Bias Reduction in Dynamic Panel Data Models by Common Recursive Mean Adjustment

Chi-Young Choi
University of Texas at Arlington

Nelson C. Mark
University of Notre Dame and NBER

Donggyu Sul*
University of Auckland

June 18, 2008

Abstract

LSDV estimation of the dominant root in dynamic panel regression is vulnerable to downward bias. This paper studies recursive mean adjustment (RMA) as a bias reduction strategy. We develop the RMA estimators for general AR(p) process under both cross sectional independence and dependence. We study its asymptotic properties as $N, T \to \infty$ jointly and find that the proposed asymptotically normal estimator exhibits nearly negligible bias when $\left(\log^2 T\right)(N/T) \to \zeta$ where ζ is a non-zero constant. The proposed method is an efficient and effective bias reduction strategy and is straightforward to implement. Our simulation experiments suggest that the RMA estimator performs quite well in terms of bias, variance and MSE reduction both when error terms are cross sectionally independent and correlated. It dominates comparable estimators particularly when T is small and/or the underlying process is highly persistent.

Keywords: Recursive Mean Adjustment, Fixed Effects, Cross Sectional Dependence.

JEL Classification Numbers: C33

---

*Corresponding author. Email: d.sul@auckland.ac.nz. We thank Peter C.B. Phillips for introducing this topic and for helpful comments on an earlier version of the paper and Chirok Han for useful and thoughtful comments.

1 Introduction

In small T samples, accurate estimation of the autocorrelation coefficient $\rho \in [0,1)$, of a stationary but persistent first-order autoregressive time series $y_{it} = \alpha_i + \rho y_{it-1} + \epsilon_{it}$, $\epsilon_{it} \overset{iid}{\sim} (0, \sigma_i^2)$, must account for downward bias induced by running the regression with a constant. To see the source of this bias, think of running least squares without a constant on deviations from the sample mean $y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$. Then for any observation $t = 1, ..., T$, the regression error ϵ_{it} is correlated with current and future values of y_{it} embedded in the sample mean component of the explanatory variable.[1] This small T bias is also present in fixed-effects estimators for panel data due to the incidental parameters problem. For fixed T as $N \to \infty$, Nickell (1981) shows that the least squares dummy variable (LSDV) estimator for the dynamic panel regression model remains substantially downward biased.

In this paper, we study and apply the recursive-mean adjustment (RMA) technique to estimate linear dynamic panel data models under cross-sectional homogeneity of the dominant root. The paper builds on the success of RMA to reduce bias in the regression context.[2] In the RMA strategy, the observations are adjusted by the common recursive mean, $(t-1)^{-1}\sum_{j=1}^{t-1} y_{ij}$ instead of the sample mean.[3] Because the recursive mean does not contain future values of y_{it}, the adjusted regressor $(y_{i,t-1} - \frac{1}{(t-1)}\sum_{j=1}^{t-1} y_{ij})$ is orthogonal to the original regression error (ϵ_{it}) and hence reduces substantively the bias for the fixed effect.[4]

We first develop the RMA strategy for general AR(p) processes under cross-sectionally independent observations and under cross-sectional dependence. We also consider estimation of dynamic panels with exogenous variables. As in Alvarez and Arellano (2003), Bai (2004) and Hahn and Kuersteiner (2002), our asymptotic analysis is based on large T and large N.[5] We then carry out a series of Monte Carlo experiments to evaluate the precision and effectiveness of the RMA estimator in reducing bias and the accuracy of the asymptotic theory for small T and moderate N sample sizes. When the observations are independent, the pooled RMA estimator is shown to deliver effective bias reduction. When the observations are cross-sectionally correlated and the dependence is generated by an underlying factor structure, we find that the RMA estimator also performs quite well particularly when T is small and/or ρ is close to unity.

The remainder of the paper is organized as follows. Section 2 develops and discusses asymptotic properties of the panel RMA estimators under cross-sectionally independent observations. In section

[1] Mariott and Pope (1954) and Kendall (1954) discuss and characterize the first-order approximation of this bias. Several bias correction strategies have been suggested in the literature, such as median unbiased estimation [Andrews (1993)], approximately median unbiased estimation [Andrews and Chen (1994)] and mean unbiased estimation [Phillips and Sul (2007)].

[2] The RMA strategy was applied to reduce bias in regression by So and Shin (1999a) and in the context of unit root testing by So and Shin (1999b, 2002).

[3] The commonality refers to the fact that the identical recursive mean is subtracted from both dependent and independent variables.

[4] As illustrated below, since the error term after the RMA adjustment now contains $-(1-\rho)\frac{1}{(t-1)}\sum_{j=1}^{t-1} y_{ij}$ which is correlated with the adjusted regressor, one can obtain an unbiased control for the fixed effect if $\rho = 1$.

[5] Other research that has addressed the bias in dynamic panel data estimation include Phillips and Sul (2007) and Sul (2007). Phillips and Sul study the mean unbiased estimator for the panel AR(1) model with cross-sectionally dependent observations, whereas Sul applies the RMA method to construct panel unit root tests for fixed N and large T.

3, we extend the asymptotic analysis to an environment with cross-sectionally correlated observations. Section 4 reports the results of Monte Carlo experiments that compare performance across alternative estimators. Section 5 concludes. The Appendix contains proofs and details of many arguments made in the text.

2 Asymptotic Properties Under Cross-Sectional Independence

In this section, we consider observations in dynamic panel data with individual-specific fixed effects. Section 2.2 considers dynamic panels with exogenous regressors. Here, we begin with the panel p-th order autoregression.

2.1 Panel AR(p)

The data are assumed to be generated by the latent model in

Assumption 1 For $i = 1, \dots N, t = 1, \dots T$, the observations $\{y_{it}\}$ have the latent model structure

$$\begin{aligned} y_{it} &= \mu_i + z_{it}, && (1) \\ z_{it} &= \sum_{j=1}^{p} \rho_j z_{i,t-j} + \epsilon_{it}, && (2) \end{aligned}$$

where $\left|\sum_{j=1}^{p} \rho_j\right| = |\rho| < 1$, $\mu_i \overset{iid}{\sim} \left(\mu, \sigma_\mu^2\right)$, $y_{i0} \overset{iid}{\sim} \left(0, \sigma_i^2 / \left(1 - \rho^2\right)\right)$, and $\epsilon_{it} \overset{iid}{\sim} \left(0, \sigma_i^2\right)$ is independent of μ_i and y_{i0}, and has finite moments up to the fourth order.[6]

The latent model formulation (1)-(2) has the observationally equivalent regression representation

$$y_{it} = \mu_i (1 - \rho) + \sum_{j=1}^{p} \rho_j y_{it-j} + \epsilon_{it}, \tag{3}$$

with initial observation $y_{i0} = \mu_i + z_{i0}$. Since most economic time series are positively serially correlated, we assume that $\rho = \sum_{j=1}^{p} \rho_j \in [0, 1)$. On occasion, it will be useful to recast (3) in the following augmented Dickey-Fuller (ADF) form

$$y_{it} = a_i + \rho y_{it-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{it-j} + \epsilon_{it}. \tag{4}$$

2.1.1 The AR(1) environment

A widely studied environment for dynamic panel data estimators sets $p = 1$ in Assumption 1,

$$y_{it} = a_i + \rho y_{i,t-1} + \epsilon_{it}, \tag{5}$$

[6] The independence and fourth-moment restrictions were also imposed by Alvarez and Arellano (2003).

where $a_i = (1-\rho)\,\mu_i$ with $\epsilon_{it} \overset{iid}{\sim} (0, \sigma_i^2)$.

LSDV estimation of the panel AR(1) model is equivalent to estimating $y_{it} - \overline{y}_{iT} = \rho\,(y_{it-1} - \overline{y}_{iT}) + e_{it}$ by least squares without a constant. Notice that the observations are deviations from the sample mean $\overline{y}_{iT} = T^{-1}\sum_{j=1}^{T} y_{ij}$ and the regression error is $e_{it} = \epsilon_{it} + a_i - (1-\rho)\overline{y}_{iT}$. For $\rho \in [0,1)$, fixed T as $N \to \infty$, Nickell (1981) shows that the LSDV estimator is downward biased due to the positive correlation between ϵ_{it} and current and future values of y_{it} contained in the $\overline{y}_{iT-1}$ component of the regressor.

The panel RMA estimator in the AR(1) case can be obtained as follows. First, form the recursive mean $c_{it-1} = (t-1)^{-1}\sum_{j=1}^{t-1} y_{ij}$ and form the adjusted variables $(y_{it} - c_{it-1})$ and $(y_{it-1} - c_{it-1})$ to run pooled least-squares without a constant on

$$(y_{it} - c_{it-1}) = \rho\,(y_{it-1} - c_{it-1}) + e_{it}, \tag{6}$$

where $e_{it} = \epsilon_{it} + (1-\rho)\,\mu_i - (1-\rho)c_{it-1} = \epsilon_{it} - (1-\rho)\,(t-1)^{-1}\sum_{j=1}^{t-1} z_{ij}$ and z_{it} is given in (2). The RMA estimator for ρ is

$$\hat{\rho}_{\text{RMA}} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it-1} - c_{it-1})\,(y_{it} - c_{it-1})}{\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it-1} - c_{it-1})^2}. \tag{7}$$

Estimates of the individual-specific effects are then given by $\widehat{a}_i = T^{-1}\sum_{t=1}^{T}(y_{it} - \hat{\rho}_{\text{RMA}}\, y_{it-1})$.

It is worth noting that RMA is able to control for the individual-specific effect because $E(1-\rho)c_{it-1} = \mu_i$ especially when $\rho = 1$. The central insight of this strategy is that the adjusted regressor $(y_{it-1} - c_{it-1})$ is orthogonal to the original error term (ϵ_{it}) because the recursive mean contains observations only up through date $t-1$. When $\rho < 1$, a correlation still exists between the adjusted regressor $(y_{it-1} - c_{it-1})$ and the new error term $(\epsilon_{it} - (1-\rho)\,(t-1)^{-1}\sum_{j=1}^{t-1} z_{ij})$ via c_{it-1} and $\sum_{j=1}^{t-1} z_{ij}$, but RMA adjustment reduces the resulting bias substantively. Since it is the identical recursive mean c_{it-1} that is subtracted from both y_{it} and y_{it-1}, the estimator is more precisely described as the 'common' RMA estimator. For the ease of exposition, however, we will drop the term 'common' and simply refer to the RMA estimator throughout the paper. The asymptotic properties of the estimator are stated in

Proposition 1 *(Asymptotic distribution in the AR(1) model with fixed effects)* Let the observations be generated by *Assumption 1 with* $p = 1$.

(i) If $(\log^2 T)\,(N/T) \to \zeta$ *as* $T, N \to \infty$*, then* $\hat{\rho}_{\text{RMA}}$ *is asymptotically distributed as*

$$\sqrt{NT}\,(\hat{\rho}_{\text{RMA}} - \rho - B\,(\rho, T)) \to^{d} \mathcal{N}\left(0, 1-\rho^2\right), \tag{8}$$

where

$$B(\rho,T) \equiv -(1-\rho)\frac{C(\rho,T)}{D(\rho,T)} > 0,\ C(\rho,T) = \sum_{t=1}^{T-1} t^{-1}\left\{2t^{-1}\sum_{h=1}^{t-1} h\gamma_h^{(z)} - \sum_{h=1}^{t-1}\gamma_h^{(z)}\right\},$$

$$D(\rho,T) = (T-1)\gamma_0^{(z)} - \sum_{t=1}^{T-1} t^{-1}\{\gamma_0^{(z)} - 2t^{-1}\sum_{h=1}^{t-1} h\gamma_h^{(z)}\},$$

and $\gamma_h^{(z)}$ *is the covariance between* z_{it} *and* z_{it+h}.

(ii) If $\left(\log^2 T\right)(N/T) \to 0$ *as* $T, N \to \infty$, *then* $\hat{\rho}_{\mathrm{RMA}}$ *is asymptotically distributed as*

$$\sqrt{NT}\left(\hat{\rho}_{\mathrm{RMA}} - \rho\right) \to^d \mathcal{N}\left(0, 1-\rho^2\right). \quad (9)$$

The proof of Proposition 1 is presented in the Appendix. Whereas Alvarez and Arellano (2003) show that LSDV is consistent if $N/T \to 0$ as $T, N \to \infty$, consistency of $\hat{\rho}_{\mathrm{RMA}}$ requires $\log T\sqrt{N/T} \to 0$ because the bias term $B(\rho,T)$ is $O\left(\frac{\log T}{T}\right)$. However, the actual bias of $\hat{\rho}_{\mathrm{RMA}}$ turns out to be so small that it can be ignored in practice. To see this, consider the explicit formula of the bias for fixed T as $N \to \infty$,

$$B(\rho,T) \equiv \mathrm{plim}_{N\to\infty}\left(\hat{\rho}_{\mathrm{RMA}} - \rho\right) = \frac{\sum_{t=2}^{T}\left[\frac{\rho}{t-1}\left(1+\rho^{t-2} - 2\left\{\frac{1+\rho}{t-1}\right\}\left\{\frac{1-\rho^{t-1}}{1-\rho^2}\right\}\right)\right]}{(T-1) - \sum_{t=2}^{T}\left[\frac{1}{t-1}\left(1 - 2\frac{\rho^{t-1}}{1-\rho} + 2\left\{\frac{\rho}{t-1}\right\}\left\{\frac{1-\rho^{t-1}}{(1-\rho)^2}\right\}\right)\right]} \geq 0.$$

Figure 1 shows the value of the bias $B(\rho,T)$ for various ρ and T. The maximum bias is 0.028 which occurs at $T = 13$ and $\rho = 0.46$. In addition, the bias gets smaller as ρ is closer to unity as predicted.

There are two special cases where $\hat{\rho}_{\mathrm{RMA}}$ is exactly unbiased for fixed T as $N \to \infty$. The first case is when $\rho = 1$. Here, it can be seen from (6) that $e_{it} = \epsilon_{it}$ which implies $\mathrm{E}(y_{it-1}\epsilon_{it}) = \mathrm{E}(c_{it-1}\epsilon_{it}) = 0$. The second case is when $T = 3$ for any $\rho \in [0,1)$. In this case,

$$\begin{aligned} B(\rho,3) &= -(1-\rho)\frac{\mathrm{E}\left[(y_{i1} - y_{i1})\,y_{i1} + \left(y_{i2} - \frac{1}{2}\{y_{i1}+y_{i2}\}\right)\frac{1}{2}\{y_{i1}+y_{i2}\}\right]}{\mathrm{E}\left[(y_{i1}-y_{i1})^2 y_{i1}^2 + \left(y_{i2} - \frac{1}{2}\{y_{i1}+y_{i2}\}\right)^2\right]} \\ &= -(1-\rho)\frac{\mathrm{E}\left[y_{i2}^2 - y_{i1}^2\right]}{\mathrm{E}\left[(y_{i2}-y_{i1})^2\right]} = 0, \end{aligned}$$

because $\mathrm{E}\left(z_{i2}^2\right) = \mathrm{E}\left(z_{i1}^2\right)$ by the covariance stationarity of y_{it}. Generally speaking, since the bias is so small, it is fair to say that $\hat{\rho}_{\mathrm{RMA}}$ is approximately consistent when N is much larger than T.

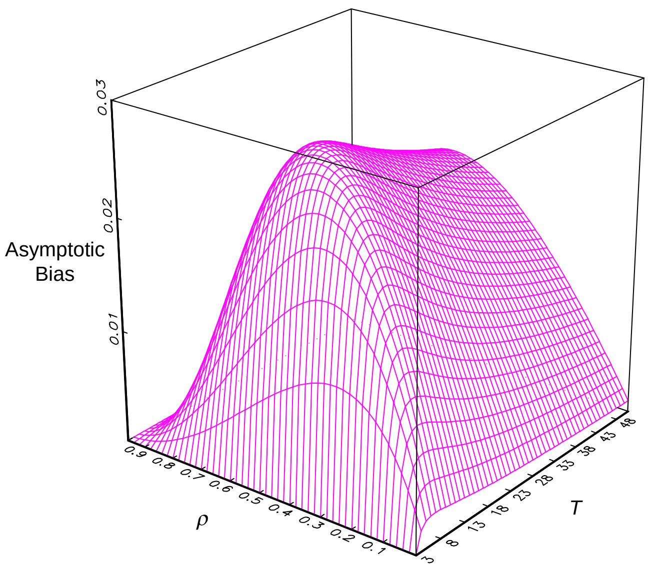

Figure 1: Asymptotic bias of RMA estimator under AR(1) and independence

2.1.2 The AR(p) environment

Accurate measurement of higher-order autoregressive terms is vital in many applied problems such as half-life estimation from impulse responses. In extending the model to the general AR(p) structure, we refer to the augmented Dickey–Fuller form of the regression (4) and employ the RMA strategy to obtain bias-corrected estimates of the ϕ_j coefficients as well as ρ. Estimation proceeds in three steps as follows.

Step 1: Estimate (4) by LSDV and call the estimated coefficients on the lagged differences $\hat{\phi}_{j,\text{LSDV}}$. Let $y_{it}^{+} = y_{it} - \sum_{j=1}^{p-1} \hat{\phi}_{j,\text{LSDV}} \Delta y_{it-j}$ and $\epsilon_{it}^{+} = \epsilon_{it} + \sum_{j=1}^{p-1} \left(\phi_j - \hat{\phi}_{j,\text{LSDV}}\right) \Delta y_{it-j}$. Then (4) can be rewritten as

$$y_{it}^{+} = a_i + \rho y_{it-1} + \epsilon_{it}^{+}. \tag{10}$$

Step 2: The RMA estimator for ρ, $\hat{\rho}_{\text{RMA}}^{p}$, is then obtained from running a pooled least-squares regression on

$$\left(y_{it}^{+} - c_{it-1}\right) = \rho \left(y_{it-1} - c_{it-1}\right) + e_{it},$$

where $c_{it-1} = (t-1)^{-1} \sum_{s=1}^{t-1} y_{is}$ and $e_{it} = -(1-\rho)(t-1)^{-1} \sum_{s=1}^{t-1} z_{is} + \epsilon_{it} + \sum_{j=1}^{p-1} \left(\phi_j - \hat{\phi}_{j,\text{LSDV}}\right) \Delta z_{it-j}$.

Step 3: The LSDV estimator $\hat{\phi}_{j,\text{LSDV}}$ from step 1 is biased by $O\left(T^{-1}\right)$. This bias can be reduced by treating $\hat{\rho}^p_{\text{RMA}}$ as the true value of ρ and running a second LSDV regression on

$$\left(y_{it} - \hat{\rho}^p_{\text{RMA}} y_{it-1}\right) = a_i + \sum_{j=1}^{p-1} \phi_j \Delta y_{it-j} + e^p_{it}, \tag{11}$$

where $e^p_{it} = \epsilon_{it} + \left(\rho - \hat{\rho}^p_{\text{RMA}}\right) y_{it-1}$. Call the resulting estimator $\hat{\phi}^p_{j,\text{RMA}}$. ■

The asymptotic distribution of $\hat{\rho}^p_{\text{RMA}}$ and $\hat{\phi}^p_{j,\text{RMA}}$ are given in,

Proposition 2: *(Asymptotic Distribution in the AR(p) model with fixed effects).* Let the observations be generated by Assumption 1.

(i) If $N \to \infty$ for fixed T, then the asymptotic bias of $\hat{\rho}^p_{\text{RMA}}$ and $\hat{\phi}^p_{j,\text{RMA}}$ are

$$\text{plim}_{N\to\infty} \left(\hat{\rho}^p_{\text{RMA}} - \rho\right) = B(\rho, T) + O\left(T^{-2}\right), \tag{12}$$

$$\text{plim}_{N\to\infty} \left(\hat{\phi}^p_{j,\text{RMA}} - \phi_j\right) = \frac{1}{2} B(\rho, T) + O\left(T^{-2}\right), \tag{13}$$

where $B(\rho, T)$ is given in Proposition 1.

(ii) If $\left(\log^2 T\right)(N/T) \to \zeta$ as $T, N \to \infty$,

$$\sqrt{NT}\left(\hat{\rho}^p_{\text{RMA}} - \rho - B(\rho, T)\right) \to^d \mathcal{N}\left(0, 1 - \rho^2\right).$$

The proof of Proposition 2 is presented in the Appendix. Note that estimation in step 3 reduces the $O(T^{-1})$ bias in $\hat{\phi}_{j,\text{LSDV}}$ to $O(T^{-2})$ if $B(\rho, T)$ is small and can be ignored. These estimators have trivial $N-$asymptotic bias for relatively large values of ρ. The asymptotic bias of $\hat{\rho}^p_{\text{RMA}}$ is only slightly larger than it is under the AR(1) case.

2.2 Panel AR(p) with exogenous regressor

We now consider dynamic panel data regressions with exogenous regressors. The observations are generated according to

Assumption 2. For $i = 1, \ldots N, t = 1, \ldots T$, the observations $\{y_{it}\}$ have the latent model structure

$$y_{it} = \mu_i + \beta w_{it} + z_{it}, \tag{14}$$

$$z_{it} = \rho z_{it-1} + \sum_{j=1}^{p-1} \phi_j \Delta z_{it-j} + \epsilon^*_{it}, \tag{15}$$

$$\epsilon^*_{it} = \gamma q_{it} + \epsilon_{it}, \tag{16}$$

where w_{it} and q_{it} are econometrically exogenous to y_{it}, $|\rho| < 1$, $\mu_i \overset{iid}{\sim} \left(\mu, \sigma_\mu^2\right)$, and $\epsilon_{it} \overset{iid}{\sim} \left(0, \sigma_i^2\right)$.

The latent form (14)-(16) has the observationally equivalent dynamic panel regression representation,

$$y_{it} = a_i + \rho y_{it-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{it-j} + \beta w_{it} + \sum_{j=1}^{p-1} \kappa_j \Delta w_{it-j} + \gamma q_{it} + \epsilon_{it}. \tag{17}$$

Balestra and Nerlove (1966) emphasize different roles of two exogenous variables: w_{it} as a level effect and q_{it} as a difference effect. In their empirical example, y_{it} is the quantity of natural gas, w_{it} is per capita income, and q_{it} is the relative price of natural gas. Per capita income affects gas demand in levels but the relative price affects gas demand in the '$\rho-$difference.'

Phillips and Sul (2007) show that for fixed T as $N \to \infty$ the LSDV estimator for β is consistent but $E\left(\hat{\kappa}_{\text{LSDV}} - \kappa\right) = O\left(T^{-1}\right)$ such that the LSDV estimators of the coefficients on the lagged w_{it} variables are biased. In general, the parameters of interest are ρ, γ, and β. However, when the observations are cross-sectionally correlated, as is discussed in the next section, the methods used to control for cross-sectional dependence are sensitive to bias in the estimated covariance matrix which in turn is sensitive to bias in the κ_j coefficients. Therefore, our treatment of the RMA estimator here pertains to obtaining bias attenuation in the estimation of all the coefficients in (17). The RMA estimator is obtained as follows:

Step 1: Estimate (17) by LSDV and call the estimated coefficients $\hat{\beta}_{\text{LSDV}}, \hat{\phi}_{j,\text{LSDV}}, \hat{\kappa}_{j,\text{LSDV}}$. Let $y_{it}^+ = y_{it} - \sum_{j=1}^{p-1} \hat{\phi}_{j,\text{LSDV}} \Delta y_{it-j} - \hat{\beta}_{\text{LSDV}} w_{it} - \sum_{j=1}^{p-1} \hat{\kappa}_{j,\text{LSDV}} \Delta w_{it-j}$ and $\epsilon_{it}^+ = \epsilon_{it} + \sum_{j=1}^{p-1} \left(\phi_j - \hat{\phi}_{j,\text{LSDV}}\right) \Delta y_{it-j}$ $+ \left(\beta - \hat{\beta}_{\text{LSDV}}\right) w_{it} + \sum_{j=1}^{p-1} \left(\kappa_j - \hat{\kappa}_{j,\text{LSDV}}\right) \Delta w_{it-j}$. Note that (17) can be rewritten as

$$y_{it}^+ = a_i + \rho y_{it-1} + \epsilon_{it}^+.$$

Step 2. Let $c_{it-1} = (t-1)^{-1} \sum_{s=1}^{t-1} y_{is}$ and run pooled least-squares on

$$\left(y_{it}^+ - c_{it-1}\right) = \rho \left(y_{it-1} - c_{it-1}\right) + e_{it},$$

which gives $\hat{\rho}_{\text{RMA}}^p$, the RMA estimator of ρ.

Step 3. Treat $\hat{\rho}_{\text{RMA}}^p$ as the true value of ρ and run LSDV on

$$y_{it} - \hat{\rho}_{\text{RMA}}^p y_{it-1} = a_i + \sum_{j=1}^{p-1} \phi_j \Delta y_{it-j} + \beta w_{it} + \sum_{j=1}^{p-1} \kappa_j \Delta w_{it-j} + e_{it}^p. \tag{18}$$

If the regression contains only the level effect ($\gamma = 0$), then stop. If a difference effect is present, then proceed to

Step 4. Let $y_{it}^{++} = y_{it} - \hat{\rho}_{\text{RMA}}^{p} y_{it-1} - \sum_{j=1}^{p-1} \hat{\phi}_{j,\text{RMA}} \Delta y_{it-j} + \hat{\beta}_{\text{RMA}} w_{it} + \sum_{j=1}^{p-1} \hat{\kappa}_{j,\text{RMA}} \Delta w_{it-j}$ and run LSDV on

$$y_{it}^{++} = a_i + \gamma q_{it} + \epsilon_{it}^{p},$$

which gives $\hat{\gamma}_{\text{RMA}}$, the RMA estimator for γ.■

It is straightforward to show that Proposition 2 holds for $\hat{\rho}_{\text{RMA}}^{p}$ and $\hat{\phi}_{j,\text{RMA}}^{p}$. With regard to $\hat{\gamma}_{\text{RMA}}$, we note that

$$\epsilon_{it}^{p} = \epsilon_{it} + \left(\rho - \hat{\rho}_{\text{RMA}}^{p}\right) y_{it-1} + \sum_{j=1}^{p-1} \left(\phi_j - \hat{\phi}_{j,\text{LSDV}}\right) \Delta y_{it-j} + \left(\beta - \hat{\beta}_{\text{LSDV}}\right) w_{it} + \sum_{j=1}^{p-1} \left(\kappa_j - \hat{\kappa}_{j,\text{LSDV}}\right) \Delta w_{it-j}.$$

Also, note that for large N and T,

$$\begin{aligned}
\left(\rho - \hat{\rho}_{\text{RMA}}^{p}\right) \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it} \tilde{y}_{it-1} &= O_p\left(\frac{1}{\sqrt{NT}} + \frac{\log T}{T}\right) O_p(1), \\
\left(\phi_j - \hat{\phi}_{j,\text{LSDV}}\right) \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it} \Delta \tilde{y}_{it-1} &= O_p\left(\frac{1}{\sqrt{NT}} + \frac{\log T}{T}\right) O_p(1), \\
\left(\beta - \hat{\beta}_{\text{LSDV}}\right) \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it} \tilde{w}_{it} &= O_p\left(\frac{1}{\sqrt{NT}}\right) O_p(1) = O_p\left(\frac{1}{NT}\right), \\
\left(\kappa_j - \hat{\kappa}_{j,\text{LSDV}}\right) \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it} \Delta \tilde{w}_{it-1} &= O_p\left(\frac{1}{\sqrt{NT}} + \frac{\log T}{T}\right) O_p(1),
\end{aligned}$$

where the affix notation '$\tilde{x}_t$' signifies that the series x_t has been demeaned. It follows that

$$\sqrt{NT}\left(\hat{\gamma}_{\text{RMA}} - \gamma\right) = \frac{\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it} \tilde{\epsilon}_{it}^{p}}{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it}^{2}} = O_p\left(\frac{1}{\sqrt{NT}} + \frac{\log T}{T}\right) + \frac{\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it} \epsilon_{it}}{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{q}_{it}^{2}}.$$

As $N, T \to \infty$,

$$\sqrt{TN}\left(\hat{\gamma}_{\text{RMA}} - \gamma\right) \to^{d} \quad \mathcal{N}\left(0, \sigma_{\gamma}^{2}\right).$$

3 Asymptotic Properties with cross-sectionally correlated observations

In most panel data environments, observations will be correlated across individuals. To maintain clarity and to avoid a proliferation of notation, we assume that the cross-sectional correlation of the regression error has a single factor representation. The results obtained here under the single-factor representation carry over to multi-factor environments at a cost of additional notational burden. The environment under consideration is again the one in which observations are weakly stationary.

In addition to dependence on past values, the idiosyncratic component ϵ_{it} and individual-specific fixed effects, however, y_{it} now may be hit by a common factor θ_t.[7] The observations are generated according to

Assumption 3 For $i = 1, \ldots N, t = 1, \ldots T$, the observations $\{y_{it}\}$ have the latent model structure

$$y_{it} = \mu_i + z_{it}, \tag{19}$$

$$z_{it} = \sum_{j=1}^{p} \rho_j z_{i,t-j} + u_{it}, \tag{20}$$

$$u_{it} = \delta_i \theta_t + \epsilon_{it}, \tag{21}$$

where the vector of idiosyncratic errors $\underline{\epsilon} = (\underline{\epsilon}_1', \ldots, \underline{\epsilon}_N')'$ and $\underline{\epsilon}_i' = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$ have $\mathrm{E}(\underline{\epsilon}) = \underline{0}$, $E(\underline{\epsilon}\underline{\epsilon}') = \mathbf{\Sigma}_\epsilon = diag[\sigma_{\epsilon,1}^2, \ldots \sigma_{\epsilon,N}^2]$, $\mathrm{E}|\underline{\epsilon}|^8 \leq M < \infty$, $E(\epsilon_s' \epsilon_t / N) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \leq M$ for all s, and $T^{-1}\sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s,t)| \leq M$. The common factor has $E\|\theta_t\|^4 < \infty$, $\mathrm{plim}_{T\to\infty} T^{-1}\sum_{t=1}^T \theta_t^2 = \sigma_\theta^2$, with factor loadings $\|\delta_i\| \leq D < \infty$, and $\mathrm{plim}_{N\to\infty} \left\|\underline{\delta}'\underline{\delta}/N - D\right\| = 0$ for some $D > 0$, where $\underline{\delta} = (\delta_1, ..., \delta_N)'$.

The above latent model representation can be recast to the common factor representation

$$y_{it} = \mu_i + \delta_i F_t + x_{it}, \tag{22}$$

where $F_t = \sum_{j=0}^{\infty} \rho^j \theta_{t-j}$ is the serially correlated common factor component and $x_{it} = \sum_{j=0}^{\infty} \rho^j \epsilon_{it-j}$ is the serially correlated but cross-sectionally independent idiosyncratic component. Since the error-covariance matrix is characterized by $2N$ unknowns, the feasibility condition for implementing a generalized least squares (GLS) correction is $T > 4$.[8]

To derive the exact bias formulae under cross-sectionally correlated y_{it}, we introduce the following notation. Let

$$\hat{\rho}_{\mathrm{F,LSDV}} = \rho + \frac{\sum_{t=2}^T \left(F_{t-1} - \bar{F}_{t-1}\right)\left(\theta_t - \bar{\theta}\right)}{\sum_{t=2}^T \left(F_{t-1} - \bar{F}_{t-1}\right)^2} \tag{23}$$

be the LSDV estimator of ρ from the regression of $F_t = k + \rho F_{t-1} + \theta_t$ where $\bar{F}_{t-1} = (t-1)^{-1}\sum_{j=1}^{t-1} F_{t-j}$, and let

$$\hat{\rho}_{\mathrm{F,RMA}} = \rho + \frac{\sum_{t=2}^T \left(F_{t-1} - \bar{F}_{t-1}\right)\left(\theta_t - (1-\rho)\bar{F}_{t-1}\right)}{\sum_{t=2}^T \left(F_{t-1} - \bar{F}_{t-1}\right)^2}, \tag{24}$$

be the RMA estimator from $F_t - \bar{F}_{t-1} = \rho\left(F_{t-1} - \bar{F}_{t-1}\right) + \varepsilon_t$ where $\varepsilon_t = \theta_t - (1-\rho)\bar{F}_{t-1}$. Let $\hat{\rho}_{\mathrm{LSDV|CSD}}$ and $\hat{\rho}_{\mathrm{RMA|CSD}}$ respectively denote the LSDV and RMA estimators applied to the cross-

[7] The factor structure has been employed by Bai and Ng (2002), Moon and Perron (2004) and Phillips and Sul (2003).

[8] Including the lagged dependent variable and the fixed effects further reduces the degrees of freedom by 2 in the panel AR(1) model. In a $K-$factor environment, for instance, there are $(K+1)N$ unknowns and the feasibility condition is $T > K + 3$ in the panel AR(1) model.

sectionally correlated data. With these definitions in hand, we can now characterize the $N-$asymptotic bias of the RMA and LSDV estimators when they are applied to observations where the error term is governed by a single-factor structure.

Proposition 3: *($N-$asymptotic bias under cross-sectional dependence).* Under Assumption 3 for fixed T as $N \to \infty$,

$$\text{plim}_{N\to\infty}(\hat{\rho}_{\text{LSDV|CSD}} - \rho) = (1-\eta)\left(\frac{-(1+\rho)}{T}\right) + \eta\underbrace{\left(\hat{\rho}_{\text{F,LSDV}} - \rho\right)}_{(23)} + o_p\left(T^{-1}\right), \quad (25)$$

$$\text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{RMA|CSD}} - \rho\right) = (1-\eta)\, B(\rho, T) + \eta\underbrace{\left(\hat{\rho}_{\text{F,RMA}} - \rho\right)}_{(24)} + o_p\left(T^{-1}\right), \quad (26)$$

where $\eta \equiv m_\delta^2\sigma_\theta^2\left(\sigma^2 + m_\delta^2\sigma_\theta^2\right)^{-1} \in [0,1]$ captures the degree of cross-sectional dependence, $m_\delta^2 = N^{-1}\sum^N \delta_i^2$, $\sigma_\theta^2 = T^{-1}\sum_{t=1}^T \theta_t^2$, $\sigma^2 = N^{-1}T^{-1}\sum_{i=1}^N\sum_{t=1}^T \epsilon_{it}^2$, and $B(\rho, T)$ is given in Proposition 1.

Since the terms $\left(\hat{\rho}_{\text{F,LSDV}} - \rho\right)$ and $\left(\hat{\rho}_{\text{F,RMA}} - \rho\right)$ depend on the realization of the common factors θ_t as seen from (23) and (24) and they are random for finite T, we say that $\hat{\rho}_{\text{LSDV|CSD}}$ and $\hat{\rho}_{\text{RMA|CSD}}$ exhibit random $N-$asymptotic bias. The random bias of $\hat{\rho}_{\text{RMA|CSD}}$ attributable to $\left(\hat{\rho}_{\text{F,RMA}} - \rho\right)$ is not troublesome because $\text{E}\left(\hat{\rho}_{\text{F,RMA}} - \rho\right)$ is inconsequential as shown by So and Shin (1999b). Consequently, the total bias in $\hat{\rho}_{\text{RMA|CSD}}$ is small even when the y_{it}s are perfectly cross-sectionally correlated ($\eta = 1$). When $\eta = 0$, the limiting values are not random since they do not depend on the factor θ_t. In this case, the $o_p\left(T^{-1}\right)$ term in (25) and (26) becomes $O\left(T^{-2}\right)$ in Proposition 3.

In light of the random asymptotic bias, it may be instructive to show their mean values. Taking expectations of the expressions in Proposition 3 with noting that $E\left(\hat{\rho}_{\text{F,OLS}} - \rho\right) = -(1+3\rho)/T$ and that $E\left(\hat{\rho}_{\text{F,RMA}} - \rho\right) - B(\rho, T)$ is a tiny negative number, this yields

$$E(\hat{\rho}_{\text{LSDV|CSD}} - \rho) = -\frac{1+\rho}{T} - \eta\frac{2\rho}{T} + O\left(T^{-2}\right),$$

$$E\left(\hat{\rho}_{\text{RMA|CSD}} - \rho\right) = B(\rho, T) - \eta\left[B(\rho, T) - E\left(\hat{\rho}_{\text{F,RMA}} - \rho\right)\right] < B(\rho, T).$$

Somewhat surprisingly, $\hat{\rho}_{\text{RMA|CSD}}$ has lower mean bias than that of $\hat{\rho}_{\text{RMA}}$ under independence. This suggests that the performance of the feasible GLS estimator will be enhanced if RMA residuals are used to estimate the error covariance matrix because it is less biased than the other estimators under consideration.

4 Monte Carlo Experiments

In this section, we conduct Monte Carlo experiments to examine the precision and the effectiveness of bias reduction achieved by the RMA estimators in small T and moderate N samples for $\rho \in [0,1)$.

We vary the environments by the autoregressive order and by the degree of cross-sectional dependence. For the economy of exposition, we are selective in tabulating the simulation results especially when $\log T\,(N/T)$ is relatively large because in this case $B\,(\rho, T)$ (and therefore the asymptotic bias) remains in the distribution of the RMA estimator. Further details of the simulation results are available at the author's web site.[9] We consider the following four cases for the simulation experiments.

Case 1: *AR(1) with cross-sectionally independent observations.* For this widely studied environment, several bias reduction methods have been proposed. We compare two of these to the RMA estimator together with LSDV. The first is a GMM estimator studied by Arellano and Bover (1995, hereafter AB) and the other is proposed by Hahn and Kuersteiner (2002, hereafter HK). The data generating process is,

$$\begin{aligned} y_{it} &= \mu_i + z_{it}, \\ z_{it} &= \rho z_{it-1} + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\mu_i \overset{iid}{\sim} \mathcal{N}\left(1, \sigma_\mu^2\right)$. As in Assumption 1, the initial observation obeys $y_{i1} = \mu_i + z_{i1} \overset{iid}{\sim} \mathcal{N}\left(1, \sigma_\mu^2 + \frac{1}{1-\rho^2}\right)$, which produces weakly stationary sequences of y_{it}. We consider sample sizes of $N \in \{50, 100, 200\}$ and $T \in \{6, 11, 21\}$ so that the total sample size used in the regression is $T_0 = T - 1 \in \{5, 10, 20\}$.

The asymptotic variance of the AB estimator depends on the nuisance parameter of $\psi = \sigma_\epsilon/\sigma_\mu = 1/\sigma_\mu$, whereas the variances of RMA, LSDV, and HK do not. To explore the potential small-sample dependence of the AB estimator on the variability of the individual-specific effect, we consider alternative values of relative variance of the individual-specific component of the error term, $\sigma_\mu \in \{1, 5, 10\}$, or $\psi \in \{1, 0.2, 0.1\}$.

Table 1 reports the bias and mean-square error of the four estimators under comparison. The RMA and AB estimators are seen to be upward biased while LSDV and HK are biased downward. RMA compares well to HK for small T. Although the relative performance of HK improves as T grows, when $T_0 = 5$, for example, the HK estimator bears substantial downward bias (-0.22 for $\rho = 0.9$) even when N is as large as 200. For relatively large T, the performance of HK is comparable to that of RMA particularly when ρ is relatively small. The GMM estimator due to AB performs well for $\psi = 1$, but its performance deteriorates substantively for $\psi = 0.2$ and $\psi = 0.1$. Even for $\psi = 1$, it is dominated by RMA for moderate values of ρ. In sum, RMA dominates HK both in terms of attenuating bias and in precision for small T and it is typically more precise than AB, whose performance is quite sensitive to ψ. The dominance of RMA over HK is particularly noticeable for small T or for highly persistent ρ when T is relatively large.

Case 2: *AR(2) with cross-sectionally independent observations.* Since it is not straightforward to correct for bias with HK or AB in the AR(2) case, we only report the performance results for RMA

[9] Full reports (MS excel format) are available at http://homes.eco.auckland.ac.nz/dsul013/working/MC_RMA.xls

in comparison to LSDV. For simplicity but without loss of generality, the DGP for this case is

$$\begin{aligned} y_{it} &= \mu_i + z_{it}, \\ z_{it} &= \rho_1 z_{it-1} + \rho_2 z_{it-2} + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} N(0,1)$, $\mu_i \overset{iid}{\sim} N\left(0, \sigma_\mu^2\right)$. We consider the lag coefficients $(\rho, \rho_2) \in \{(0.9, 0.2), (0.5, 0.2), (0.9, 0.3), (0.5, 0.3), (0.9, 0.4), (0.5, 0.4)\}$ where $\rho = \rho_1 + \rho_2$. Here we report the results for $(\rho, \rho_2) \in \{(0.9, 0.2), (0.5, 0.2)\}$ only because the results of the other cases are largely similar.

Table 2 reports the bias, variance and mean-square error of RMA and LSDV. We note that the bias of LSDV in the AR(2) case is much more serious than that in AR(1) environment. For example, when $\rho = 0.9$ and $N = 50$, the LSDV bias for ρ is -0.62, -0.32, -0.16 for $T_0 = 5, 10$, and 20 respectively whereas the biases were -0.47, -0.25, and -0.12 for the corresponding values of T_0 in AR(1) case. As in AR(1) case, the bias hinges upon T rather than N. Although the variance of RMA is slightly larger than that of LSDV, the MSE of RMA is consistently much smaller than that of LSDV estimator.

Case 3: *AR(1) and AR(2) with exogenous regressor.* Since Phillips and Sul (2007) show that $\hat{\beta}_{\text{LSDV}}$ is asymptotically unbiased when the exogenous variable enters with a level effect, we concentrate on the following DGP that allows an exogenous variable to enter in difference form,

$$\begin{aligned} y_{it} &= \mu_i + z_{it}, \\ z_{it} &= \rho_1 z_{it-1} + \rho_2 z_{it-2} + \gamma q_{it} + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} N(0,1)$, $\mu_i \overset{iid}{\sim} N(0,1)$ and $q_{it} \overset{iid}{\sim} N(0,1)$. We set $\gamma = 1$ and report the size of the t-test under the null hypothesis of $\gamma = 1$. Among the various values for the parameters considered, we report in Table 3 only the results for $\rho = 0.9$ for the AR(1) case and $\rho = 0.9$ and $\rho_2 = 0.2$ for the AR(2) case.

Several features of Table 3 are noteworthy. First, the LSDV estimator for γ is biased downward in both the AR(1) and AR(2) cases and the bias directly distorts the size of the t-test. While the size distortion for RMA is relatively small and remains fairly constant, the distortion of the LSDV based t-test increases with N. At the nominal size of 0.05, the size of t-test based on LSDV estimator is as large as 0.94 when $N = 200$ given $T_0 = 5$ in AR(1) case, whereas the corresponding size of t-test based on RMA is merely 0.08. Second, RMA reduces the bias and variance significantly for all cases considered. However, in the AR(2) case for very small T, there is some size distortion in the RMA based t-test when N is relatively large, mainly driven by the large second order bias of RMA. The size distortion, however, diminishes quickly as T increases.

Case 4: *AR(1) with cross-sectionally dependent observations.* The DGP for this case is

$$\begin{aligned} y_{it} &= a_i + \rho y_{it-1} + u_{it}, \\ u_{it} &= \delta_i \theta_t + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\delta_i \overset{iid}{\sim} \mathcal{N}(1,1)$ and $\theta_t \overset{iid}{\sim} \mathcal{N}(0,1)$. We consider exactly the same simulation environment with that of Case 1 except that the error term is now serially correlated as reflected by the common factor θ_t. Table 4 reports the bias and MSE of the four estimators under comparison. As in Table 1, we present the AB estimator for three different values of relative variance of the individual-specific component of the error term.

We first note from Table 4 that the bias of RMA estimator is smaller for moderate ρ and/or large T than when the observations are cross-sectionally independent. By stark contrast, the biases of the competing estimators are larger under cross sectional dependence. This result is consistent with the predictions of Proposition 3 that the bias of RMA (LSDV) is smaller (larger) under cross sectional dependence. Consequently, the performance of RMA estimator stands out even when the observations are cross-sectionally dependent. The dominance of RMA estimator is particularly noticeable when N and T are moderate. Unlike the case of cross sectional independence, the RMA estimator continues to dominate the alternative estimators even when N and T are large. Take $N = 200$ and $T = 20$ for instance, the HK estimator had a comparable performance to the RMA estimator under cross sectional independence, but the bias of RMA estimator is now much smaller when observations are cross sectionally correlated. The story remains much the same in terms of MSE. Although the MSE of RMA estimator is larger than when the observations are cross-sectionally independent mainly due to the increased variance, it decrease more rapidly than the alternative estimators as N and T grow. As a result, the RMA estimator has smaller MSE than the other estimators especially when the underlying processes are highly persistent. To sum, our simulation results suggest that the finite sample performance of the RMA estimator is appealing even when the observations are cross-sectionally correlated.

5 Conclusions

In this paper, we extend the idea of recursive mean adjustment as a bias reduction strategy to estimating the dominant root in dynamic panel data regressions. Specifically we develop the RMA estimators under general AR(p) process under both cross sectional independence and dependence. We show that the RMA estimator delivers effective bias reduction when the observations are independent across individuals. When the observations are correlated across individuals and when this dependence arises from an underlying factor structure, we find that effective bias reduction still can be achieved by using the RMA estimator. Our simulation results based on small T and larger N suggest that the RMA estimator dominates comparable estimators in terms of bias, variance and MSE reduction when error terms are cross sectionally independent. This finding still holds in the presence of exogenous regressors especially in terms of t-test performance. Overall our method is efficient and effective in reducing bias and more importantly is straightforward to implement. In light of the fact that mean and median unbiased estimators are generally unavailable for higher ordered panel autoregression models, the recursive mean adjustment procedure advocated in this study is believed to fill an important gap in the dynamic panel literature.

References

[1] Alvarez J. and Arellano, M. (2003), "The Time Series and Cross-Section Asymptotics of Dynamic Data Estimators," *Econometrica*, 71, 1121–1159.

[2] Andrews, D.W.K. (1993), "Exactly Median-unbiased Estimation of First Order Au- toregressive/Unit root Models, " *Econometrica*, 61, 139–165.

[3] Andrews, D.W.K. and Chen, H., (1994). "Approximately median-unbiased estimation of autoregressive models," *Journal of Business and Economic Statistics*, 12, 187–204.

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

[5] Bai, J. (2003), "Inference on Factor Models of Large Dimensions," *Econometrica,* 71, 135–172.

[6] Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70, 191–221.

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

[8] Hahn, J., and Kuersteiner, G. (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica* 70(4), 1639–1657.

[9] Kendall, M.G., (1954), "Note on Bias in the Estimation of Autocorrelation," *Biometrika* 41, 403–404.

[10] Marriott, F. H. C., and Pope, J. A.. (1954), "Bias in the Estimation of Autocorrelations," *Biometrika* 41, 393–403.

[11] Moon, H.R., and Perron, B. (2004), "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics* 122, 82–126.

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

[13] Phillips, P.C.B. and Sul, D. (2003), "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *The Econometrics Journal* 6, 217-259.

[14] Phillips, P.C.B. and Sul, D. (2007), "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence," *Journal of Econometrics,* Vol. 137 (1), 162-188.

[15] So, B.S. and Shin, D.W. (1999a), "Recursive Mean Adjustment in Time-Series Inferences," *Statistics & Probability Letters*, 43, pp 65–73

[16] So, B.S. and Shin, D.W. (1999b), "Cauchy Estimators for Autoregressive Processes with Application to Unit-Root Tests And Confidence Intervals," *Econometric Theory*, Vol. 15, pp. 165-176.

[17] So, B.S. and Shin, D.W. (2002), "Recursive Mean Adjustment and Tests for Nonstationarities," *Economics Letters*, Vol 75, pp. 203–208.

[18] Sul, D. (2007), "Panel Unit Root Tests under Cross Section Dependence with Recursive Mean Adjustment," *mimeo*, University of Auckland.

6 Appendix

Lemma 1: $\mathrm{E}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\varepsilon_{it+1}=0$

Lemma 2: $-\left(1-\rho\right)\mathrm{E}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\bar{z}_{it} \quad = \quad N\sum_{t=1}^{T}\frac{1}{t}\left\{\frac{2}{t}\sum_{h=1}^{t}h\gamma_h^{(z)}-\sum_{h=1}^{t}\gamma_h^{(z)}\right\}$
where $\gamma_h^{(z)}\equiv\mathrm{E}z_{it}z_{it-h}$.

Lemma 3: $\frac{1}{NT}\mathrm{E}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^2=\frac{\sigma^2}{1-\rho^2}+O\left(\frac{\log T}{T}\right)$

Proof of Lemma 1 & 2: It is straightforward, hence omitted. Note that for AR(1) case, we have

$$
\begin{aligned}
-\left(1-\rho\right)\mathrm{E}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\bar{z}_{it} &= N\sum_{t=1}^{T}\frac{\rho}{t}\left(1+\rho^{t+1}-\frac{2}{t}\left(\frac{1-\rho^t}{1-\rho}\right)\right) \\
&= NT\left\{\rho\log T+O\left(1\right)\right\}
\end{aligned} \tag{A1}
$$

Proof of Lemma 3: To prove Lemma 3, we use convergence in mean square. Note that for general AR(p) case, we have

$$
\begin{aligned}
&\mathrm{E}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\frac{1}{t}\sum_{s=1}^{t}z_{is}\right)^2 \\
=\ &\gamma_0^{(z)}+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left\{\gamma_0^{(z)}+2\sum_{h=1}^{t}\gamma_h^{(z)}+\frac{2}{t}\sum_{h=1}^{t}h\gamma_h^{(z)}\right\}-2\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left\{\sum_{h=0}^{t}\gamma_h^{(z)}\right\} \\
=\ &\gamma_0^{(z)}-\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left\{\gamma_0^{(z)}-\frac{2}{t}\sum_{h=1}^{t-1}h\gamma_h^{(z)}\right\}
\end{aligned}
$$

Hence for AR(1) case, we have

$$
\mathrm{E}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\frac{1}{t}\sum_{s=1}^{t}z_{is}\right)^2=\frac{\sigma^2}{1-\rho^2}+O\left(\frac{\log T}{T}\right)\to\frac{\sigma^2}{1-\rho^2}. \tag{A2}
$$

Proof of Proposition 1 Firstly, from a standard central limit theorem for panel autoregressive processes, we have

$$
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}\varepsilon_{it+1}\to^{d}\quad\mathcal{N}\left(0,\frac{\sigma^4}{1-\rho^2}\right) \tag{A3}
$$

From (A2), we have

$$
\sqrt{NT}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}\varepsilon_{it+1}\right)\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^2\right]^{-1}\quad\to^{d}\quad\mathcal{N}\left(0,1-\rho^2\right) \tag{A4}
$$

Now note that

$$\sqrt{NT}\left(\hat{\rho}_{\mathrm{RMA}}-\rho\right)=\frac{\sqrt{NT}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}\varepsilon_{it+1}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^{2}}-(1-\rho)\frac{\sqrt{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\bar{z}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^{2}} \tag{A5}$$

From (A1), (A2), and direct calculation, we have

$$\begin{aligned}-(1-\rho)\operatorname{plim}_{N\to\infty}\frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\bar{z}_{it}}{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^{2}} &= \frac{\sum_{t=1}^{T}\left[\frac{\rho}{t}\left(1+\rho^{t+1}-\left(\frac{2}{t}\right)\left(\frac{1-\rho^{t}}{1-\rho}\right)\right)\right]}{T-\sum_{t=1}^{T}\left[\frac{1}{t}\left(1-\left(\frac{2\rho^{t}}{1-\rho}\right)+\frac{2\rho}{t}\left(\frac{1-\rho^{t}}{(1-\rho)^{2}}\right)\right)\right]}\\ &= \frac{\rho\log T}{T}+O\left(T^{-2}\right)\end{aligned}$$

For calculating its variance, first consider

$$\mathrm{E}\left(\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\bar{z}_{it}\right)^{2}=\mathrm{E}\left(\sum_{t=1}^{T}z_{it}\left(\frac{1}{t}\sum_{s=1}^{t}z_{is}\right)-\sum_{t=1}^{T}\left(\frac{1}{t}\sum_{s=1}^{t}z_{is}\right)^{2}\right)^{2}$$

From tedious calcualtion, we have

$$\mathrm{E}\left(\sum_{t=1}^{T}z_{it}\left(\frac{1}{t}\sum_{s=1}^{t}z_{is}\right)-\sum_{t=1}^{T}\left(\frac{1}{t}\sum_{s=1}^{t}z_{is}\right)^{2}\right)^{2}=\sum_{t=2}^{T}\frac{1}{t}\sigma^{4}+O(1)=O(\log T)$$

Hence the variance of the second term in (A5) is given by

$$Var\left(-(1-\rho)\frac{\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\bar{z}_{it}}{\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^{2}}\right)=O\left(\frac{\log T}{T^{2}}\right)$$

Therefore

$$\sqrt{NT}\left(\hat{\rho}_{\mathrm{RMA}}-\rho\right)=O_{p}\left(\log T\sqrt{\frac{N}{T}}\right)+O_{p}\left(\sqrt{\frac{N}{T^{3}}}\right)+\frac{\sqrt{NT}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}\varepsilon_{it+1}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^{2}} \tag{A6}$$

so that as $T,N\to\infty$ but $\log T\sqrt{\frac{N}{T}}\to 0$, we have

$$\sqrt{NT}\left(\hat{\rho}_{\mathrm{RMA}}-\rho\right)\to^{d}\quad\mathcal{N}\left(0,1-\rho^{2}\right) \tag{A7}$$

If $N/T^{3}\to 0$ but $\log T\sqrt{\frac{N}{T}}\to\zeta$ where ζ is a constant, then we have

$$\sqrt{NT}\left(\hat{\rho}_{\mathrm{RMA}}-\rho-B(\rho,T)\right)\to^{d}\quad\mathcal{N}\left(0,1-\rho^{2}\right) \tag{A8}$$

or

$$\sqrt{NT}\left(\hat{\rho}_{\text{RMA}}-\rho\right)\rightarrow^{d}\quad\mathcal{N}\left(\zeta B\left(\rho,T\right),1-\rho^{2}\right)$$

where $B\left(\rho,T\right)$ is defined as

$$B\left(\rho,T\right)=\frac{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\rho}{t}\left(1+\rho^{t+1}-\left(\frac{2}{t}\right)\left(\frac{1-\rho^{t}}{1-\rho}\right)\right)\right]}{1-\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{t}\left(1-\left(\frac{2\rho^{t}}{1-\rho}\right)+\frac{2\rho}{t}\left(\frac{1-\rho^{t}}{(1-\rho)^{2}}\right)\right)\right]}. \tag{A9}$$

Proof of Proposition 2

Proof of (i) For ease of reference, we restate (10) here as

$$y_{it}^{+}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}=\rho\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)+e_{it},$$

where

$$e_{it}=-\left(1-\rho\right)\frac{1}{t-1}\sum_{s=1}^{t-1}z_{is}+\epsilon_{it}+\sum_{j=1}^{p-1}\left(\phi_{j}-\hat{\phi}_{j,\text{LSDV}}\right)\Delta z_{it-j}.$$

Noting that $E\sum_{t=p}^{T}\Delta z_{it-j}\bar{z}_{it-1}=O\left(1\right)$ and $\text{plim}_{N\rightarrow\infty}\left(\hat{\phi}_{j,\text{LSDV}}-\phi_{j}\right)=O\left(T^{-1}\right)$, we have

$$\begin{aligned}
\text{plim}_{N\rightarrow\infty}\left(\hat{\rho}_{\text{RMA}}^{p}-\rho\right) &= -\left(1-\rho\right)\text{plim}_{N\rightarrow\infty}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)\frac{1}{t-1}\sum_{s=1}^{t-1}z_{is}\right]}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)^{2}}\\
&\quad+\text{plim}_{N\rightarrow\infty}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)\sum_{j=1}^{p-1}\left(\phi_{j}-\hat{\phi}_{j,\text{LSDV}}\right)\Delta z_{it-j}\right]}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)^{2}}\\
&= B\left(\rho,T\right)+\frac{O\left(T^{-1}\right)}{D\left(\rho,T\right)}=B\left(\rho,T\right)+O\left(T^{-2}\right),
\end{aligned} \tag{A10}$$

which establishes (12) in the text.

Next, we obtain (13). For concreteness we will consider the AR(2) case and then show how the logic generalizes to an AR(p). Consider the regression error

$$\epsilon_{it}^{\dagger}=\epsilon_{it}-\left(\hat{\rho}_{\text{RMA}}^{p}-\rho\right)y_{it-1}.$$

Using the fact that $\left(y_{it-1}-\mu_{i}\right)-\left(y_{it-2}-\mu_{i}\right)=\left(y_{it-1}-y_{it-2}\right)$ where $\mu_{i}=E\left(y_{it}\right)$, the bias of the pooled estimator $\hat{\phi}^p_{\text{RMA}}$ can be written as

$$\hat{\phi}^p_{\text{RMA}} - \phi = -\left(\hat{\rho}^p_{\text{RMA}} - \rho\right) \underbrace{\frac{\sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)\left(y_{it-1} - y_{i\cdot-1}\right)}{\sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)^2}}_{\text{A}} + \frac{\sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)\left(\epsilon_{it} - \epsilon_{i\cdot}\right)}{\sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)^2},$$

where $y_{i\cdot-1}$ and $\epsilon_{i\cdot}$ are time-series averages. As $N \to \infty$, the term labeled A above has the limiting value

$$\operatorname*{plim}_{N\to\infty} \frac{\sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)\left(y_{it-1} - y_{i\cdot-1}\right)}{\sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)^2} = \frac{\sum_{t=1}^T \left(\gamma_0^{(z)} - \gamma_1^{(z)}\right)}{2\sum_{t=1}^T \left(\gamma_0^{(z)} - \gamma_1^{(z)}\right)} = \frac{1}{2}.$$

It follows that

$$\operatorname*{plim}_{N\to\infty} \left(\hat{\phi}^p_{\text{RMA}} - \phi\right) = -\frac{1}{2} \operatorname*{plim}_{N\to\infty} \left(\hat{\rho}^p_{\text{RMA}} - \rho\right) + O\left(T^{-1}\right) \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)\left(\epsilon_{it} - \epsilon_{i\cdot}\right).$$

Noting that the AR(2) model has the representation

$$y_{it} = \frac{c_1}{1 - \lambda_1 L}\epsilon_{it} + \frac{c_2}{1 - \lambda_2 L}\epsilon_{it},$$

where λ_1 and λ_2 are the roots of $\left(1 - \rho_1 z - \rho_2 z^2\right)$, and because

$$\begin{aligned} & E\frac{1}{T}\left(\sum_{j=0}^{\infty} c_s \lambda_s^j u_{it-j}\right)\left(\sum_{t=1}^{T} \epsilon_{it}\right) - E\frac{1}{T}\left(\sum_{j=0}^{\infty} c_s \lambda_s^j u_{it-j-1}\right)\left(\sum_{t=1}^{T} \epsilon_{it}\right) \\ = \; & c_s\left(1 - \lambda_1\right) - c_s\left(\frac{T-1}{T}\right)\left(1 - \lambda_1\right) + O\left(T^{-2}\right) = O\left(T^{-1}\right) \quad \text{for } s = 1, 2, \end{aligned}$$

where

$$c_1 = \lambda_1\left(\lambda_1 - \lambda_2\right) \text{ and } c_2 = -\lambda_2/\left(\lambda_1 - \lambda_2\right).$$

It follows that

$$\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\right)\left(\epsilon_{it} - \epsilon_{i\cdot}\right) = -\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \left(y_{it-1} - y_{it-2}\epsilon_{i\cdot}\right) = O\left(T^{-1}\right),$$

or equivalently

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

It is apparent that this logic goes through in the AR(p) case. We can therefore say

$$\begin{aligned}\underset{N\to\infty}{\text{plim}}\left(\hat{\underline{\phi}}^p_{\text{RMA}}-\underline{\phi}\right) &= -\frac{1}{2}\underset{N\to\infty}{\text{plim}}\left(\hat{\rho}^p_{\text{RMA}}-\rho\right)+\underset{N\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^N\sum_{t=1}^T z_{it}\tilde{u}_{it}}{\sum_{i=1}^N\sum_{t=1}^T z_{it}^2}\right)\\ &= -\frac{1}{2}\underset{N\to\infty}{\text{plim}}\left(\hat{\rho}^p_{\text{RMA}}-\rho\right)+O\left(T^{-2}\right).\end{aligned}$$

It follows that the residual bias in $\hat{\underline{\phi}}^p_{\text{RMA}}$ is inconsequential.

Proof of (ii) From (A10), we have

$$\sqrt{NT}\left(\hat{\rho}^p_{\text{RMA}}-\rho\right)=\sqrt{NT}\frac{C_{1,NT}}{D_{NT}}+\sqrt{NT}\frac{C_{2,NT}}{D_{NT}},$$

where

$$\begin{aligned}C_{1,NT} &= -(1-\rho)\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left[\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)\frac{1}{t-1}\sum_{s=1}^{t-1}z_{is}\right]\\ &= B(\rho,T)+O_p\left(\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

and

$$C_{2,NT}=\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left[\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)\sum_{j=1}^{p-1}\left(\phi_j-\hat{\phi}_{j,\text{LSDV}}\right)\Delta z_{it-j}\right]=O_p\left(\sqrt{\frac{1}{NT^3}}\right),$$

since

$$\left(\phi_j-\hat{\phi}_{j,\text{LSDV}}\right)=O_p\left(\frac{1}{\sqrt{NT}}\right)+O\left(\frac{1}{T}\right),$$

and

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(y_{it-1}-\frac{1}{t-1}\sum_{s=1}^{t-1}y_{is}\right)\Delta z_{it-j}=O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Finally we have

$$\sqrt{NT}\left(\hat{\rho}^p_{\text{RMA}}-\rho\right)=\sqrt{NT}\frac{C_{1,NT}}{D_{NT}}+O_p\left(\log T\sqrt{\frac{N}{T}}\right)+O_p\left(\frac{1}{T}\right)$$

Hence as $N,T\to\infty$ but , $\log T\sqrt{\frac{N}{T}}\to\zeta$, we have

$$\sqrt{NT}\left(\hat{\rho}^p_{\text{RMA}}-\rho-B(\rho,T)\right)\to^d \quad \mathcal{N}\left(0,1-\rho^2\right) \qquad \text{(A11)}$$

or

$$\sqrt{NT}\left(\hat{\rho}^{p}_{\text{RMA}}-\rho\right)\to^{d}\quad \mathcal{N}\left(\zeta B\left(\rho,T\right),1-\rho^{2}\right).$$

Proof of Proposition 3: For LSDV, begin by representing the estimator as

$$\hat{\rho}_{\text{LSDV}}-\rho=\frac{A^{C}_{NT}}{B^{C}_{NT}},$$

where

$$\begin{aligned} A^{C}_{NT} &\equiv \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1}-T^{-1}\sum_{t=1}^{T}y_{it-1})(u_{it}-T^{-1}\sum_{t=1}^{T}u_{it}), \\ B^{C}_{NT} &\equiv \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1}-T^{-1}\sum_{t=1}^{T}y_{it-1})^{2}. \end{aligned}$$

In the one factor ($K=1$) case, the latent model representation is

$$y_{it}=\mu_{i}+z_{it},\ z_{it}=\rho z_{it-1}+u_{it},\quad u_{it}=\delta_{i}\theta_{t}+\varepsilon_{it}, \tag{A12}$$

with

$$z_{it}=\delta_{i}\sum_{j=0}^{\infty}\rho^{j}\theta_{t-j}+\sum_{j=0}^{\infty}\rho^{j}\varepsilon_{it-j}=\delta_{i}F_{t}+x_{it}. \tag{A13}$$

Since $y_{it}-T^{-1}\sum_{t=1}^{T}y_{it}=x_{it}-T^{-1}\sum_{i=1}^{N}x_{it}$, we have

$$\begin{aligned} & \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1}-T^{-1}\sum_{t=1}^{T}y_{it-1})(u_{it}-T^{-1}\sum_{t=1}^{T}u_{it}) \\ = \ & \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}z_{it-1}u_{it}-\frac{1}{T}\sum_{t=1}^{T}z_{it-1}\sum_{t=1}^{T}u_{it}\right] \\ = \ & \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(\delta_{i}F_{t-1}+x_{it-1}\right)\left(\delta_{i}\theta_{t}+\varepsilon_{it}\right)-\frac{1}{T}\sum_{t=1}^{T}\left(\delta_{i}F_{t-1}+x_{it-1}\right)\sum_{t=1}^{T}\left(\delta_{i}\theta_{t}+\varepsilon_{it}\right)\right] \\ = \ & -\sigma^{2}A_{\text{LSDV}}\left(\rho,T\right)+\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_{i}^{2}\left[\sum_{t=1}^{T}F_{t-1}\theta_{t}-\frac{1}{T}\sum_{t=1}^{T}F_{t-1}\sum_{s=1}^{T}\theta_{s}\right], \end{aligned} \tag{A14}$$

where

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

Let $\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\delta_{i}^{2}\equiv m_{\delta}^{2}$, then

$$\text{plim}_{N\to\infty}\ \frac{1}{N}A^{C}_{NT}=-\sigma^{2}A_{\text{LSDV}}\left(\rho,T\right)+m_{\delta}^{2}\sum_{t=1}^{T}\left(F_{t-1}-T^{-1}\sum_{t=1}^{T}F_{t-1}\right)\left(\theta_{t}-\bar{\theta}\right). \tag{A15}$$

Dealing with the denominator in a similar manner, we get

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

Note that

$$\sigma^2 B\left(\rho, T\right) = \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})^2 - \frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})\right)^2\right] = \sigma^2\frac{1}{1-\rho^2}T + O\left(1\right). \tag{A16}$$

Combining the two results gives

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

Let $T\to\infty$ to have

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

so that

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

It follows that

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\sum_{t=1}^{T}\left(F_{t-1} - T^{-1}\sum_{t=1}^{T}F_{t-1}\right)^2 = Tm_\delta^2\left[\frac{\sigma_\theta^2}{1-\rho^2} + O_p\left(T^{-1/2}\right)\right],\ \text{as } T\to\infty. \tag{A17}$$

Taking the limit as $N\to\infty$ followed by an expansion as $T\to\infty$ gives

$$\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^{C}} = \frac{-\sigma^2 A_{\text{LSDV}}\left(\rho, T\right)\left[\sum_{t=1}^{T}\left(F_{t-1} - T^{-1}\sum_{t=1}^{T}F_{t-1}\right)^2\right]^{-1} + m_\delta^2\left(\hat{\rho} - \rho\right)}{\sigma^2 B_{\text{LSDV}}\left(\rho, T\right)\left[\sum_{t=1}^{T}\left(F_{t-1} - T^{-1}\sum_{t=1}^{T}F_{t-1}\right)^2\right]^{-1} + m_\delta^2},$$

where

$$\left(\hat{\rho} - \rho\right) = \frac{\sum_{t=1}^{T}\left(F_{t-1} - T^{-1}\sum_{t=1}^{T}F_{t-1}\right)\left(\theta_t - \bar{\theta}\right)}{\sum_{t=1}^{T}\left(F_{t-1} - T^{-1}\sum_{t=1}^{T}F_{t-1}\right)^2}.$$

This implies

$$\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^{C}} = -\frac{\sigma^2 A_{\text{LSDV}}(\rho,T)\,T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}+O_p\left(T^{-1/2}\right)\right]^{-1}-m_\delta^2(\hat{\rho}-\rho)}{\sigma^2 B_{\text{LSDV}}(\rho,T)\,T^{-1}\left[\frac{\sigma_\theta^2}{1-\rho^2}+O_p\left(T^{-1/2}\right)\right]^{-1}+m_\delta^2}$$

$$= -\frac{\sigma^2 A_{\text{LSDV}}(\rho,T)\,T^{-1}\left[\frac{1}{1-\rho^2}+O_p\left(T^{-1/2}\right)\right]^{-1}-\sigma_\theta^2 m_\delta^2(\hat{\rho}-\rho)}{\sigma^2\left[\frac{T}{1-\rho^2}+O(1)\right]T^{-1}\left[\frac{1}{1-\rho^2}+O_p\left(T^{-1/2}\right)\right]^{-1}+\sigma_\theta^2 m_\delta^2}$$

$$= -\frac{\sigma^2 A_{\text{LSDV}}(\rho,T)\,T^{-1}\left[\frac{1}{1-\rho^2}+O_p\left(T^{-1/2}\right)\right]^{-1}-\sigma_\theta^2 m_\delta^2(\hat{\rho}-\rho)}{\sigma^2\left[1+O\left(T^{-1}\right)\right]\left[1+O_p\left(T^{-1/2}\right)\right]^{-1}+\sigma_\theta^2 m_\delta^2}.$$

Noting that

$$\left[1+O_p\left(T^{-\gamma}\right)\right]^{-1} = \frac{1}{1+O_p\left(T^{-\gamma}\right)} = 1-O_p\left(T^{-\gamma}\right),$$

we have

$$\frac{\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^{C}} = -\frac{\sigma^2 A_{\text{LSDV}}(\rho,T)\,T^{-1}\left[\frac{1}{1-\rho^2}+O_p\left(T^{-1/2}\right)\right]^{-1}-\sigma_\theta^2 m_\delta^2(\hat{\rho}-\rho)}{\sigma^2\left[1+O\left(T^{-1}\right)\right]\left[1+O_p\left(T^{-1/2}\right)\right]^{-1}+\sigma_\theta^2 m_\delta^2}$$

$$= -\frac{\sigma^2\frac{1+\rho}{T}-\sigma_\theta^2 m_\delta^2(\hat{\rho}-\rho)+O_p\left(T^{-3/2}\right)}{\sigma^2+\sigma_\theta^2 m_\delta^2}$$

$$= (1-\eta)\frac{1+\rho}{T}+\eta(\hat{\rho}-\rho)+o_p\left(T^{-1}\right).$$

Next, for RMA-LS, define

$$\frac{\text{plim}_{N\to\infty}\frac{1}{N}C_{NT}^{C}}{\text{plim}_{N\to\infty}\frac{1}{N}D_{NT}^{C}} = \frac{\sigma^2 C(\rho,T)+m_\delta^2\sum_{t=1}^{T}\left(F_{t-1}-(t-1)^{-1}\sum_{s=1}^{t-1}F_s\right)\left[-(1-\rho)(t-1)^{-1}\sum_{s=1}^{t-1}F_s+\theta_t\right]}{\sigma^2 D(\rho,T)+m_\delta^2\left[\sum_{t=1}^{T}\left(F_{t-1}-(t-1)^{-1}\sum_{s=1}^{t-1}F_s\right)^2\right]}.$$

Noting that

$$D(\rho,T) = \frac{T}{1-\rho^2}+O(1),$$

and as $T\to\infty$, without loss of generality, we have

$$T^{-1}\sum_{t=1}^{T}\left(F_{t-1}-(t-1)^{-1}\sum_{s=1}^{t-1}F_s\right)^2 = \frac{\sigma_\theta^2}{1-\rho^2}+O_p\left(T^{-1}\right).$$

It follows that

$$\begin{aligned}
\frac{\operatorname{plim}_{N\to\infty}\frac{1}{N}C_{NT}^{C}}{\operatorname{plim}_{N\to\infty}\frac{1}{N}D_{NT}^{C}} &= (1-\eta)\,\frac{\rho\ln T}{T}+\eta\,(\hat{\rho}_{\mathrm{R}}-\rho)+O_p\left(T^{-3/2}\right) \\
&= (1-\eta)\,E\,(\hat{\rho}_{\text{RMA-LS}}-\rho)+\eta\,(\hat{\rho}_{\mathrm{R}}-\rho)+O_p\left(T^{-3/2}\right).
\end{aligned}$$

Table 1: Comparison of Alternative Estimators in Panel AR(1) case

			Bias						MSE ×100					
T_0	N	ρ	AB1	AB2	AB3	LSDV	HK	RMA	AB1	AB2	AB3	LSDV	HK	RMA
5	50	0.3	0.11	0.70	0.70	-0.28	-0.11	0.03	1.70	48.62	48.98	8.08	1.71	0.97
5	50	0.5	0.03	0.50	0.50	-0.33	-0.14	0.03	0.62	24.54	24.97	11.63	2.58	0.94
5	50	0.9	-0.01	0.05	0.09	-0.47	-0.23	0.01	0.26	0.39	0.92	22.33	5.89	0.75
5	100	0.3	0.11	0.70	0.70	-0.27	-0.10	0.03	1.49	48.64	48.98	7.78	1.38	0.54
5	100	0.5	0.04	0.50	0.50	-0.33	-0.14	0.03	0.40	24.56	24.98	11.26	2.20	0.54
5	100	0.9	0.00	0.05	0.09	-0.47	-0.23	0.01	0.13	0.34	0.91	21.91	5.46	0.39
5	200	0.3	0.11	0.70	0.70	-0.27	-0.10	0.03	1.39	48.64	48.98	7.67	1.23	0.30
5	200	0.5	0.04	0.50	0.50	-0.33	-0.14	0.03	0.29	24.57	24.97	11.12	2.03	0.32
5	200	0.9	0.00	0.05	0.09	-0.46	-0.22	0.01	0.06	0.31	0.90	21.67	5.22	0.19
10	50	0.3	0.05	0.70	0.70	-0.14	-0.03	0.03	0.55	48.57	48.97	2.04	0.33	0.39
10	50	0.5	0.01	0.49	0.50	-0.16	-0.04	0.04	0.27	24.49	24.97	2.88	0.42	0.40
10	50	0.9	-0.01	0.04	0.09	-0.25	-0.10	0.01	0.17	0.29	0.90	6.21	1.11	0.19
10	100	0.3	0.06	0.70	0.70	-0.14	-0.03	0.03	0.44	48.61	48.98	1.94	0.21	0.23
10	100	0.5	0.02	0.50	0.50	-0.16	-0.04	0.04	0.15	24.53	24.97	2.76	0.29	0.26
10	100	0.9	0.00	0.05	0.09	-0.24	-0.09	0.01	0.08	0.26	0.90	6.06	0.98	0.10
10	200	0.3	0.06	0.70	0.70	-0.13	-0.03	0.03	0.39	48.63	48.98	1.87	0.14	0.16
10	200	0.5	0.02	0.50	0.50	-0.16	-0.04	0.04	0.09	24.55	24.97	2.69	0.23	0.20
10	200	0.9	0.00	0.05	0.09	-0.24	-0.09	0.02	0.04	0.25	0.90	6.00	0.92	0.07
20	50	0.3	0.02	0.70	0.70	-0.07	-0.01	0.02	0.19	48.51	48.97	0.54	0.11	0.17
20	50	0.5	0.00	0.49	0.50	-0.08	-0.01	0.03	0.11	24.39	24.97	0.72	0.11	0.20
20	50	0.9	-0.01	0.03	0.09	-0.12	-0.04	0.02	0.08	0.14	0.87	1.53	0.19	0.07
20	100	0.3	0.03	0.70	0.70	-0.07	-0.01	0.02	0.13	48.56	48.97	0.49	0.06	0.11
20	100	0.5	0.01	0.49	0.50	-0.08	-0.01	0.03	0.06	24.46	24.96	0.67	0.06	0.14
20	100	0.9	0.00	0.03	0.09	-0.12	-0.04	0.02	0.04	0.13	0.88	1.48	0.16	0.05
20	200	0.3	0.03	0.70	0.70	-0.07	-0.01	0.02	0.10	48.61	48.98	0.47	0.03	0.08
20	200	0.5	0.01	0.49	0.50	-0.08	-0.01	0.03	0.03	24.50	24.97	0.64	0.04	0.12
20	200	0.9	0.00	0.03	0.09	-0.12	-0.04	0.02	0.02	0.12	0.88	1.46	0.14	0.04

Notes: AB1, AB2, and AB3, respectively represent the Arellano and Bover (1995) estimators using $\sigma_\mu = 1, 5$ and 10. Entries are obtained from 10,000 replications. DGP is

$$\begin{aligned} y_{it} &= \mu_i + z_{it}, \\ z_{it} &= \rho z_{it-1} + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\mu_i \overset{iid}{\sim} \mathcal{N}\left(1, \sigma_\mu^2\right)$.

Table 2: Comparison of LSDV and RMA in Panel AR(2) case

				Bias				Variance × 100				MSE × 100			
				$\hat{\rho}^p_{\mathrm{RMA}}$		$\hat{\rho}_2$		$\hat{\rho}^p_{\mathrm{RMA}}$		$\hat{\rho}_2$		$\hat{\rho}^p_{\mathrm{RMA}}$		$\hat{\rho}_2$	
T_0	N	ρ	ρ_2	LSDV	RMA	LSDV	RMA	LSDV	RMA	LSDV	RMA	LSDV	RMA	LSDV	RMA
5	50	0.9	0.2	-0.62	-0.20	-0.19	0.02	0.92	1.82	0.53	0.84	39.70	5.99	4.07	0.88
5	100	0.9	0.2	-0.62	-0.20	-0.19	0.02	0.47	0.93	0.28	0.44	38.77	4.92	3.78	0.49
5	200	0.9	0.2	-0.62	-0.20	-0.19	0.02	0.23	0.46	0.13	0.21	38.40	4.37	3.62	0.27
10	50	0.9	0.2	-0.32	-0.07	-0.11	0.02	0.28	0.37	0.23	0.29	10.82	0.79	1.46	0.32
10	100	0.9	0.2	-0.32	-0.06	-0.11	0.02	0.14	0.19	0.12	0.15	10.46	0.56	1.31	0.19
10	200	0.9	0.2	-0.32	-0.06	-0.11	0.02	0.07	0.09	0.06	0.07	10.40	0.46	1.26	0.12
20	50	0.9	0.2	-0.16	-0.01	-0.06	0.01	0.08	0.09	0.10	0.11	2.63	0.10	0.48	0.13
20	100	0.9	0.2	-0.16	-0.01	-0.06	0.01	0.04	0.04	0.05	0.06	2.54	0.05	0.42	0.08
20	200	0.9	0.2	-0.16	-0.01	-0.06	0.01	0.02	0.02	0.03	0.03	2.50	0.03	0.39	0.05
5	50	0.5	0.2	-0.58	-0.31	-0.24	-0.11	1.15	2.23	0.50	0.73	34.50	11.70	6.31	1.87
5	100	0.5	0.2	-0.57	-0.30	-0.24	-0.10	0.55	1.08	0.25	0.36	33.31	10.16	5.95	1.43
5	200	0.5	0.2	-0.57	-0.30	-0.24	-0.10	0.28	0.54	0.12	0.18	32.85	9.47	5.77	1.22
10	50	0.5	0.2	-0.28	-0.11	-0.13	-0.04	0.39	0.58	0.21	0.25	8.28	1.69	1.84	0.41
10	100	0.5	0.2	-0.28	-0.10	-0.13	-0.04	0.20	0.30	0.10	0.13	7.92	1.32	1.70	0.27
10	200	0.5	0.2	-0.28	-0.10	-0.13	-0.04	0.10	0.15	0.05	0.06	7.80	1.16	1.64	0.20
20	50	0.5	0.2	-0.13	-0.03	-0.06	-0.01	0.16	0.21	0.10	0.11	1.94	0.28	0.51	0.12
20	100	0.5	0.2	-0.13	-0.03	-0.06	-0.01	0.08	0.10	0.05	0.06	1.82	0.17	0.45	0.07
20	200	0.5	0.2	-0.13	-0.03	-0.06	-0.01	0.04	0.05	0.03	0.03	1.77	0.12	0.42	0.04

Notes: DGP is

$$\begin{aligned} y_{it} &= \mu_i + z_{it}, \\ z_{it} &= \rho_1 z_{it-1} + \rho_2 z_{it-2} + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} N(0,1)$, $\mu_i \overset{iid}{\sim} N\left(0, \sigma_\mu^2\right)$.

Table 3: Comparison of LSDV and RMA with Exogeneous Variable.

			Bias		Variance × 100		MSE × 100		Rej. of t-test	
			AR(1) case							
T_0	N	ρ	LSDV	RMA	LSDV	RMA	LSDV	RMA	LSDV	RMA
5	50	0.9	-0.12	0.00	0.52	0.51	1.91	0.51	0.47	0.08
5	100	0.9	-0.12	0.00	0.26	0.25	1.65	0.25	0.72	0.08
5	200	0.9	-0.12	0.00	0.13	0.13	1.50	0.13	0.94	0.08
10	50	0.9	-0.05	0.00	0.23	0.22	0.48	0.22	0.22	0.06
10	100	0.9	-0.05	0.01	0.12	0.11	0.36	0.12	0.36	0.07
10	200	0.9	-0.05	0.01	0.06	0.06	0.30	0.06	0.60	0.07
20	50	0.9	-0.02	0.00	0.11	0.11	0.14	0.11	0.10	0.06
20	100	0.9	-0.02	0.00	0.05	0.05	0.09	0.05	0.14	0.06
20	200	0.9	-0.02	0.00	0.03	0.03	0.06	0.03	0.22	0.07
			AR(2) case: $\rho_2 = 0.2$							
T_0	N	ρ	LSDV	RMA	LSDV	RMA	LSDV	RMA	LSDV	RMA
5	50	0.9	-0.13	-0.05	0.52	0.49	2.11	0.71	0.51	0.16
5	100	0.9	-0.12	-0.04	0.26	0.25	1.82	0.45	0.77	0.22
5	200	0.9	-0.13	-0.05	0.13	0.12	1.70	0.33	0.96	0.33
10	50	0.9	-0.06	-0.01	0.24	0.23	0.54	0.24	0.25	0.08
10	100	0.9	-0.05	-0.01	0.11	0.11	0.41	0.12	0.41	0.07
10	200	0.9	-0.05	-0.01	0.06	0.06	0.36	0.07	0.68	0.09
20	50	0.9	-0.02	0.00	0.11	0.11	0.15	0.11	0.11	0.05
20	100	0.9	-0.02	0.00	0.05	0.05	0.10	0.05	0.17	0.06
20	200	0.9	-0.02	0.00	0.03	0.03	0.07	0.03	0.28	0.06

Notes: DGP is

$$\begin{aligned} y_{it} &= \mu_i + z_{it}, \\ z_{it} &= \rho_1 z_{it-1} + \rho_2 z_{it-2} + \gamma q_{it} + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} N(0,1)$, $\mu_i \overset{iid}{\sim} N(0,1)$, $q_{it} \overset{iid}{\sim} N(0,1)$ and $\gamma = 1$.

Table 4: Comparison of Alternative Estimators in Panel AR(1) under Cross Sectional Dependence

			Bias						MSE ×100					
T_0	N	ρ	AB1	AB2	AB3	LSDV	HK	RMA	AB1	AB2	AB3	LSDV	HK	RMA
5	50	0.3	-0.04	0.69	0.70	-0.30	-0.13	-0.01	9.71	47.82	48.93	15.84	11.20	14.63
5	50	0.5	-0.09	0.49	0.50	-0.37	-0.18	-0.02	10.46	23.60	24.92	20.75	13.04	14.00
5	50	0.9	-0.10	-0.02	0.08	-0.52	-0.29	-0.06	5.83	2.68	0.87	34.28	18.39	12.58
5	100	0.3	-0.03	0.69	0.70	-0.29	-0.13	0.00	9.44	47.90	48.94	15.22	10.53	13.72
5	100	0.5	-0.09	0.49	0.50	-0.36	-0.17	-0.01	10.04	23.69	24.93	20.03	12.29	13.12
5	100	0.9	-0.09	-0.02	0.08	-0.52	-0.29	-0.06	5.58	2.51	0.86	33.84	17.88	12.15
5	200	0.3	-0.02	0.69	0.70	-0.29	-0.12	0.00	9.22	47.94	48.94	14.91	10.36	13.46
5	200	0.5	-0.08	0.49	0.50	-0.36	-0.17	-0.01	9.83	23.74	24.92	19.50	11.93	12.76
5	200	0.9	-0.09	-0.01	0.08	-0.51	-0.28	-0.06	5.37	2.42	0.85	33.18	17.30	11.55
10	50	0.3	-0.04	0.69	0.70	-0.16	-0.05	0.00	4.99	47.66	48.92	6.19	4.76	5.76
10	50	0.5	-0.08	0.48	0.50	-0.19	-0.08	0.00	5.20	23.31	24.91	7.36	4.85	5.05
10	50	0.9	-0.11	-0.06	0.08	-0.29	-0.14	-0.04	4.34	2.34	0.79	11.19	5.36	3.25
10	100	0.3	-0.04	0.69	0.70	-0.15	-0.05	0.01	4.76	47.78	48.93	5.92	4.52	5.53
10	100	0.5	-0.07	0.48	0.50	-0.19	-0.07	0.00	4.91	23.46	24.92	7.02	4.53	4.76
10	100	0.9	-0.11	-0.05	0.08	-0.29	-0.14	-0.03	4.08	2.14	0.79	10.79	4.99	2.91
10	200	0.3	-0.04	0.69	0.70	-0.15	-0.05	0.00	4.61	47.85	48.93	5.83	4.34	5.28
10	200	0.5	-0.08	0.48	0.50	-0.19	-0.07	0.00	4.76	23.53	24.92	6.95	4.37	4.57
10	200	0.9	-0.11	-0.05	0.08	-0.29	-0.14	-0.04	3.92	2.00	0.77	10.84	4.93	2.88
20	50	0.3	-0.03	0.69	0.70	-0.08	-0.02	0.01	2.49	47.24	48.90	2.63	2.24	2.57
20	50	0.5	-0.05	0.47	0.50	-0.10	-0.03	0.01	2.37	22.60	24.88	2.74	2.05	2.14
20	50	0.9	-0.09	-0.07	0.07	-0.15	-0.07	-0.01	2.14	1.51	0.61	3.21	1.50	0.92
20	100	0.3	-0.02	0.69	0.70	-0.08	-0.02	0.01	2.35	47.46	48.92	2.49	2.12	2.46
20	100	0.5	-0.04	0.48	0.50	-0.10	-0.03	0.01	2.23	22.87	24.90	2.62	1.94	2.05
20	100	0.9	-0.09	-0.06	0.07	-0.15	-0.07	-0.01	2.06	1.42	0.62	3.14	1.43	0.87
20	200	0.3	-0.03	0.69	0.70	-0.08	-0.02	0.01	2.25	47.60	48.92	2.42	2.01	2.32
20	200	0.5	-0.04	0.48	0.50	-0.10	-0.03	0.01	2.14	23.03	24.91	2.55	1.85	1.94
20	200	0.9	-0.09	-0.06	0.07	-0.15	-0.07	-0.01	1.98	1.36	0.63	3.08	1.37	0.82

Notes: See footnotes in Table 1. DGP is

$$\begin{aligned} y_{it} &= a_i + \rho y_{it-1} + u_{it}, \\ u_{it} &= \delta_i \theta_t + \epsilon_{it}, \end{aligned}$$

where $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\delta_i \overset{iid}{\sim} \mathcal{N}(1,1)$ and $\theta_t \overset{iid}{\sim} \mathcal{N}(0,1)$.