.022	-0.024	0.030	0.032	0.040	-0.009	0.029	0.032	0.033
(50,20)	0.095	0.023	0.051	0.107	-0.004	0.019	0.019	0.020	-0.020	0.026	0.028	0.034	-0.008	0.026	0.028	0.029
(5,50)	0.064	0.047	0.127	0.142	-0.030	0.044	0.072	0.078	-0.231	0.074	0.199	0.305	-0.122	0.062	0.175	0.213
(10,50)	0.076	0.033	0.095	0.122	-0.015	0.029	0.034	0.037	-0.101	0.043	0.076	0.126	-0.044	0.041	0.069	0.082
(20,50)	0.086	0.023	0.072	0.112	-0.009	0.019	0.021	0.023	-0.048	0.028	0.035	0.059	-0.019	0.027	0.033	0.039
(30,50)	0.090	0.019	0.060	0.109	-0.005	0.016	0.016	0.017	-0.032	0.022	0.025	0.040	-0.012	0.022	0.024	0.027
(40,50)	0.094	0.016	0.054	0.108	-0.004	0.013	0.014	0.014	-0.023	0.019	0.021	0.031	-0.009	0.019	0.021	0.022
(50,50)	0.095	0.014	0.048	0.107	-0.004	0.012	0.012	0.013	-0.019	0.017	0.018	0.026	-0.007	0.016	0.017	0.019
Experiment 2: $\rho = 0.8$, $\theta = 0.4$, $\omega = 1$, $\sigma_s^2 = 3$ and $\gamma \sim i.i.d.U(0.5, 2.1)$																
(5,20)	0.025	0.056	0.105	0.108	-0.029	0.056	0.076	0.082	-0.281	0.106	0.233	0.365	-0.164	0.081	0.223	0.277
(10,20)	0.032	0.039	0.072	0.079	-0.011	0.035	0.039	0.040	-0.123	0.057	0.104	0.161	-0.047	0.050	0.095	0.106
(20,20)	0.040	0.027	0.053	0.066	-0.007	0.023	0.025	0.025	-0.057	0.035	0.049	0.075	-0.018	0.033	0.043	0.046
(30,20)	0.044	0.022	0.044	0.062	-0.005	0.019	0.019	0.019	-0.036	0.027	0.034	0.050	-0.010	0.026	0.030	0.032
(40,20)	0.046	0.019	0.038	0.059	-0.003	0.016	0.016	0.017	-0.027	0.023	0.026	0.038	-0.007	0.022	0.025	0.026
(50,20)	0.047	0.017	0.034	0.058	-0.003	0.014	0.014	0.015	-0.022	0.020	0.023	0.032	-0.006	0.020	0.022	0.022
(5,50)	0.029	0.035	0.095	0.100	-0.026	0.035	0.061	0.066	-0.282	0.067	0.213	0.354	-0.164	0.052	0.204	0.261
(10,50)	0.034	0.024	0.066	0.074	-0.010	0.022	0.027	0.029	-0.121	0.036	0.088	0.150	-0.044	0.032	0.075	0.087
(20,50)	0.040	0.017	0.048	0.063	-0.006	0.015	0.016	0.017	-0.054	0.022	0.036	0.065	-0.015	0.021	0.029	0.033
(30,50)	0.044	0.014	0.040	0.059	-0.003	0.012	0.012	0.013	-0.035	0.017	0.024	0.042	-0.009	0.016	0.020	0.022
(40,50)	0.046	0.012	0.036	0.058	-0.003	0.010	0.010	0.011	-0.026	0.015	0.018	0.031	-0.006	0.014	0.016	0.017
(50,50)	0.048	0.011	0.032	0.058	-0.002	0.009	0.009	0.009	-0.021	0.013	0.015	0.026	-0.005	0.012	0.014	0.015

Continued on next page

Continued from previous page

	POLSn				POLSi				CCEPu				CCEPr			
(T, N)	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 3: $\rho = 0.4$, $\theta = 0.4$, $\omega = 1$, $\sigma_s^2 = 3$ and $\gamma \sim i.i.d.U(0.5, 2.1)$																
(5,20)	0.096	0.086	0.173	0.198	-0.036	0.080	0.101	0.107	-0.176	0.124	0.211	0.275	-0.101	0.109	0.192	0.217
(10,20)	0.112	0.060	0.128	0.170	-0.019	0.052	0.057	0.060	-0.082	0.075	0.095	0.126	-0.044	0.072	0.096	0.106
(20,20)	0.126	0.042	0.096	0.158	-0.011	0.035	0.036	0.038	-0.041	0.049	0.054	0.067	-0.022	0.049	0.055	0.059
(30,20)	0.131	0.034	0.082	0.154	-0.008	0.028	0.028	0.029	-0.027	0.039	0.041	0.049	-0.014	0.039	0.042	0.044
(40,20)	0.134	0.030	0.071	0.152	-0.005	0.024	0.025	0.025	-0.020	0.034	0.036	0.041	-0.010	0.033	0.036	0.038
(50,20)	0.136	0.027	0.064	0.150	-0.005	0.022	0.022	0.022	-0.017	0.030	0.031	0.035	-0.009	0.030	0.031	0.033
(5,50)	0.100	0.054	0.156	0.185	-0.032	0.050	0.077	0.083	-0.179	0.079	0.180	0.254	-0.097	0.069	0.155	0.183
(10,50)	0.116	0.038	0.120	0.167	-0.018	0.033	0.038	0.042	-0.082	0.047	0.069	0.107	-0.043	0.046	0.067	0.080
(20,50)	0.127	0.027	0.091	0.156	-0.011	0.022	0.023	0.026	-0.041	0.031	0.035	0.054	-0.021	0.031	0.036	0.041
(30,50)	0.132	0.022	0.076	0.152	-0.007	0.018	0.018	0.019	-0.027	0.025	0.026	0.037	-0.013	0.025	0.027	0.030
(40,50)	0.136	0.019	0.068	0.152	-0.005	0.015	0.016	0.017	-0.020	0.021	0.023	0.030	-0.010	0.021	0.023	0.025
(50,50)	0.136	0.017	0.061	0.150	-0.004	0.014	0.014	0.015	-0.016	0.019	0.019	0.025	-0.008	0.019	0.020	0.021
Experiment 4: $\rho = 0.6$, $\theta = 0.8$, $\omega = 1$, $\sigma_s^2 = 3$ and $\gamma \sim i.i.d.U(0.5, 2.1)$																
(5,20)	0.117	0.068	0.133	0.177	-0.116	0.084	0.128	0.173	-0.318	0.122	0.227	0.390	-0.207	0.101	0.204	0.290
(10,20)	0.131	0.047	0.103	0.166	-0.059	0.053	0.066	0.088	-0.155	0.070	0.105	0.187	-0.093	0.066	0.103	0.138
(20,20)	0.145	0.033	0.082	0.166	-0.032	0.035	0.039	0.050	-0.075	0.045	0.053	0.092	-0.046	0.044	0.054	0.071
(30,20)	0.151	0.027	0.071	0.167	-0.022	0.028	0.030	0.037	-0.049	0.035	0.039	0.063	-0.030	0.035	0.040	0.050
(40,20)	0.156	0.023	0.062	0.168	-0.016	0.024	0.025	0.030	-0.036	0.030	0.033	0.049	-0.022	0.030	0.033	0.040
(50,20)	0.158	0.020	0.057	0.168	-0.014	0.022	0.022	0.026	-0.030	0.027	0.028	0.041	-0.018	0.026	0.029	0.034
(5,50)	0.122	0.043	0.121	0.172	-0.109	0.053	0.107	0.153	-0.341	0.077	0.204	0.397	-0.215	0.065	0.170	0.275
(10,50)	0.133	0.030	0.097	0.165	-0.057	0.033	0.048	0.075	-0.165	0.045	0.086	0.186	-0.098	0.042	0.076	0.124
(20,50)	0.145	0.021	0.077	0.164	-0.031	0.022	0.026	0.040	-0.079	0.028	0.037	0.087	-0.047	0.028	0.036	0.060
(30,50)	0.152	0.017	0.066	0.166	-0.021	0.018	0.019	0.028	-0.052	0.022	0.026	0.058	-0.031	0.022	0.026	0.041
(40,50)	0.156	0.014	0.060	0.168	-0.016	0.015	0.016	0.023	-0.038	0.019	0.021	0.044	-0.023	0.019	0.021	0.032
(50,50)	0.159	0.013	0.055	0.168	-0.013	0.014	0.014	0.019	-0.031	0.017	0.018	0.036	-0.019	0.017	0.018	0.026

Continued on next page

Continued from previous page

	POLSn				POLSi				CCEPu				CCEPr			
(T,N)	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 5: $\rho=0.6$, $\theta=0$, $\omega=1$, $\sigma_s^2=3$ and $\gamma\sim i.i.d.U(0.5,2.1)$																
(5,20)	-0.028	0.081	0.166	0.169	-0.005	0.066	0.078	0.078	-0.146	0.113	0.200	0.248	-0.063	0.095	0.190	0.200
(10,20)	-0.023	0.057	0.122	0.124	-0.001	0.043	0.046	0.046	-0.065	0.066	0.090	0.111	-0.012	0.062	0.083	0.084
(20,20)	-0.013	0.040	0.089	0.090	-0.001	0.030	0.030	0.030	-0.033	0.043	0.049	0.059	-0.005	0.042	0.047	0.047
(30,20)	-0.010	0.033	0.075	0.076	-0.001	0.024	0.024	0.024	-0.021	0.034	0.037	0.043	-0.003	0.034	0.036	0.036
(40,20)	-0.008	0.028	0.066	0.067	-0.000	0.020	0.021	0.021	-0.016	0.029	0.031	0.035	-0.002	0.029	0.031	0.031
(50,20)	-0.006	0.025	0.059	0.059	-0.001	0.018	0.018	0.018	-0.014	0.026	0.027	0.031	-0.002	0.026	0.027	0.027
(5,50)	-0.025	0.051	0.155	0.157	-0.002	0.042	0.058	0.058	-0.145	0.071	0.180	0.231	-0.051	0.060	0.159	0.167
(10,50)	-0.020	0.036	0.114	0.116	0.001	0.027	0.030	0.030	-0.061	0.042	0.068	0.091	-0.008	0.039	0.056	0.057
(20,50)	-0.012	0.025	0.084	0.085	0.000	0.019	0.019	0.019	-0.030	0.027	0.033	0.045	-0.002	0.027	0.030	0.030
(30,50)	-0.010	0.021	0.070	0.071	0.000	0.015	0.015	0.015	-0.020	0.022	0.024	0.031	-0.001	0.021	0.023	0.023
(40,50)	-0.007	0.018	0.062	0.062	0.000	0.013	0.013	0.013	-0.015	0.019	0.020	0.025	-0.001	0.018	0.019	0.019
(50,50)	-0.007	0.016	0.055	0.055	0.000	0.012	0.012	0.012	-0.012	0.017	0.017	0.021	-0.001	0.016	0.017	0.017
Experiment 6: $\rho=0.6$, $\theta=4$, $\omega=2$, $\sigma_s^2=3$ and $\gamma\sim i.i.d.U(0.5,2.1)$																
(5,20)	0.077	0.072	0.176	0.192	-0.024	0.061	0.077	0.081	-0.234	0.118	0.224	0.324	-0.130	0.098	0.204	0.242
(10,20)	0.094	0.050	0.125	0.156	-0.011	0.038	0.042	0.043	-0.104	0.068	0.099	0.143	-0.048	0.064	0.095	0.107
(20,20)	0.112	0.035	0.091	0.144	-0.006	0.026	0.027	0.028	-0.051	0.044	0.051	0.072	-0.022	0.043	0.051	0.055
(30,20)	0.118	0.028	0.077	0.141	-0.005	0.021	0.021	0.021	-0.033	0.035	0.038	0.050	-0.014	0.034	0.038	0.040
(40,20)	0.122	0.024	0.066	0.139	-0.003	0.018	0.018	0.018	-0.024	0.030	0.032	0.040	-0.010	0.029	0.032	0.034
(50,20)	0.125	0.022	0.059	0.138	-0.003	0.016	0.016	0.016	-0.020	0.026	0.028	0.035	-0.008	0.026	0.028	0.029
(5,50)	0.079	0.045	0.163	0.181	-0.020	0.038	0.059	0.062	-0.233	0.074	0.200	0.307	-0.123	0.062	0.175	0.214
(10,50)	0.097	0.031	0.119	0.154	-0.010	0.024	0.028	0.030	-0.102	0.043	0.076	0.127	-0.045	0.041	0.069	0.082
(20,50)	0.112	0.022	0.087	0.142	-0.006	0.016	0.017	0.018	-0.049	0.028	0.035	0.060	-0.020	0.027	0.034	0.039
(30,50)	0.118	0.018	0.073	0.139	-0.004	0.013	0.013	0.014	-0.032	0.022	0.025	0.040	-0.012	0.022	0.025	0.027
(40,50)	0.123	0.015	0.064	0.139	-0.003	0.011	0.011	0.012	-0.023	0.019	0.021	0.031	-0.009	0.019	0.021	0.022
(50,50)	0.126	0.014	0.058	0.138	-0.003	0.010	0.010	0.010	-0.019	0.017	0.018	0.026	-0.008	0.016	0.018	0.019

Continued on next page

Continued from previous page

(T,N)	POLSn				POLSi				CCEPu				CCEPr			
	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 7: $\rho = 0.6$, $\theta = 0$, $\omega = 0.5$, $\sigma_s^2 = 3$ and $\gamma \sim i.i.d.U(0.5, 2.1)$																
(5,20)	0.044	0.076	0.113	0.121	-0.046	0.079	0.105	0.115	-0.224	0.117	0.221	0.315	-0.123	0.098	0.202	0.237
(10,20)	0.051	0.053	0.083	0.097	-0.023	0.051	0.058	0.062	-0.099	0.068	0.096	0.138	-0.044	0.064	0.093	0.103
(20,20)	0.058	0.037	0.061	0.084	-0.012	0.035	0.036	0.038	-0.047	0.044	0.051	0.069	-0.019	0.043	0.050	0.054
(30,20)	0.061	0.031	0.051	0.079	-0.008	0.028	0.029	0.030	-0.031	0.035	0.038	0.049	-0.011	0.034	0.038	0.039
(40,20)	0.063	0.026	0.045	0.077	-0.006	0.024	0.024	0.025	-0.023	0.030	0.032	0.039	-0.009	0.029	0.032	0.033
(50,20)	0.064	0.024	0.041	0.076	-0.005	0.021	0.022	0.023	-0.019	0.026	0.029	0.035	-0.007	0.026	0.029	0.030
(5,50)	0.048	0.048	0.098	0.109	-0.042	0.050	0.085	0.095	-0.231	0.074	0.200	0.305	-0.124	0.062	0.177	0.216
(10,50)	0.054	0.034	0.074	0.092	-0.020	0.032	0.041	0.046	-0.100	0.043	0.076	0.126	-0.043	0.041	0.070	0.082
(20,50)	0.061	0.024	0.055	0.082	-0.011	0.022	0.024	0.026	-0.048	0.028	0.035	0.059	-0.019	0.027	0.034	0.039
(30,50)	0.061	0.019	0.047	0.077	-0.007	0.018	0.019	0.020	-0.031	0.022	0.026	0.040	-0.011	0.022	0.025	0.027
(40,50)	0.063	0.017	0.041	0.076	-0.006	0.015	0.016	0.017	-0.023	0.019	0.021	0.031	-0.009	0.019	0.020	0.022
(50,50)	0.064	0.015	0.038	0.074	-0.005	0.014	0.014	0.015	-0.019	0.017	0.018	0.026	-0.007	0.016	0.018	0.019

Moving on to the infeasable POLS. For small T, the experiments report a significant bias. Experiment 5 forms the exception. Introducing inertia into the data generating process of F_t, thus causes the POLSi estimator to become biased for fixed T. However, contrary to the naive POLS, the bias does disappear as T increases. This is in line with our findings in proposition 2. The size of N is irrelevant for the bias. Higher degrees of inertia in F_t, such as in experiment 4, render a higher bias. Experiments 2 and 6 demonstrate that increasing the persistency of y_{it} or the degree of error cross-sectional dependence, lowers the bias. Comparing the two POLS estimators, the infeasable POLS outperforms the naive POLS in terms of both bias and efficiency. For each experiment, the Root Mean Squared Error (RMSE) of the infeasable POLS is smaller than the one of the naive POLS.

Next, turning to the unrestricted CCEP. We notice a significant small sample bias throughout all experiments. For relatively small T, the unrestricted CCEP performs even worse than the naive POLS in terms of RMSE. Contrasted to our benchmark, the infeasable POLS, the small sample bias is larger. This fits with our previous findings expressed in (18), which implies that for relevant range of parameter values θ and ρ, the unrestricted CCEP has a larger asymptotic bias than the infeasable POLS. The small sample bias reduces as T becomes larger, whereas the size of N only has a marginal impact. Experiment 5 reveals that, in contrast to the POLS estimators, the small sample bias of CCEPu for fixed T, persists even when the common factors are randomly drawn from a normal distribution. Therefore, the bias also directly depends on ρ.

Finally, the simulations indicate that the restricted CCEP estimator exhibits a bias in a small sample, which diminishes for greater T. Again, the size of N does not seem to matter. Setting both CCEP estimators side by side, we notice that the restricted CCEP outperforms its unrestricted counterpart in terms of bias for small T and N. This is in accordance with our earlier statement, that the restricted CCEP is preferable over the unrestricted version. Higher or lower degrees of error cross section dependence, as in experiment 6 and 7, have only a minor effect on the performances of both CCEP estimators. Increasing the persistence of the common factors or y_{it}, on the other hand, worsens the asymptotic bias. Experiments 2-5 demonstrate this.

5 Conclusion

This papers examines the effects of error cross-sectional dependence on POLS and CCEP estimators in a linear dynamic panel data model. The naive POLS turns out to be inconsistent for $N \to \infty$. In accordance with Phillips and Sul, we find that, for fixed T, the asymptotic bias depends on the particular realisation of F_t. The inconsistency remains, even for both T and $N \to \infty$, due to persistence in the common factors. The infeasable POLS estimator, which we use as a point of reference, is inconsistent for fixed T and $N \to \infty$. Just as for the naive POLS estimator, the inertia in F_t plays an important role in the inconsistency. As $T \to \infty$, the asymptotic bias disappears. The unrestricted CCEP estimator, as proposed by Pesaran (2006) is no longer consistent for fixed T and $N \to \infty$. In addition, the size of the inconsistency for fixed T is bigger than the one of the infeasable POLS estimator. For both T and $N \to \infty$, this estimator becomes consistent. Contrary to the standard, unrestricted CCEP estimator, the restricted CCEP takes into account the restrictions on the individual-specific factor loadings as implied by the derivation of the specification of the model augmented with cross-sectional averages. This restricted version is, like the unrestricted CCEP, consistent for both T and $N \to \infty$. However, its asymptotic bias is similar to that of the infeasible POLS. Therefore, when faced with error cross-sectional dependence, the restricted CCEP is the more appropriate estimator.

Future work will extend the current framework by introducing individual effects, an additional variable x_{it} and allowing for multiple factors. It is then necessary to apply an estimation procedure which can deal with both error cross-sectional dependence due to common factors and individual effects in a linear dynamic panel data model.

References

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Review of Economic Studies 58 (2), 277–297.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87 (1), 115–143.

Bun, M., Carree, M., 2005. Bias-corrected estimation in dynamic panel data models. Journal of Business & Economic Statistics 23 (2), 200–210.

Coakley, J., Fuertes, A.-M., Smith, R., 2002. A principal components approach to cross-section dependence in panels. Unpublished Manuscript, Birkbeck College, University of London.

Everaert, G., Pozzi, L., 2007. Bootstrap-based bias correction for dynamic panels. Journal of Economic Dynamics & Control 31 (4), 1160–1184.

Kiviet, J., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68 (1), 53–78.

Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49 (6), 1417–1426.

Pesaran, M., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74 (4), 967–1012.

Phillips, P. C. B., Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. Econometrics Journal 6 (1), 217–259.

Phillips, P. C. B., Sul, D., 2007. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. Journal Of Econometrics 137 (1), 162–188.

Sarafidis, V., Robertson, D., 2009. On the impact of error cross-sectional dependence in short dynamic panel estimation. Econometric Journal 12 (1), 62–81.

Sarafidis, V., Yamagata, T., Robertson, D., 2009. A test of cross section dependence for a linear dynamic panel model with regressors. Journal of Econometrics 148 (2), 149–161.

Serlenga, L., Shin, Y., 2007. Gravity models of intra-eu trade: Application of the ccep-ht estimation in heterogeneous panels with unobserved common time-specific factors. Journal of Applied Econometrics 22 (2), 361–381.

Appendices

Appendix A Proofs

Lemma A-1.

Under Assumption A1 we have from (4)

$$\begin{aligned}
\lambda_0 &= E\left(F_t^+\right)^2 = (\rho+\theta)\,\lambda_1 - \rho\theta\lambda_2 + \sigma_\mu^2, \\
\lambda_1 &= E\left(F_t^+ F_{t-1}^+\right) = (\rho+\theta)\,\lambda_0 - \rho\theta\lambda_1, \\
\lambda_s &= E\left(F_t^+ F_{t-s}^+\right) = (\rho+\theta)\,\lambda_{s-1} - \rho\theta\lambda_{s-2}, \qquad \forall s \geq 2
\end{aligned} \tag{A-1}$$

which can be solved to obtain

$$\begin{aligned}
\lambda_0 &= \frac{1+\theta\rho}{(1-\theta\rho)(1-\theta-\rho+\theta\rho)(1+\theta+\rho+\theta\rho)}\sigma_\mu^2, \\
&= \frac{1+\theta\rho}{(1-\theta\rho)(1-\rho^2)(1-\theta^2)}\sigma_\mu^2, \qquad \text{(A-2)} \\
\lambda_1 &= \frac{\theta+\rho}{(1-\theta\rho)(1-\rho^2)(1-\theta^2)}\sigma_\mu^2. \qquad \text{(A-3)}
\end{aligned}$$

Next, using that $F_t = F_t^+ - \rho F_{t-1}^+$ we have

$$E\left(F_t F_{t-1}^+\right) = E\left(F_t^+ F_{t-1}^+\right) - \rho E\left(F_{t-1}^+\right)^2 = \lambda_1 - \rho\lambda_0 = \frac{\theta}{(1-\theta^2)(1-\theta\rho)}\sigma_\mu^2. \tag{A-4}$$

Proof of Proposition 1. Suppose Assumptions A1-A2 hold, then:

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}}\left(\widehat{\rho}_{\text{POLSn}} - \rho\right) &= \underset{N\to\infty}{\text{plim}} \frac{(1/NT)\sum_{i=1}^N y_{i,-1}'\left(\gamma_i F + \varepsilon_i\right)}{(1/NT)\sum_{i=1}^N y_{i,-1}' y_{i,-1}}, \\
&= \frac{\underset{N\to\infty}{\text{plim}}\,(1/NT)\sum_{i=1}^N\sum_{t=1}^T\left(y_{i,t-1}^+ + \gamma_i F_{t-1}^+\right)\left(\gamma_i F_t + \varepsilon_{it}\right)}{\underset{N\to\infty}{\text{plim}}\,(1/NT)\sum_{i=1}^N\sum_{t=1}^T\left(y_{i,t-1}^+ + \gamma_i F_{t-1}^+\right)^2}, \\
&= \frac{\frac{1}{T}\sum_{t=1}^T \underset{N\to\infty}{\text{plim}}\,\frac{1}{N}\sum_{i=1}^N\left(y_{i,t-1}^+ + \gamma_i F_{t-1}^+\right)\left(\gamma_i F_t + \varepsilon_{it}\right)}{\frac{1}{T}\sum_{t=1}^T \underset{N\to\infty}{\text{plim}}\,\frac{1}{N}\sum_{i=1}^N\left(y_{i,t-1}^+ + \gamma_i F_{t-1}^+\right)^2}, \\
&= \frac{\frac{1}{T}\sum_{t=1}^T \underset{N\to\infty}{\text{plim}}\,\frac{1}{N}\sum_{i=1}^N \gamma_i^2 F_t F_{t-1}^+}{\frac{1}{T}\sum_{t=1}^T \underset{N\to\infty}{\text{plim}}\,\frac{1}{N}\sum_{i=1}^N\left(\left(y_{i,t-1}^+\right)^2 + \gamma_i^2\left(F_{t-1}^+\right)^2\right)},
\end{aligned}$$

$$= \frac{m_\gamma^2 \frac{1}{T} \sum_{t=1}^T F_t F_{t-1}^+}{\frac{\sigma_\varepsilon^2}{1-\rho^2} + m_\gamma^2 \frac{1}{T} \sum_{t=1}^T \left(F_{t-1}^+\right)^2}. \tag{A-5}$$

Letting $T \to \infty$ and using Lemma A-1, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^+\right)^2 = E\left(F_{t-1}^+\right)^2 + O_p\left(\frac{1}{\sqrt{T}}\right) = \frac{1+\theta\rho}{(1-\theta\rho)(1-\rho^2)(1-\theta^2)}\sigma_\mu^2 + O_p\left(\frac{1}{\sqrt{T}}\right),$$
$$\frac{1}{T}\sum_{t=1}^{T} F_t F_{t-1}^+ = E\left(F_t F_{t-1}^+\right) + O_p\left(\frac{1}{\sqrt{T}}\right) = \frac{\theta}{(1-\theta^2)(1-\theta\rho)}\sigma_\mu^2 + O_p\left(\frac{1}{\sqrt{T}}\right),$$

Thus, taking limits as $N \to \infty$ followed by an expansion as $T \to \infty$, (7) is given by

$$\operatorname*{plim}_{N\to\infty}\left(\widehat{\rho}_{\text{POLSn}} - \rho\right) = \frac{m_\gamma^2\left(\frac{\theta}{(1-\theta^2)(1-\theta\rho)}\sigma_\mu^2 + o_p(1)\right)}{\frac{\sigma_\varepsilon^2}{1-\rho^2} + m_\gamma^2\left(\frac{1+\theta\rho}{(1-\theta\rho)(1-\theta^2)(1-\rho^2)}\sigma_\mu^2 + o_p(1)\right)} \qquad \text{as } T \to \infty,$$
$$= \frac{\left(1-\rho^2\right)\theta}{(1+\theta\rho) + (1-\theta^2)(1-\theta\rho)\frac{\sigma_\varepsilon^2}{m_\gamma^2\sigma_\mu^2}} + o_p(1) \qquad \text{as } T \to \infty.$$

Proof of Proposition 2. First, the probability limit as $N \to \infty$ of the numerator of equation (9) is given by

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}' M_F \varepsilon_i = \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\left(\varepsilon_{it} - \sum_{s=1}^{T}\tau_{st}\varepsilon_{is}\right),$$
$$= \frac{1}{T}\sum_{t=1}^{T}\left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}\varepsilon_{it} - \sum_{s=1}^{T}\tau_{st}\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}\varepsilon_{is}\right), \tag{A-6}$$

where $\tau_{st} = F_s F_t \Big/ \sum_{t=1}^T F_t^2$. Using that

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}\varepsilon_{i,t-s} = \sigma_\varepsilon^2 \rho^{s-1} \qquad \forall s \geq 1, \tag{A-7}$$
$$= 0 \qquad \forall s < 1, \tag{A-8}$$

equation (A-6) can be written as

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}' M_F \varepsilon_i = -\sigma_\varepsilon^2 \frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\tau_{st}\rho^{s+t-1} = -\sigma_\varepsilon^2 \frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\tau_{t,t-s}\rho^{s-1},$$
$$= -\sigma_\varepsilon^2 \frac{1}{T}\left(\sum_{t=2}^{T}\tau_{t,t-1} + \rho\sum_{t=3}^{T}\tau_{t,t-2} + \rho^2\sum_{t=4}^{T}\tau_{t,t-3} + \ldots + \rho^{T-2}\tau_{T,1}\right),$$
$$= -\sigma_\varepsilon^2 \frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\tau_{s,s-t} = -\sigma_\varepsilon^2 \frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1} g_{F,t}, \tag{A-9}$$

where $g_{F,t} = \sum_{s=t+1}^{T}\tau_{s,s-t} = \sum_{s=t+1}^{T} F_s F_{s-t} \Big/ \sum_{t=1}^T F_t^2$.

Second, the probability limit as $N \to \infty$ of the denominator of equation (9) is given by

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}} \frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}' M_F y_{i,-1} &= \underset{N\to\infty}{\text{plim}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\left(y_{i,t-1} - \sum_{s=1}^{T}\tau_{st} y_{i,s-1}\right), \\
&= \underset{N\to\infty}{\text{plim}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^2 - \underset{N\to\infty}{\text{plim}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{st} y_{i,t-1} y_{i,s-1}, \\
&= \frac{1}{T}\sum_{t=1}^{T}\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}^2 - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{st}\,\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1} y_{i,s-1}. \qquad \text{(A-10)}
\end{aligned}$$

Since

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1} y_{i,s-1} &= \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\left(y_{i,t-1}^{+} + \gamma_i F_{t-1}^{+}\right)\left(y_{i,s-1}^{+} + \gamma_i F_{s-1}^{+}\right), \\
&= \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}^{+} y_{i,s-1}^{+} + \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\gamma_i^2 F_{t-1}^{+} F_{s-1}^{+}, \\
&= \frac{\rho^{|t-s|}}{1-\rho^2}\sigma_\varepsilon^2 + m_\gamma^2 F_{t-1}^{+} F_{s-1}^{+}, \qquad \text{(A-11)}
\end{aligned}$$

where $m_\gamma^2 = \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\gamma_i^2$, equation (A-10) can be written as

$$\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sigma_\varepsilon^2}{1-\rho^2} + m_\gamma^2\left(F_{t-1}^{+}\right)^2\right) - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}\left(\frac{\rho^{|t-s|}}{1-\rho^2}\sigma_\varepsilon^2 + m_\gamma^2 F_{t-1}^{+} F_{s-1}^{+}\right), \\
&= \frac{\sigma_\varepsilon^2}{1-\rho^2}\left(1 - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}\rho^{|t-s|}\right) + m_\gamma^2\left(\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{st} F_{t-1}^{+} F_{s-1}^{+}\right), \\
&= \frac{\sigma_\varepsilon^2}{1-\rho^2}\left(1 - \left(\frac{1}{T}\sum_{t=1}^{T}\tau_{tt} + 2\rho\frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\tau_{t,t-s}\rho^{s-1}\right)\right) \\
&\qquad + m_\gamma^2\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2\left(1 - \frac{\sum_{t=1}^{T}\sum_{s=1}^{T} F_t F_s F_{t-1}^{+} F_{s-1}^{+}}{\sum_{t=1}^{T} F_t^2 \sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2}\right), \\
&= \frac{\sigma_\varepsilon^2}{1-\rho^2}\left(1 - \frac{1}{T}\left(1 + 2\rho\sum_{t=1}^{T-1}\rho^{t-1} g_{F,t}\right)\right) + m_\gamma^2 h_F, \qquad \text{(A-12)}
\end{aligned}$$

where

$$h_F = \frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2\left(1 - \frac{\left(\frac{1}{T}\sum_{t=1}^{T} F_t F_{t-1}^{+}\right)^2}{\frac{1}{T}\sum_{t=1}^{T} F_t^2 \frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2}\right).$$

Deviding (A-9) by (A-12) yields the result in equation (10).

Next, letting $T \to \infty$ and using Lemma A-1 we have

$$
\begin{aligned}
g_{F,t} &= \frac{T-t}{T} \frac{E\left(F_s F_{s-t}\right)}{E\left(F_t^2\right)} + O_p\left(\frac{1}{\sqrt{T}}\right), \\
&= \frac{T-t}{T} \theta^t + O_p\left(\frac{1}{\sqrt{T}}\right), \\
h_F &= E\left(\left(F_{t-1}^{+}\right)^2\right)\left(1 - \frac{E\left(F_t F_{t-1}^{+}\right)^2}{E\left(F_t^2\right) E\left(\left(F_{t-1}^{+}\right)^2\right)}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \\
&= \frac{1}{\left(1-\theta\rho\right)^2\left(1-\rho^2\right)} \sigma_\mu^2 + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

Hence, as $T \to \infty$ we have

$$
\begin{aligned}
-\sigma_\varepsilon^2 \frac{1}{T} \sum_{t=1}^{T-1} \rho^{t-1} g_{F,t} &= -\sigma_\varepsilon^2 \frac{1}{T} \sum_{t=1}^{T-1} \rho^{t-1} \left(\frac{T-t}{T} \theta^t + o_p(1)\right), \\
&= -\sigma_\varepsilon^2 \left(\theta \frac{1}{T} \sum_{t=1}^{T-1} \left(\rho\theta\right)^{t-1} - \frac{1}{\rho} \frac{1}{T^2} \sum_{t=1}^{T-1} t \left(\rho\theta\right)^t\right) + o_p\left(\frac{1}{T}\right), \\
&= -\sigma_\varepsilon^2 \left(\theta \frac{1}{T} \frac{1-\left(\theta\rho\right)^{T-1}}{1-\theta\rho} - \frac{1}{\rho} \frac{1}{T^2} \left(\theta\rho \frac{1-\left(\theta\rho\right)^{T-1}}{\left(1-\theta\rho\right)^2} - \left(T-1\right) \frac{\left(\theta\rho\right)^T}{1-\theta\rho}\right)\right) + o_p\left(\frac{1}{T}\right), \\
&= -\sigma_\varepsilon^2 \frac{1}{T} \frac{\theta}{1-\theta\rho} \left[1 - \frac{1}{T} \frac{1-\theta^T \rho^T}{1-\theta\rho}\right] + o_p\left(\frac{1}{T}\right), \\
&= -\sigma_\varepsilon^2 \frac{1}{T} \frac{\theta}{1-\theta\rho} + o_p\left(\frac{1}{T}\right), \qquad \text{(A-13)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\sigma_\varepsilon^2}{1-\rho^2}\left(1 - \frac{1}{T}\left(1 + 2\rho \sum_{t=1}^{T-1} \rho^{t-1} g_{F,t}\right)\right) + m_\gamma^2 h_F, \\
&\quad = \frac{\sigma_\varepsilon^2}{1-\rho^2}\left(1 - \frac{1}{T}\left(1 + 2\rho \frac{\theta}{1-\theta\rho}\left[1 - \frac{1}{T} \frac{1-\theta^T\rho^T}{1-\theta\rho}\right]\right)\right) + m_\gamma^2 \frac{\sigma_\mu^2}{\left(1-\theta\rho\right)^2\left(1-\rho^2\right)} + o_p(1), \\
&\quad = \frac{\sigma_\varepsilon^2}{1-\rho^2} + m_\gamma^2 \frac{\sigma_\mu^2}{\left(1-\theta\rho\right)^2\left(1-\rho^2\right)} + o_p(1). \qquad \text{(A-14)}
\end{aligned}
$$

Thus, taking limits as $N \to \infty$ followed by an expansion as $T \to \infty$, equation (10) is given by

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty} \left(\widehat{\rho}_{\text{POLSi}} - \rho\right) &= \left(-\sigma_\varepsilon^2 \frac{1}{T} \frac{\theta}{1-\theta\rho} + o_p\left(\frac{1}{T}\right)\right)\left(\frac{\sigma_\varepsilon^2}{1-\rho^2} + \frac{m_\gamma^2 \sigma_\mu^2}{\left(1-\theta\rho\right)^2\left(1-\rho^2\right)} + o_p(1)\right)^{-1}, \\
&= -\frac{1}{T} \frac{\theta\left(1-\rho^2\right)}{1-\theta\rho}\left(1 + \frac{m_\gamma^2}{\left(1-\theta\rho\right)^2} \frac{\sigma_\mu^2}{\sigma_\varepsilon^2}\right)^{-1} + o_p\left(\frac{1}{T}\right). \qquad \text{(A-15)}
\end{aligned}
$$

Proof of Theorem 1. First note that

$$\underset{N\to\infty}{\text{plim}}\ \overline{\varepsilon}_t = 0, \tag{A-16}$$

$$\underset{N\to\infty}{\text{plim}}\ \overline{y}_t = \underset{N\to\infty}{\text{plim}}\left(\overline{y}_t^+ + \overline{\gamma}F_t^+\right) = \overline{\gamma}F_t^+, \tag{A-17}$$

such that the probability limit as $N\to\infty$ of the numerator of equation (15) is given by

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}' M_G \widetilde{\varepsilon}_i = \underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}' M_H \varepsilon_i, \tag{A-18}$$

with $M_H = I_T - H\left(H'H\right)^{-1}H'$ and $H = \left(F^+, F_{-1}^+\right)$. Defining $\tau_{s,t}^+$ to be the $(s,t)\,th$ element in $H\left(H'H\right)^{-1}H'$ and using (A-7)-(A-8), (A-18) can be written as

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}' M_G \widetilde{\varepsilon}_i &= \underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\left(\varepsilon_{it} - \sum_{s=1}^{T}\tau_{st}^+\varepsilon_{is}\right), \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}\varepsilon_{it} - \sum_{s=1}^{T}\tau_{st}^+\ \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}\varepsilon_{is}\right), \\
&= -\sigma_\varepsilon^2\frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\tau_{st}^+\rho^{s+t-1} = -\sigma_\varepsilon^2\frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\tau_{t,t-s}^+\rho^{s-1}, \\
&= -\sigma_\varepsilon^2\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\tau_{s,s-t}^+ = -\sigma_\varepsilon^2\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}^+,
\end{aligned} \tag{A-19}$$

where $g_{F,t}^+ = \sum_{s=t+1}^{T}\tau_{s,s-t}^+$.

Second, as (A-17) implies that for $N\to\infty$ $M_G F^+ = 0$ such that $M_G y_{i,-1} = M_G y_{i,-1}^+$, the probability limit as $N\to\infty$ of the denominator of equation (15) is given by

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}' M_G y_{i,-1} &= \underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N} y_{i,-1}^{+'} M_G y_{i,-1}^+, \\
&= \underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^+\left(y_{i,t-1}^+ - \sum_{s=1}^{T}\tau_{st}^+ y_{i,s-1}^+\right), \\
&= \underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1}^+\right)^2 - \underset{N\to\infty}{\text{plim}}\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{st}^+ y_{i,t-1}^+ y_{i,s-1}^+, \\
&= \frac{1}{T}\sum_{t=1}^{T}\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N}\left(y_{i,t-1}^+\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{st}^+\ \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}^+ y_{i,s-1}^+, \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sigma_\epsilon^2}{1-\rho^2}\right) - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}^+\left(\frac{\rho^{|t-s|}}{1-\rho^2}\sigma_\varepsilon^2\right), \\
&= \frac{\sigma_\epsilon^2}{1-\rho^2}\left(1 - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}^+\rho^{|t-s|}\right), \\
&= \frac{\sigma_\varepsilon^2}{1-\rho^2}\left(1 - \left(\frac{1}{T}\sum_{t=1}^{T}\tau_{tt}^+ + 2\rho\frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\tau_{t,t-s}^+\rho^{s-1}\right)\right),
\end{aligned}$$

$$= \frac{\sigma_\varepsilon^2}{1-\rho^2}\left(1 - \frac{2}{T}\left(1+\rho\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}^{+}\right)\right). \tag{A-20}$$

Deviding (A-19) by (A-20) yields the result in equation (16).

In order to derive the the expression for $T \to \infty$, first note that

$$\tau_{st}^{+} = \frac{1}{\alpha_0}\left(\alpha_1 F_t^{+}F_s^{+} - \alpha_2 F_t^{+}F_{s-1}^{+} + \alpha_3 F_{t-1}^{+}F_{s-1}^{+} - \alpha_2 F_{t-1}^{+}F_s^{+}\right), \tag{A-21}$$

with $\alpha_0 = \alpha_1\alpha_3 - \alpha_2^2$, $\alpha_1 = \sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2$, $\alpha_2 = \sum_{t=1}^{T} F_t^{+}F_{t-1}^{+}$ and $\alpha_3 = \sum_{t=1}^{T}\left(F_t^{+}\right)^2$. Using Lemma A-1 and letting $T \to \infty$ we therefore have

$$g_{F,t}^{+} = \frac{T-t}{T}\frac{2\omega_t - \omega_1\omega_{t+1} - \omega_1\omega_{t-1}}{1-\omega_1^2} + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{A-22}$$

with $\omega_t = \lambda_t / \lambda_0$. Hence as $T \to \infty$, the numerator of (16) is given by

$$\begin{aligned}
-\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}^{+} &= -\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}\left(\frac{T-t}{T}\frac{2\omega_t - \omega_1\omega_{t+1} - \omega_1\omega_{t-1}}{1-\omega_1^2} + o_p(1)\right), \\
&= -\frac{(1+\theta\rho)^2}{(1-\theta^2)(1-\rho^2)}\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}\left(1-\frac{t}{T}\right)\left((2-(\theta+\rho)\omega_1)\omega_t - (1-\theta\rho)\omega_1\omega_{t-1}\right) + o_p\left(\frac{1}{T}\right), \\
&= -(1+\theta\rho)\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}\left(1-\frac{t}{T}\right)\left(\frac{\omega_t - \rho\omega_{t-1}}{1-\rho^2} + \frac{\omega_t - \theta\omega_{t-1}}{1-\theta^2}\right) + o_p\left(\frac{1}{T}\right), \\
&= -\frac{1}{T}\left(\frac{\theta}{1-\theta\rho} + \frac{\rho}{1-\rho^2}\right) + o_p\left(\frac{1}{T}\right),
\end{aligned} \tag{A-23}$$

where use is made of $\omega_t = (\theta+\rho)\omega_{t-1} - \theta\rho\omega_{t-2}\ \forall t \geq 2$ and

$$\begin{aligned}
\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t &= (\theta+\rho)\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_{t-1} - \theta\rho\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_{t-2}, \\
&= (\theta+\rho)\left(\left(1-\frac{2}{T}\right)\rho\omega_1 + \rho\left(\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t - \frac{1}{T}\sum_{t=2}^{T-1}\rho^{t-1}\omega_t\right)\right), \\
&\quad - \theta\rho\left(\left(1-\frac{2}{T}\right)\rho + \left(1-\frac{3}{T}\right)\rho^2\omega_1 + \rho^2\left(\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t - \frac{2}{T}\sum_{t=2}^{T-1}\rho^{t-1}\omega_t\right)\right), \\
&= (\theta+\rho)\left(\rho\omega_1 + \rho\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t\right), \\
&\qquad - \theta\rho\left(\rho + \rho^2\omega_1 + \rho^2\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t\right) + O\left(\frac{1}{T}\right), \\
&= \frac{(\theta+\rho)\rho\omega_1 - \theta\rho\left(\rho+\rho^2\omega_1\right)}{1-(\theta+\rho)\rho + \theta\rho^3} + O\left(\frac{1}{T}\right), \\
&= \rho\frac{\theta\left(1-\rho^2\right)(\theta+\rho) + \rho^2\left(1-\theta^2\right)}{(1-\rho^2)(1-\theta^2\rho^2)} + O\left(\frac{1}{T}\right),
\end{aligned}$$

such that

$$\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t=\omega_1+\sum_{t=2}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t=\frac{\theta\left(1-\rho^2\right)+\rho\left(1-\theta^2\rho^2\right)}{\left(1-\rho^2\right)\left(1-\theta^2\rho^2\right)}+O\left(\frac{1}{T}\right),$$

and

$$\begin{aligned}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_{t-1}&=1+\rho\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho^{t-1}\omega_t-\frac{1}{T}\sum_{t=1}^{T}\rho^{t-1}\omega_{t-1},\\&=1+\rho\frac{\omega_1-\theta\rho^2}{\left(1-\rho^2\right)\left(1-\theta\rho\right)}+O\left(\frac{1}{T}\right)=\frac{\left(1-\theta^2\rho^2\right)+\theta\rho\left(1-\rho^2\right)}{\left(1-\rho^2\right)\left(1-\theta^2\rho^2\right)}+O\left(\frac{1}{T}\right).\end{aligned}$$

Similarly, the denominator of (16) as $T\to\infty$ is given by

$$\frac{1}{1-\rho^2}\left(1-\frac{2}{T}\left(1+\rho\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}^{+}\right)\right)=\frac{1}{1-\rho^2}+O\left(\frac{1}{T}\right).\tag{A-24}$$

Thus, taking limits as $N\to\infty$ followed by an expansion as $T\to\infty$, equation (17) is given by

$$\begin{aligned}\operatorname*{plim}_{N\to\infty}\left(\widehat{\rho}_{\text{CCEPu}}-\rho\right)&=-\frac{1}{T}\left[\frac{\theta}{1-\theta\rho}+\frac{\rho}{\left(1-\rho^2\right)}\right]\left(1-\rho^2\right)+o_p\left(\frac{1}{T}\right),\\&=-\frac{1}{T}\left(\frac{\theta(1-\rho^2)}{1-\theta\rho}+\rho\right)+o_p\left(\frac{1}{T}\right).\end{aligned}\tag{A-25}$$

Proof of Theorem 2. Letting $\left(\rho^0,F^0\right)$ denote the true parameter ρ and the true factor F respectively such that, after centering, the objective function in (19) is given by

$$\begin{aligned}S_{NT}\left(\rho,F\right)&=\frac{1}{NT}\sum_{i=1}^{N}\left(y_i-\rho y_{i,-1}\right)'M_F\left(y_i-\rho y_{i,-1}\right)-\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'M_{F^0}\varepsilon_i,\\&=\frac{1}{NT}\sum_{i=1}^{N}\left(\left(\rho^0-\rho\right)y_{i,-1}+\lambda_iF^0+\varepsilon_i\right)'M_F\left(\left(\rho^0-\rho\right)y_{i,-1}+\lambda_iF^0+\varepsilon_i\right)-\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'M_{F^0}\varepsilon_i,\\&=\widetilde{S}_{NT}\left(\rho,F\right)+2\frac{\left(\rho^0-\rho\right)}{NT}\sum_{i=1}^{N}y_{i,-1}'M_F\varepsilon_i+2\frac{1}{NT}\sum_{i=1}^{N}\lambda_iF^{0'}M_F\varepsilon_i+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'\left(M_F-M_{F^0}\right)\varepsilon_i,\end{aligned}$$

where

$$\widetilde{S}_{NT}\left(\rho,F\right)=\frac{\left(\rho^0-\rho\right)^2}{NT}\sum_{i=1}^{N}y_{i,-1}'M_Fy_{i,-1}+\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^2F^{0'}M_FF^0+2\frac{\left(\rho^0-\rho\right)}{NT}\sum_{i=1}^{N}y_{i,-1}'M_FF^0\lambda_i.$$

Using that for $N,T\to\infty$

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}y_{i,-1}'M_F\varepsilon_i&=\frac{1}{NT}\sum_{i=1}^{N}y_{i,-1}'\varepsilon_i-\frac{1}{N}\sum_{i=1}^{N}\frac{y_{i,-1}'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\varepsilon_i}{T}=o_p\left(1\right),\\\frac{1}{NT}\sum_{i=1}^{N}\lambda_iF^{0'}M_F\varepsilon_i&=\frac{1}{NT}\sum_{i=1}^{N}\lambda_iF^{0'}\varepsilon_i-\frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_iF^{0'}F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\varepsilon_i}{T}=o_p\left(1\right),\end{aligned}$$

$$\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'\left(M_F-M_{F^0}\right)\varepsilon_i=\frac{1}{N}\sum_{i=1}^{N}\frac{\varepsilon_i'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\frac{F^{0\prime}\varepsilon_i}{T}-\frac{1}{N}\sum_{i=1}^{N}\frac{\varepsilon_i'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\varepsilon_i}{T}=o_p\left(1\right).$$

we have

$$S_{NT}\left(\rho,F\right)=\widetilde{S}_{NT}\left(\rho,F\right)+o_p\left(1\right), \tag{A-26}$$

uniformly over ρ and F.

First note that as $M_{F^0}F^0=0$, $\widetilde{S}_{NT}\left(\rho^0,F^0\right)=0$. Second, we show that for any $(\rho,F)\neq\left(\rho^0,F^0\right)$, $\widetilde{S}_{NT}\left(\rho,F\right)>0$; thus $\widetilde{S}_{NT}\left(\rho^0,F^0\right)$ attains its unique minimum value at $(\rho,F)=\left(\rho^0,F^0\right)$. Define

$$A=\frac{1}{NT}\sum_{i=1}^{N}y_{i,-1}'M_Fy_{i,-1};\quad B=\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^2;\quad C=\frac{1}{NT}\sum_{i=1}^{N}\lambda_iM_Fy_{i,-1}.$$

Then

$$\begin{aligned}
\widetilde{S}_{NT}\left(\rho,F\right)&=\left(\rho^0-\rho\right)^2A+F^{0\prime}M_FBM_FF^0+2\left(\rho^0-\rho\right)C'M_FF^0,\\
&=\left(\rho^0-\rho\right)^2\left(A-C'B^{-1}C\right)+\left(F^{0\prime}M_F+\left(\rho^0-\rho\right)C'B^{-1}\right)B\left(M_FF^0+B^{-1}C\left(\rho^0-\rho\right)\right),\\
&=\left(\rho^0-\rho\right)^2D\left(F\right)+\theta'B\theta,\\
&\geq 0,
\end{aligned}$$

since $D\left(F\right)=A-C'B^{-1}C$ and B are both positive definite, where $\theta=M_FF^0+B^{-1}C\left(\rho^0-\rho\right)$. Note that $\widetilde{S}_{NT}\left(\rho,F\right)>0$ if either $\rho\neq\rho^0$ or $F\neq F^0$. This implies that $\widehat{\rho}_{\text{CCEPr}}$ is consistent for ρ.

Next, letting $N\to\infty$ and using the same derivations as to obtain (10) together with (A-16) and (A-17), we have from (21)

$$\underset{N\to\infty}{\text{plim}}\left(\widehat{\rho}_{\text{CCEPr}}-\rho\right)=-\frac{\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}g_{\widehat{F},t}}{\frac{1}{1-\rho^2}\left(1-\frac{1}{T}\left(1+2\rho\sum_{t=1}^{T-1}\rho^{t-1}g_{\widehat{F},t}\right)\right)+\frac{m_\gamma^2}{\sigma_\varepsilon^2}h_{\widehat{F}}}, \tag{A-27}$$

where

$$\begin{aligned}
g_{\widehat{F},t}&=\sum_{s=t+1}^{T}\widehat{F}_s\widehat{F}_{s-t}/\sum_{t=1}^{T}\widehat{F}_t^2,\\
h_{\widehat{F}}&=\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2\left(1-\frac{\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{F}_tF_{t-1}^{+}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\widehat{F}_t^2\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2}\right),
\end{aligned}$$

with $\widehat{F}_t=\overline{y}_t-\widehat{\rho}_{\text{CCEPr}}\overline{y}_{t-1}=\left(\rho_0-\widehat{\rho}_{\text{CCEPr}}\right)\overline{y}_{t-1}+\overline{\gamma}F_t+o_p\left(1\right)$. Letting also $T\to\infty$, we have

$$g_{\widehat{F},t}=\frac{\sum_{s=t+1}^{T}F_sF_{s-t}}{\sum_{t=1}^{T}F_t^2}+o_p(1)=g_{F,t}+o_p(1), \tag{A-28}$$

$$h_{\widehat{F}} = \frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2\left(1-\frac{\left(\frac{1}{T}\sum_{t=1}^{T}F_tF_{t-1}^{+}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}F_t^2\frac{1}{T}\sum_{t=1}^{T}\left(F_{t-1}^{+}\right)^2}\right)+o_p\left(\frac{1}{T}\right) = h_F + o_p\left(\frac{1}{T}\right), \quad \text{(A-29)}$$

where use is made of the consistency of $\widehat{\rho}_{\text{CCEPr}}$ which implies that $(\rho_0 - \widehat{\rho}_{\text{CCEPr}}) = o_p(1)$, such that

$$\widehat{F} = \overline{\gamma}F_t + o_p\left(1\right).$$

Substituting (A-28)-(A-29) in (A-27) implies that for $N \to \infty$ followed by an expansion as $T \to \infty$

$$\underset{N\to\infty}{\text{plim}}\left(\widehat{\rho}_{\text{CCEPr}} - \rho\right) = -\frac{\frac{1}{T}\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}}{\frac{1}{1-\rho^2}\left(1-\frac{1}{T}\left(1+2\rho\sum_{t=1}^{T-1}\rho^{t-1}g_{F,t}\right)\right)+\frac{m_\gamma^2}{\sigma_\varepsilon^2}h_F} + o_p\left(\frac{1}{T}\right), \quad \text{(A-30)}$$

which from using Proposition 2 implies

$$\underset{N\to\infty}{\text{plim}}\left(\widehat{\rho}_{\text{CCEPr}} - \widehat{\rho}_{\text{POLSi}}\right) = o_p\left(\frac{1}{T}\right). \quad \text{(A-31)}$$