Bias-Corrected FE Estimator with Exogenous Variables and Heteroskedasticity in Dynamic Panel Models

Rui Sun*
University of Connecticut

January 2020

Abstract

This paper proposes a biased-corrected FE estimator for dynamic panel models under a large sample size. The proposed bias-corrected estimator has several advantages compared to other dynamic panel estimators: first of all, it works for panel autoregressive coefficient $\rho \in (-1, 1]$. Secondly, it is more efficient than the bias-corrected first-difference estimator. Lastly, unlike most existing dynamic panel estimators, the consistency of the proposed bias-corrected estimator does not depend on the stationarity of the initial condition. This paper further extends the model to include exogenous variables and heteroskedasticity. Based on the asymptotic distributions of the FE estimator for $|\rho| < 1$ and $\rho = 1$ under large n and T, the bias-corrected estimators for the models with exogenous variables and heteroskedasticity are proposed. According to the Monte Carlo simulations, the proposed bias-corrected estimator outperforms the GMM-type estimators in a large sample size.

Keywords: dynamic panel data model; fixed effects estimator; large T; bias correction; unit roots; initial condition; exogenous variables; heteroskedasticity

JEL classification: C01 C13 C23

*Department of Economics, University of Connecticut, 341 Mansfield Road, Unit 1063, Storrs, CT 06269-1063; Email: rui.sun@uconn.edu

1 Introduction

Many economic issues involve modeling individual behaviors overtime, for example, the dynamic wage equation, employment model, and investment of firms overtime. Panel data offers empirical researchers great opportunities to understand this dynamic behaviors over time. However, using panel data to model this dynamic also creates challenges, one of which is measuring the individual specific effects in conjunction with the dynamic behavioral effects. People tend to consider the fixed effects approach to treat the unobserved effects as parameters to be estimated, therefore allows arbitrary correlation between the unobserved effects and the regressors. However, estimation of the unobserved effects also brings in the incidental parameters problem as first pointed out by Neyman and Scott (1948), where a finite number of observations is used to estimate the incidental parameters. In the case of dynamic panel data model, when the number of observations for the time series is fixed at T, it is well known that the fixed effects (FE) estimator is inconsistent due to the incidental parameter problem, and so is the first-difference (FD) estimator. A nice review on the dynamic panel data model is available in Baltagi (2013). The standard solution to this inconsistency problem in dynamic panel model is to implement the instrumental variable (IV) estimation which was first proposed by Anderson and Hsiao (1981, 1982). This strategy is performed based on the first-differenced model, which is free of the individual effects, and estimating the autoregressive coefficient using the lagged dependent variable, either in the form of level or first-diferenced, as the instrument. They show that this IV estimator is consistent under $n \to \infty$ or $T \to \infty$ or both. In light of this Anderson-Hsiao estimator, several generalized method of moments (GMM) estimators were proposed, such as Arellano and Bond (1991); Arellano and Bover (1995); Blundell and Bond (1998) where they use all the available lagged dependent variables as instruments, rather than just one lagged dependent variable. However, these GMM estimators have their drawbacks. First of all, it is well known that these GMM estimators suffer from weak instrument problem when the true value of the autoregressive coefficient is close to unity. In the extreme case where the true value equals to one, the instruments become totally irrelevant. Secondly, for these GMM estimators, once we allow the time dimension to grow as the cross-sectional dimension, the number of instruments $\frac{T(T-1)}{2} \to \infty$ as $T \to \infty$[1]. Hence the heavy computation burden of these GMM estimators is a drawback for dynamic panel data with large T. Thirdly, when both n and T tend to infinity, Alvarez and Arellano (2003) showed that GMM estimators are consistent but asymptotically biased of order $1/n$.

Apart from the GMM estimators, a variety of papers suggest bias-corrected estimators. For example, Han and Phillips (2010) proposed a first-differenced least squares (FDLS) estimator for a simple dynamic panel data model where no exogenous variables were included. As we showed in Section 2, this FDLS estimator could be viewed as a bias-corrected FD estimator. Later on, Han, Phillips and Sul (2014) further proposed a panel fully aggregated (PFAE) estimator, which is also called X-differencing estimator, to improve efficiency. Both

[1] In this case, the GMM estimator wouldn't suffer from the many IV problem (see Bekker (1994)), as pointed in Alvarez and Arellano (2003) that large T dimension is desriedble as the "endoegeneity bias" vanishes asymptotically.

FDLS and PFAE estimators can accommodate the unit root case where $\rho = 1$. However, FDLS and PFAE estimator crucially rely on the stationary initial condition assumption. The importance of the initial condition has been studied by previous research such as Kiviet (2005); Chao, Kim and Sul (2014); Hsiao and Zhou (2018). As Chao, Kim and Sul (2014) pointed out when the initial condition is non-stationary, both FDLS and PFAE estimator can be inconsistent. Besides the bias-corrected FD estimator, existing research also considers to correct the bias of the FE estimator. When T is fixed, the FE estimator is inconsistent due to the incidental parameters problem. To recover the consistency, Bun and Kiviet (2001) constructs the bias-corrected FE estimator for the dynamic panel model under fixed T. This bias-corrected FE estimator has been further extended by a series of papers such as Bun (2003); Bruno (2005); Bun and Carree (2005, 2006), to name a few. As shown in Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003), once T is allowed to go to infinity as n, we are able to recover the consistency of the dynamic panel FE estimator without correcting the "endogeneity bias". The intuition is similar to the explanations provided by Alvarez and Arellano (2003); Baltagi, Kao and Liu (2008, 2012, 2017); Kao (1999) that the "endogeneity bias" vanishes completely as $T \to \infty$. However, the asymptotic bias remains and is of order $1/T$. Hahn and Kuersteiner (2002) suggested a bias-corrected FE estimator under large T asymptotics, but for $|\rho| < 1$ only.

In this paper, we first extend the bias-corrected FE estimator in Hahn and Kuersteiner (2002) for a simple dynamic panel data model with no additional regressors to the case of $\rho = 1$. In order to bridge the cases for $|\rho| < 1$ and $\rho = 1$, following the idea of Perron and Yabu (2009a,b), we suggest a new bias-corrected FE estimator for all the ρ values within $(-1, 1]$. This idea of bridge estimator has been applied by Baltagi, Kao and Liu (2014, 2019) in case of panel data with stationary or non-stationary disturbances. We derive the asymptotic distribution of the suggested bias-corrected FE estimator. The proposed bias-corrected has several advantages compared other dynamic panel estimators: first of all, it works for all $\rho \in (-1, 1]$, which means applying this bias-corrected estimator can avoid implementing any pre-testing procedures or having a prior knowledge about the true value of the autoregressive coefficient. Secondly, the bias-corrected estimator we propose is more efficient than the bias-corrected first-difference estimator which also works for all ρ values. Lastly, unlike most existing dynamic panel estimators, the consistency of the proposed bias-corrected estimator does not depend on the stationarity of the initial condition. Therefore, it is more robust to the non-stationary initial condition than other estimators. One restriction of the proposed bias-corrected FE estimator, including FDLS and PFAE estimators, is that it only works for a simple dynamic panel model where no additional regressors are included, which limits its applications in empirical work. Therefore, we further generalize the idea of bias-correction to the model with additional strictly exogenous variables. We first derive the asymptotic distributions for the FE estimator under both $|\rho| < 1$ and $\rho = 1$ with large T asymptotics and show that under both cases the FE estimator is asymptotically biased. Unlike the case in the simple dynamic model where the asymptotic bias only depends on the size of time dimension, T, and the true value of ρ, with exogenous variables being added, the asymptotic bias depends not only on the sample size T and value of ρ, but also the variance of the error and the asymptotic variance-covariance matrix of the within-transformed regressors.

Based on the asymptotic results, we propose the bias-corrected estimator for the model with exogenous variables and derive the corresponding limiting distribution. In addition, we also consider the heteroskedasticity for the bias-corrected estimator. As Phillips and Sul (2007) points out that the particular form for the "endogeneity bias" of the FE estimator does not depend on the mild cross-sectional heteroskedasticity under fixed T. We show that this result holds valid if we allow T to go to infinity and the particular form of the asymptotic bias remains unchanged. However, if one allows for time-series heteroskedasticity, the form of the asymptotic bias does change. In the paper, following Bun and Carree (2006), we allow both cross-sectional and time-series heterscedasticity for the model and construct the bias-corrected estimator accordingly.

The article is organized as follows: Section 2 introduces the model and reviews the bias-corrected FD estimators in Han and Phillips (2010) and Han, Phillips and Sul (2014). We suggest a bias-corrected FE estimator that works for all ρ values and derive the asymptotic distribution of the proposed estimator, and show that it is more efficient than the bias-corrected FD estimator. We extend the simple dynamic panel model to a more general model where exogenous variables are included in Section 3. We derive the limiting distributions of the FE estimator under both $|\rho| < 1$ and $\rho = 1$ cases and construct the bias-corrected estimator accordingly. We further extend the model in Section 4 to allow heteroskedasticity. Simulation results are presented in Section 5 and Section 6 provides the concluding remarks. Mathematical proofs are contained in the supplemental appendix. A few words on notation. We use $(n, T) \to \infty$ to denote the joint limit. Convergence in probability and distribution are denoted as $\overset{p}{\to}$ and $\overset{d}{\to}$, respectively.

2 The bias-corrected FE estimator for simple dynamic panel data model

We consider the following dynamic panel model as suggested by Han and Phillips (2010) and Han, Phillips and Sul (2014):

$$y_{it} = \alpha_i + u_{it}, \tag{1}$$

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \tag{2}$$

for $i = 1, \cdots, n$, $t = 1, \cdots, T$, where α_i is unobservable individual heterogeneity and $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$. The panel autoregressive coefficient $\rho \in (-1, 1]$. This model could be rewritten into

$$y_{it} = \rho y_{i,t-1} + \lambda_i + \varepsilon_{it} \tag{3}$$

where $\lambda_i = (1 - \rho)\alpha_i$. The parameter of interest is the panel autoregressive coefficient ρ, and λ_i are the nuisance parameters to be estimated. When $\rho = 1$, this model is free of individual heterogeneity and when $|\rho| < 1$ the model we consider here, the usual dynamic panel model are not distinguishable. In order to eliminate the individual heterogeneity, we normally perform within-transformation by subtracting the time-series mean from the variables. The FE estimator, which is actually the same as the quasi-maximum likelihood estimator (QMLE), has the following form

$$\hat{\rho}_{FE} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{i,t-1}-\bar{y}_{i,-1}\right)\left(y_{it}-\bar{y}_{i}\right)}{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{i,t-1}-\bar{y}_{i,-1}\right)^{2}} \tag{4}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$. Unfortunately, to eliminate the unobservable individual heterogeneity by demeaning the variable of interest $y_{i,t-1}$, the FE estimator introduces new endogeneity to the transformed regressor, which causes it to be inconsistent under fixed T asymptotics. To see this, let us rewrite Equation (4) into

$$\hat{\rho}_{FE} - \rho = \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{i,t-1}-\bar{y}_{i,-1}\right)\left(\varepsilon_{it}-\bar{\varepsilon}_{i}\right)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{i,t-1}-\bar{y}_{i,-1}\right)^{2}}$$

Hahn and Kuersteiner (2002) showed that if $|\rho| < 1$,

$$\sqrt{nT}\left(\hat{\rho}_{FE} - \rho + \frac{1+\rho}{T}\right) \xrightarrow{d} N\left(0, 1-\rho^2\right), \tag{5}$$

as $(n,T) \to \infty$, which implies a bias of $-\left(1+\rho\right)/T$ in $\hat{\rho}_{FE}$ asymptotically. Hahn and Kuersteiner (2002) hence suggested a bias-corrected estimator of ρ as

$$\hat{\rho}_{HK} = \hat{\rho}_{FE} + \frac{1+\hat{\rho}_{FE}}{T} \tag{6}$$

and its asymptotic distribution is

$$\sqrt{nT}\left(\hat{\rho}_{HK} - \rho\right) \xrightarrow{d} N\left(0, 1-\rho^2\right)$$

as $(n,T) \to \infty$. As we can see, this bias-corrected estimator is asymptotically unbiased, however it is worth pointing out that this estimator only works for the case when $|\rho| < 1$. If $\rho = 1$, Theorem 4 in Hahn and Kuersteiner (2002) shows that

$$\sqrt{n}T\left(\hat{\rho}_{FE} - 1 + \frac{3}{T}\right) \xrightarrow{d} N\left(0, \frac{51}{5}\right) \tag{7}$$

as $(n,T) \to \infty$, which implies a bias of $-3/T$ in $\hat{\rho}_{FE}$.[2] In this case, the bias-corrected estimator $\hat{\rho}_{HK}$ in Equation (6) does not work properly any more. In order to bridge the cases of $|\rho| < 1$ and $\rho = 1$, following the idea of Perron and Yabu (2009*a*,*b*); Baltagi, Kao and Liu (2014, 2019), we can suggest a bias-corrected (BC) estimator of ρ as follows:

$$\hat{\rho}_{BC} = \begin{cases} \hat{\rho}_{FE} + \dfrac{1+\hat{\rho}_{FE}}{T} & \text{if } \hat{\rho}_{FE} < 1 - \dfrac{3}{T} \\ 1 & \text{if } \hat{\rho}_{FE} \geq 1 - \dfrac{3}{T} \end{cases} \tag{8}$$

[2]This result for the case $\rho = 1$ has also been derived in Theorem 2 in Kao (1999), which discusses the asymptotic result in the context of the spurious panel data model.

Essentially, we compare two values, i.e., $1-\hat{\rho}_{FE}$ and $3/T$, to determine what value to assign to the bias-corrected estimator. $1-\hat{\rho}_{FE}$ is the difference between one and the estimated value of FE estimator. $3/T$ is the theoretical bias when $\rho=1$. The intuition of this bias-corrected is the following: we first assume that the true value of ρ is one, if the estimated bias is less than the theoretical bias, i.e., $3/T$, we set the value of bias-corrected estimator to be one. However, if the estimated bias is greater than the theoretical bias of ρ being one, we set the bias-corrected estimator to be $\hat{\rho}_{HK}$ in Equation (6), which is the bias-corrected estimator for the case $|\rho|<1$. Consider the case when the number of time series T is large with a small theoretical bias of $3/T$. If the estimated bias is even smaller than the small theoretical bias of $\rho=1$, this is probably a strong evidence supporting that the true value of the parameter is one. In contrast, if the estimated bias is greater than its theoretical bias, this turns out to favor the fact that the true value is not one. Following Hahn and Kuersteiner (2002), we assume the followings:

Assumption 1 *(i)* $\varepsilon_{it} \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)$*; (ii)* $0<\lim_{(n,T)\to\infty}\frac{n}{T}=\kappa<\infty$*; (iii)* $\frac{1}{n}\sum_{i=1}^n y_{i0}^2 = O_p(1)$*; (iv)* $\frac{1}{n}\sum_{i=1}^n \alpha_i^2 = O_p(1)$*.*

The asymptotic results for $\hat{\rho}_{BC}$ are given in the following theorem:

Theorem 1 *As* $(n,T)\to\infty$*, under Assumption 1, we have*

1. *If* $|\rho|<1$*,*
$$\sqrt{nT}\left(\hat{\rho}_{BC}-\rho\right) \xrightarrow{d} N\left(0,1-\rho^2\right).$$
2. *If* $\rho=1$*,*
$$\sqrt{n}T\left(\hat{\rho}_{BC}-1\right) \xrightarrow{p} 0.$$

As an important remark, Theorem 1 does not rely on the stationary initial condition as in Han and Phillips (2010) and Han, Phillips and Sul (2014). For example, if $u_{i0} \overset{iid}{\sim} N(m_1,m_2)$ for all i, where $m_2>0$, Section 3.1 in Chao, Kim and Sul (2014) showed that the consistency of FE estimator still holds. This means, the $\hat{\rho}_{FE}$ has the same asymptotic bias whether or not the initial observation satisfies $m_2=\sigma_\varepsilon^2/\left(1-\rho^2\right)$. Therefore, a big advantage of our suggested bias-corrected FE estimator is that it is robust to different initial conditions. We illustrate it by Monte Carlo simulations in the Section 5.

3 Dynamic panel data model with exogenous variables

In this section, we extend the simple dynamic panel data model in Section 2 to the model with K additional exogenous variables. We assume that

$$y_{it} = \alpha_i + u_{it}, \tag{9}$$
$$u_{it} = \rho u_{i,t-1} + X_{it}'\beta + \varepsilon_{it}, \tag{10}$$

for $i = 1, \cdots, n$, $t = 1, \cdots, T$, where u_{it} is a latent variable that determines the dynamic structure of the dependent variable. The panel autoregressive coefficient is allowed to be in $(-1, 1]$. The model specified above can be expressed as the conventional dynamic panel data model as follows,

$$y_{it} = \rho y_{i,t-1} + X_{it}{'}\beta + \lambda_i + \varepsilon_{it}, \tag{11}$$

where $\lambda_i = (1-\rho)\alpha_i$. In this model, the dependent variable y_{it} is determined by the one-period lagged dependent variable $y_{i,t-1}$, a $K \times 1$ vector of explanatory variables X_{it}, an individual heterogeneity α_i, and an idiosyncratic error term ε_{it}. Unlike the simple dynamic panel data model in Equation (1) and (2), where no explanatory variables are included, this model has wide applications. For example, Balestra and Nerlove (1966) studies the demand for natural gas where the exogenous variables are included in addition to the lagged dependent variable, and also the panel analysis of growth convergence where other covariates are included to study the dynamic. We first formalize the assumptions for the dynamic panel model with additional explanatory variables as follows,

Assumption 2 *(i)* $(X_{i1}, \cdots, X_{iT}, \varepsilon_{i1}, \cdots, \varepsilon_{iT})$ *are independent and identically distributed over* $i = 1, \cdots, n$*; (ii)* $E[\varepsilon_{it}|X_{i1}, \cdots, X_{iT}, y_{i0}] = 0$*; (iii)* $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[X_{it}'X_{it}]$ *is nonsingular; (iv)* $E[\varepsilon_{it}\varepsilon_{is}|X_{i1}, \cdots, X_{iT}, y_{i0}] = 0$ *for* $t \neq S$*.*

Assumption 2 states that (i) both X_{it} and ε_{it} are *i.i.d.* over i, (ii) the explanatory variables X_{it} are strictly exogenous, (iii) there is no multicollinearity in X_{it}, and (iv) the error terms are conditional serially uncorrelated. In regards to the error behavior, we assume that

Assumption 3 $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ *for* $i = 1, \cdots, n$, $t = 1, \cdots, T$*.*

In contrast to the preview sections where the normality assumption is not necessary, in this section we assume the error term ε_{it} is normally distributed with mean 0 and constant variance σ_ε^2. To proceed, we stack T time periods observations for the i^{th} individual based on Equation (11) yields

$$y_i = \rho y_{i,-1} + X_i\beta + \lambda_i 1_T + \varepsilon_i,$$

where $y_i = (y_{i1}, \cdots, y_{iT})'$, $y_{i,-1} = (y_{i0}, \cdots, y_{i,T-1})'$, $X_i = (X_{i1}{'}, \cdots, X_{iT}{'})'$, $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$, and $1_T = (1, \cdots, 1)'$. Furthermore, we stack all the observations over time and across individuals, we have

$$y = Z\delta + (I_n \otimes 1_T)\lambda + \varepsilon$$

where $\delta = (\rho, \beta')'$ collects $K+1$ parameters, y and Z are $nT \times 1$ and $nT \times (K+1)$ matrices of stacked observations with $Z = [y_{-1} \vdots X]$, $\lambda = (\lambda_1, \cdots, \lambda_n)'$, and ε is the $nT \times 1$ vector of errors. Using the standard regression results, the FE estimator for δ can be expressed as

$$\begin{aligned} \hat{\delta}_{FE} = \begin{pmatrix} \hat{\rho}_{FE}^k \\ \hat{\beta}_{FE}' \end{pmatrix} &= (Z'AZ)^{-1} Z'Ay \\ &= \delta + (Z'AZ)^{-1} Z'A\varepsilon \\ &= \delta + (Z'AZ)^{-1} \begin{pmatrix} y_{-1}'A\varepsilon \\ X'A\varepsilon \end{pmatrix}, \end{aligned} \tag{12}$$

where A is a $nT \times nT$ idempotent within-transformation matrix, i.e., $A = I_n \otimes A_T$ with $A_T = I_T - \frac{1}{T} 1_T 1_T'$. To distinguish the FE estimator in Section 2, we use the superscript K in $\hat{\rho}^k_{FE}$ to indicate the fact that K exogenous variables are included. The strict exogeneity in Assumption 2 (ii) implies that $E[X' A \varepsilon] = 0$, however due to the non-zero correlation between the within-transformed one-period lagged dependent variable and the within-transformed error term, i.e., $E[y_{-1}' A\varepsilon] \neq 0$, we have $E[Z' A\varepsilon] \neq 0$. Here the expectation refers to the expectation conditional on X_{it} and y_{i0}. As shown in Nickell (1981)[3], when $|\rho| < 1$

$$E[y_{-1}' A\varepsilon] = -\sigma_\varepsilon^2 \frac{n}{1-\rho} \left(1 - \frac{1}{T} \frac{1-\rho^T}{1-\rho}\right).$$

It can be shown that when $\rho = 1$, the inclusion of additional exogenous variables will not change the particular form of the source of inconsistency as well. Specifically, when $\rho = 1$

$$E[y_{-1}' A\varepsilon] = -\sigma_\varepsilon^2 \frac{n(T-1)}{2}.$$

The fact that the source of inconsistency remains uncorrelated with the introduction of the K additional variables is due to the assumed strict exogeneity of X_{it} with respect to the error term. The FE estimator, regardless of the ρ values, is consistent under large T, however remains to be asymptotically biased of $O(1/T)$.

Our goal is to first characterize the asymptotic bias based on the limiting distributions of the FE estimator under large T for both $|\rho| < 1$ and $\rho = 1$, and then construct the asymptotic unbiased estimators based on the limiting distributions. Once we have the asymptotic unbiased estimators, we employ the same idea of the bridge estimator in Section 2 to build the bias-corrected estimator that works for $\rho \in (-1, 1]$. To characterize the limiting behaviors of $Z'AZ$ and $Z'A\varepsilon$ as in Equation (12), following Kiviet (1995) we introduce the following notations: for any stochastic or non-stochastic $w_1 \times w_2$ matrix W, we define $\bar{W} = E[W]$ and $\tilde{W} = W - E[W]$, hence $W = \bar{W} + \tilde{W}$, where $\bar{W}$ is nonrandom and $E[\tilde{W}] = 0$. Since the stochastic nature of $Z = [y_{-1} \vdots X]$ comes from the inclusion of the lagged dependent variable, i.e., y_{-1}, we have

$$\bar{Z} = [\bar{y}_{-1} \vdots X], \quad \tilde{Z} = [\tilde{y}_{-1} \vdots 0].$$

We first focus on the stochastic component of $AZ = A\bar{Z} + A\tilde{Z}$, i.e., $A\tilde{Z}$. More specifically, we focus on $A\tilde{y}_{-1}$.

I. Stationary case ($|\rho| < 1$):

Starting from the stationary case where $|\rho| < 1$, since λ_i is time invariant and based on equation (9), we have

$$A\tilde{y}_{-1} = A\tilde{u}_{-1}, \tag{13}$$

where u_{-1} is a $nT \times 1$ vector that stacks all u_{it} for $i = 1, \cdots n, t = 0, \cdots, T-1$. From the equation above, we can see that the stochastic nature of AZ originates from $\tilde{u}_{-1}$, which

[3]Nickell (1981) points out in Equation (27) that the particular form of the source of inconsistency remains unchanged after introducing the exogenous variables.

contains the idiosyncratic errors determine $\tilde{y}_{-1}$. Next, we relate $\tilde{u}_{-1}$ to the error term ε. We proceed by writing the stochastic version of Equation (10). Since $\tilde{X}_{it} = 0$ and $\tilde{\varepsilon}_{it} = \varepsilon_{it}$ as $E[\varepsilon_{it}] = 0$, we have

$$\tilde{u}_{it} = \rho \tilde{u}_{i,t-1} + \varepsilon_{it}.$$

Stacking T observations of the above equation for the i^{th} individual gives us

$$\tilde{u}_i - \rho \tilde{u}_{i,-1} = \varepsilon_i. \tag{14}$$

Define the lag operator matrix L_T which is a $T \times T$ matrix with first lower subdiagonal being ones and all others elements being zeros, i.e.,

$$L_T = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

L_T enables us to relate $\tilde{u}_{i,-1}$ to $\tilde{u}_i$, i.e., $\tilde{u}_{i,-1} = L_T \tilde{u}_i$. As $\bar{u}_{i0} = u_{i0}$ and $\tilde{u}_{i0} = 0$, Equation (14) can be rewritten as

$$(I_T - \rho L_T)\tilde{u}_i = \varepsilon_i,$$

which relates $\tilde{u}_i$ to the error term ε_i. We further write it as $\tilde{u}_i = \Gamma_{\mathrm{T}} \varepsilon_i$, where $\Gamma_{\mathrm{T}} = (I_T - \rho L_T)^{-1}$. Consequently, given $\tilde{u}_{i,-1} = L_T \tilde{u}_i$, we have

$$L_T \Gamma_{\mathrm{T}} \varepsilon_i = \tilde{u}_{i,-1},$$

where

$$L_T \Gamma_T = \begin{bmatrix} 0 & \cdot & \cdot & \cdots & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdots & \cdot & \cdot & \cdot \\ \rho & 1 & 0 & \cdots & \cdot & \cdot & \cdot \\ \rho^2 & \rho & 1 & \cdots & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \rho^{T-2} & 0 & 0 & \cdots & \rho & 1 & 0 \end{bmatrix}. \tag{15}$$

Stacking all the observations for n individuals yields

$$L \Gamma \varepsilon = \tilde{u}_{-1},$$

where $L = I_n \otimes L_T$ and $\Gamma = I_n \otimes \Gamma_{\mathrm{T}}$. We plug the above equation, where $\tilde{u}_{-1}$ is related to ε, into Equation (13) so that the stochastic component of the endogenous variable, i.e., $\tilde{y}_{-1}$, is expressed in terms of the error term ε. In particular,

$$A \tilde{y}_{-1} = \Pi \varepsilon,$$

where $\Pi = I_n \otimes \Pi_T$ with $\Pi_T = A_T L_T \Gamma_T$. As $\tilde{Z} = [\tilde{y}_{-1} \vdots 0]$, we have

$$A\tilde{Z} = A\tilde{y}_{-1}e'_{K+1} = \Pi\varepsilon e'_{K+1} \tag{16}$$

where $e_{K+1} = (1, 0, \cdots, 0)'$ has $K + 1$ elements. Essentially, Equation (16) relates the within-transformed regressors $A\tilde{Z}$ including y_{-1} to the within-transformed error term. This equation shows the source of inconsistency of the FE estimator under fixed T. In what follows, we make use of Equation (16) to characterize the asymptotic bias term and the limiting distribution of the FE estimator under large T. A series of literature including Bun and Kiviet (2001), Kiviet (1995), and Bun and Carree (2005) study the issues of bias-correction for the FE estimator for the dynamic panel data model with exogenous variables under the fixed T asymptotics and when $|\rho| < 1$. We generalize their results to the large T (and large n) case and derive the limiting distribution of the FE estimator when $\rho = 1$. Based on the large T asymptotic results, we construct the bias-corrected estimator for the stationary case, which is considered as the first step to construct the bias-corrected estimator that bridges the stationary and non-stationary cases.

To show that the FE estimator $\hat{\delta}_{FE}$ under large T is asymptotically biased, we need to first show that $Z'A\varepsilon$ is asymptotically normal but not centered at zero, and then characterize the asymptotic bias term for the estimator $\hat{\delta}_{FE}$. Since $E[\bar{Z}'A\varepsilon] = 0$, using (16) we obtain

$$E[Z'A\varepsilon] = E[\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon] = e_{K+1}E[\varepsilon'\Pi'\varepsilon] = e_{K+1}E[\varepsilon'(I_n \otimes \Pi'_T)\varepsilon] = n\sigma_\varepsilon^2 tr(\Pi_T)e_{K+1}$$

where

$$tr(\Pi_T) = -\frac{1}{1-\rho}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right),$$

and the variance[4]

$$var(Z'A\varepsilon) = \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4 tr(\Pi_T^2)e_{K+1}e'_{K+1} \tag{17}$$

where

$$tr(\Pi_T^2) = -\frac{1+2\rho^{T-1}}{(1-\rho)^2} + \frac{2(1-\rho^T)}{T(1-\rho)^3} + \frac{(1-\rho^T)^2}{T^2(1-\rho)^4}.$$

Therefore, as $(n, T) \to \infty$ we have

$$\lim_{(n,T)\to\infty} \frac{1}{nT} var(Z'A\varepsilon) = \sigma_\varepsilon^2 Q_{ZZ} \tag{18}$$

with

$$Q_{ZZ} \equiv \underset{(n,T)\to\infty}{\text{plim}} \frac{1}{nT} Z'AZ = \underset{(n,T)\to\infty}{\text{plim}} \frac{1}{nT} \begin{bmatrix} y'_{-1}Ay_{-1} & y'_{-1}AX \\ X'Ay_{-1} & X'AX \end{bmatrix} = \begin{bmatrix} \sigma_{y_{-1}}^2 & \Sigma'_{Xy_{-1}} \\ \Sigma_{Xy_{-1}} & \Sigma_{XX} \end{bmatrix} \tag{19}$$

[4]See Appendix B for the derivation.

By central limit theorem, as $(n,T)\to\infty$, we have

$$\frac{1}{\sqrt{nT}}\left(Z'A\varepsilon - E[Z'A\varepsilon]\right) \xrightarrow{d} N\left(0,\sigma_\varepsilon^2 Q_{ZZ}\right).$$

Due to the fact that $E[Z'A\varepsilon]\neq 0$, $\frac{1}{\sqrt{nT}}Z'A\varepsilon$ converges to a normal random variable centered at a nonzero value. In particular, as $(n,T)\to\infty$ with $\frac{n}{T}\to\kappa$,

$$\lim_{(n,T)\to\infty}\frac{1}{\sqrt{nT}}E[Z'A\varepsilon] = \lim_{(n,T)\to\infty}\sqrt{\frac{n}{T}}\sigma_\varepsilon^2 tr(\Pi_T)e_{K+1} = -\sqrt{\kappa}\frac{\sigma_\varepsilon^2}{1-\rho}e_{K+1}.$$

Consequently,

$$\frac{1}{\sqrt{nT}}Z'A\varepsilon \xrightarrow{d} N\left(-\sqrt{\kappa}\frac{\sigma_\varepsilon^2}{1-\rho}e_{K+1},\sigma_\varepsilon^2 Q_{ZZ}\right).$$

We have the following theorem for the FE estimator under stationary case, i.e., $|\rho|<1$.

Theorem 2 *When $|\rho|<1$, under Assumption 1 (ii)-(iv), 2, and 3, as $(n,T)\to\infty$, we have*

$$\sqrt{nT}(\hat{\delta}_{FE}-\delta) \xrightarrow{d} N\left(B_\delta^s,\Omega_\delta^s\right)$$

where $B_\delta^s = -\sqrt{\kappa}\frac{\sigma_\varepsilon^2}{1-\rho}Q_{ZZ}^{-1}e_{K+1}$ and $\Omega_\delta^s=\sigma_\varepsilon^2 Q_{ZZ}^{-1}$.

An obvious implication of Theorem 2 is that

$$\sqrt{nT}(\hat{\delta}_{FE}-\delta^s-\delta) \xrightarrow{d} N\left(0,\Omega_\delta^s\right).$$

where $\delta^s = -\frac{\sigma_\varepsilon^2}{T(1-\rho)}Q_{ZZ}^{-1}e_{K+1}$. Based on this result, a bias-corrected estimator for the stationary case could be suggested as

$$\hat{\delta}_{BC}^s = \hat{\delta}_{FE} + \frac{\sigma_\varepsilon^2}{T(1-\rho)}Q_{ZZ}^{-1}e_{K+1} \tag{20}$$

and it is easily seen that

$$\sqrt{nT}(\hat{\delta}_{BC}^s-\delta) \xrightarrow{d} N\left(0,\Omega_\delta^s\right) \tag{21}$$

where $\Omega_\delta^s = \sigma_\varepsilon^2 Q_{ZZ}^{-1}$. Under large T asymptotics, the FE estimator $\hat{\delta}_{FE}$ is $\sqrt{nT}$-consistent. Hence, ρ, σ_ε^2, and Q_{ZZ} can all be consistently estimated. As a consequence, the corresponding bias-corrected estimator for ρ is,

$$\hat{\rho}_{BC}^k = \hat{\rho}_{FE}^k + \frac{\sigma_\varepsilon^2}{T(1-\rho)}e_{K+1}'Q_{ZZ}^{-1}e_{K+1}.$$

From the joint asymptotic normality of $\hat{\delta}_{BC}^s$ in (21), we can obtain the marginal limiting distribution of $\hat{\rho}_{BC}^k$ which is

$$\sqrt{nT}(\hat{\rho}_{BC}^k-\rho) \xrightarrow{d} N(0,\Omega_\rho^s),$$

where $\Omega_\rho^s = e'_{K+1}\Omega_\delta^s e_{K+1}$. Notice that this result is a generalization of the result that obtained in Hahn and Kuersteiner (2002) where they consider the case with no exogenous variables. To see this, consider the bias-corrected estimator as in (20), when $K = 0$ we have

$$\hat{\rho}^k_{BC} = \hat{\rho}^k_{FE} + \frac{\sigma_\varepsilon^2}{T(1-\hat{\rho}_{FE})\sigma^2_{y_{-1}}}$$

where $\sigma^2_{y_{-1}} = \text{plim}_{(n,T)\to\infty} \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{i,-1})^2$. As we have showed that $\sigma^2_{y_{-1}} = \frac{\sigma_\varepsilon^2}{1-\rho^2}$ when no exogenous variable is included. It turns out that

$$\hat{\rho}^k_{BC} = \hat{\rho}^k_{FE} + \frac{1+\hat{\rho}_{FE}}{T}$$

which is the same as in (13). To employ the same idea to construct the bis-corrected estimator to bridge the cases of $|\rho| < 1$ and $\rho = 1$, we need to derive the asymptotic bias term for the non-stationary case which can then be used as the threshold to determine the value of the new bias-corrected estimator.

II. Non-stationary case ($\rho = 1$):

For the non-stationary case, we employ the same method as for the stationary case. We start with the model (10), when $\rho = 1$ it can be simplified as

$$y_{it} = y_{i,t-1} + X_{it}{}'\beta + \varepsilon_{it},$$

where the individual heterogeneity λ_i in this case is zero. When T is fixed, the FE estimator of ρ is inconsistent due to the elimination of an unknown constant from each observation, which yields the correlation between the within-transformed $y_{i,t-1}$ and within-transformed ε_{it}. Maintaining the same notation we use to distinguish the stochastic and non-stochastic parts of a matrix, we have the following equation for the non-stationary case,

$$\tilde{y}_{it} = \tilde{y}_{i,t-1} + \varepsilon_{it}.$$

This is due to the fact that $\tilde{X}_{it} = 0$ and $\tilde{\varepsilon}_{it} = \varepsilon_{it}$. Now, having $\rho = 1$ we can decompose the vector $\tilde{y}_{i,-1}$, which stacks time T observations for the i^{th} individual as

$$\tilde{y}_{i,-1} = C_T\varepsilon_i,$$

where C_T is the $T \times T$ matrix

$$C_T = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix}. \tag{22}$$

After applying the within-transformation matrix A_T, we have

$$A_T\tilde{y}_{i,-1} = \Phi_T\varepsilon_i,$$

where $\Phi_T = A_T C_T$. Again, we stack all the observations for n individuals, we have

$$A\tilde{y}_{-1} = \Phi\varepsilon,$$

where $\Phi = I_n \otimes \Phi_T$. Like in the stationary case, the equation above allows us to relate the within-transformed lagged dependent variable to the within-transformed error term. As a result,

$$A\tilde{Z} = \Phi\varepsilon e'_{K+1}, \tag{23}$$

which parallels Equation (16) for the stationary case and will be used to derive the limiting behavior of the FE estimator under $\rho = 1$. For the non-stationary case, it has been shown that the rate of convergence for the lagged dependent variable is faster than the usual rate $\sqrt{nT}$ due to the stronger time series signal in the regressor of the unit root case (see Hahn and Kuersteiner (2002) and Harris and Tzavalis (1999)). However, for the remaining exogenous variables the rates of convergence stay the same as the usual rate. Therefore, to accommodate the difference in the rate of convergence, we introduce the following matrix: $\Lambda = \text{diag}(r_1, \cdots, r_{K+1})$, with $r_1 = \sqrt{n}T$ and $r_i = \sqrt{nT}$ for $i = 2, \cdots, K+1$. Our goal is to derive the limiting distribution of

$$\Lambda(\hat{\delta}_{FE} - \delta) = \left(\Lambda^{-1}Z'AZ\Lambda^{-1}\right)^{-1}\left(\Lambda^{-1}Z'A\varepsilon\right)$$

For the denominator, with the right rates we assume that

$$Q_{ZZ} \equiv \underset{(n,T)\to\infty}{\text{plim}} \Lambda^{-1}Z'AZ\Lambda^{-1} = \underset{(n,T)\to\infty}{\text{plim}} \begin{bmatrix} n^{-1}T^{-2}y'_{-1}Ay_{-1} & n^{-1}T^{-3/2}y'_{-1}AX \\ n^{-1}T^{-3/2}X'Ay_{-1} & n^{-1}T^{-1}X'AX \end{bmatrix} = \begin{bmatrix} \sigma^2_{y_{-1}} & \Sigma'_{Xy_{-1}} \\ \Sigma_{Xy_{-1}} & \Sigma_{XX} \end{bmatrix} \tag{24}$$

For the numerator $\Lambda^{-1}Z'A\varepsilon$, making use of Equation (23), we obtain the following results: first of all,

$$E[Z'A\varepsilon] = E[\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon] = n\sigma^2_\varepsilon tr(\Phi_T)e_{K+1} \tag{25}$$

where

$$tr(\Phi_T) = -\frac{T-1}{2}.$$

The variance[5]

$$var(Z'A\varepsilon) = \sigma^2_\varepsilon E[Z'AZ] + n\sigma^4_\varepsilon tr(\Phi^2_T)e_{K+1}e'_{K+1} \tag{26}$$

where

$$tr(\Phi^2_T) = -\frac{1}{12}T^2 + \frac{1}{2}T - \frac{5}{12}.$$

[5]See Appendix B for derivation.

Consequently, the asymptotic variance can be obtained by

$$\lim_{(n,T)\to\infty} var(\Lambda^{-1}Z'A\varepsilon) = \sigma_\varepsilon^2 Q_{ZZ} - \frac{\sigma_\varepsilon^4}{12} e_{K+1}e'_{K+1}. \tag{27}$$

By central limit theorem, as $(n,T)\to\infty$

$$\Lambda^{-1}\left(Z'A\varepsilon - E[Z'A\varepsilon]\right) \xrightarrow{d} N\left(0, \sigma_\varepsilon^2 Q_{ZZ} - \frac{\sigma_\varepsilon^4}{12} e_{K+1}e'_{K+1}\right).$$

However, as we have derived in (25) that

$$E[Z'A\varepsilon] = -\sigma_\varepsilon^2 \frac{n(T-1)}{2} e_{K+1}.$$

We have the following limiting distribution for the fix effected estimator under the non-stationary case:

Theorem 3 *When $\rho = 1$, under Assumption 1 (iii), 2, and 3, as $(n,T)\to\infty$, we have*

$$\Lambda(\hat{\delta}_{FE} - \delta^n - \delta) \xrightarrow{d} N(0, \Omega_\delta^n)$$

where $\delta^n = -\frac{\sigma_\varepsilon^2}{2T} Q_{ZZ}^{-1} e_{K+1}$ and $\Omega_\delta^n = (\sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^4}{12} q + \frac{\sigma_\varepsilon^8}{180} q^3) Q_{ZZ}^{-1}$ with $q = Q_{ZZ}^{-1} e_{K+1} e'_{K+1}$.

Notice that for the non-stationary case, we do not impose the rate condition for n and T. The bias-corrected estimator for $\rho = 1$ case can be defined as

$$\hat{\delta}_{BC}^n = \hat{\delta}_{FE} + \frac{\sigma_\varepsilon^2}{2T} Q_{ZZ}^{-1} e_{K+1} \tag{28}$$

In the special case of $K = 0$, the asymptotic bias term is given by $\delta^n = -\frac{\sigma_\varepsilon^2}{2T}\left(\sigma_{y_{-1}}^2\right)^{-1}$ and asymptotic variance is $\Omega_\delta^n = \sigma_\varepsilon^2\left(\sigma_{y_{-1}}^2\right)^{-1} - \frac{\sigma_\varepsilon^4}{12}\left(\sigma_{y_{-1}}^2\right)^{-2} + \frac{\sigma_\varepsilon^8}{180}\left(\sigma_{y_{-1}}^2\right)^{-4}$. We have showed that when $\rho = 1$, $\sigma_{y_{-1}}^2 = \frac{\sigma_\varepsilon^2}{6}$. Thus, directly apply Theorem 3 gives us

$$\sqrt{n}T\left(\hat{\rho}_{FE}^k - \rho + \frac{3}{T}\right) \xrightarrow{d} N\left(0, \frac{51}{5}\right),$$

which is consistent with the result obtained by Harris and Tzavalis (1999), Hahn and Kuersteiner (2002), and Moon, Perron and Phillips (2015).

Apply the same idea of the bias-corrected estimator as we propose in (8), the bias-corrected estimator of ρ for the model with exogenous variables can be suggested as

$$\hat{\rho}_{BC}^k = \begin{cases} \hat{\rho}_{FE}^k + \dfrac{\sigma_\varepsilon^2}{T(1-\rho)} e'_{K+1} Q_{ZZ}^{-1} e_{K+1} & \text{if } \hat{\rho}_{FE}^k < 1 - \dfrac{\sigma_\varepsilon^2}{2T} e'_{K+1} Q_{ZZ}^{-1} e_{K+1} \\ 1 & \text{if } \hat{\rho}_{FE}^k \geq 1 - \dfrac{\sigma_\varepsilon^2}{2T} e'_{K+1} Q_{ZZ}^{-1} e_{K+1} \end{cases}$$

Theorem 4 *As $(n, T) \to \infty$, under Assumption 1 – 3, we have*

1. *If $|\rho| < 1$,*
$$\sqrt{nT}\left(\hat{\rho}_{BC}^{k} - \rho\right) \xrightarrow{d} N\left(0, \Omega_{\rho}^{s}\right).$$

2. *If $\rho = 1$,*
$$\sqrt{n}T\left(\hat{\rho}_{BC}^{k} - 1\right) \xrightarrow{p} 0.$$

In many empirical applications, a researcher might be interested in estimating the coefficient of exogenous variables β, as the lagged dependent variable can be only served as a control variable to eliminate some potential bias of β. For example, Donohue III and Levitt (2001) studies the effect of legalized abortion on crimes in the USA. They recognize that the source of possible bias in the analysis of the relationship between abortion and crime could be due to a dynamic misspecification. To control for the cumulative and persistent effects from abortion to crimes, one needs to include the lagged crime rates in the specification. From the Theorem 2, when $|\rho| < 1$ we have

$$\sqrt{nT}(\hat{\delta}_{FE} - \delta^{s} - \delta) \xrightarrow{d} N\left(0, \Omega_{\delta}^{s}\right),$$

where $\delta^{s} = -\frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} Q_{ZZ}^{-1} e_{K+1}$. From Theorem 3 we know, when $\rho = 1$

$$\Lambda(\hat{\delta}_{FE} - \delta^{n} - \delta) \xrightarrow{d} N(0, \Omega_{\delta}^{n}),$$

where $\delta^{n} = -\frac{\sigma_{\varepsilon}^{2}}{2T} Q_{ZZ}^{-1} e_{K+1}$. As we can see, the whole vector of $\hat{\delta}_{FE}$ is asymptotically biased regardless of ρ which would result in misleading inference on $\hat{\beta}_{FE}$ if we do not correct the bias for $\hat{\beta}_{FE}$. To construct the bias-corrected estimator for $\hat{\beta}_{FE}$ that bridges between the two types of the limiting distributions above, we can use the same decision rule in constructing $\hat{\rho}_{BC}^{k}$. If the bias-correction under stationarity were used in $\hat{\rho}_{BC}^{k}$, that is $\hat{\rho}_{FE}^{k} < 1 - \dfrac{\sigma_{\varepsilon}^{2}}{2T} e_{K+1}' Q_{ZZ}^{-1} e_{K+1}$, we use the asymptotic bias term of $\hat{\beta}_{FE}$ under stationary to correct the bias, otherwise we use the asymptotic bias term under the unit root to correct the bias. Define $E_{K} \equiv [0_{K} \vdots I_{K}]$ as a $K \times (K+1)$ matrix that extracts the second to last element from a $(K+1) \times 1$ vector where $0_{K} \equiv (0, \cdots, 0)'$. The corresponding bias-corrected estimator for $\hat{\beta}_{FE}$ can be suggested as follows

$$\hat{\beta}_{BC} = \begin{cases} \hat{\beta}_{FE} + \dfrac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_{K} Q_{ZZ}^{-1} e_{K+1} & \text{if } \hat{\rho}_{FE}^{k} < 1 - \dfrac{\sigma_{\varepsilon}^{2}}{2T} e_{K+1}' Q_{ZZ}^{-1} e_{K+1} \\ \hat{\beta}_{FE} + \dfrac{\sigma_{\varepsilon}^{2}}{2T} E_{K} Q_{ZZ}^{-1} e_{K+1} & \text{if } \hat{\rho}_{FE}^{k} \geq 1 - \dfrac{\sigma_{\varepsilon}^{2}}{2T} e_{K+1}' Q_{ZZ}^{-1} e_{K+1} \end{cases}$$

where $\hat{\rho}_{FE}^{k}$ and $\hat{\beta}_{FE}$ are defined in (12).

Theorem 5 *As $(n, T) \to \infty$, under Assumption 1 – 3, we have*

1. *If* $|\rho| < 1$,
$$\sqrt{nT}\left(\hat{\beta}_{BC} - \beta\right) \xrightarrow{d} N\left(0, \Omega_{\beta}^{s}\right).$$
where $\Omega_{\beta}^{s} = \sigma_{\varepsilon}^{2} E_K Q_{ZZ}^{-1} E_K'$.

2. *If* $\rho = 1$,
$$\sqrt{nT}\left(\hat{\beta}_{BC} - \beta\right) \xrightarrow{d} N\left(0, \Omega_{\beta}^{n}\right).$$
where $\Omega_{\beta}^{s} = E_K(\sigma_{\varepsilon}^{2} - \frac{\sigma_{\varepsilon}^{4}}{12} q + \frac{\sigma_{\varepsilon}^{8}}{180} q^3) Q_{ZZ}^{-1} E_K'$ *with* $q = Q_{ZZ}^{-1} e_{K+1} e_{K+1}'$.

Unlike the bias-corrected estimator for ρ, $\hat{\rho}_{BC}^{k}$, the limiting distribution of $\hat{\beta}_{BC}$ is asymptotically normal at the same rate. This enables us to construct a robust $t-$statistic for bias-corrected estimator $\hat{\beta_{BC}}$ which converges in distribution a standard normal random variable for all $\rho \in (-1, 1]$.

4 The bias-corrected estimator with heteroskedasticity

For the heterscedastic errors, it has been shown in Phillips and Sul (2007) that the particular form of the endogeneity bias of the FE estimator does not depend on the cross-sectional heteroskedasticity under fixed T. We show that this result holds valid if we allow T goes to infinity. However, Bun and Carree (2006) shows that if one allows for time-series heteroskedasticity, the form of the inconsistency does change. Following Bun and Carree (2006), we allow for both time-series and cross-sectional heteroskedasticity. In particular, we make the following assumption for the errors,

Assumption 4 *For* $i = 1, \cdots, n, t = 1, \cdots, T$, *we assume that* $\varepsilon_{it} \sim N(0, \sigma_{it}^2)$ *with* $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Sigma_i = \Sigma_T = diag(\sigma_t^2)$, *where* $\Sigma_i = diag(\sigma_{it}^2)$, *and* $\lim_{T\to\infty} \frac{1}{T} tr(\Sigma_T) = \sigma_{\infty}^2 < \infty$.

In the case with cross-sectional heteroskedasticity only, i.e., $\Sigma_T = \sigma_{\varepsilon}^2 I_T$, to assess whether or not the asymptotic bias term under large T changes from the case with homoskedasticity, we need to consider the particular form of $E[y_{-1}' A\varepsilon]$, i.e., the source of asymptotic bias, which is
$$E[y_{-1}' A\varepsilon] = \sum_{i=1}^{n} E[y_{i,-1}' A_T \varepsilon_i].$$

When $|\rho| < 1$,
$$\sum_{i=1}^{n} E[y_{i,-1}' A_T \varepsilon_i] = \sum_{i=1}^{n} tr(\Pi_T \Sigma_i). \tag{29}$$

where $\Pi_T = A_T L_T \Gamma_T$, with $A_T = I_T - \frac{1}{T} 1_T 1_T'$ and $L_T\Gamma_T$ defined in (15). Therefore, as $(n,T) \to \infty$ with $\frac{n}{T} \to \kappa$, we have

$$\lim_{(n,T)\to\infty} \frac{1}{\sqrt{nT}} E[y_{-1}' A\varepsilon] = \sqrt{\kappa} \lim_{T\to\infty} tr\left(\Pi_T \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Sigma_i\right) = \sqrt{\kappa} \lim_{T\to\infty} tr\left(\Pi_T \Sigma_T\right),$$

Note that in this case, if $\Sigma_T = \sigma_\varepsilon^2 I_T$, i.e., there's only cross-sectional heteroskedasticity, it is easily seen that $tr\left(\Pi_T\Sigma_T\right) = \sigma_\varepsilon^2 tr\left(\Pi_T\right)$, which means the particular form of the asymptotic bias does not depend on the cross-sectional heteroskedasticity. However, once we allow the time-series heteroskedasticity, we have

$$tr\left(\Pi_T\Sigma_T\right) = -\frac{1}{T}\sum_{t=0}^{T-1} \sigma_{T-1-t}^2 \sum_{j=0}^{t} \rho^j = -\frac{1}{1-\rho}\left(\frac{1}{T}\sum_{t=0}^{T-1}\sigma_t^2 - \frac{1}{T}\sum_{t=0}^{T-1}\sigma_t^2\rho^{T-t}\right)$$

Having the assumption that $\lim_{T\to\infty} \frac{1}{T} tr(\Sigma_T) = \sigma_\infty^2 < \infty$, we have the following limiting value of $tr\left(\Pi_T\Sigma_T\right)$

$$\lim_{T\to\infty} tr\left(\Pi_T\Sigma_T\right) = \frac{a_\infty - \sigma_\infty^2}{1-\rho},$$

where $a_\infty \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \sigma_t^2 \rho^{T-t}$. As we can see that the specific form of the asymptotic bias term does change if you allow for time-series heteroskedasticity, and it depends on the value of a_T. If $a_\infty = 0$, the asymptotic bias for $|\rho| < 1$ case is the same as homoscedasticity and cross-sectional heterscedasticity case. As a result, when $|\rho| < 1$,

$$\lim_{(n,T)\to\infty} \frac{1}{\sqrt{nT}} E[Z'A\varepsilon] = \sqrt{\kappa}\frac{a_\infty - \sigma_\infty^2}{1-\rho} e_{K+1},$$

and the asymptotic bias term of the FE estimator can be defined accordingly as

$$\delta_h^s = \frac{a_\infty - \sigma_\infty^2}{T(1-\rho)} Q_{ZZ}^{-1} e_{K+1}$$

where the subscript h of δ_h^s denotes for the heteroskedasticity and Q_{ZZ} is defined in (19). For the asymptotic variance of $Z'A\varepsilon$ under heteroskedasticity[6], we have

$$\begin{aligned}\lim_{(n,T)\to\infty} \frac{1}{nT} var(Z'A\varepsilon) &= \lim_{(n,T)\to\infty} \frac{1}{nT} E[Z'A\varepsilon\varepsilon'AZ] + \lim_{(n,T)\to\infty} \frac{1}{nT}\sum_{i=1}^{n} tr[(\Pi_T'\Sigma_i)^2] e_{K+1}e_{K+1}' \\ &= \Sigma_{ZZ} + \tilde{a}_\infty e_{K+1}e_{K+1}', \qquad (30)\end{aligned}$$

where $\Sigma_{ZZ} \equiv \text{plim}_{(n,T)\to\infty} \frac{1}{nT} Z'A\varepsilon\varepsilon'AZ$ and $\tilde{a}_\infty \equiv \lim_{(n,T)\to\infty} \frac{1}{nT}\sum_{i=1}^{n} tr\left[(\Pi_T'\Sigma_i)^2\right]$. It is obvious to see that under Assumption 1 (ii) - (iv), 2, and 4, as $(n,T)\to\infty$, we have

$$\sqrt{nT}(\hat{\delta}_{FE} - \delta_h^s - \delta) \xrightarrow{d} N\left(0, \Omega_{\delta,h}^s\right)$$

[6]See Appendix E for derivation.

where $\Omega^s_{\delta,h} = Q^{-1}_{ZZ}\Sigma_{ZZ}Q^{-1}_{ZZ} + \tilde{a}_\infty Q^{-1}_{ZZ}e_{K+1}e'_{K+1}Q^{-1}_{ZZ}$.

When $\rho = 1$, the corresponding term of (29) has the following expression:

$$E[y'_{-1}A\varepsilon] = \sum_{i=1}^{n} tr(\Phi_T\Sigma_i),$$

where $\Phi_T = A_T C_T$ with C_T defined in (22) and

$$tr\left(\Phi_T\Sigma_T\right) = \frac{1}{T}\sum_{t=1}^{T-1} t\sigma_t^2.$$

To define the asymptotic bias term under non-stationary case with heteroskedasticity, we examine the following expression as $n \to \infty$:

$$\lim_{n\to\infty}\frac{1}{nT}E[y'_{-1}A\varepsilon] = \frac{1}{T}tr\left(\Pi_T \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Sigma_i\right) = \frac{1}{T}tr\left(\Phi_T\Sigma_T\right).$$

Based on this result, we define that

$$b_\infty \equiv \lim_{T\to\infty}\frac{1}{T}tr\left(\Phi_T\Sigma_T\right).$$

The asymptotic bias term of the FE estimator when $\rho = 1$ can be defined as

$$\delta^n_h = \frac{b_\infty}{T}Q^{-1}_{ZZ}e_{K+1}$$

Notice that, just as in the homoscedasticity case, as $(n, T) \to \infty$, $\Lambda(\hat{\delta}_{FE} - \delta^n_H - \delta)$ converges to a normal random variable the centered at zero. Therefore, based on the results we have so far, we are able to construct the bias-corrected estimator for ρ under heteroskedastic error assumption,

$$\hat{\rho}^h_{BC} = \begin{cases} \hat{\rho}^k_{FE} - \dfrac{a_\infty - \sigma^2_\varepsilon}{T(1-\hat{\rho}^k_{FE})}e'_{K+1}Q^{-1}_{ZZ}e_{K+1} & \text{if } \hat{\rho}^k_{FE} < 1 + \dfrac{b_\infty}{T}e'_{K+1}Q^{-1}_{ZZ}e_{K+1} \\ 1 & \text{if } \hat{\rho}^k_{FE} \geq 1 + \dfrac{b_\infty}{T}e'_{K+1}Q^{-1}_{ZZ}e_{K+1} \end{cases}$$

Again, all the terms in this bias-corrected estimator including $\hat{\rho}^k_{FE}$, a_∞, b_∞, σ^2_ε, and Q_{ZZ} can be consistently estimated. We have the following result for the bias-corrected estimator of ρ under heteroskedasticity:

Theorem 6 *Under Assumption 1 (ii) - (iv), 2, and 4, as $(n, T) \to \infty$, we have*

1. *If $|\rho| < 1$,*

$$\sqrt{nT}\left(\hat{\rho}^h_{BC} - \rho\right) \xrightarrow{d} N\left(0, \Omega^s_{\rho,h}\right)$$

where $\Omega^s_{\rho,h} = e'_{K+1}\Omega^s_{\delta,h}e_{K+1}$ with $\Omega^s_{\delta,h} = Q^{-1}_{ZZ}\Sigma_{ZZ}Q^{-1}_{ZZ} + \tilde{a}_\infty Q^{-1}_{ZZ}e_{K+1}e'_{K+1}Q^{-1}_{ZZ}$;

2. If $\rho = 1$,

$$\sqrt{n}T\left(\hat{\rho}^h_{BC} - 1\right) \xrightarrow{p} 0.$$

The corresponding bias-corrected estimator for $\hat{\beta}_{FE}$ under heteroskedasticity can be suggested as follows

$$\hat{\beta}^h_{BC} = \begin{cases} \hat{\beta}_{FE} + \dfrac{a_\infty - \sigma^2_\infty}{T(1-\rho)} E_K Q^{-1}_{ZZ} e_{K+1} & \text{if } \hat{\rho}^k_{FE} < 1 + \dfrac{b_\infty}{T} e'_{K+1} Q^{-1}_{ZZ} e_{K+1} \\ \\ \hat{\beta}_{FE} + \dfrac{b_\infty}{T} E_K Q^{-1}_{ZZ} e_{K+1} & \text{if } \hat{\rho}^k_{FE} \geq 1 + \dfrac{b_\infty}{T} e'_{K+1} Q^{-1}_{ZZ} e_{K+1} \end{cases}$$

5 Monte Carlo simulation

In this section, we report the finite sample performance of different dynamic panel data estimators based on the Monte Carlo simulation. We consider the three data generating processes. In the first design, we consider homoskedastic errors with no exogenous variables. In the second, we consider the model with strictly exogenous variables and homoskedastic errors. At last, we incorporate the heteroskedasticity in the model with exogenous variables. **DGP1:** simple dynamic panel data model,

$$y_{it} = \alpha_i + u_{it},$$
$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it},$$

where $\alpha_i \overset{iid}{\sim} N(2,1)$ and $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$. For DGP1, we compare our bias-corrected estimator as defined in (8) with several existing dynamic panel estimators. In particular, we compare the biases and RMSEs among the following estimators: The true ρ we consider are $0, 0.3, 0.6, 0.9,$

$\hat{\rho}_{FE}$	FE estimator
$\hat{\rho}_{FD}$	FD estimator
$\hat{\rho}_{DIF}$	Difference GMM estimator as in Arellano and Bond (1991)
$\hat{\rho}_{SYS}$	System GMM estimator as in Blundell and Bond (1998)
$\hat{\rho}_{FDLS}$	Bias-corrected FD estimator as in Han and Phillips (2010)
$\hat{\rho}_{PFAE}$	X-differencing estimator as in Han, Phillips and Sul (2014)
$\hat{\rho}_{BC}$	Bias-corrected FE estimator

and 1. Although the time dimension of the panel data is getting larger and larger now in empirical work, but in most of the cases they are still smaller than the cross-sectional dimension. Therefore, for all the designs, we consider $n \in \{100, 200\}$ and $T \in \{5, 10, 20, 50\}$. For a better evaluation of the performance of each estimator, we consider two cases of the initial conditions for DGP1, i.e., stationary and non-stationary, respectively. For the stationary initial condition, following Han, Phillips and Sul (2014), the data generating processes are initialized at $t = -100$ such that $u_{i,-100} = 0$ for all i and then drop the observations for $t < 0$. For the non-stationary initial condition, following Chao, Kim and

Sul (2014), the initial values are generated as $u_{i0} \overset{iid}{\sim} N(5,1)$ for all i. The X-differencing estimator is obtained by direct calculation based

DGP2: dynamic panel model with one exogenous variable,

$$y_{it} = \alpha_i + u_{it}$$
$$u_{it} = \rho u_{i,t-1} + \beta x_{it} + \varepsilon_{it}$$

where $\alpha_i \overset{iid}{\sim} N(2,1)$ and $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$ with the exogenous variable x_{it} being generated as

$$x_{it} = \gamma x_{i,t-1} + \xi_{it}$$

where $\xi_{it} \overset{iid}{\sim} N(0,1)$.

DGP3: heteroskedastic errors with the same data generating process of DGP2, however, we assume ε_{it} is independently distributed as $N(0, \sigma_{it}^2)$ where we consider three designs regarding σ_{it}^2:

i. cross-sectional heteroskedasticity: $\sigma_{it}^2 = \sigma_i^2 \sim \chi^2(1)$, where $\frac{1}{n}\sum_{i=1}^n \sigma_i^2 \approx 1$ is ensured.

ii. time-series heteroskedasticity: $\sigma_{it}^2 = \sigma_t^2 = 0.9995 - 0.5T + 0.001t$, where $\frac{1}{T}\sum_{t=1}^T \sigma_t^2 = 1$ is ensured.

iii. general heteroskedasticity where the combination of the first two has been used.

Since both $\hat{\rho}_{FDLS}$ and $\hat{\rho}_{PFAE}$ do not work for the model with exogenous variables, for DGP2 and DGP3, we compare our bias-corrected estimator with $\hat{\rho}_{FE}$ and $\hat{\rho}_{SYS}$. In addition to the values of ρ we consider for the DGP1, we choose $\gamma = 0.8$ and $\beta = 1$. With only one additional regressor, the FE estimator of ρ has the following expression,

$$\hat{\rho}_{FE}^k = \frac{\sum_{i=1}^n x_i' A_T x_i \sum_{i=1}^n y_{i,-1}' A_T y_i - \sum_{i=1}^n y_{i,-1}' A_T x_i \sum_{i=1}^n x_i' A_T y_i}{\sum_{i=1}^n y_{i,-1}' A_T y_{i,-1} \sum_{i=1}^n x_i' A_T x_i - \left(\sum_{i=1}^n y_{i,-1}' A_T x_i\right)^2} \tag{31}$$

For the bias-corrected estimator with only one exogenous variable under homoscedasticity, the estimates can be calculated based on

$$\hat{\rho}_{BC}^k = \begin{cases} \hat{\rho}_{FE}^k + \dfrac{\hat{\sigma}_\varepsilon^2}{T(1-\hat{\rho}_{FE}^k)\hat{\sigma}_{y_{-1}}^2} & \text{if } \hat{\rho}_{FE}^k < 1 - \dfrac{\hat{\sigma}_\varepsilon^2}{2T\hat{\sigma}_{y_{-1}}^2} \\ 1 & \text{if } \hat{\rho}_{FE}^k \geq 1 - \dfrac{\hat{\sigma}_\varepsilon^2}{2T\hat{\sigma}_{y_{-1}}^2} \end{cases} \tag{32}$$

where $\hat{\sigma}_{y_{-1}}^2$ is the sample counterpart of $\sigma_{y_{-1}}^2$, which can be consistently estimated by $\frac{1}{n(T-1)}\sum_{i=1}^n \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{i,-1})^2$. For the DGP3 with heteroskedasticity, the bias-corrected estimator can be expressed as

$$\hat{\rho}_{BC}^h = \begin{cases} \hat{\rho}_{FE}^k - \dfrac{\hat{a}_\infty - \hat{\sigma}_\infty^2}{T(1-\hat{\rho}_{FE}^k)\hat{\sigma}_{y_{-1}}^2} & \text{if } \hat{\rho}_{FE}^k < 1 + \dfrac{\hat{b}_\infty}{T\hat{\sigma}_{y_{-1}}^2} \\ 1 & \text{if } \hat{\rho}_{FE}^k \geq 1 + \dfrac{\hat{b}_\infty}{T\hat{\sigma}_{y_{-1}}^2} \end{cases} \tag{33}$$

where $\hat{a}_\infty = \frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\sigma}_\infty^2\left(\hat{\rho}_{FE}^k\right)^{T-t}$, $\hat{b}_\infty = \frac{1}{T(T-1)}\sum_{t=1}^{T-1}t\hat{\sigma}_\infty^2$, and $\hat{\sigma}_\infty^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_t^2$ with $\hat{\sigma}_t^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_{it}^2$. All the results are based on $10,000$ replications.

Under the first design of the data generating process, when the initial condition is stationary, as in Table 1, we can see that FE and FD are largely biased. Both DIF and SYS perform well when ρ is small. However, when ρ is close to 1, their biases increase sharply due to the weak instrument problem. Both FDLS and PFAE perform well in all situations. When $T = 5$, their performances are similar. When T increases to 10, 20 and 50, PFAE outperforms FDLS. For our proposed BC estimator, its bias is larger than FDLS an PFAE when $T = 5$. When T increases to 10, 20 and 50, its bias reduces very quickly. In Table 3, when n and T are large, we can see that PFAE and BC have the smallest RMSE for cases of $\rho = 0$, 0.3 or 0.6. PFAE outperforms best when $\rho = 0.9$ or 1. BC is a close second best. When the initial condition is non-stationary, the same as reported previously for the case of stationary initial condition, largely bias of FE and FD are found in Table 2. DIF and SYS still have large bias when ρ is close to 1. However, in the case of non-stationary initial condition, the overall performances of FDLS and PFAE in terms of bias drop substantially, which are much larger than BC in magnitude. In Table 4, generally speaking, FDLS and PFAE have much larger RMSE than BC unless $\rho = 1$. Overall, our suggested BC estimator performs relatively stable as oppose to other dynamic panel estimators. Especially, when the initial condition is non-stationary and T is large, our estimator outperforms most of the estimators in terms of bias and RMSE. When the initial condition is stationary, BC is only slightly worse than PFAE. When the initial condition is non-stationary, BC dominates other estimators. DIF and SYS have large bias when ρ is close to 1. FDLS and PFAE have large bias when the initial condition is non-stationary. When additional exogenous variable is added, as in Table 5 and Table 6, similar observations could be drawn: FE estimator is largely biased but the degree of bias decreases as the sample size T increases. The SYS estimator works worse in large T cases than in moderate time dimension. The proposed bias-corrected estimator works well in all cases compared the other two estimators in terms of bias and RMSE, especially when the time dimension is relatively large.

6 Conclusion

In this paper, we propose a bias-corrected FE estimator for dynamic panel model that accommodates the cases when $|\rho| < 1$ and $\rho = 1$. We derive its asymptotic distribution and show that it is more efficient than the bias-corrected FD estimator. More importantly, a nice feature of this bias-correct FE estimator is that its consistency does not depend on the stationarity of the initial condition. As we can see from the Monte Carlo simulations, the performance of bias-corrected FE estimator dominates the performances of other existing dynamic panel estimators, especially when the initial condition is non-stationary. We further apply the idea of this bias-corrected estimator to a more general dynamic panel model where exogenous variables are included and heteroskedasticity is allowed. We derive the asymptotic distributions for the FE estimator under large T for the cases when $|\rho| < 1$ and $\rho = 1$, respectively. Based on these two asymptotic results, we construct the bias-corrected estimator for the model with exogenous variables.

Table 1: Bias under DGP1 with stationary initial condition

T	ρ	FE	FD	DIF	SYS	FDLS	PFAE	BC
					$n = 100$			
5	0	-0.2003	-0.4993	-0.0179	0.0034	0.0013	-0.0014	-0.0404
	0.3	-0.2757	-0.6497	-0.0345	0.0021	0.0006	-0.0021	-0.0709
	0.6	-0.3638	-0.8002	-0.0708	-0.0042	-0.0003	-0.0024	-0.1166
	0.9	-0.4670	-0.9515	-0.1794	-0.0188	-0.0030	-0.0042	0.0063
	1	-0.5040	-1.0020	-0.7998	-0.0098	-0.0040	-0.0047	-0.0097
10	0	-0.1001	-0.4986	-0.0155	0.0019	0.0029	-0.0002	-0.0102
	0.3	-0.1359	-0.6488	-0.0276	-0.0029	0.0024	-0.0012	-0.0195
	0.6	-0.1801	-0.7992	-0.0515	-0.0112	0.0016	-0.0018	-0.0381
	0.9	-0.2453	-0.9494	-0.1129	-0.0295	0.0011	-0.0015	-0.0731
	1	-0.2746	-0.9995	-0.4679	-0.0135	0.0009	-0.0014	-0.0273
20	0	-0.0504	-0.4995	-0.0124	0.0019	0.0010	-0.0004	-0.0029
	0.3	-0.0671	-0.6498	-0.0194	-0.0059	0.0005	-0.0008	-0.0055
	0.6	-0.0865	-0.8000	-0.0313	-0.0174	0.0001	-0.0009	-0.0108
	0.9	-0.1204	-0.9500	-0.0609	-0.0395	0.0000	-0.0002	-0.0315
	1	-0.1433	-0.9999	-0.2265	-0.0382	0.0002	-0.0002	-0.0206
50	0	-0.0194	-0.4994	-0.0245	-0.0050	0.0013	0.0006	0.0002
	0.3	-0.0258	-0.6494	-0.0379	-0.0282	0.0013	0.0004	-0.0004
	0.6	-0.0328	-0.7994	-0.0491	-0.0439	0.0012	0.0002	-0.0014
	0.9	-0.0441	-0.9495	-0.0562	-0.0523	0.0011	-0.0002	-0.0070
	1	-0.0594	-0.9996	-0.1051	-0.0568	0.0009	-0.0007	-0.0119
					$n = 200$			
5	0	-0.2018	-0.5007	-0.0079	0.0001	-0.0015	-0.0021	-0.0421
	0.3	-0.2765	-0.6509	-0.0156	-0.0009	-0.0018	-0.0024	-0.0718
	0.6	-0.3634	-0.8009	-0.0339	-0.0033	-0.0018	-0.0017	-0.1161
	0.9	-0.4636	-0.9504	-0.0889	-0.0070	-0.0007	0.0000	0.0475
	1	-0.4998	-0.9999	-0.8202	-0.0050	0.0001	0.0006	-0.0010
10	0	-0.0997	-0.4994	-0.0073	0.0013	0.0011	0.0004	-0.0096
	0.3	-0.1349	-0.6493	-0.0123	0.0010	0.0015	-0.0002	-0.0184
	0.6	-0.1789	-0.7991	-0.0239	-0.0015	0.0017	-0.0007	-0.0368
	0.9	-0.2430	-0.9488	-0.0592	-0.0097	0.0024	0.0006	-0.0761
	1	-0.2731	-0.9992	-0.5023	-0.0068	0.0016	-0.0004	-0.0098
20	0	-0.0494	-0.4994	-0.0054	0.0013	0.0012	0.0006	-0.0019
	0.3	-0.0659	-0.6495	-0.0087	-0.0017	0.0011	0.0004	-0.0042
	0.6	-0.0853	-0.7994	-0.0150	-0.0072	0.0012	0.0003	-0.0096
	0.9	-0.1195	-0.9493	-0.0337	-0.0204	0.0015	0.0004	-0.0304
	1	-0.1429	-0.9992	-0.2168	-0.0156	0.0017	0.0001	-0.0149
50	0	-0.0197	-0.4998	-0.0093	0.0001	0.0003	0.0003	-0.0001
	0.3	-0.0260	-0.6498	-0.0136	-0.0081	0.0005	0.0002	-0.0005
	0.6	-0.0329	-0.7997	-0.0181	-0.0142	0.0007	0.0001	-0.0016
	0.9	-0.0444	-0.9496	-0.0230	-0.0200	0.0008	-0.0002	-0.0072
	1	-0.0593	-0.9998	-0.1013	-0.0419	0.0004	-0.0004	-0.0104

Note: 1. DGP: $y_{it} = \alpha_i + u_{it}$ and $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$ where $\alpha_i \overset{iid}{\sim} N(2,1)$ and $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$.
2. "FE" refers to FE estimator defined in Equation (4); "FD" refers to first-difference estimator; "DIF" refers to Arrellano-Bond GMM estimator; "SYS" refers to Blundell-Bond GMM estimator; "FDLS" refers to the bias-corrected first-difference estimator; "PFAE" refers to the X-Differencing estimator; "BC" refers to the bias-corrected FE estimator.
3. For stationary initial condition, $u_{i,-100}$ is initialized at 0 for all i and all u_{it} with $t < 0$ are dropped.

Table 2: Bias under DGP1 with non-stationary initial condition

T	ρ	FE	FD	DIF	SYS	FDLS	PFAE	BC
					$n = 100$			
	0	-0.0334	-0.1208	-0.0015	0.0002	0.7584	0.0010	0.1600
	0.3	-0.0488	-0.2038	-0.0019	0.0003	0.8924	1.1584	0.2014
5	0.6	-0.0860	-0.3622	-0.0029	-0.0006	0.8755	0.9371	0.4000
	0.9	-0.3335	-0.8263	-0.0209	-0.0031	0.2475	0.2403	0.0997
	1	-0.5034	-1.0001	-0.8062	-0.0026	-0.0003	-0.0028	-0.0086
	0	-0.0289	-0.2088	-0.0023	0.0002	0.5824	-0.0012	0.0682
	0.3	-0.0396	-0.3295	-0.0032	-0.0007	0.6410	0.6240	0.0865
10	0.6	-0.0568	-0.5175	-0.0049	-0.0022	0.5650	0.5781	0.0975
	0.9	-0.1585	-0.8649	-0.0183	-0.0079	0.1703	0.1687	0.0947
	1	-0.2732	-0.9994	-0.4706	-0.0149	0.0012	-0.0002	-0.0240
	0	-0.0230	-0.3016	-0.0049	0.0002	0.3968	-0.0017	0.0259
	0.3	-0.0305	-0.4455	-0.0066	-0.0029	0.4090	0.3239	0.0330
20	0.6	-0.0400	-0.6361	-0.0085	-0.0060	0.3278	0.3517	0.0380
	0.9	-0.0764	-0.9017	-0.0161	-0.0135	0.0966	0.1107	0.0157
	1	-0.1429	-0.9996	-0.2303	-0.0522	0.0009	0.0003	-0.0207
	0	-0.0138	-0.3984	-0.0073	-0.0010	0.2031	-0.0007	0.0060
	0.3	-0.0180	-0.5519	-0.0096	-0.0045	0.1962	0.1330	0.0077
50	0.6	-0.0227	-0.7275	-0.0118	-0.0085	0.1449	0.1649	0.0089
	0.9	-0.0324	-0.9306	-0.0176	-0.0162	0.0388	0.0571	0.0049
	1	-0.0591	-1.0002	-0.1172	-0.0791	-0.0004	-0.0001	-0.0112
					$n = 200$			
	0	-0.0334	-0.1218	-0.0010	0.0001	0.7564	-0.0007	0.1599
	0.3	-0.0486	-0.2051	-0.0010	0.0005	0.8897	1.1543	0.2017
5	0.6	-0.0854	-0.3638	-0.0011	0.0005	0.8724	0.9353	0.4000
	0.9	-0.3324	-0.8278	-0.0088	-0.0005	0.2443	0.2398	0.1000
	1	-0.5017	-1.0007	-0.8238	-0.0002	-0.0015	-0.0019	-0.0007
	0	-0.0293	-0.2098	-0.0018	-0.0006	0.5805	-0.0026	0.0678
	0.3	-0.0398	-0.3306	-0.0021	-0.0007	0.6388	0.6221	0.0863
10	0.6	-0.0567	-0.5187	-0.0027	-0.0010	0.5626	0.5774	0.0976
	0.9	-0.1580	-0.8655	-0.0093	-0.0022	0.1689	0.1689	0.0988
	1	-0.2730	-0.9999	-0.5097	-0.0039	0.0001	0.0000	-0.0118
	0	-0.0229	-0.3015	-0.0030	-0.0004	0.3970	-0.0016	0.0260
	0.3	-0.0303	-0.4454	-0.0038	-0.0018	0.4092	0.3243	0.0331
20	0.6	-0.0397	-0.6361	-0.0047	-0.0032	0.3279	0.3522	0.0383
	0.9	-0.0760	-0.9016	-0.0086	-0.0073	0.0968	0.1110	0.0153
	1	-0.1427	-0.9998	-0.2132	-0.0214	0.0004	0.0000	-0.0148
	0	-0.0135	-0.3983	-0.0041	-0.0003	0.2035	-0.0003	0.0062
	0.3	-0.0178	-0.5518	-0.0053	-0.0026	0.1965	0.1333	0.0079
50	0.6	-0.0224	-0.7274	-0.0065	-0.0047	0.1451	0.1652	0.0091
	0.9	-0.0321	-0.9305	-0.0088	-0.0081	0.0390	0.0574	0.0052
	1	-0.0591	-1.0003	-0.1047	-0.0542	-0.0005	-0.0003	-0.0100

Note: 1. For non-stationary initial condition, $u_{i0} \overset{iid}{\sim} N(5, 1)$ for all i.
2. See note 1 and 2 of Table 1.

Table 3: RMSE under DGP1 with stationary initial condition

T	ρ	FE	FD	DIF	SYS	FDLS	PFAE	BC
					$n = 100$			
	0	0.2051	0.5007	0.0899	0.0758	0.0744	0.0577	0.0666
	0.3	0.2797	0.6509	0.1177	0.0825	0.0814	0.0634	0.0904
5	0.6	0.3670	0.8014	0.1680	0.0834	0.0886	0.0674	0.1302
	0.9	0.4697	0.9528	0.3070	0.0836	0.0984	0.0725	0.1575
	1	0.5065	1.0033	0.8973	0.0524	0.1018	0.0742	0.0580
	0	0.1049	0.4992	0.0536	0.0484	0.0486	0.0351	0.0357
	0.3	0.1395	0.6494	0.0648	0.0512	0.0546	0.0360	0.0398
10	0.6	0.1827	0.7998	0.0848	0.0518	0.0602	0.0357	0.0509
	0.9	0.2470	0.9500	0.1441	0.0549	0.0655	0.0339	0.0892
	1	0.2761	1.0001	0.5087	0.0299	0.0671	0.0325	0.0634
	0	0.0550	0.4997	0.0310	0.0286	0.0329	0.0232	0.0232
	0.3	0.0705	0.6500	0.0356	0.0306	0.0369	0.0231	0.0235
20	0.6	0.0888	0.8002	0.0441	0.0352	0.0403	0.0212	0.0236
	0.9	0.1216	0.9502	0.0699	0.0491	0.0434	0.0170	0.0360
	1	0.1439	1.0002	0.2360	0.0478	0.0439	0.0151	0.0376
	0	0.0242	0.4995	0.0366	0.0236	0.0211	0.0148	0.0148
	0.3	0.0293	0.6495	0.0494	0.0411	0.0241	0.0141	0.0141
50	0.6	0.0349	0.7995	0.0589	0.0553	0.0266	0.0122	0.0123
	0.9	0.0449	0.9496	0.0630	0.0595	0.0284	0.0084	0.0111
	1	0.0598	0.9997	0.1089	0.0636	0.0293	0.0062	0.0179
					$n = 200$			
	0	0.2043	0.5015	0.0657	0.0534	0.0538	0.0420	0.0571
	0.3	0.2786	0.6516	0.0861	0.0579	0.0594	0.0467	0.0827
5	0.6	0.3651	0.8016	0.1171	0.0568	0.0647	0.0494	0.1235
	0.9	0.4649	0.9510	0.1838	0.0533	0.0696	0.0505	0.1327
	1	0.5010	1.0006	0.9179	0.0361	0.0713	0.0513	0.0180
	0	0.1020	0.4997	0.0358	0.0326	0.0351	0.0246	0.0259
	0.3	0.1367	0.6496	0.0419	0.0342	0.0387	0.0250	0.0305
10	0.6	0.1801	0.7994	0.0530	0.0336	0.0418	0.0242	0.0435
	0.9	0.2438	0.9491	0.0876	0.0319	0.0457	0.0227	0.0809
	1	0.2737	0.9995	0.5455	0.0200	0.0475	0.0228	0.0369
	0	0.0520	0.4995	0.0204	0.0199	0.0231	0.0171	0.0171
	0.3	0.0679	0.6496	0.0230	0.0215	0.0264	0.0169	0.0173
20	0.6	0.0865	0.7996	0.0268	0.0231	0.0295	0.0151	0.0178
	0.9	0.1200	0.9494	0.0416	0.0291	0.0323	0.0120	0.0327
	1	0.1433	0.9993	0.2257	0.0210	0.0332	0.0111	0.0312
	0	0.0222	0.4999	0.0167	0.0130	0.0145	0.0103	0.0103
	0.3	0.0277	0.6498	0.0197	0.0168	0.0167	0.0098	0.0098
50	0.6	0.0340	0.7997	0.0227	0.0203	0.0185	0.0084	0.0086
	0.9	0.0447	0.9496	0.0255	0.0230	0.0198	0.0058	0.0094
	1	0.0595	0.9998	0.1039	0.0451	0.0202	0.0043	0.0161

Note: See note 1, 2, and 3 of Table 1.

Table 4: RMSE under DGP1 with non-stationary initial condition

T	ρ	FE	FD	DIF	SYS	FDLS	PFAE	BC
					$n = 100$			
	0	0.0389	0.1230	0.0232	0.0229	0.7599	0.1570	0.1618
	0.3	0.0535	0.2056	0.0252	0.0250	0.8941	1.1619	0.2031
5	0.6	0.0897	0.3638	0.0303	0.0293	0.8781	0.9385	0.4000
	0.9	0.3364	0.8277	0.0806	0.0406	0.2658	0.2486	0.1002
	1	0.5059	1.0014	0.9139	0.0370	0.1008	0.0721	0.0546
	0	0.0336	0.2097	0.0217	0.0207	0.5837	0.0670	0.0708
	0.3	0.0434	0.3302	0.0225	0.0212	0.6425	0.6260	0.0886
10	0.6	0.0596	0.5181	0.0224	0.0211	0.5673	0.5787	0.0994
	0.9	0.1602	0.8654	0.0392	0.0256	0.1810	0.1706	0.0984
	1	0.2746	0.9999	0.5087	0.0297	0.0631	0.0313	0.0593
	0	0.0273	0.3020	0.0174	0.0169	0.3980	0.0353	0.0302
	0.3	0.0338	0.4459	0.0180	0.0173	0.4104	0.3253	0.0365
20	0.6	0.0422	0.6364	0.0178	0.0171	0.3300	0.3522	0.0405
	0.9	0.0774	0.9019	0.0227	0.0204	0.1056	0.1113	0.0228
	1	0.1436	0.9998	0.2404	0.0605	0.0454	0.0156	0.0376
	0	0.0179	0.3986	0.0157	0.0141	0.2043	0.0177	0.0132
	0.3	0.0211	0.5520	0.0165	0.0145	0.1976	0.1340	0.0136
50	0.6	0.0246	0.7276	0.0169	0.0147	0.1471	0.1653	0.0132
	0.9	0.0332	0.9307	0.0208	0.0196	0.0471	0.0574	0.0087
	1	0.0594	1.0003	0.1223	0.0855	0.0276	0.0061	0.0173
					$n = 200$			
	0	0.0365	0.1229	0.0167	0.0163	0.7572	0.1116	0.1609
	0.3	0.0511	0.2060	0.0182	0.0176	0.8906	1.1560	0.2026
5	0.6	0.0874	0.3646	0.0217	0.0204	0.8737	0.9360	0.4000
	0.9	0.3339	0.8286	0.0555	0.0263	0.2541	0.2441	0.1000
	1	0.5028	1.0013	0.9310	0.0238	0.0691	0.0481	0.0146
	0	0.0319	0.2102	0.0153	0.0147	0.5811	0.0497	0.0692
	0.3	0.0419	0.3310	0.0158	0.0150	0.6396	0.6232	0.0875
10	0.6	0.0582	0.5190	0.0158	0.0148	0.5638	0.5777	0.0987
	0.9	0.1589	0.8658	0.0264	0.0158	0.1746	0.1699	0.0996
	1	0.2737	1.0002	0.5533	0.0152	0.0468	0.0226	0.0409
	0	0.0252	0.3017	0.0119	0.0118	0.3976	0.0253	0.0283
	0.3	0.0321	0.4456	0.0123	0.0120	0.4099	0.3250	0.0349
20	0.6	0.0409	0.6362	0.0119	0.0115	0.3290	0.3524	0.0396
	0.9	0.0765	0.9017	0.0143	0.0131	0.1011	0.1114	0.0182
	1	0.1431	0.9999	0.2224	0.0261	0.0309	0.0113	0.0309
	0	0.0159	0.3983	0.0106	0.0097	0.2040	0.0128	0.0106
	0.3	0.0195	0.5518	0.0110	0.0103	0.1972	0.1338	0.0114
50	0.6	0.0234	0.7275	0.0106	0.0098	0.1463	0.1654	0.0115
	0.9	0.0325	0.9305	0.0109	0.0104	0.0436	0.0575	0.0072
	1	0.0593	1.0003	0.1074	0.0576	0.0202	0.0044	0.0159

Note: See note 1 and 2 of Table 2.

Table 5: Bias under DGP2 with one exogenous variable

		$n = 100$			$n = 200$		
T	ρ	FE	SYS	BC	FE	SYS	BC
5	0	-0.1410	0.0912	-0.0490	-0.1413	0.0453	-0.0494
	0.3	-0.1629	0.0988	-0.0499	-0.1622	0.0553	-0.0491
	0.6	-0.1607	0.0741	-0.0183	-0.1613	0.0454	-0.0208
	0.9	-0.1280	0.0319	0.0970	-0.1271	0.0188	0.0972
	1	-0.0389	0.0097	0.3931	-0.0386	0.0097	0.3852
10	0	-0.0585	0.1685	-0.0183	-0.0576	0.0936	-0.0174
	0.3	-0.0597	0.1779	-0.0164	-0.0603	0.1069	-0.0172
	0.6	-0.0537	0.1335	-0.0087	-0.0532	0.0849	-0.0082
	0.9	-0.0376	0.0504	0.0291	-0.0373	0.0356	0.0293
	1	-0.0094	0.0094	0.2373	-0.0093	0.0094	0.2240
20	0	-0.0262	0.2788	-0.0092	-0.0268	0.1665	-0.0098
	0.3	-0.0265	0.2685	-0.0099	-0.0262	0.1728	-0.0096
	0.6	-0.0206	0.1896	-0.0056	-0.0205	0.1281	-0.0056
	0.9	-0.0122	0.0635	0.0049	-0.0123	0.0487	0.0048
	1	-0.0025	0.0089	0.1237	-0.0024	0.0089	0.1327
50	0	-0.0103	0.4628	-0.0044	-0.0104	0.3063	-0.0046
	0.3	-0.0100	0.3925	-0.0046	-0.0101	0.2813	-0.0047
	0.6	-0.0074	0.2526	-0.0030	-0.0075	0.1919	-0.0031
	0.9	-0.0035	0.0755	0.0000	-0.0036	0.0631	-0.0000
	1	-0.0004	0.0077	0.0550	-0.0004	0.0077	0.0662

Note: 1. DGP: $y_{it} = \alpha_i + u_{it}$ and $u_{it} = \rho u_{i,t-1} + \beta x_{it} + \varepsilon_{it}$ with $x_{it} = \gamma x_{i,t-1} + \xi_{it}$, where $\alpha_i \overset{iid}{\sim} N(2,1)$ and $\xi_{it} \overset{iid}{\sim} N(0,1)$.
2. Stationary initial condition is assumed.
3. Homoscedastic error: $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$.
4. $\beta = 1$ and $\gamma = 0.8$.
5. "FE" refers to the FE estimator which can be calculated based on equation (31); "SYS" refers to the Blundell-Bond GMM estimator; "BC" refers to the bias-corrected estimator that can be calculated based on Equation (32).

Table 6: RMSE under DGP2 with one exogenous variable

		$n = 100$			$n = 200$		
T	ρ	FE	SYS	BC	FE	SYS	BC
5	0	0.0215	0.0208	0.0044	0.0208	0.0085	0.0035
	0.3	0.0282	0.0218	0.0047	0.0271	0.0093	0.0035
	0.6	0.0272	0.0116	0.0026	0.0268	0.0054	0.0016
	0.9	0.0174	0.0018	0.0134	0.0166	0.0008	0.0114
	1	0.0018	0.0001	0.2174	0.0016	0.0001	0.1822
10	0	0.0040	0.0358	0.0010	0.0036	0.0124	0.0006
	0.3	0.0041	0.0379	0.0008	0.0039	0.0153	0.0006
	0.6	0.0032	0.0209	0.0005	0.0030	0.0092	0.0002
	0.9	0.0015	0.0028	0.0011	0.0015	0.0015	0.0010
	1	0.0001	0.0001	0.1016	0.0001	0.0001	0.0684
20	0	0.0010	0.0837	0.0004	0.0009	0.0308	0.0002
	0.3	0.0009	0.0756	0.0003	0.0008	0.0325	0.0002
	0.6	0.0005	0.0371	0.0001	0.0005	0.0176	0.0001
	0.9	0.0002	0.0041	0.0001	0.0002	0.0025	0.0000
	1	0.0000	0.0001	0.0398	0.0000	0.0001	0.0377
50	0	0.0002	0.2170	0.0001	0.0002	0.0962	0.0001
	0.3	0.0002	0.1554	0.0001	0.0001	0.0806	0.0001
	0.6	0.0001	0.0642	0.0000	0.0001	0.0373	0.0000
	0.9	0.0000	0.0057	0.0000	0.0000	0.0040	0.0000
	1	0.0000	0.0001	0.0176	0.0000	0.0001	0.0279

Note: see notes of Table 5.

Table 7: Bias under DGP3 with heteroskedasticity

		Cross-section and time-series			Cross-section			Time-series		
T	ρ	FE	SYS	BC	FE	SYS	BC	FE	SYS	BC
5	0	-0.1424	0.0820	-0.0472	-0.1390	0.0844	-0.0452	-0.1440	0.0889	-0.0473
	0.3	-0.1691	0.0898	-0.0602	-0.1628	0.0931	-0.0548	-0.1696	0.0927	-0.0593
	0.6	-0.1648	0.0777	-0.0487	-0.1587	0.0784	-0.0451	-0.1697	0.0758	-0.0518
	0.9	-0.1293	0.0311	-0.0270	-0.1249	0.0302	-0.0253	-0.1324	0.0312	-0.0286
	1	-0.0385	0.0097	0.0218	-0.0380	0.0097	0.0321	-0.0406	0.0098	0.0033
10	0	-0.0594	0.1675	-0.0185	-0.0569	0.1693	-0.0165	-0.0588	0.1661	-0.0172
	0.3	-0.0617	0.1757	-0.0193	-0.0611	0.1740	-0.0195	-0.0627	0.1805	-0.0200
	0.6	-0.0542	0.1362	-0.0143	-0.0537	0.1335	-0.0148	-0.0564	0.1393	-0.0161
	0.9	-0.0375	0.0514	-0.0050	-0.0373	0.0503	-0.0051	-0.0381	0.0505	-0.0043
	1	-0.0097	0.0094	0.0070	-0.0093	0.0094	-0.0023	-0.0099	0.0094	-0.0031
20	0	-0.0271	0.2731	-0.0099	-0.0264	0.2753	-0.0095	-0.0268	0.2729	-0.0092
	0.3	-0.0268	0.2626	-0.0102	-0.0265	0.2610	-0.0102	-0.0273	0.2650	-0.0105
	0.6	-0.0215	0.1857	-0.0074	-0.0214	0.1838	-0.0076	-0.0214	0.1889	-0.0069
	0.9	-0.0126	0.0631	-0.0016	-0.0123	0.0626	-0.0015	-0.0128	0.0636	-0.0015
	1	-0.0025	0.0089	0.0003	-0.0024	0.0089	-0.0065	-0.0026	0.0089	0.0002
50	0	-0.0104	0.4592	-0.0044	-0.0105	0.4617	-0.0046	-0.0106	0.4560	-0.0045
	0.3	-0.0097	0.3917	-0.0042	-0.0096	0.3917	-0.0042	-0.0104	0.3890	-0.0048
	0.6	-0.0076	0.2512	-0.0032	-0.0075	0.2506	-0.0032	-0.0077	0.2516	-0.0032
	0.9	-0.0035	0.0759	-0.0006	-0.0035	0.0755	-0.0006	-0.0037	0.0755	-0.0006
	1	-0.0004	0.0077	-0.0003	-0.0004	0.0077	-0.0008	-0.0004	0.0077	-0.0000

Note: 1. $n = 100$.
2. For cross-sectional heteroskedasticity: $\sigma_{it}^2 = \sigma_i^2 \sim \chi^2(1)$; for time-series heteroskedasticity: $\sigma_{it}^2 = \sigma_\infty^2 = 0.9995 - 0.5T + 0.001t$; for general heteroskedasticity where the combination of the first two has been used.
3. "BC" refers to the bias-corrected estimator under heteroskedasticity which can be calculated baed on Equation (33).
4. See note 1 ,2 ,and 4 of Table 5.

Table 8: RMSE under DGP3 with heteroskedasticity

		Cross-section and time-series			Cross-section			Time-series		
T	ρ	FE	SYS	BC	FE	SYS	BC	FE	SYS	BC
5	0	0.0225	0.0191	0.0048	0.0228	0.0215	0.0061	0.0224	0.0202	0.0042
	0.3	0.0306	0.0197	0.0061	0.0294	0.0202	0.0064	0.0305	0.0197	0.0055
	0.6	0.0289	0.0120	0.0048	0.0277	0.0127	0.0056	0.0304	0.0119	0.0048
	0.9	0.0179	0.0017	0.0077	0.0171	0.0018	0.0167	0.0186	0.0017	0.0052
	1	0.0018	0.0001	0.1138	0.0017	0.0001	0.1708	0.0019	0.0001	0.0698
10	0	0.0043	0.0352	0.0012	0.0045	0.0367	0.0016	0.0041	0.0349	0.0009
	0.3	0.0044	0.0374	0.0010	0.0047	0.0366	0.0014	0.0044	0.0389	0.0009
	0.6	0.0033	0.0216	0.0006	0.0034	0.0210	0.0007	0.0035	0.0223	0.0006
	0.9	0.0015	0.0029	0.0003	0.0016	0.0028	0.0006	0.0016	0.0028	0.0002
	1	0.0001	0.0001	0.0856	0.0001	0.0001	0.0659	0.0001	0.0001	0.0077
20	0	0.0011	0.0802	0.0004	0.0012	0.0817	0.0006	0.0010	0.0804	0.0004
	0.3	0.0010	0.0727	0.0004	0.0011	0.0719	0.0005	0.0010	0.0738	0.0003
	0.6	0.0006	0.0359	0.0002	0.0006	0.0353	0.0002	0.0006	0.0369	0.0002
	0.9	0.0002	0.0041	0.0000	0.0002	0.0040	0.0000	0.0002	0.0041	0.0000
	1.0	0.0000	0.0001	0.0044	0.0000	0.0001	0.0214	0.0000	0.0001	0.0006
50	0	-0.0104	0.4592	-0.0044	0.0003	0.2159	0.0002	0.0002	0.2107	0.0001
	0.3	-0.0097	0.3917	-0.0042	0.0002	0.1548	0.0001	0.0002	0.1526	0.0001
	0.6	-0.0076	0.2512	-0.0032	0.0001	0.0632	0.0001	0.0001	0.0637	0.0000
	0.9	-0.0035	0.0759	-0.0006	0.0000	0.0057	0.0000	0.0000	0.0057	0.0000
	1	-0.0004	0.0077	-0.0003	0.0000	0.0001	0.0052	0.0000	0.0001	0.0000

Note: see notes of Table 7

References

Alvarez, Javier, and Manuel Arellano. 2003. "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators." *Econometrica*, 71(4): 1121–1159.

Anderson, Theodore Wilbur, and Cheng Hsiao. 1981. "Estimation of Dynamic Models with Error Components." *Journal of the American Statistical Association*, 76(375): 598–606.

Anderson, Theodore Wilbur, and Cheng Hsiao. 1982. "Formulation and estimation of dynamic models using panel data." *Journal of Econometrics*, 18(1): 47–82.

Arellano, Manuel, and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68(1): 29–51.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies*, 58(2): 277–297.

Balestra, Pietro, and Marc Nerlove. 1966. "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas." *Econometrica*, 34(3): 585–612.

Baltagi, Badi H. 2013. *Econometric analysis of panel data.* John Wiley & Sons.

Baltagi, Badi H., Chihwa Kao, and Long Liu. 2008. "Asymptotic properties of estimators for the linear panel regression model with random individual effects and serially correlated errors: the case of stationary and non-stationary regressors and residuals." *The Econometrics Journal*, 11(3): 554–572.

Baltagi, Badi H., Chihwa Kao, and Long Liu. 2012. "On the estimation and testing of fixed effects panel data models with weak instruments." In *30th Anniversary Edition.* 199–235. Emerald Group Publishing Limited.

Baltagi, Badi H., Chihwa Kao, and Long Liu. 2014. "Test of Hypotheses in a Time Trend Panel Data Model with Serially Correlated Error Component Disturbances." In *Essays in Honor of Peter C. B. Phillips.* Vol. 33 of *Advances in Econometrics*, 347–394. Emerald Publishing Ltd.

Baltagi, Badi H., Chihwa Kao, and Long Liu. 2017. "Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term." *Econometric Reviews*, 36(1-3): 85–102.

Baltagi, Badi H., Chihwa Kao, and Long Liu. 2019. "Testing for Shifts in a Time Trend Panel Data Model with Serially Correlated Error Component Disturbance." *Econometric Reviews.* forthcoming.

Bekker, Paul A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica*, 62(3): 657–681.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1): 115–143.

Bruno, Giovanni SF. 2005. "Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models." *Economics letters*, 87(3): 361–366.

Bun, Maurice JG. 2003. "Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix." *Econometric Reviews*, 22(1): 29–58.

Bun, Maurice J.G., and Jan F. Kiviet. 2001. "The Accuracy of Inference in Small Samples of Dynamic Panel Data Models." Tinbergen Institute Tinbergen Institute Discussion Papers 01-006/4.

Bun, Maurice JG, and Martin A Carree. 2005. "Bias-corrected estimation in dynamic panel data models." *Journal of Business & Economic Statistics*, 23(2): 200–210.

Bun, Maurice JG, and Martin A Carree. 2006. "Bias-corrected estimation in dynamic panel data models with heteroscedasticity." *Economics Letters*, 92(2): 220–227.

Chao, John, Myungsup Kim, and Donggyu Sul. 2014. "Mean Average Estimation of Dynamic Panel Models with Nonstationary Initial Condition." *Essays in Honor of Peter C. B. Phillips*, Chapter 8, 241–279. Emerald Group Publishing Limited.

Donohue III, John J., and Steven D. Levitt. 2001. "The Impact of Legalized Abortion on Crime." *The Quarterly Journal of Economics*, 116(2): 379–420.

Hahn, Jinyong, and Guido Kuersteiner. 2002. "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T Are Large." *Econometrica*, 70(4): 1639–1657.

Hahn, Jinyong, and Hyungsik Roger Moon. 2006. "Reducing bias of MLE in a dynamic panel model." *Econometric Theory*, 22(3): 499–512.

Han, Chirok, and Peter C. B. Phillips. 2010. "GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity." *Econometric Theory*, 26(1): 119–151.

Han, Chirok, Peter C. B. Phillips, and Donggyu Sul. 2014. "X-Differencing and Dynamic Panel Model Estimation." *Econometric Theory*, 30(1): 201–251.

Harris, Richard D.F., and Elias Tzavalis. 1999. "Inference for unit roots in dynamic panels where the time dimension is fixed." *Journal of Econometrics*, 91(2): 201 – 226.

Hsiao, Cheng, and Qiankun Zhou. 2018. "Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models." *Journal of Econometrics*, 207(1): 114–128.

Kao, Chihwa. 1999. “Spurious regression and residual-based tests for cointegration in panel data.” *Journal of Econometrics*, 90(1): 1–44.

Kiviet, Jan F. 1995. “On bias, inconsistency, and efficiency of various estimators in dynamic panel data models.” *Journal of Econometrics*, 68(1): 53 – 78.

Kiviet, Jan F. 2005. “Judging contending estimators by simulation: tournaments in dynamic panel data models.” In *The refinement of econometric estimation and test procedures.* 282–318. Cambridge: Cambridge University Press.

Moon, Hyungsik Roger, Benoit Perron, and Peter C. B. Phillips. 2015. “Incidental Parameters and Dynamic Panel Modeling.”

Neyman, J., and Elizabeth L. Scott. 1948. “Consistent Estimates Based on Partially Consistent Observations.” *Econometrica*, 16(1): 1–32.

Nickell, Stephen. 1981. “Biases in Dynamic Models with Fixed Effects.” *Econometrica*, 49(6): 1417–1426.

Perron, Pierre, and Tomoyoshi Yabu. 2009*a*. “Estimating deterministic trends with an integrated or stationary noise component.” *Journal of Econometrics*, 151(1): 56–69.

Perron, Pierre, and Tomoyoshi Yabu. 2009*b*. “Testing for shifts in trend with an integrated or stationary noise component.” *Journal of Business & Economic Statistics*, 27(3): 369–396.

Phillips, Peter C.B., and Donggyu Sul. 2007. “Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence.” *Journal of Econometrics*, 137(1): 162 – 188.

A Proof of Theorem 1

Proof. Let $S = \{T(1-\hat{\rho}_{FE}) - 3 > 0\}$ and $\bar{S} = \{T(1-\hat{\rho}_{FE}) - 3 \leq 0\}$.

When $|\rho| < 1$, it suffices to show that

$$\begin{aligned}&\sqrt{nT}\left(\hat{\rho}_{BC}-\rho\right)-\sqrt{nT}\left(\hat{\rho}_{FE}-\rho+\frac{1+\hat{\rho}_{FE}}{T}\right)\\=&\sqrt{nT}\left(\hat{\rho}_{BC}-\hat{\rho}_{FE}-\frac{1+\hat{\rho}_{FE}}{T}\right)\xrightarrow{p}0.\end{aligned}$$

We have

$$\begin{aligned}&\lim_{(n,T)\to\infty}\Pr\left(\left|\sqrt{nT}\left(\hat{\rho}_{BC}-\hat{\rho}_{FE}-\frac{1+\hat{\rho}_{FE}}{T}\right)\right|>\epsilon\right)\\&=\lim_{(n,T)\to\infty}\Pr\left(\left|\sqrt{nT}\left(\hat{\rho}_{BC}-\hat{\rho}_{FE}-\frac{1+\hat{\rho}_{FE}}{T}\right)\right|>\epsilon|S\right)\Pr\left(S\right)\\&+\lim_{(n,T)\to\infty}\Pr\left(\left|\sqrt{nT}\left(\hat{\rho}_{BC}-\hat{\rho}_{FE}-\frac{1+\hat{\rho}_{FE}}{T}\right)\right|>\epsilon|\bar{S}\right)\Pr\left(\bar{S}\right).\end{aligned}$$

The first term is zero given that, if S is true, we have $\hat{\rho}_{BC} = \hat{\rho}_{FE} + \frac{1+\hat{\rho}_{FE}}{T}$ so that

$$\Pr\left(\left|\sqrt{nT}\left(\hat{\rho}_{BC}-\hat{\rho}_{FE}-\frac{1+\hat{\rho}_{FE}}{T}\right)\right|>\epsilon|S\right)=0.$$

The second term is zero since $\sqrt{nT}\left(\hat{\rho}_{FE}-\rho+\frac{1+\rho}{T}\right)=O_p(1)$ implies

$$\begin{aligned}&T\left(1-\hat{\rho}_{FE}\right)-3\\&=T\left(1-\rho\right)-\sqrt{\frac{T}{n}}\left[\sqrt{nT}\left(\hat{\rho}_{FE}-\rho+\frac{1+\rho}{T}\right)\right]-3+\left(1+\rho\right)\\&=T\left(1-\rho\right)+O_p\left(\sqrt{\frac{T}{n}}\right)\xrightarrow{p}\infty\end{aligned}$$

and hence $\Pr\left(\bar{S}\right)\to 0$ as $T\to\infty$. Therefore, $\Pr\left(\left|\sqrt{nT}\left(\hat{\rho}_{BC}-\hat{\rho}_{FE}-\frac{1+\hat{\rho}_{FE}}{T}\right)\right|>\epsilon\right)\to 0$ as $(n,T)\to\infty$.

When $\rho = 1$, we have

$$\begin{aligned}&\lim_{(n,T)\to\infty}\Pr\left(\left|\sqrt{n}T\left(\hat{\rho}_{BC}-1\right)\right|>\epsilon\right)\\&=\lim_{(n,T)\to\infty}\Pr\left(\left|\sqrt{n}T\left(\hat{\rho}_{BC}-1\right)\right|>\epsilon|S\right)\Pr\left(S\right)\\&+\lim_{(n,T)\to\infty}\Pr\left(\left|\sqrt{n}T\left(\hat{\rho}_{BC}-1\right)\right|>\epsilon|\bar{S}\right)\Pr\left(\bar{S}\right).\end{aligned}$$

Now the fact that $\sqrt{n}T\left(\hat{\rho}_{FE}-1+\frac{3}{T}\right)=O_p(1)$ implies $\Pr(S)\to 0$ as $(n,T)\to\infty$, so that the first term is zero. For the second term, if $\bar{S}$ is true, $\hat{\rho}_{BC}=1$ so that $\Pr\left(\left|\sqrt{n}T\left(\hat{\rho}_{BC}-1\right)\right|>\epsilon|\bar{S}\right)=0$. Thus, $\Pr\left(\left|\sqrt{n}T\left(\hat{\rho}_{BC}-1\right)\right|>\epsilon\right)\to 0$ as $(n,T)\to\infty$. ■

B Derivation of Equation (18) and (27)

First of all, it is well known that for general non-stochastic $m \times m$ matrices Q_1 and Q_2 and a $m-$element vector $\xi \sim N(\mu_\xi, \Sigma_\xi)$, we have the following result

$$\begin{aligned} E[\xi' Q_1 \xi \xi' Q_2 \xi] &= tr\left[Q_1 \Sigma_\xi (Q_2 + Q_2') \Sigma_\xi\right] + \mu'(Q_1 + Q_1') \Sigma_\xi (Q_2 + Q_2') \mu_\xi \\ &\quad + \left[tr(Q_1 \Sigma_\xi) + \mu_\xi' Q_1 \mu_\xi\right] \left[tr(Q_2 \Sigma_\xi) + \mu_\xi' Q_2 \mu_\xi\right] \end{aligned} \tag{34}$$

With homoscedasticity, where $E[\varepsilon_{it}^2] = \sigma_\varepsilon^2$, for $i = 1, \cdots, n; t = 1, \cdots T$, we have the following results for the variance of $Z'A\varepsilon$ when $|\rho| < 1$:

$$\begin{aligned} var(Z'A\varepsilon) &= E[(Z'A\varepsilon)(\varepsilon' AZ)] - E[Z'A\varepsilon]E[Z'A\varepsilon]' \\ &= E[(\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon)(\varepsilon' A\bar{Z} + \varepsilon' A\tilde{Z})] - [\sigma_\varepsilon^2 ntr(\Pi_T)]^2 e_{K+1} e_{K+1}' && (35) \\ &= \bar{Z}'AE[\varepsilon\varepsilon']A\bar{Z} + E[\tilde{Z}'A\varepsilon\varepsilon' A\tilde{Z}] - [\sigma_\varepsilon^2 ntr(\Pi_T)]^2 e_{K+1} e_{K+1}' && (36) \\ &= \sigma_\varepsilon^2 \bar{Z}'A\bar{Z} + e_{K+1} E[\varepsilon' \Pi' \varepsilon\varepsilon' \Pi\varepsilon] e_{K+1}' - [n\sigma_\varepsilon^2 tr(\Pi_T)]^2 e_{K+1} e_{K+1}' \end{aligned}$$

By (34), we have

$$E[\varepsilon' \Pi' \varepsilon\varepsilon' \Pi\varepsilon] = \sigma_\varepsilon^4 tr\left[\Pi'(\Pi + \Pi')\right] + 2\sigma_\varepsilon^2 tr(\Pi) = \sigma_\varepsilon^4 \left[tr\left(\Pi'\Pi\right) + tr\left(\Pi^2\right)\right] + \sigma_\varepsilon^4 [tr(\Pi)]^2.$$

Therefore,

$$\begin{aligned} var(Z'A\varepsilon) &= \sigma_\varepsilon^2 \bar{Z}'A\bar{Z} + n\sigma_\varepsilon^4 [tr(\Pi_T'\Pi_T) + tr(\Pi_T^2)] e_{K+1} e_{K+1}' \\ &= \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4 tr(\Pi_T^2) e_{K+1} e_{K+1}' \end{aligned}$$

with

$$tr(\Pi_T^2) = -\frac{1 + 2\rho^{T-1}}{(1-\rho)^2} + \frac{2(1-\rho^T)}{T(1-\rho)^3} + \frac{(1-\rho^T)^2}{T^2(1-\rho)^4}$$

From (35) to (36) is due the fact that the third moment of a normal random variable is zero. The last step is because $\bar{Z}'A\bar{Z} = E[Z'AZ] - n\sigma_\varepsilon^2 tr(\Pi_T'\Pi_T) e_{K+1} e_{K+1}'$. Similarly, when $\rho = 1$, we have

$$var(Z'A\varepsilon) = \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4 tr(\Phi_T^2) e_{K+1} e_{K+1}'$$

with

$$tr(\Phi_T^2) = -\frac{1}{12}T^2 + \frac{1}{2}T - \frac{5}{12}$$

C Proof of Theorem 2

Proof.

For $|\rho| < 1$, first of all, as $(n,T) \to \infty$ with $\frac{n}{T} \to \sqrt{\kappa}^2$ we have

$$\lim_{(n,T)\to\infty} \frac{1}{\sqrt{nT}} E[Z'A\varepsilon] = \lim_{(n,T)\to\infty} \sqrt{\frac{n}{T}} \sigma_\varepsilon^2 tr(\Pi_T) e_{K+1} = -\sqrt{\kappa} \frac{\sigma_\varepsilon^2}{1-\rho} e_{K+1}$$

and

$$\lim_{(n,T)\to\infty} \frac{1}{nT} var(Z'A\varepsilon) = \sigma_\varepsilon^2 Q_{ZZ}.$$

Therefore,

$$\frac{1}{\sqrt{nT}} Z'A\varepsilon \xrightarrow{d} N\left(-\sqrt{\kappa}\frac{\sigma_\varepsilon^2}{1-\rho} e_{K+1}, \sigma_\varepsilon^2 Q_{ZZ}\right).$$

Now,

$$\sqrt{nT}(\hat{\delta}_{FE} - \delta) = \left(\frac{1}{nT} Z'AZ\right)^{-1} \left(\frac{1}{\sqrt{nT}} Z'A\varepsilon\right) \xrightarrow{d} N\left(-\sqrt{\kappa}\frac{\sigma_\varepsilon^2}{1-\rho} Q_{ZZ}^{-1} e_{K+1}, \sigma_\varepsilon^2 Q_{ZZ}^{-1}\right)$$

■

D Proof of Theorem 3

Proof.

For $\rho = 1$, note that

$$\hat{\delta}_{FE} - \delta = (Z'AZ)^{-1} (Z'A\varepsilon)$$

Now,

$$\hat{\delta}_{FE} - \delta - \delta^n = (Z'AZ)^{-1} (Z'A\varepsilon - (Z'AZ)\delta^n)$$

Thus,

$$\Lambda\left(\hat{\delta}_{FE} - \delta - \delta^n\right) = \left(\Lambda^{-1} Z'AZ\Lambda^{-1}\right)^{-1} \Lambda^{-1} (Z'A\varepsilon - (Z'AZ)\delta^n).$$

We first study

$$\Lambda^{-1} (Z'A\varepsilon - (Z'AZ)\delta^n) = \Lambda^{-1}\left(Z'A\varepsilon - \Lambda(\Lambda^{-1} Z'AZ\Lambda^{-1})\Lambda\delta^n\right)$$

As $(n,T)\to\infty$, given that $\delta^n = -\frac{\sigma_\varepsilon^2}{2T}Q_{ZZ}^{-1}e_{K+1}$, we have

$$\begin{aligned}
&\lim_{(n,T)\to\infty}\Lambda^{-1}E\left[\left(Z'A\varepsilon - \Lambda(\Lambda^{-1}Z'AZ\Lambda^{-1})\Lambda\delta^n\right)\right] \\
&= \lim_{(n,T)\to\infty}\Lambda^{-1}E[Z'A\varepsilon] - \Lambda^{-1}\Lambda\left(\lim_{(n,T)\to\infty}E[\Lambda^{-1}Z'AZ\Lambda^{-1}]\right)\Lambda\delta^n \\
&= \lim_{(n,T)\to\infty}\Lambda^{-1}E[Z'A\varepsilon] - Q_{ZZ}\Lambda\left(-\frac{\sigma_\varepsilon^2}{2T}Q_{ZZ}^{-1}e_{K+1}\right) \\
&= \lim_{(n,T)\to\infty}\Lambda^{-1}E[Z'A\varepsilon] + Q_{ZZ}\sqrt{n}T\frac{\sigma_\varepsilon^2}{2T}Q_{ZZ}^{-1}e_{K+1} \\
&= -\frac{\sigma_\varepsilon^2\sqrt{n}}{2}e_{K+1} + \frac{\sigma_\varepsilon^2\sqrt{n}}{2}e_{K+1} \\
&= 0
\end{aligned}$$

Next, we study the variance of $(Z'A\varepsilon - Z'AZ\delta^n)$, that is

$$\begin{aligned}
&var\,(Z'A\varepsilon - Z'AZ\delta^n) \\
&= E\left[(Z'A\varepsilon - Z'AZ\delta^n)(Z'A\varepsilon - Z'AZ\delta^n)'\right] - E[Z'A\varepsilon - Z'AZ\delta^n]E[Z'A\varepsilon - Z'AZ\delta^n]' \\
&= E[Z'A\varepsilon\varepsilon'AZ] - E[Z'A\varepsilon]E[Z'A\varepsilon]' + E[Z'AZ\delta^n\delta^{n'}Z'AZ] - E[Z'AZ\delta^n]E[Z'AZ\delta^n]' \\
&+ E[Z'AZ\delta^n]E[Z'A\varepsilon]' + E[Z'A\varepsilon]E[Z'AZ\delta^n]' - E[Z'AZ\delta^n\varepsilon'AZ] - E[Z'A\varepsilon\delta^{n'}Z'AZ]
\end{aligned}$$

We approach the variance using the same techniques where we decompose Z into $\bar{Z}+\tilde{Z}$ and we have that $A\tilde{Z} = \Phi\varepsilon e_{K+1}'$. First of all,

$$\begin{aligned}
&E[Z'A\varepsilon\varepsilon'AZ] - E[Z'A\varepsilon]E[Z'A\varepsilon]' \\
&= E[(\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon)(\varepsilon'A\bar{Z} + \varepsilon'A\tilde{Z})] - E[(\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon)]E[(\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon)]' \\
&= E[\bar{Z}'A\varepsilon\varepsilon'A\bar{Z}] + E[\tilde{Z}'A\varepsilon\varepsilon'A\bar{Z}] + E[\bar{Z}'A\varepsilon\varepsilon'A\tilde{Z}] + E[\tilde{Z}'A\varepsilon\varepsilon'A\tilde{Z}] - E[\tilde{Z}'A\varepsilon]E[\tilde{Z}'A\varepsilon]'
\end{aligned}$$

The second and third terms are zeros since they involve the third moment of normal random variable. Now,

$$\begin{aligned}
E[Z'A\varepsilon\varepsilon'AZ] - E[Z'A\varepsilon]E[Z'A\varepsilon]' &= \sigma_\varepsilon^2\bar{Z}'A\bar{Z} + n\sigma_\varepsilon^4[tr(\Phi_T'\Phi_T) + tr(\Phi_T^2)]e_{K+1}e_{K+1}' \\
&= \sigma_\varepsilon^2E[Z'AZ] + n\sigma_\varepsilon^4\left[tr(\Phi_T^2)\right]e_{K+1}e_{K+1}' \qquad (37)
\end{aligned}$$

The last equality is due the fact that $\bar{Z}'A\bar{Z} = E[Z'AZ] - n\sigma_\varepsilon^2 tr(\Phi_T'\Phi_T)e_{K+1}e_{K+1}'$. Secondly,

$$
\begin{aligned}
&E[Z'AZ\delta^n\delta^{n'}Z'AZ] - E\left[Z'AZ\delta^n\right]E\left[Z'AZ\delta^n\right]' \\
&= E\left[\left(\bar{Z}'A\bar{Z}\delta^n + \tilde{Z}'A\tilde{Z}\delta^n\right)\left(\delta^{n'}\bar{Z}'A\bar{Z} + \delta^{n'}\tilde{Z}'A\tilde{Z}\right)\right] \\
&- E\left[\bar{Z}'A\bar{Z}\delta^n + \tilde{Z}'A\tilde{Z}\delta^n\right]E\left[\bar{Z}'A\bar{Z}\delta^n + \tilde{Z}'A\tilde{Z}\delta^n\right]' \\
&= \bar{Z}'A\bar{Z}\delta^n\delta^{n'}\bar{Z}'A\bar{Z} + \bar{Z}'A\bar{Z}\delta^n\delta^{n'}E\left[\tilde{Z}'A\tilde{Z}\right] + E\left[\tilde{Z}'A\tilde{Z}\right]\delta^n\delta^{n'}\bar{Z}'A\bar{Z} \\
&+ E\left[\tilde{Z}'A\tilde{Z}\delta^n\delta^{n'}\tilde{Z}'A\tilde{Z}\right] - \left(\bar{Z}'A\bar{Z}\delta^n + E\left[\tilde{Z}'A\tilde{Z}\right]\delta^n\right)\left(\delta^{n'}\bar{Z}'A\bar{Z} + \delta^{n'}E\left[\tilde{Z}'A\tilde{Z}\right]\right) \\
&= E\left[\tilde{Z}'A\tilde{Z}\delta^n\delta^{n'}\tilde{Z}'A\tilde{Z}\right] - E\left[\tilde{Z}'A\tilde{Z}\right]\delta^n\delta^{n'}E\left[\tilde{Z}'A\tilde{Z}\right] \\
&= E[e_{K+1}\varepsilon'\Phi'\Phi\varepsilon e_{K+1}'\delta^n\delta^{n'}e_{K+1}\varepsilon'\Phi'\Phi\varepsilon e_{K+1}'] - E\left[e_{K+1}\varepsilon'\Phi'\Phi\varepsilon e_{K+1}'\right]\delta^n\delta^{n'}E\left[e_{K+1}\varepsilon'\Phi'\Phi\varepsilon e_{K+1}'\right] \\
&= \sigma_\varepsilon^4\left(n^2\left[tr\left(\Phi_T'\Phi_T\right)\right]^2 + 2ntr\left[\left(\Phi_T'\Phi_T\right)^2\right]\right)e_{K+1}e_{K+1}'\delta^n\delta^{n'}e_{K+1}e_{K+1}' \\
&- [n\sigma_\varepsilon^2 tr(\Phi_T'\Phi_T)]^2 e_{K+1}e_{K+1}'\delta^n\delta^{n'}e_{K+1}e_{K+1}' \\
&= 2n\sigma_\varepsilon^4 tr\left[\left(\Phi_T'\Phi_T\right)^2\right]e_{K+1}e_{K+1}'\delta^n\delta^{n'}e_{K+1}e_{K+1}'
\end{aligned} \tag{38}
$$

Thirdly,

$$
\begin{aligned}
&E[Z'AZ\delta^n\varepsilon'AZ] \\
&= E\left[\left(\bar{Z}'A\bar{Z} + \tilde{Z}'A\tilde{Z}\right)\delta^n\left(\varepsilon'A\bar{Z} + \varepsilon'A\tilde{Z}\right)\right] \\
&= E\left[\bar{Z}'A\bar{Z}\delta^n\varepsilon'A\bar{Z} + \tilde{Z}'A\tilde{Z}\delta^n\varepsilon'A\bar{Z} + \bar{Z}'A\bar{Z}\delta^n\varepsilon'A\tilde{Z} + \tilde{Z}'A\tilde{Z}\delta^n\varepsilon'A\tilde{Z}\right] \\
&= \bar{Z}'A\bar{Z}\delta^n E\left[\varepsilon'A\tilde{Z}\right] + E\left[\tilde{Z}'A\tilde{Z}\delta^n\varepsilon'A\tilde{Z}\right] \\
&= \bar{Z}'A\bar{Z}\delta^n E[\varepsilon'\Phi\varepsilon e_{K+1}'] + E[e_{K+1}\varepsilon'\Phi'\Phi\varepsilon e_{K+1}'\delta^n\varepsilon'\Phi\varepsilon e_{K+1}'] \\
&= n\sigma_\varepsilon^2 tr(\Phi_T)\bar{Z}'A\bar{Z}\delta^n e_{K+1}' + \sigma_\varepsilon^4\left(n^2 tr(\Phi_T'\Phi_T)tr(\Phi_T) + 2ntr(\Phi_T'\Phi_T^2)\right)e_{K+1}e_{K+1}'\delta^n e_{K+1}' \\
&= n\sigma_\varepsilon^2 tr(\Phi_T)E[Z'AZ]\delta^n e_{K+1}' + 2n\sigma_\varepsilon^4 tr(\Phi_T'\Phi_T^2)e_{K+1}e_{K+1}'\delta^n e_{K+1}'
\end{aligned} \tag{39}
$$

Similarly,

$$\begin{aligned}
&E[Z'A\varepsilon\delta^{n'}Z'AZ] \\
&= E\left[\left(\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon\right)\delta^{n'}\left(\bar{Z}'A\bar{Z} + \tilde{Z}'A\tilde{Z}\right)\right] \\
&= E\left[\bar{Z}'A\varepsilon\delta^{n'}\bar{Z}'A\bar{Z} + \bar{Z}'A\varepsilon\delta^{n'}\tilde{Z}'A\tilde{Z} + \tilde{Z}'A\varepsilon\delta^{n'}\bar{Z}'A\bar{Z} + \tilde{Z}'A\varepsilon\delta^{n'}\tilde{Z}'A\tilde{Z}\right] \\
&= E\left[\tilde{Z}'A\varepsilon\delta^{n'}\bar{Z}'A\bar{Z}\right] + E\left[\tilde{Z}'A\varepsilon\delta^{n'}\tilde{Z}'A\tilde{Z}\right] \\
&= E\left[e_{K+1}\varepsilon'\Phi'\varepsilon\right]\delta^{n'}\bar{Z}'A\bar{Z} + E\left[e_{K+1}\varepsilon'\Phi'\varepsilon\delta^{n'}e_{K+1}\varepsilon'\Phi'\Phi\varepsilon e'_{K+1}\right] \\
&= n\sigma_\varepsilon^2 tr(\Phi_T)e_{K+1}\delta^{n'}\bar{Z}'A\bar{Z} + e_{K+1}E\left[\varepsilon'\Phi'\varepsilon\delta^{n'}e_{K+1}\varepsilon'\Phi'\Phi\varepsilon\right]e'_{K+1} \\
&= n\sigma_\varepsilon^2 tr(\Phi_T)e_{K+1}\delta^{n'}\left(E[Z'AZ] - \sigma_\varepsilon^2 n tr(\Phi'_T\Phi_T)e_{K+1}e'_{K+1}\right) + \sigma_\varepsilon^4\left[n^2 tr(\Phi'_T)tr(\Phi'_T\Phi_T)\right. \\
&\left. + n tr(\Phi'_T\Phi'_T\Phi_T + \Phi_T\Phi'_T\Phi_T)\right]e_{K+1}\delta^{n'}e_{K+1}e'_{K+1} \\
&= n\sigma_\varepsilon^2 tr(\Phi_T)e_{K+1}\delta^{n'}E[Z'AZ] + n\sigma_\varepsilon^4 tr(\Phi'_T\Phi_T^2)e_{K+1}\delta^{n'}e_{K+1}e'_{K+1} \\
&+ n\sigma_\varepsilon^4 tr(\Phi_T\Phi'_T\Phi_T)e_{K+1}\delta^{n'}e_{K+1}e'_{K+1} \\
&= n\sigma_\varepsilon^2 tr(\Phi_T)e_{K+1}\delta^{n'}E[Z'AZ] + 2n\sigma_\varepsilon^4 tr(\Phi'_T\Phi_T^2)e_{K+1}\delta^{n'}e_{K+1}e'_{K+1} \qquad (40)
\end{aligned}$$

Lastly, we have

$$E[Z'AZ\delta^n]E[Z'A\varepsilon]' = E[Z'AZ]\delta^n E[Z'A\varepsilon]' = n\sigma_\varepsilon^2 tr(\Phi_T)E[Z'AZ]\delta^n e'_{K+1} \qquad (41)$$

and

$$E[Z'A\varepsilon]E[Z'AZ\delta^n]' = n\sigma_\varepsilon^2 tr(\Phi_T)e_{K+1}\delta^{n'}E[Z'AZ] \qquad (42)$$

We rewrite the variance of $Z'A\varepsilon - Z'AZ\delta^n$ using (37) – (42), which gives us

$$\begin{aligned}
&var\left(Z'A\varepsilon - Z'AZ\delta^n\right) \\
&= \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4\left[tr(\Phi_T^2)\right]e_{K+1}e'_{K+1} + 2n\sigma_\varepsilon^4 tr\left[(\Phi'_T\Phi_T)^2\right]e_{K+1}e'_{K+1}\delta^n\delta^{n'}e_{K+1}e'_{K+1} \\
&- n\sigma_\varepsilon^2 tr(\Phi_T)E[Z'AZ]\delta^n e'_{K+1} - 2n\sigma_\varepsilon^4 tr(\Phi'_T\Phi_T^2)e_{K+1}e'_{K+1}\delta^n e'_{K+1} - n\sigma_\varepsilon^2 tr(\Phi_T)e_{K+1}\delta^{n'}E[Z'AZ] \\
&- 2n\sigma_\varepsilon^4 tr(\Phi'_T\Phi_T^2)e_{K+1}\delta^{n'}e_{K+1}e'_{K+1} + n\sigma_\varepsilon^2 tr(\Phi_T)E[Z'AZ]\delta^n e'_{K+1} + n\sigma_\varepsilon^2 tr(\Phi_T)e_{K+1}\delta^{n'}E[Z'AZ] \\
&= \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4\left[tr(\Phi_T^2)\right]e_{K+1}e'_{K+1} + 2n\sigma_\varepsilon^4 tr\left[(\Phi'_T\Phi_T)^2\right]e_{K+1}e'_{K+1}\delta^n\delta^{n'}e_{K+1}e'_{K+1} \\
&- 2n\sigma_\varepsilon^4 tr(\Phi'_T\Phi_T^2)e_{K+1}e'_{K+1}\delta^n e'_{K+1} - 2n\sigma_\varepsilon^4 tr(\Phi'_T\Phi_T^2)e_{K+1}\delta^{n'}e_{K+1}e'_{K+1} \\
&= \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4\left[tr(\Phi_T^2)\right]e_{K+1}e'_{K+1} + 2n\sigma_\varepsilon^4 tr\left[(\Phi'_T\Phi_T)^2\right]e_{K+1}e'_{K+1}\delta^n\delta^{n'}e_{K+1}e'_{K+1} \\
&- 4n\sigma_\varepsilon^4 tr(\Phi'_T\Phi_T^2)e_{K+1}e'_{K+1}\delta^n e'_{K+1}
\end{aligned}$$

The last step is because that $e_{K+1}e'_{K+1}\delta^n e'_{K+1} = \left(-\frac{\sigma_\varepsilon^2}{2T}\right)e_{K+1}e'_{K+1}Q_{ZZ}^{-1}e_{K+1}e'_{K+1} = e_{K+1}\delta^{n'}e_{K+1}e'_{K+1}$, given that $\delta^n = -\frac{\sigma_\varepsilon^2}{2T} Q_{ZZ}^{-1} e_{K+1}$ and Q_{ZZ}^{-1} is symmetric. Now, we can rewrite the variance as

$$
\begin{aligned}
& var\left(Z'A\varepsilon - Z'AZ\delta^n\right) \\
& = \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4 \left[tr(\Phi_T^2)\right] e_{K+1}e_{K+1}' + 2n\sigma_\varepsilon^4 tr\left[\left(\Phi_T'\Phi_T\right)^2\right] e_{K+1}e_{K+1}'\delta^n\delta^{n\prime} e_{K+1}e_{K+1}' \\
& - 4n\sigma_\varepsilon^4 tr(\Phi_T'\Phi_T^2) e_{K+1}e_{K+1}'\delta^n e_{K+1}' \\
& = \sigma_\varepsilon^2 E[Z'AZ] + n\sigma_\varepsilon^4 \left[tr(\Phi_T^2)\right] e_{K+1}e_{K+1}' \\
& + \frac{n\sigma_\varepsilon^8}{2T^2} tr\left[\left(\Phi_T'\Phi_T\right)^2\right] e_{K+1}e_{K+1}' Q_{ZZ}^{-1} e_{K+1}e_{K+1}' Q_{ZZ}^{-1} e_{K+1}e_{K+1}' \\
& + \frac{2n\sigma_\varepsilon^6}{T} tr(\Phi_T'\Phi_T^2) e_{K+1}e_{K+1}' Q_{ZZ}^{-1} e_{K+1}e_{K+1}'
\end{aligned}
$$

Lemma 1 *According to Hahn and Kuersteiner (2002), we have*

$$
\begin{aligned}
tr(\Phi_T^2) &= -\frac{1}{12}T^2 + \frac{1}{2}T - \frac{5}{12}; \\
tr\left[\left(\Phi_T'\Phi_T\right)^2\right] &= \frac{1}{90}T^4 + \frac{1}{36}T^2 - \frac{7}{180}; \\
tr(\Phi_T'\Phi_T^2) &= -\frac{1}{12}T^2 + \frac{1}{12}.
\end{aligned}
$$

Therefore, by Lemma 1

$$
\begin{aligned}
& \lim_{(n,T)\to\infty} var\left[\Lambda^{-1}(Z'A\varepsilon - Z'AZ\delta^n)\right] \\
& = \sigma_\varepsilon^2 Q_{ZZ} - \frac{\sigma_\varepsilon^4}{12} e_{K+1}e_{K+1}' + \frac{\sigma_\varepsilon^8}{180} e_{K+1}e_{K+1}' Q_{ZZ}^{-1} e_{K+1}e_{K+1}' Q_{ZZ}^{-1} e_{K+1}e_{K+1}'
\end{aligned}
$$

Thus, by central limit theorem,

$$
\Lambda\left(\hat{\delta}_{FE} - \delta^s - \delta\right) \xrightarrow{d} N\left(0, \Omega_\delta^n\right)
$$

where

$$
\Omega_\delta^n = \sigma_\varepsilon^2 Q_{ZZ}^{-1} - \frac{\sigma_\varepsilon^4}{12} Q_{ZZ}^{-1} e_{K+1}e_{K+1}' Q_{ZZ}^{-1} + \frac{\sigma_\varepsilon^8}{180}\left(Q_{ZZ}^{-1} e_{K+1}e_{K+1}'\right)^3 Q_{ZZ}^{-1}
$$

■

E Derivation of Equation (30)

With the heteroskedasticity, where $E[\varepsilon_{it}^2] = \sigma_{it}^2$, for $|\rho| < 1$ we have

$$E[Z'A\varepsilon] = E[\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon] = \bar{Z}'AE[\varepsilon] + E[\tilde{Z}'A\varepsilon] = e_{K+1}E[\varepsilon'\Pi'\varepsilon] = \left[\sum_{i=1}^{n} tr(\Pi_T\Sigma_i)\right] e_{K+1}$$

Therefore, we have

$$\begin{aligned} var(Z'A\varepsilon) &= E[(Z'A\varepsilon)(\varepsilon'AZ)] - E[Z'A\varepsilon]E[Z'A\varepsilon]' \\ &= E[(\bar{Z}'A\varepsilon + \tilde{Z}'A\varepsilon)(\varepsilon'A\bar{Z} + \varepsilon'A\tilde{Z})] - \left[\sum_{i=1}^{n} tr(\Pi_T\Sigma_i)\right]^2 e_{K+1}e_{K+1}' \\ &= \bar{Z}'AE[\varepsilon\varepsilon']A\bar{Z} + E[\tilde{Z}'A\varepsilon\varepsilon'A\tilde{Z}] - \left[\sum_{i=1}^{n} tr(\Pi_T\Sigma_i)\right]^2 e_{K+1}e_{K+1}' \\ &= \bar{Z}'AE[\varepsilon\varepsilon']A\bar{Z} + E[e_{K+1}\varepsilon'\Pi'\varepsilon\varepsilon'\Pi\varepsilon e_{K+1}'] - \left[\sum_{i=1}^{n} tr(\Pi_T\Sigma_i)\right]^2 e_{K+1}e_{K+1}' \\ &= \bar{Z}'AE[\varepsilon\varepsilon']A\bar{Z} + E[\varepsilon'\Pi'\varepsilon\varepsilon'\Pi\varepsilon]e_{K+1}e_{K+1}' - \left[\sum_{i=1}^{n} tr(\Pi_T\Sigma_i)\right]^2 e_{K+1}e_{K+1}' \end{aligned}$$

By the result in (34), we have

$$\begin{aligned} E[\varepsilon'\Pi'\varepsilon\varepsilon'\Pi\varepsilon] &= tr\left[\Pi'\Sigma(\Pi + \Pi')\Sigma\right] + tr(\Pi'\Sigma)tr(\Pi\Sigma) \\ &= tr\left(\Pi'\Sigma\Pi\Sigma\right) + tr[(\Pi'\Sigma)^2] + [tr(\Pi\Sigma)]^2 . \end{aligned}$$

With the symmetry of Σ, we have $tr(\Pi'\Sigma) = tr(\Pi\Sigma)$. Now,

$$var(Z'A\varepsilon) = E[Z'A\varepsilon\varepsilon'AZ] + \sum_{i=1}^{n} tr[(\Pi_T'\Sigma_i)^2]e_{K+1}e_{K+1}',$$

since $E[Z'A\varepsilon\varepsilon'AZ] = \bar{Z}'A\Sigma A\bar{Z} + \left[\sum_{i=1}^{n} tr\left(\Pi_T'\Sigma_i\Pi_T\Sigma_i\right)\right] e_{K+1}e_{K+1}'$.

F Proof of Theorem 5

Proof. Let

$$S = \left\{2T(1-\hat{\rho}_{FE}^{k}) - \sigma_{\varepsilon}^{2} e_{K+1}' Q_{ZZ}^{-1} e_{K+1} > 0\right\}$$

and

$$\bar{S} = \left\{2T(1-\hat{\rho}_{FE}^{k}) - \sigma_{\varepsilon}^{2} e_{K+1}' Q_{ZZ}^{-1} e_{K+1} \leq 0\right\}.$$

When $|\rho| < 1$, it suffices to show that

$$\begin{aligned}
&\sqrt{nT}\left(\hat{\beta}_{BC} - \beta\right) - \sqrt{nT}\left(\hat{\beta}_{FE} - \beta + \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}\right) \\
&= \sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}\right) \xrightarrow{p} 0.
\end{aligned}$$

We have

$$\begin{aligned}
&\lim_{(n,T)\to\infty} \Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon\right) \\
&= \lim_{(n,T)\to\infty} \Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon | S\right) \Pr(S) \\
&+ \lim_{(n,T)\to\infty} \Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon | \bar{S}\right) \Pr\left(\bar{S}\right).
\end{aligned}$$

The first term is zero given that, if S is true, we have $\hat{\beta}_{BC} = \hat{\beta}_{FE} + \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}$ so that

$$\Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon | S\right) = 0.$$

We want to show that the second term shrinks to zero. $\sqrt{nT}\left(\hat{\rho}_{FE}^{k} - \rho + \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} e_{K+1}' Q_{ZZ}^{-1} e_{K+1}\right) = O_p(1)$ implies

$$\begin{aligned}
&2T(1-\hat{\rho}_{FE}^{k}) - \sigma_{\varepsilon}^{2} e_{K+1}' Q_{ZZ}^{-1} e_{K+1} \\
&= 2T(1-\rho) - 2\sqrt{\frac{T}{n}}\left[\sqrt{nT}\left(\hat{\rho}_{FE}^{k} - \rho + \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} e_{K+1}' Q_{ZZ}^{-1} e_{K+1}\right)\right] + \frac{\sigma_{\varepsilon}^{2}(1+\rho)}{1-\rho} e_{K+1}' Q_{ZZ}^{-1} e_{K+1} \\
&= 2T(1-\rho) + O_p\left(\sqrt{\frac{T}{n}}\right) \xrightarrow{p} \infty,
\end{aligned}$$

and hence $\Pr\left(\bar{S}\right) \to 0$ as $T \to \infty$. Therefore,

$$\Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_{\varepsilon}^{2}}{T(1-\rho)} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon\right) \to 0$$

as $(n,T) \to \infty$.

When $\rho = 1$, it suffices to show that

$$\begin{aligned} &\sqrt{nT}\left(\hat{\beta}_{BC} - \beta\right) - \sqrt{nT}\left(\hat{\beta}_{FE} - \beta + \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}\right) \\ &= \sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}\right) \xrightarrow{p} 0. \end{aligned}$$

We have

$$\begin{aligned} &\lim_{(n,T)\to\infty} \Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon\right) \\ &= \lim_{(n,T)\to\infty} \Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon | S\right) \Pr\left(S\right) \\ &+ \lim_{(n,T)\to\infty} \Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon | \bar{S}\right) \Pr\left(\bar{S}\right). \end{aligned}$$

For the second term, if $\bar{S}$ is true, we have $\hat{\beta}_{BC} = \hat{\beta}_{FE} + \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}$ so that

$$\Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon | \bar{S}\right) = 0.$$

For the first term, $\sqrt{n}T(\hat{\rho}_{FE}^k - 1 + \frac{\sigma_\varepsilon^2}{2T} e_{K+1}' Q_{ZZ}^{-1} e_{K+1}) = O_p(1)$ implies

$$\begin{aligned} &2T(1 - \hat{\rho}_{FE}^k) - \sigma_\varepsilon^2 e_{K+1}' Q_{ZZ}^{-1} e_{K+1} \\ &= -\frac{2}{\sqrt{n}}\left(\sqrt{n}T(\hat{\rho}_{FE}^k - 1 + \frac{\sigma_\varepsilon^2}{2T} e_{K+1}' Q_{ZZ}^{-1} e_{K+1})\right) \xrightarrow{p} 0, \end{aligned}$$

and hence $\Pr(S) \to 0$ as $(n, T) \to \infty$, so that the first term is $o(1)$. Therefore, we have

$$\Pr\left(\left|\sqrt{nT}\left(\hat{\beta}_{BC} - \hat{\beta}_{FE} - \frac{\sigma_\varepsilon^2}{2T} E_K Q_{ZZ}^{-1} e_{K+1}\right)\right| > \epsilon\right) \to 0$$

as $(n, T) \to \infty$, as desired. ■