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 Texas at Dallas

May 25, 2009

Abstract

The within-group estimator (same as the least squares dummy variable estimator) of the dominant root in dynamic panel regression is known to be biased downwards. This paper studies recursive mean adjustment (RMA) as a strategy to reduce this bias for AR(p) processes that may exhibit cross-sectional dependence. Asymptotic properties for $N, T \to \infty$ jointly are developed. When $\left(\log^2 T\right)(N/T) \to \zeta$ where ζ is a non-zero constant, the estimator exhibits nearly negligible inconsistency. Simulation experiments demonstrate that the RMA estimator performs well in terms of reducing bias, variance and mean square error both when error terms are cross-sectionally independent and when they are not. RMA dominates comparable estimators when T is small and/or when the underlying process is persistent.

Keywords: Recursive Mean Adjustment, Fixed Effects, Cross-sectional Dependence.

JEL Classification Numbers: C33

*Corresponding author. Email: d.sul@utdallas.edu. We are grateful to two anonymous referees for constructive comments that helped to improve the paper. We also thank Peter C.B. Phillips for suggesting the topic and for helpful comments on an earlier draft and Chirok Han for useful and thoughtful comments.

1 Introduction

In small T samples, the least squares estimator of the autocorrelation coefficient $\rho \in [0,1)$, of a stationary but persistent first-order autoregressive time series $y_t = \alpha + \rho y_{t-1} + \epsilon_t$, $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$, is biased downwards when the constant is also estimated.[1] To illustrate the source of the bias, suppose one runs least squares without a constant but on deviations from the sample mean. That is, regress $(y_t - \bar{y})$ on $(y_{t-1} - \bar{y})$. The regression error ϵ_t is correlated with current and future values of y_t, and these current and future values of y_t are embedded in the $\bar{y}$ component of the explanatory variable. It follows that the error term is correlated with the regressor $(y_{t-1} - \bar{y})$. In the panel data context, this small T bias is also present in fixed-effects estimators. For fixed T as $N \to \infty$, Nickell (1981) shows that the within group (WG, which is equivalent to the least squares dummy variable method) estimator for the dynamic panel regression model is substantially inconsistent.

This paper studies and applies the recursive-mean adjustment (RMA, henceforth) technique to reduce bias in the estimation of linear dynamic panel data models when the dominant root is homogeneous across individuals. The paper builds on work by So and Shin (1999a) who show that the RMA strategy is useful in reducing bias in univariate regression and in the context of unit-root testing [So and Shin (1999b, 2002)]. In RMA, the constant is dealt with by adjusting observations with the common recursive mean $\bar{y}_{t-1} = (t-1)^{-1} \sum_{s=1}^{t-1} y_s$, instead of the sample mean. As a result, the adjusted regressor $(y_{t-1} - \bar{y}_{t-1})$ is orthogonal to the regression error ϵ_t because the recursive mean does not contain any future values of y_t. The RMA strategy, however, does not completely eliminate the bias because the error term after RMA adjustment contains $-(1-\rho)\bar{y}_{t-1}$ which is correlated with the adjusted regressor. But when $\rho = 1$ in the univariate time series context, the RMA method completely eliminates the small sample bias which explains why it has been used in unit-root testing by Taylor (2002), Phillips, Park and Chang (2004), and Sul (2008).

We find that the RMA estimator effectively alleviates the bias problem in a finite sample compared to the WG estimator. Similar results are obtained in a more general AR(p) model using a simple two-step approach to reducing the bias. The performance of the RMA method is notable particularly when ρ is near unity in which alternative bias reduction methods based on GMM/IV estimators may not properly work due to the weak moment conditions [Brundell and Bond (1998)]. The context in which this study is conducted is relevant to the empirical studies based on a data set with larger N and small T, such as firm level analyses on the estimation of dynamic labor demand equation [e.g. Blundell and Bond (1998)] or production function [e.g., Blundell, Bond and Windmeijer (2000)] or dynamics of macroeconomic variables at the regional level [e.g. Campbell and Lapham (2004), Bun and Carree (2005)].

The remainder of the paper is organized as follows. In the next section, we begin by discussing the

[1] If there is no constant term, the bias is $-2\rho/T$ which is trivial for even moderate T. 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)].

asymptotic properties of the panel RMA estimator for a general AR(p) model under cross-sectionally independent observations. This section also discusses extensions of the framework to environments with incidental trends, local-to-unity observations, and issues involved when the dominant root is heterogeneous across individuals. Section 3 considers an environment where the observations are cross-sectionally correlated and can be represented by a common factor structure. As in Alvarez and Arellano (2003), Bai (2003), and Hahn and Kuersteiner (2002), our asymptotic analysis here is based on large T and large N. Section 4 reports results 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. Section 5 concludes. An appendix contains proofs and details of many arguments made in the text.

Before proceeding, a few words on the notations might be helpful. Throughout the paper, 'T' denotes the span of time-series and 'N'' is written as the cross-section dimension of panel, and the symbols '$\xrightarrow{d}$' and '$\text{plim}_{N\to\infty}$' respectively indicate the weak convergence in distribution and probability limit as $N \to \infty$. Also, an estimator is said to be 'inconsistent' if the probability limit of an estimator is not equal to its true value as $N \to \infty$ with fixed T. For example, the WG estimator for the first autocorrelation coefficient is inconsistent as $N \to \infty$ with fixed T, whereas it becomes consistent as $T \to \infty$ regardless of N. An estimator is 'asymptotically biased' when we compare the magnitude of the inconsistency of an estimator for small T with that for large T.

2 Asymptotic properties when the observations are cross-sectionally independent

The data are assumed to be generated by the following latent model.

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

$$y_{it} = \mu_i + \beta w_{it} + z_{it}, \tag{1}$$

$$z_{it} = \sum_{j=1}^{p} \rho_j z_{it-j} + \epsilon_{it}, \tag{2}$$

where w_{it} is strictly exogenous, $E w_{it} z_{is} = 0$ for all t and s, and all roots of lag polynomial of z_{it} lie outside the unit circle, $\mu_i \overset{iid}{\sim} \left(\mu, \sigma_\mu^2\right)$, $z_{i1} \overset{iid}{\sim} \left(0, \sigma_i^2 / \left(1 - \rho^2\right)\right)$ for $\rho = \sum_{j=1}^{p} \rho_j$, and $\epsilon_{it} \overset{iid}{\sim} \left(0, \sigma_i^2\right)$ is independent of μ_i and y_{i1}, and has finite moments up to the fourth order.[2]

[2] Assumption 1 admits the level of an exogenous regressor w_{it}. In their study on natural gas demand, Balestra and Nerlove (1966) distinguish between level and quasi-difference exogenous regressors. They argue that per capita income affects gas demand in levels but relative price affects gas demand in the quasi-difference. In their regression model, y_{it} is the quantity of natural gas, w_{it} is the per capita income, and q_{it} is the relative price of natural gas. The

This data generating process (DGP) assumes that the initial condition is stationary, as in Blundell and Bond (1998).[3] The latent model (1)-(2) is observationally equivalent to the dynamic panel regression representation

$$y_{it} = a_i + \rho y_{it-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{it-j} + \beta w_{it} + \sum_{l=1}^{p} \kappa_l w_{it-l} + \epsilon_{it}. \tag{3}$$

where $a_i = (1-\rho)\mu_i$, $\rho = \sum_{j=1}^{p} \rho_j$, $\phi_j = \rho_j + \phi_{j-1}$ for $j = 2, ..., p-1$ with $\phi_1 = -\sum_{k=2}^{p} \rho_k$, and $\kappa_l = -\beta \rho_l$ for $l = 1, ..., p$. Note that so long as w_{it} is strictly exogenous, both the WG estimator for β in (1) - which is the static version - and in (3) - which is the dynamic version - are consistent, but the latter is more efficient than the former for a large T.

The RMA estimator for this model is obtained from the following steps.

Step 1: (RMA Estimator for ρ) Estimate (3) with the WG estimator. Let $\hat{x}_{it} = \sum_{j=1}^{p-1} \hat{\phi}_{j,\mathrm{WG}} \Delta y_{it-j} + \hat{\beta}_{\mathrm{WG}} w_{it} + \sum_{j=1}^{p} \hat{\kappa}_{j,\mathrm{WG}} w_{it-j}$, and $y_{it}^{+} = y_{it} - \hat{x}_{it}$. This allows (3) to be rewritten as

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

where $\epsilon_{it}^{+} = \epsilon_{it} + (x_{it} - \hat{x}_{it})$. Then using pooled least squares, regress

$$\left(y_{it}^{+} - \bar{y}_{it-1}\right) = \rho\left(y_{it-1} - \bar{y}_{it-1}\right) + e_{it} \tag{4}$$

where $\bar{y}_{it-1} = (t-1)^{-1} \sum_{s=1}^{t-1} y_{is}$, $e_{it} = -(1-\rho)\bar{z}_{it-1} + \epsilon_{it} + (x_{it} - \hat{x}_{it})$, and $\bar{z}_{it-1} = (t-1)^{-1} \sum_{s=1}^{t-1} z_{is}$. This gives $\hat{\rho}_{\mathrm{RMA}}^{p}$, the RMA estimator of ρ.

Step 2: (RMA Estimator for ϕ_j and κ_l) For the other coefficients, $\hat{\beta}_{\mathrm{WG}}$ is consistent but $\hat{\phi}_{j,\mathrm{WG}}$ and $\hat{\kappa}_{l,\mathrm{WG}}$ are not. The inconsistency of $\hat{\phi}_{j,\mathrm{WG}}$ and $\hat{\kappa}_{l,\mathrm{WG}}$, however, can be reduced by running an additional WG regression,

$$\left(y_{it} - \hat{\rho}_{\mathrm{RMA}}^{p} y_{it-1} - \hat{\beta}_{\mathrm{WG}} w_{it}\right) = a_i + \sum_{j=1}^{p-1} \phi_j \Delta y_{it-j} + \sum_{l=1}^{p} \kappa_l w_{it-l} + e_{it}^{p}, \tag{5}$$

quasi-difference exogenous regressor can be introduced by letting $\epsilon_{it} = \gamma q_{it} + \epsilon_{it}^{o}$ in (3). According to Phillips and Sul (2007), the WG estimator for β in (1) is consistent, while the WG estimator for γ is inconsistent for fixed T as $N \to \infty$. In this case, we can utilize the simple bias correction method for γ provided by Phillips and Sul (2007) for a moderately large N but small T.

[3] Assumption 1 does not include the case of unit-root because RMA estimator becomes consistent when $\rho = 1$. For the unit-root case, however, the initial condition can be set as $z_{i1} = O_p(1)$. See Kiviet (1995) for the impact of a nonstationary initial condition on the inconsistency.

where $e_{it}^{p} = \epsilon_{it} + \left(\rho - \hat{\rho}_{\text{RMA}}^{p}\right) y_{it-1} + \left(\beta - \hat{\beta}_{\text{WG}}\right) w_{it}$. Let us call the resulting estimator $\hat{\phi}_{j,\text{RMA}}^{p}$ and $\hat{\kappa}_{l,\text{RMA}}^{p}$.[4]

We have the following asymptotic properties of the RMA estimators for ρ and ϕ.

Proposition 1: **(Asymptotic Properties of the RMA Estimators)**. *Let the observations be generated by Assumption 1.*

(i) For fixed T, as $N \to \infty$, $\hat{\rho}_{\text{RMA}}^{p}$ and $\hat{\phi}_{j,\text{RMA}}^{p}$ are inconsistent where

$$\text{plim}_{N\to\infty} \left(\hat{\rho}_{\text{RMA}}^{p} - \rho\right) = B(\rho, T) + O\left(T^{-2}\right), \tag{6}$$

$$\text{plim}_{N\to\infty} \left(\hat{\phi}_{j,\text{RMA}}^{p} - \phi_j\right) = \frac{1}{2} B(\rho, T) + O\left(T^{-2}\right), \tag{7}$$

$$B(\rho, T) \equiv -(1-\rho)\frac{C(\rho, T)}{D(\rho, T)} = O\left(\frac{\ln T}{T}\right) > 0,$$

$$C(\rho, T) = \sum_{t=1}^{T-1} t^{-1} \left\{ 2t^{-1} \sum_{h=1}^{t} h\gamma_h^{(z)} - \sum_{h=1}^{t} \gamma_h^{(z)} \right\},$$

$$D(\rho, T) = (T-1)\gamma_0^{(z)} - \sum_{t=1}^{T_1} t^{-1} \{\gamma_0^{(z)} - 2t^{-1} \sum_{h=1}^{t-1} h\gamma_h^{(z)}\},$$

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

(ii) If $\left(\log^2 T\right)(N/T) \to \zeta$ where ζ is a non-zero constant, then *as $T, N \to \infty$,*

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

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

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

The proof is given in Appendix A. As mentioned earlier, the WG estimator for β is consistent and thus its asymptotic properties are omitted from Proposition 1.

[4] In the case of quasi-difference exogenous regressor, we need an additional step to reduce the asymptotic bias for γ. To be specific, let $y_{it}^{++} = y_{it} - \hat{\rho}_{\text{RMA}}^{p} y_{it-1} - \sum_{j=1}^{p-1} \hat{\phi}_{j,\text{RMA}} \Delta y_{it-j} + \hat{\beta}_{\text{WG}} w_{it} + \sum_{l=1}^{p} \hat{\kappa}_{l,\text{RMA}} w_{it-l}$ and running WG on $y_{it}^{++} = a_i + \gamma q_{it} + \epsilon_{it}^{p}$ gives $\hat{\gamma}_{\text{RMA}}$, the RMA estimator for γ. See Case 3 in Section 4 for the relevant Monte Carlo simulation results.

Remark 1 (Exact Inconsistency Formula for AR(1) Case) The explicit formula of the inconsistency for fixed T as $N \to \infty$ is given as

$$B(\rho, T) \equiv \text{plim}_{N\to\infty} \left(\hat{\rho}_{\text{RMA}} - \rho\right) = \frac{\rho \log T + B_{1,T} - (1-\rho) B_{2,T}}{(T-1) - \log T + \frac{2}{1-\rho} B_{1,T} - B_{2,T}} = \frac{\rho \log T}{T} + O\left(T^{-2}\right), \quad (9)$$

where

$$B_{1,T} = \sum_{t=1}^{T-1} \frac{1}{t} \rho^t = O(1), \quad B_{2,T} = 2\rho \sum_{t=1}^{T-1} \frac{1-\rho^t}{t^2 (1-\rho)^2} = O(1).$$

See Appendix A1 for the detailed proof. To evaluate the inconsistency exactly, we plot the inconsistency in Figure 1. A couple of interesting features emerge from the plot. First, the maximum inconsistency is 0.028 which occurs when $T-1 = 13$ and $\rho = 0.46$. Second, as predicted, the inconsistency diminishes as ρ gets closer to unity. However, it is important to note that this small inconsistency would affect the statistical inference when $N > T$.

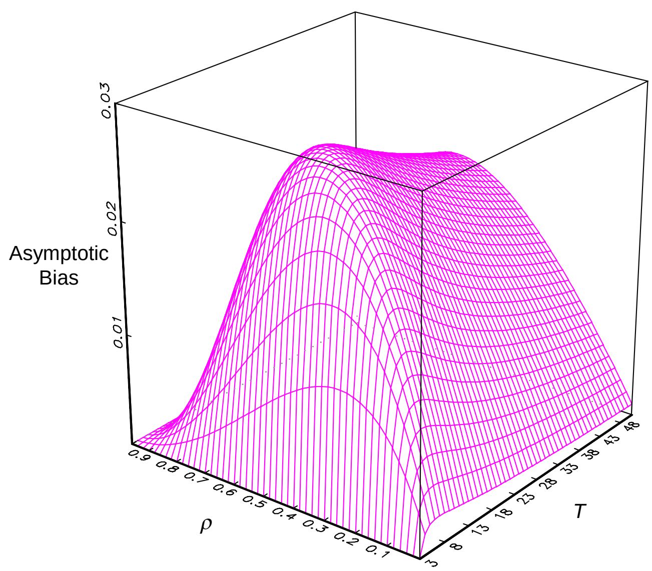


Figure 1: Inconsistency of RMA estimator under AR(1) and independence

Remark 2 (Consistent Cases) There are two special cases when $\hat{\rho}_{\text{RMA}}$ becomes consistent for fixed T with $N \to \infty$. The first case is when $\rho = 1$, which implies $e_{it} = \epsilon_{it}$ in (4) and hence $\text{E}(y_{it-1} e_{it})$ $=\text{E}(\bar{y}_{it-1} e_{it}) = 0$. The second case is when $T-1 = 2$, or $T = 3$, for any $\rho \in [0, 1)$. In this case,

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

$$= -(1-\rho)\frac{\mathrm{E}\left[y_{i2}^2 - y_{i1}^2\right]}{\mathrm{E}\left[(y_{i2}-y_{i1})^2\right]} = 0,$$

because $\mathrm{E}\left(y_{i2}^2\right) = \mathrm{E}\left(y_{i1}^2\right)$ by the covariance stationarity of y_{it}.

Remark 3 (Incidental Linear Trends) When the DGP in (1)-(2) contains an incidental linear trend, the suggested RMA procedure is no longer valid.[5] While the case of incidental trend is not much studied in the literature of dynamic panel regressions, it is an important issue in the literature of panel unit-root testing.[6] In our treatment of the trend, we consider the latent AR(1) model of

$$y_{it} = a_i + b_i t + z_{it},$$
$$z_{it} = \rho z_{it-1} + \epsilon_{it},$$

which is observationally equivalent to

$$y_{it} = \alpha_i + \beta_i t + \rho y_{it-1} + \epsilon_{it},$$

where $\alpha_i = a_i(1-\rho) + b_i\rho$ and $\beta_i = b_i(1-\rho)$.

Here we briefly review Sul's (2008) detrending method before suggesting a new detrending approach to reducing the bias in finite sample. Let $2\bar{y}_{it-1}$ be the common mean adjustment component. Then, $2\bar{y}_{it-1} = 2a_i + b_i t + 2\bar{z}_{it-1}$ so that we have

$$y_{it} - 2\bar{y}_{it-1} = -a_i + \frac{1}{2}b_i + \rho(y_{it-1} - 2\bar{y}_{it-1}) + e_{it}^{\tau}, \tag{10}$$

where $e_{it}^{\tau} = -2(1-\rho)\bar{z}_{it-1} + \epsilon_{it}$. The result is a trendless regression with fixed-effects of $-a_i + \frac{1}{2}b_i$. The RMA estimator for ρ, $\hat{\rho}_{\mathrm{RMA}}^{\tau}$, is obtained from running WG in (10) and its inconsistency is given by

$$\mathrm{plim}_{N\to\infty}\left(\hat{\rho}_{\mathrm{RMA}}^{\tau} - \rho\right) = G(\rho,T) = O\left(T^{-1}\ln T\right), \tag{11}$$

[5] According to the exact inconsistency formula of the WG estimator shown by Phillips and Sul (2007), the first order inconsistency for panel AR(1) under incidental trend case is $-2(1+\rho)T^{-1}$, which is twice as large as the inconsistency under fixed effects, $-(1+\rho)T^{-1}$.

[6] Shin and So (1999a,1999b), for instance, suggest a recursive detrending method. Their method, however, does not completely eliminate the incidental trend components in univariate context as noted by Sul, Phillips and Choi (2005). Taylor (2002) and Phillips, Park and Chang (2004) propose alternative detrending methods that are effective under unit-root case but are substantially upward biased when $\rho < 1$. Sul (2008) also suggests a double recursive mean adjustment method that yields much smaller bias when $\rho < 1$.

where $G(\rho, T)$ is presented in Appendix B2. Although this double recursive demeaning method produces smaller upward bias than the alternative methods adopted in the panel unit-root literature, the bias is still non-negligible when $\rho < 1$.

Alternatively, we suggest the following bias reduction method. First, use the WG estimator to obtain an initial estimate of the trend coefficient $\hat{b}_i$ from the latent model representation, and construct the detrended observations $y_{it}^{\dagger} = y_{it} - \hat{b}_i t$. Next, use pooled least squares to regress $\left(y_{it}^{\dagger} - \bar{y}_{it-1}^{\dagger}\right)$ on $\left(y_{it-1}^{\dagger} - \bar{y}_{it-1}^{\dagger}\right)$, where $\bar{y}_{it-1}^{\dagger} = (t-1)^{-1}\sum_{s=1}^{t-1} y_{is}^{\dagger}$ and the regression error $e_{it}^{\dagger} = -(1-\rho)\bar{z}_{it-1} + \epsilon_{it} - b_i^* \rho - \frac{1}{2} b_i^* (1-\rho) t$ where $b_i^* = \hat{b}_i - b_i$. Call the estimate $\hat{\rho}_{\text{IRD}}$. Notice that the trend is not yet completely removed because the remnants of the trend now sit in the regression error. To deal with this, update the trend coefficients by WG estimation of $y_{it} - \hat{\rho}_{\text{IRD}} y_{it-1} = a_i + \beta_i t + \varepsilon_{it}^{\dagger}$ and call the estimated trend coefficient $\hat{\beta}_{i,\text{WG}}'$. Update the trend estimate $\hat{b}_i' = \hat{\beta}_{i,\text{WG}}' / (1 - \hat{\rho}_{\text{IRD}})$. If $\hat{\rho}_{\text{IRD}} \geq 1$, set $\hat{b}_i' = \hat{b}_i$. Continue until convergence. For $\hat{\rho}_{\text{IRD}}$, we have the weak convergence result

$$\text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{IRD}} - \rho\right) = B(\rho, T) + \frac{1}{2}\frac{(1-2\rho)\left(1-\rho^2\right)}{T} + O\left(T^{-2}\right), \tag{12}$$

where $B(\rho, T)$ is given in Proposition 1. The proof is presented in Appendix B1. In (12), the second term has a faster diminishing speed than that of $B(\rho, T)$, but its magnitude is larger than that of $B(\rho, T)$ even for a moderately large T. This inconsistency, however, is far smaller than that of the WG estimator, as illustrated in Figure 2.

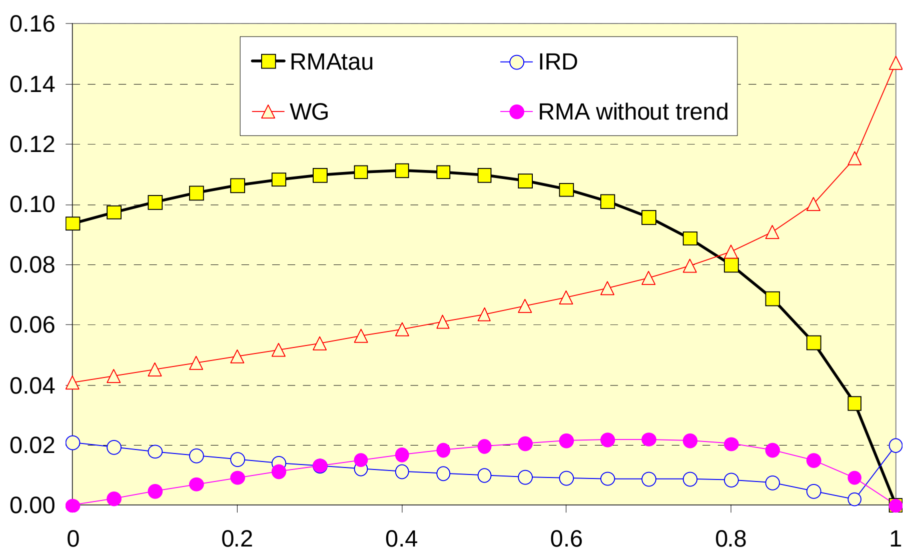


Figure 2: Comparison of the absolute asymptotic biases for incidental linear trend case $(T = 50, N = 1000)$

Figure 2 plots the absolute asymptotic bias of $\hat{\rho}_{\text{IRD}}$, $\hat{\rho}_{\text{RMA}}^{\tau}$ and $\hat{\rho}_{\text{WG}}$ for a large N under incidental linear trends. As mentioned earlier, $\hat{\rho}^{\tau}_{\text{RMA}}$ produces no asymptotic bias when $\rho = 1$, but significantly upward biased when $\rho < 1$. Consequently it is even dominated by $\hat{\rho}_{\text{WG}}$ when $\rho < 0.8$ as displayed in Figure 2. By contrast, the asymptotic bias of $\hat{\rho}_{\text{IRD}}$ is rather modest over the entire range of ρ. Whilst the asymptotic bias of $\hat{\rho}_{\text{IRD}}$ varies slightly with T and ρ, the size of the asymptotic bias is far smaller than those of $\hat{\rho}^{\tau}_{\text{RMA}}$ and $\hat{\rho}_{\text{WG}}$ except when $\rho = 1$. This is what the asymptotic theory in (12) predicts. Note that the asymptotic bias of $\hat{\rho}_{\text{IRD}}$ declines with ρ until it is kinked when ρ is near unity. This kink is due to the truncation of $\hat{\rho}_{\text{IRD}}$ in the iterative procedure. The figure also compares the asymptotic bias of $\hat{\rho}_{\text{IRD}}$ with that of $\hat{\rho}_{\text{RMA}}$ when there is no trend. $\hat{\rho}_{\text{IRD}}$ also outperforms $\hat{\rho}_{\text{RMA}}$ in the no trend case in most regions of ρ.

Remark 4 (Weakly integrated and local-to-unity processes) Since the RMA estimator becomes consistent when $\rho \to 1$ as shown above, one should anticipate that the estimator exhibits only modest inconsistency under local-to-unity. To confirm this guess, let $\rho = 1 - c/T^{\alpha}$ for $0 < \alpha \leq 1$ and some $c > 0$. When $0 < \alpha < 1$, the process is said to be nearly stationary (Giraitis and Phillips, 2006) or weakly integrated (Park, 2003). When $\alpha = 1$, the process is said to be local-to-unity. In either case, we have

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

It can be seen that the inconsistency is of order $O\left(\ln T/T\right)$ so that the first order inconsistency is larger than that of WG estimator. However, as we will show shortly, the overall inconsistency of the RMA estimator is much smaller than that of WG estimator.

Remark 5 (GLS Demeaning Procedure) One might wonder how the generalized least squares (GLS) adjustment proposed by Elliott, Rothenberg and Stock (1996), originally designed for efficient unit-root tests, would compare to the RMA estimator if it were employed to reduce the bias in the local-to-unity environment. To explore this, let us assume

$$\rho = 1 - \frac{c}{T}.$$

The GLS correction requires a quasi-demeaning of the observations using the factor of $1 - 7/T$. Define

$$y^{g}_{it} = \begin{cases} y_{it} - \left(1 - \frac{7}{T}\right) y_{it-1} & \text{if } t > 1 \\ y_{i1} & \text{if } t = 1 \end{cases}, \quad Z_t = \begin{cases} 1 - \frac{7}{T} & \text{if } t > 1 \\ 1 & \text{if } t = 1 \end{cases}.$$

Next, let

$$u_{it} = y_{it} - \frac{\sum_{t=2}^{T} y^{g}_{it} Z_t}{\sum_{t=2}^{T} Z_t^2}.$$

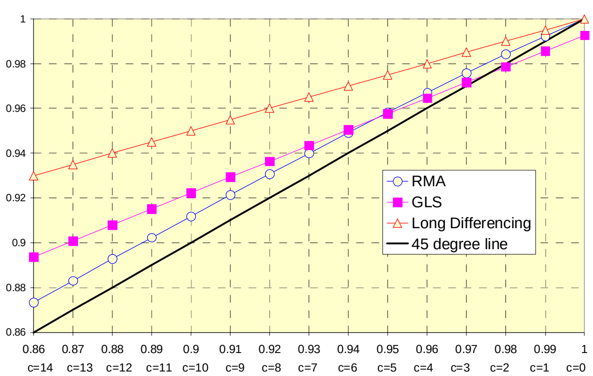
Then the estimator $\hat{\rho}_{\text{GLS}}$ is obtained by running pooled least squares on $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, where

$$\hat{\rho}_{\text{GLS}} = \hat{\rho}_{\text{long}} + O_p\left(T^{-1}\right)$$

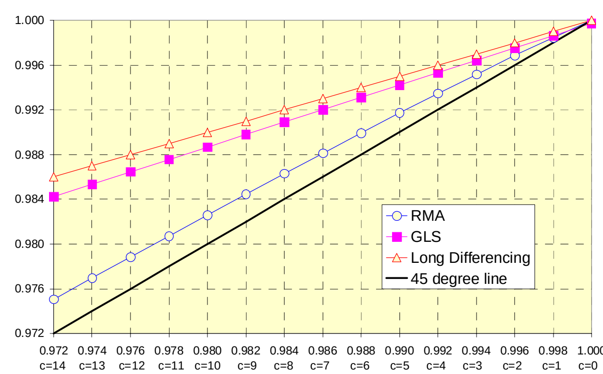
and $\hat{\rho}_{\text{long}}$ is the pooled least-squares estimator from regressing $(y_{it} - y_{i1})$ on $(y_{it-1} - y_{i1})$. For a local parameter value of $c = 7$, the inconsistency in the GLS-corrected estimator becomes

$$\text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{GLS}} - \rho\right) = \frac{35}{T} + O\left(T^{-2}\right). \tag{13}$$

As shown in Appendix C, the second order term in (13) includes a large component such as $49/T^2$, which causes $\hat{\rho}_{\text{GLS}}$ to be inconsistent for moderately T as large as 100. A quick evaluation of the asymptotic bias in $\hat{\rho}_{\text{RMA}}$ in comparison to $\hat{\rho}_{\text{GLS}}$ and $\hat{\rho}_{\text{long}}$ is presented in Figure 3 under the simulation environment of $N = 2000$ and $c = [0, 14]$ for $T \in \{100, 500\}$. As can be seen from Panel A, the inconsistency of $\hat{\rho}_{\text{RMA}}$ is in general smaller than that of $\hat{\rho}_{\text{GLS}}$ except when $1 < c < 5$. Moreover, the dominance of $\hat{\rho}_{\text{RMA}}$ over $\hat{\rho}_{\text{GLS}}$ gets stronger with larger T. As exhibited in Panel B, when T =500, the simulated inconsistency of $\hat{\rho}_{\text{RMA}}$ is much smaller than that of $\hat{\rho}_{\text{GLS}}$ for all c values. In sum, this simulation experiment suggests that the bias reduction provided by RMA is superior to that of GLS estimator when ρ is near unity.[7]



Panel A: $T = 100, N = 2000$



Panel B: $T = 500, N = 2000$

Figure 3: Comparison of Inconsistencies among $\hat{\rho}_{\text{RMA}}$, $\hat{\rho}_{\text{GLS}}$, and $\hat{\rho}_{\text{long}}$

Remark 6 (Heterogeneous Panels) In the current study, the dominant root, ρ, is assumed to be homogeneous across the cross-section. In practice, the homogeneity restriction can be tested using formal inference techniques [e.g. Pesaran and Yamagata (2008)]. If the homogeneity restriction

[7] More results on the simulation experiment are available at http://www3.uta.edu/choi/research.htm.

is rejected, one may consider the pooled mean group RMA estimator. Pesaran and Smith (1995) study the pooled mean group (PMG) estimator for nonstationary panel data, and Pesaran, Shin and Smith (1999) extend the analysis to dynamic panel regressions under cross-sectional independence. Pesaran (2006) also develops the PMG estimator under cross-sectional dependence. In general, the consistency of the PMG estimator requires the sequential limit of $T \to \infty$ first, and $N \to \infty$ next. For fixed T and $N \to \infty$, however, the PMG estimator is inconsistent just like the WG estimator. Since both univariate and pooled RMA estimators are consistent as $T \to \infty$, the asymptotic property of the PMG-RMA estimator can be derived under sequential limits. A more interesting but also more challenging environment is the case for fixed T and $N \to \infty$ in which the asymptotic properties of the PMG-RMA estimator are yet known. We leave this important issue for future research.

3 Asymptotic Properties with Cross-sectionally Correlated Observations

In this section we adapt the RMA estimator to dynamic panel data models with cross-sectionally dependent observations where the dependence arises from a common factor specification. The environment under consideration is again the one in which observations are weakly stationary and are generated by the following assumption.

Assumption 2. *For $i = 1, \dots N$ and $t = 1, \dots T$, the observations $\{y_{it}\}$ are generated by the latent model*

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

$$z_{it} = \sum_{j=1}^{p} \rho_j z_{i,t-j} + u_{it}, \tag{15}$$

$$u_{it} = \sum_{s=1}^{K} \delta_{si}\theta_{st} + \epsilon_{it} = \delta_i'\theta_t + \epsilon_{it}, \tag{16}$$

where the idiosyncratic term, ϵ_{it}, is assumed to be iid $\left(0, \sigma_i^2\right)$ and $\sigma_i^2 < \infty$ for all i. Also, θ_{st} and ϵ_{it} are assumed to be independent of each other. The common factor has $E\|\theta_{st}\|^4 < \infty$, $plim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\theta_{st}^2 = \sigma_{s\theta}^2$ with factor loadings of $\|\delta_{si}\| \le D < \infty$, and $plim_{N\to\infty}\|\delta'\delta/N - M_\delta\| = 0$ with $\delta = (\delta_1, ..., \delta_N)'$ for some $M_\delta > 0$, where $\delta_i = (\delta_{1i,} ..., \delta_{Ki})'$.

This latent model has the observationally equivalent factor representation

$$y_{it} = \mu_i + \delta_i' F_t + z_{it}, \tag{17}$$

where $F_t = (F_{1t}, ..., F_{kt})'$, $F_{st} = \sum_{j=0}^{\infty} \rho^j \theta_{st-j}$ and $z_{it} = \sum_{j=0}^{\infty} \rho^j \epsilon_{it-j}$. Phillips and Sul (2007) show that the $N-$asymptotic bias of the WG estimator becomes 'random' under the cross-sectional dependence. To derive the asymptotic 'random' bias formula under cross-sectional dependence of y_{it}, we first introduce the following notation. Let

$$\hat{\rho}_{\mathrm{F,WG}} = \rho + \frac{\sum_{t=2}^{T} \left(F_{st-1} - \bar{F}_s\right) \left(\theta_{st} - \bar{\theta}_s\right)}{\sum_{t=2}^{T} \left(F_{st-1} - \bar{F}_s\right)^2} \tag{18}$$

be the WG estimator of ρ from the regression of $F_{st} = m_s + \rho F_{st-1} + \theta_{st}$ where $\bar{F}_s = (T-1)^{-1} \sum_{t=1}^{T-1} F_{st-1}$ for $s = 1, ..., K$. According to Phillips and Sul (2007), the WG estimator $\hat{\rho}_{\mathrm{WG|CSD}}$ in (14)-(16) can be decomposed into

$$\hat{\rho}_{\mathrm{WG|CSD}} = (1 - \eta) \hat{\rho}_{\mathrm{WG|CSI}} + \eta \hat{\rho}_{\mathrm{F,WG}} + o_p \left(T^{-1}\right)$$

where $\hat{\rho}_{\mathrm{WG|CSI}}$ represents the WG estimator under cross-sectional independence, $\eta = m_\delta^2 \sigma_\theta^2 \left(\sigma^2 + m_\delta^2 \sigma_\theta^2\right)^{-1}$, $m_\delta^2 = (NK)^{-1} \sum_{i=1}^{N} \sum_{s=1}^{K} \delta_{si}^2$, $\sigma_\theta^2 = (KT)^{-1} \sum_{s=1}^{K} \sum_{t=1}^{T} \theta_{st}^2$, and $\sigma^2 = N^{-1} T^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{it}^2$. Here η can be interpreted as the degree of cross-sectional dependence, such that $\eta = 1$ for perfect cross-sectional dependence and $\eta = 0$ for cross-sectional independence. Since $\hat{\rho}_{\mathrm{F,WG}}$ does not depend on the dimension of cross-section (N), the $N-$asymptotic bias for $\hat{\rho}_{\mathrm{WG|CSD}}$ now depends on the inconsistency of $\hat{\rho}_{\mathrm{F,WG}}$. That is, as shown by Phillips and Sul (2007),

$$\mathrm{plim}_{N \to \infty} (\hat{\rho}_{\mathrm{WG|CSD}} - \rho) = -(1 - \eta) \frac{1 + \rho}{T} + \eta \left(\hat{\rho}_{\mathrm{F,WG}} - \rho\right) + o_p \left(T^{-1}\right),$$

as $N \to \infty$.

Similarly, we can directly apply this approach to the RMA estimator by letting

$$\hat{\rho}_{\mathrm{F,RMA}} = \rho + \frac{\sum_{t=2}^{T} \left(F_{st-1} - \bar{F}_{st-1}\right) \left(\theta_{st} - (1 - \rho) \bar{F}_{st-1}\right)}{\sum_{t=2}^{T} \left(F_{st-1} - \bar{F}_{st-1}\right)^2} \tag{19}$$

be the RMA estimator from $F_{st} - \bar{F}_{st-1} = \rho \left(F_{st-1} - \bar{F}_{st-1}\right) + \varepsilon_{st}$ where $\varepsilon_{st} = \theta_{st} - (1 - \rho) \bar{F}_{st-1}$ and $\bar{F}_{st-1} = (t - 1)^{-1} \sum_{j=1}^{t-1} F_{sj}$. And let $\hat{\rho}_{\mathrm{RMA|CSD}}$ be the RMA estimator under cross-sectional dependence. Then the $N-$asymptotic bias of the RMA estimator can be shown as follows.

Proposition 2: **($N-$asymptotic bias under cross-sectional dependence)**. *Under Assumption 2, for $N \to \infty$ first and then $T \to \infty$, the probability limit of $\hat{\rho}_{\mathrm{RMA|CSD}}$ is given by*

$$\mathrm{plim}_{N \to \infty} \left(\hat{\rho}_{\mathrm{RMA|CSD}} - \rho\right) = (1 - \eta) B(\rho, T) + \eta \left(\hat{\rho}_{\mathrm{F,RMA}} - \rho\right) + o_p \left(T^{-1}\right), \tag{20}$$

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

The proof is presented in Appendix D. Note that the asymptotic bias of $\hat{\rho}_{\mathrm{RMA|CSD}}$ now depends on three components: (i) the degree of cross-sectional dependence, η; (ii) the inconsistency of $\hat{\rho}_{\mathrm{RMA|CSI}}$, the RMA estimator under cross-sectional independence; and (iii) the bias of the time-series estimator $\hat{\rho}_{\mathrm{F,RMA}}$ ($K = 1$ case). Obviously, Proposition 1 applies when $\eta = 0$, while the inconsistency becomes purely random when $\eta = 1$. As $K \to \infty$, however, the asymptotic bias expression is identical to that in Proposition 1 regardless of the value of η. But the direction of the inconsistency is not clear because $\hat{\rho}_{\mathrm{F,RMA}}$ is downward biased for a small T while $B(\rho, T)$ is always positive. Unfortunately the exact finite sample bias formula for $\hat{\rho}_{\mathrm{F,RMA}}$ is available yet.

Figure 4 demonstrates the 'randomness' of the asymptotic bias by plotting the empirical distributions of $\hat{\rho}_{\mathrm{WG|CSD}}$ and $\hat{\rho}_{\mathrm{RMA|CSD}}$ for $N = 2000$ and $T = 20$ with various values of $K = \{1, 10, 50\}$. When $K = 1$, the empirical mean of $\hat{\rho}_{\mathrm{RMA|CSD}}$ is very close to the true value of $\rho = 0.5$. This is the case when the asymptotic positive bias from pooling is offset by the negative univariate bias of $\hat{\rho}_{\mathrm{F,RMA}}$. As K increases, the empirical distribution of $\hat{\rho}_{\mathrm{RMA|CSD}}$ becomes tighter but the mean value also increases slightly. By contrast, $\hat{\rho}_{\mathrm{WG|CSD}}$ has a noticeable downward asymptotic bias for all K although it tends to decrease as K grows. Overall, $\hat{\rho}_{\mathrm{RMA|CSD}}$ dominates $\hat{\rho}_{\mathrm{WG|CSD}}$ and the dominance stands out when the factor number is relatively small.

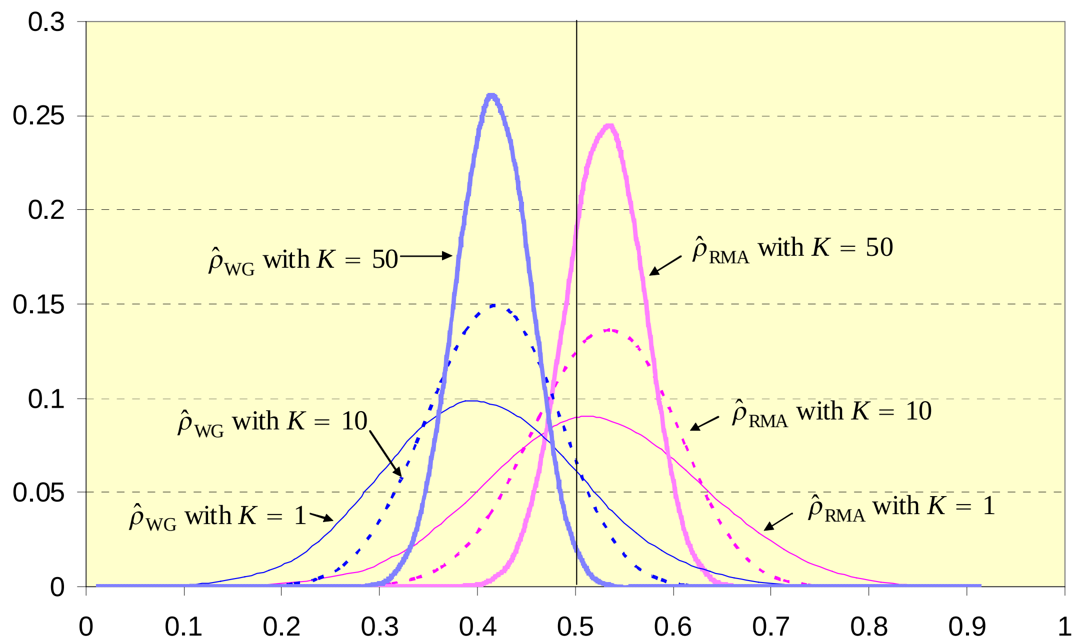


Figure 4: Random Asymptotic Bias under Cross-sectional Dependence, $T = 20$, $N = 2000$, $\rho = 0.5$
$\delta_{is} \sim iidN(0, 1)$, $\theta_{st} \sim iidN(0, 1)$, $\epsilon_{it} \sim iidN(0, 1)$

4 Monte Carlo Experiments

In this section, we report the results of Monte Carlo experiments designed 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. To economize on space, we have been selective in terms of which results to report especially when $\log T\,(N/T)$ is relatively large because in this case $B(\rho, T)$ remains in the distribution of the RMA estimator. An extensive set of simulation results are available at the author's web site.[8] We consider four cases here.

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 RMA and WG estimators. The first is the GMM estimator studied by Arellano and Bover (1995, hereafter AB) and the other is the estimator proposed by Hahn and Kuersteiner (2002, hereafter HK). The data generating process is,

$$y_{it} = \mu_i + z_{it},$$
$$z_{it} = \rho z_{it-1} + \epsilon_{it},$$

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 time series observations used in the regression are $T_0 = T - 1 \in \{5, 10, 20\}$.

The asymptotic variance of the AB estimator depends on the nuisance parameter $\psi = \sigma_\epsilon/\sigma_\mu = 1/\sigma_\mu$, whereas the variances of RMA, WG, 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 WG and HK are biased downwards. 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

[8] Full reports (MS excel format) are available at http://www3.uta.edu/choi/research.htm.

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

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

where $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\mu_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\mu^2\right)$. 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 (MSE) of RMA and WG. We note that the bias of WG 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 WG 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 WG, the MSE of RMA is consistently much smaller than that of WG estimator.

Case 3: *AR(1) and AR(2) with exogenous regressor.* Since Phillips and Sul (2007) show that $\hat{\beta}_{\text{WG}}$ 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,

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

where $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\mu_i \overset{iid}{\sim} \mathcal{N}(0,1)$ and $q_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$. 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 (again $\rho = \rho_1 + \rho_2$).

Several features of Table 3 are noteworthy. First, the WG estimator for γ is biased downwards in both the AR(1) and AR(2) cases and the bias directly distorts the size of the t-test. Whereas the size distortion for RMA is relatively small and remains fairly constant, the distortion of the WG based t-test increases with N. At the nominal size of 0.05, the size of t-test based on WG estimator is as large as 0.94 when $N = 200$ and $T_0 = 5$ in AR(1) case, while 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 due to 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

$$y_{it} = a_i + \rho y_{it-1} + u_{it},$$
$$u_{it} = \delta_i F_t + \epsilon_{it},$$

where $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\delta_i \overset{iid}{\sim} \mathcal{N}(1,1)$ and $F_t \overset{iid}{\sim} \mathcal{N}(0,1)$. We consider exactly the same simulation environment with that of Case 1 except that the error term is now cross-sectionally correlated as reflected by the common factor F_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. This contrasts to the competing estimators whose biases are larger under cross-sectional dependence. This result is consistent with the predictions of Proposition 2 that the bias of RMA (WG) 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 the 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 decreases more rapidly than the alternative estimators as N and T grow. As a result, the RMA estimator has smaller MSE than the other estimators particularly when the underlying processes are highly persistent.

To summarize, our simulation results suggest that the finite sample performance of the RMA estimator is appealing especially when the observations are cross-sectionally correlated.

5 Conclusion

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 both when error terms are cross-sectionally independent and when they are cross-sectionally correlated. 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 Autoregressive/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] Blundell, R. and S. Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

[9] Blundell, R., S. Bond, and F. Windmeijer (2000), "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimators," Working Paper 00-12, The Institute for Fiscal Studies.

[10] Bun, M. and Carree, M. (2005), "Bias-corrected Estimation in Dynamic Panel Data Models," *Journal of Business and Economic Statistics*, 23, 200-210.

[11] Campbell, J. and Lapham, B. (2004), "Real Exchange Rate Fluctuations And the Dynamics of Retail Trade Industries on the US-Canada Border," *American Economic Review,* 94, 1194-1206.

[12] Elliott, G., T.J. Rothenberg, and J.H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–836.

[13] Giraitis, L. and P.C.B. Phillips (2006), "Uniform Limit Theory for Stationary Autoregression," *Journal of Time Series Analysis*, 27(1), 51–60.

[14] 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.

[15] Hahn, J., J. Hausman, and G. Kuersteiner (2007). Long Difference Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects. *Journal of Econometrics*, 140(2), 574–617.

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

[17] Kiviet, J. F. (1995). On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics*, 68, 53–78.

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

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

[20] Park, J. Y. (2003): "Weak Unit Roots," *mimeo*, Rice University.

[21] Pesaran, H. (2006), "Estimation and Inference in Large Heterogeneous Panels with Multifactor Error Structure," *Econometrica*, 74 (4) 967–1012.

[22] Pesaran, H., Shin, Y and Smith, R. (1999), "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels," *Journal of the American Statistical Association*, Vol. 94, 621–634.

[23] Pesaran, H. and Smith, R. (1995), "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, Vol. 68, 79–113.

[24] Pesaran, H. and Yamagata, T. (2008), "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics*, Vol 142, 50–93.

[25] Phillips, P.C.B., J. Park, and Y. Chang (2004), "Nonlinear Instrumental Variable Estimation of an Autoregression,"*Journal of Econometrics* 118, 219–246.

[26] 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.

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

[28] 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.

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

[30] Sul, D. (2008), "Panel Unit Root Tests under Cross Section Dependence with Recursive Mean Adjustment," *mimeo*, University of Auckland.

[31] Sul, D., P.C.B. Phillips, and C.Y. Choi, (2005), "Prewhitening bias in HAC estimation." *Oxford Bulletin of Economics and Statistics* 67, 517-546.

[32] Taylor, R.A.M. (2002), "Regression-Based Unit Root Tests with Recursive Mean Adjustment for Seasonal and Nonseasonal Time Series," *Journal of Business and Economic Statistics*, 20, 269–281.

Technical Appendix

Define $\bar{z}_{it-1} = t_1^{-1} \sum_{s=1}^{t_1} z_{is}$ and $\bar{y}_{it-1} = t_1^{-1} \sum_{s=1}^{t_1} y_{is}$ where $t_1 = t - 1$ for the notational convenience.

Appendix A: Proof of Proposition 1

First we consider AR(1) case, and then use the result to establish the proof of Proposition 1. Note that for general AR(p) case, we have

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

Hence for AR(1) case, we have

$$\mathrm{E}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{it}-\bar{z}_{it})^2 = \frac{\sigma^2}{1-\rho^2} + O\left(\frac{\log T}{T}\right) \to \frac{\sigma^2}{1-\rho^2} \text{ as } T\to\infty, \tag{A-1}$$

where $\sigma^2 = N^{-1}\sum_{i=1}^{N}\sigma_i^2$ and $\sigma_i^2 = T^{-1}\sum_{t=1}^{T}\epsilon_{it} < \infty$ for all i.

Appendix A1: Proof of Remark 1 From a standard central limit theorem for panel autoregressive processes, we have

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

From (A-1), we have

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

Now note that

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

From direct calculation, we have

$$B(\rho, T) = -(1-\rho)\,\text{plim}_{N\to\infty} \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}(z_{it-1}-\bar{z}_{it-1})\,\bar{z}_{it-1}}{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}(z_{it-1}-\bar{z}_{it-1})^2}$$
$$= \frac{\rho \log T + B_{1T} - (1-\rho) B_{2T}}{T - 1 - \log T + \frac{2}{1-\rho} B_{1T} - B_{2T}} = \frac{\rho \log T}{T} + O\left(T^{-2}\right)$$

where

$$B_{1T} = \sum_{t=1}^{T-1} \frac{1}{t}\rho^t = O(1), \quad B_{2T} = 2\rho \sum_{t=1}^{T-1} \frac{1-\rho^t}{t^2(1-\rho)^2} = O(1).$$

For calculating its variance, first consider

$$\mathrm{E}\left(\sum_{t=2}^{T} z_{it-1}\bar{z}_{it-1} - \sum_{t=2}^{T} \bar{z}_{it-1}^2\right)^2 = \sum_{t=1}^{T-1}\frac{1}{t}\sigma_z^2 + O(1) = O(\log T).$$

Hence the variance of the second term in (A-2) is given by

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

Therefore,

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

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

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

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

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

or

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

Appendix A2: Proof of (6) in Proposition 1 For ease of reference, we restate (4) here as

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

where

$$e_{it} = -\left(1-\rho\right)\bar{z}_{it-1} + \epsilon_{it} + \sum_{j=1}^{p-1}\left(\phi_j - \hat{\phi}_{j,\mathrm{WG}}\right)\Delta z_{it-j}.$$

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

$$\begin{aligned}
\mathrm{plim}_{N\to\infty}\left(\hat{\rho}_{\mathrm{RMA}}^{p} - \rho\right) &= -\left(1-\rho\right)\mathrm{plim}_{N\to\infty}\frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\left(y_{it-1} - \bar{y}_{it-1}\right)\bar{z}_{it-1}\right]}{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \bar{y}_{it-1}\right)^2} \\
&\quad + \mathrm{plim}_{N\to\infty}\frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\left(y_{it-1} - \bar{y}_{it-1}\right)\sum_{j=1}^{p-1}\left(\phi_j - \hat{\phi}_{j,\mathrm{WG}}\right)\Delta z_{it-j}\right]}{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \bar{y}_{it-1}\right)^2} \\
&= B\left(\rho, T\right) + \frac{O\left(T^{-1}\right)}{D\left(\rho, T\right)} = B\left(\rho, T\right) + O\left(T^{-2}\right),
\end{aligned}$$

which establishes (6) in the text.

Appendix A3: Proof of (7) in Proposition 1 For the simplicity of analysis, we consider an AR(2) case and then show how the logic can be generalized to an AR(p) case. First, consider the regression error of

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

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 inconsistency of the pooled estimator $\hat{\phi}_{\mathrm{RMA}}^{p}$ can be written as

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

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 of

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

It follows that

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

Noting that the AR(2) model has the representation of

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

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

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

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

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

Appendix A4: Proof of (8) in Proposition 1 Let

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

where

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

$$\frac{C_{1,NT}}{D_{NT}} = B_{NT}(\rho,T) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$

note that $\text{plim}_{N\to\infty} B_{NT}(\rho,T) = B(\rho,T)$, and

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

Since

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

and

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it-1}-\bar{y}_{it-1})\,\Delta z_{it-j} = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

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

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

or

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

Appendix B: Proof of Remark 3 (Linear Trend Case)

Appendix B1: Proof of Iterative Recursive Detrending in (12) We address $\hat{\rho}_{\text{IRD}}$ by considering the AR(1) case with incidental trends and then discuss how it can be generalized to the AR(p) case. We work with the latent model and whether z_{it} is observable does not matter for this analysis. The point estimate $\hat{b}_i = \beta_{i,\text{WG}}/(1-\hat{\rho}_{\text{WG}})$ becomes equivalent to the point estimate from the regression for moderately large T,

$$y_{it} - \hat{\rho}_{\text{WG}} z_{it-1} = \mu_i + b_i t + \epsilon_{it} + (\rho - \hat{\rho}_{\text{WG}})\, z_{it-1}.$$

It follows that

$$\hat{b}_i - b_i = \frac{\sum_{t=1}^{T-1}\left(t - T_1^{-1}\sum_{t=1}^{T-1} t\right)\left[\left(\epsilon_{it} - T_1^{-1}\sum_{t=2}^{T}\epsilon_{it}\right) + (\rho - \hat{\rho}_{\mathrm{WG}})\left(z_{it-1} - T_1^{-1}\sum_{t=2}^{T} z_{it-1}\right)\right]}{\sum_{t=1}^{T-1}\left(t - T_1^{-1}\sum_{t=1}^{T-1} t\right)^2},$$

and by direct calculation,

$$E\left(\hat{b}_i - b_i\right)^2 = 12\frac{1}{T^3}\sigma_\epsilon^2 + O\left(T^{-2}\right).$$

Now let $b_i^* = \left(\hat{b}_i - b_i\right)$, then $\hat{\rho}_{\mathrm{IRD}}$ can be obtained either from

$$y_{it}^{\dagger} = y_{it} - \hat{b}_i t = \mu_i + \left(b_i - \hat{b}_i\right) t + z_{it} = \mu_i - b_i^* t + z_{it}$$

or from

$$y_{it}^{\dagger} = \mu_i (1-\rho) - b_i^*\rho - b_i^*(1-\rho) t + \rho y_{it-1}^{\dagger} + \epsilon_{it}.$$

Using the fact that

$$\bar{y}_{it-1}^{\dagger} = \frac{1}{t_1}\sum_{s=1}^{t_1} y_{is}^{\dagger} = \mu_i + \bar{z}_{it-1} - \frac{1}{2}b_i^* t,$$

it follows that

$$y_{it}^{\dagger} - \bar{y}_{it-1}^{\dagger} = \rho\left(y_{it-1}^{\dagger} - \bar{y}_{it-1}^{\dagger}\right) + \left\{-(1-\rho)\bar{z}_{it-1} + \epsilon_{it} - b_i^*\rho - \frac{1}{2}b_i^*(1-\rho) t\right\},$$

$$y_{it-1}^{\dagger} - \bar{y}_{it-1}^{\dagger} = -\frac{1}{2}b_i^* t + b_i^* + (z_{it} - \bar{z}_{it-1}).$$

Note that

$$E\sum_{t=1}^{T}\left(y_{it-1}^{\dagger} - \bar{y}_{it-1}^{\dagger}\right)^2 = D(\rho, T) + O(1).$$

Using these results, we have

$$\begin{aligned}
&\hat{\rho}_{\mathrm{IRD}} - \rho \\
&= \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}^{\dagger} - \bar{y}_{it-1}^{\dagger}\right)^2\right)^{-1} \\
&\times \sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}^{\dagger} - \bar{y}_{it-1}^{\dagger}\right)\left\{-(1-\rho)\bar{z}_{it-1} + \epsilon_{it} + b_i^*\rho + \frac{1}{2}b_i^*(1-\rho) t\right\}.
\end{aligned}$$

From the direct calculation, we have

$$E\sum_{t=2}^{T}\left(z_{it-1}-\bar{z}_{it-1}\right)b_i^*t=O\left(T^{-1}\right),\quad Eb_i^*\sum_{t=2}^{T}\left(t-2\right)\bar{z}_{it-1}=O\left(T^{-1}\right),\ E\left(b_i^*\right)^2\sum_{t=2}^{T}\left(t-2\right)=O\left(T^{-1}\right),$$

but since

$$E\left(b_i^*\right)^2\sum_{t=2}^{T}\left(t-2\right)t=4\sigma_\epsilon^2+O\left(T^{-1}\right),$$
$$Eb_i^*\sum_{t=2}^{T}\left(t-2\right)\epsilon_{it}=\sigma_\epsilon^2+O\left(T^{-1}\right),$$

it follows that

$$\left(1-\rho\right)\frac{1}{4}E\sum_{i=1}^{N}\left(b_i^*\right)^2\sum_{t=2}^{T}\left(t-2\right)t=\left(1-\rho\right)\sigma_\epsilon^2+O\left(T^{-1}\right),$$
$$\frac{1}{2}Eb_i^*\sum_{t=2}^{T}\left(t-2\right)\epsilon_{it}=\left(\frac{\sum_{t=2}^{T}\left(t-\bar{t}\right)\left(\epsilon_{it}-\bar{\epsilon}_i\right)}{\sum_{t=2}^{T}\left(t-\bar{t}\right)^2}\right)\sum_{t=2}^{T}\left(t-2\right)\epsilon_{it}=\frac{1}{2}\sigma_\epsilon^2+O\left(T^{-1}\right).$$

The sum of the $O\left(1\right)$ terms is

$$\left(1-\rho\right)\sigma_\epsilon^2-\frac{1}{2}\sigma_\epsilon^2=\frac{1}{2}\left(1-2\rho\right)\sigma_\epsilon^2,$$

while the denominator is

$$D\left(\rho,T\right)=\frac{T}{1-\rho^2}\sigma_\epsilon^2+O\left(1\right).$$

It follows that the second stage inconsistency of the IRD estimator is

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

Appendix B2: Proof of Double Recursive Detrending in (11) From direct calculation, we obtain the inconsistency of $\hat{\rho}_{\text{RMA}}^{\tau}$ as

$$\text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{RMA}}^{\tau}-\rho\right)=G\left(\rho,T\right)=\frac{C_T-F_T/T_1}{A_T-B_T/T_1}-\rho$$

where

$$A_T=T_1+4\sum_{t=2}^{T}\left(\frac{1}{t_1}\right)^2\frac{1-\rho^2}{\left(1-\rho\right)^2}\left(t_1-2\rho\frac{1-\rho^{t_1}}{1-\rho^2}\right)-4\sum_{t=2}^{T}\left(\frac{1}{t_1}\right)\left(\frac{1-\rho^{t_1}}{1-\rho}\right),$$

$$B_T = \left(T_1 + \frac{2\rho}{1-\rho}\sum_{k=1}^{T-2}(1-\rho^k)\right) + 4\sum_{j=1}^{T-1}\left(\sum_{s=j}^{T-1}\frac{1}{s}\right)^2 + 8\sum_{k=2}^{T-1}\rho^{k-1}\sum_{j=1}^{T-k}\left(\sum_{s=j}^{T-1}\frac{1}{s}\right)\left(\sum_{s=j+(k-1)}^{T-1}\frac{1}{s}\right)$$
$$-4T_1 - 4\left(\sum_{k=1}^{T-2}\left[\sum_{j=1}^{T-k-1}\rho^j\sum_{s=k+j}^{T-1}\frac{1}{s}\right]\right) - 4\left(\sum_{k=1}^{T-2}\left[\sum_{j=1}^{T-k-1}\rho^{T-k-j}\sum_{s=j}^{T-1}\frac{1}{s}\right]\right),$$

$$C_T = T_1\rho + 4\left(\sum_{t=2}^{T}\left(\frac{1}{t_1}\right)^2\frac{1-\rho^2}{(1-\rho)^2}\left(t_1 - 2\rho\frac{1-\rho^{t_1}}{1-\rho^2}\right)\right)$$
$$-2\left(\sum_{t=2}^{T}\left(\frac{1}{t_1}\right)\left(\frac{1-\rho^{t_1}}{1-\rho}\right) + \rho\sum_{t=2}^{T}\left(\frac{1}{t_1}\right)\left(\frac{1-\rho^{t_1}}{1-\rho}\right)\right),$$

$$F_T = \left(T_1 + 2\sum_{j=2}^{T-1}\sum_{t=1}^{T-j}\rho^t + \rho^{T-1} - 1\right)$$
$$+4\left(\sum_{j=1}^{T-1}\left(\sum_{s=j}^{T-1}\frac{1}{s}\right)^2 + 2\sum_{k=2}^{T-1}\rho^{k-1}\sum_{j=1}^{T-k}\left(\sum_{s=j}^{T-1}\frac{1}{s}\right)\left(\sum_{s=j+(k-1)}^{T-1}\frac{1}{s}\right)\right)$$
$$-2(1+\rho)\left(T_1 + \left(\sum_{k=1}^{T-2}\left(\sum_{j=1}^{T-k-1}\rho^j\sum_{s=k+j}^{T-1}\frac{1}{s}\right)\right) + \left(\sum_{k=1}^{T-2}\left(\sum_{j=1}^{T-k-1}\rho^{T-k-j}\sum_{s=j}^{T-1}\frac{1}{s}\right)\right)\right)$$
$$-2\left(1-\rho^2\right)\sum_{j=0}^{T-3}\rho^j\sum_{i=j+2}^{T-1}\sum_{s=i}^{T-1}\frac{1}{s}.$$

Appendix C: Proof of Remark 5 (GLS-Demeaning)

Assume the true DGP is given by

$$y_{it} = a_i + z_{it}, \quad z_{it} = \rho_T z_{it-1} + \epsilon_{it}, \quad \rho_T = 1 - \frac{c}{T}.$$

Define

$$y_{it}^g = \begin{cases} y_{it} - \left(1 - \frac{c}{T}\right) y_{it-1} & \text{if } t > 1 \\ y_{i1} & \text{if } t = 1 \end{cases}, \quad Z_t = \begin{cases} 1 - \frac{c}{T} & \text{if } t > 1 \\ 1 & \text{if } t = 1 \end{cases},$$

and

$$u_{it} = y_{it} - \frac{\sum_{t=2}^{T} y_{it}^g Z_t}{\sum_{t=2}^{T} Z_t^2}.$$

Note that

$$\sum_{t=1}^{T} Z_t^2 = 1 + \sum_{t=2}^{T}(1-a)^2 = 1 + \sum_{t=2}^{T}\frac{49}{T^2} = \frac{T^2 + 49T - 49}{T^2}.$$

From direct calculation, we have

$$\begin{aligned}
u_{it} &= y_{it} - \frac{\sum_{t=2}^{T} y_{it}^{g} Z_t}{\sum_{t=2}^{T} Z_t^2} \\
&= y_{it} - a_i - \frac{z_{i1}T^2 + 7\left(7-c\right)T\left(\frac{1}{T}\sum_{t=2}^{T} z_{it-1}\right) + 7T^2\left(\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}\right)}{T^2 + 49T_1} \\
&= z_{it} - z_{i1} + \Delta_{iT}
\end{aligned}$$

where

$$\Delta_{iT} = \frac{49T_1}{T^2+49T_1} z_{i1} - \frac{7\left(7-c\right)T\left(\frac{1}{T}\sum_{t=2}^{T} z_{it-1}\right) + 7T^2\left(\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}\right)}{T^2+49T_1}.$$

Hence as $T \to \infty$, we may say that

$$\Delta_{iT} = \frac{49}{T} z_{i1} - \frac{7\left(7-c\right)}{T}\left(\frac{1}{T}\sum_{t=2}^{T} z_{it-1}\right) - 7\left(\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}\right) + O_p\left(T^{-2}\right).$$

Let

$$\hat{\rho}_{\mathrm{GLS}} = \frac{\sum^{N}\sum^{T} u_{it-1}u_{it}}{\sum^{N}\sum^{T} u_{it-1}^2} = \frac{C_{NT}}{D_{NT}}.$$

Now evaluate at $c = 7$. Then we have

$$\Delta_{iT} = \frac{49}{T} z_{i1} - 7\left(\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}\right) + O_p\left(T^{-2}\right)$$

and

$$\begin{aligned}
\mathrm{plim}_{N\to\infty}\frac{1}{N}C_{NT} &= E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it}-z_{i1}\right)\left(z_{it-1}-z_{i1}\right)\right] + E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta_{iT}^2\right] \\
&\quad - E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta_{iT}\left(z_{it-1}-z_{i1}\right)\right] - E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta_{iT}\left(z_{it}-z_{i1}\right)\right] \\
&= I_1 + II + III + IV
\end{aligned}$$

where

$$\begin{aligned}
II = E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta_{iT}^2\right] &= E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\frac{49}{T}z_{i1} - 7\left(\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}\right)\right)^2\right] \\
&= \frac{49^2\sigma_z^2}{T} + 49\frac{\sigma_\epsilon^2}{T} + O\left(T^{-2}\right) = O\left(T^{-1}\right),
\end{aligned}$$

$$
\begin{aligned}
III &= -E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\frac{49}{T}z_{i1} - 7\left(\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}\right) + O_p\left(T^{-2}\right)\right)\left(z_{it-1} - z_{i1}\right)\right] \\
&= 49\sigma_z^2 - \frac{49}{T}\sigma_z^2\left(\frac{1-\rho^{T-1}}{1-\rho}\right) + O\left(T^{-1}\right) = 49\sigma_z^2 + O\left(T^{-1}\right),
\end{aligned}
$$

$$
\begin{aligned}
IV &= -E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\frac{49}{T}z_{i1} - 7\left(\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}\right) + O_p\left(T^{-2}\right)\right)\left(z_{it} - z_{i1}\right)\right] \\
&= 49\sigma_z^2 + 7\sigma_\epsilon^2 + O\left(T^{-1}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{plim}_{N\to\infty}\frac{1}{N}D_{NT} &= E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it-1} - z_{i1}\right)^2\right] + E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta_{iT}^2\right] \\
&- E\left[\frac{2}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta_{iT}\left(z_{it-1} - z_{i1}\right)\right] \\
&= I_2 + II + 2\times III.
\end{aligned}
$$

Note that the two terms, I_1 and I_2 will be given in (A-3) and (A-4), respectively.

Now consider

$$
\mathrm{plim}_{N\to\infty}\frac{1}{N}\frac{C_{NT}}{D_{NT}} = \frac{I_1 + 98\sigma_z^2 + 7\sigma_\epsilon^2 + O\left(T^{-1}\right)}{I_2 + 98\sigma_z^2 + O\left(T^{-1}\right)}.
$$

Note that we can define the long-differencing regression as

$$
z_{it} - z_{i1} = \rho\left(z_{it-1} - z_{i1}\right) - (1-\rho)\, z_{i1} + \epsilon_{it},
$$

so that we have

$$
\hat{\rho}_{\mathrm{long}} = \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it} - z_{i1}\right)\left(z_{it-1} - z_{i1}\right)}{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it-1} - z_{i1}\right)^2} = \rho - (1-\rho)\frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it-1} - z_{i1}\right) z_{i1}}{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it-1} - z_{i1}\right)^2},
$$

and

$$
\mathrm{plim}_{N\to\infty}\hat{\rho}_{\mathrm{long}} = \rho + \frac{1-\rho}{2},
$$

since with a moderate large T

$$
E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it-1} - z_{i1}\right) z_{i1}\right] = -\sigma_z^2\left[(T-1) - \frac{1-\rho^{T-1}}{1-\rho}\right], \tag{A-3}
$$

$$E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it-1}-z_{i1}\right)^{2}\right]=2\sigma_z^2\left[(T-1)-\frac{1-\rho^{T-1}}{1-\rho}\right]. \tag{A-4}$$

Therefore, for a large T we have

$$\begin{aligned}\text{plim}_{N\to\infty}\frac{C_{NT}}{D_{NT}} &= \frac{2\sigma_z^2 T\rho+(1-\rho)\,\sigma_z^2 T+98\sigma_z^2+7\sigma_\epsilon^2}{2\sigma_z^2 T+98\sigma_z^2}+O\left(T^{-2}\right)\\ &=\rho+(1-\rho)\,\frac{1}{2}+\frac{49+7\,(1+\rho)}{2T+98}+O\left(T^{-2}\right).\end{aligned}$$

Plugging $\rho=1-7/T$ yields

$$\begin{aligned}\text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{GLS}}-\rho\right) &= (1-\rho)\,\frac{1}{2}+\frac{49+7\,(1+\rho)}{2T+98}+O\left(T^{-2}\right)\\ &=\frac{35}{T}+O\left(T^{-2}\right).\end{aligned}$$

Appendix D: Proof of Proposition 2

According to Phillips and Sul (2007), the WG estimator $\hat{\rho}_{\text{WG}|\text{CSD}}$ of the model (14)-(16) can be decomposed into

$$\hat{\rho}_{\text{WG}|\text{CSD}}=(1-\eta)\,\hat{\rho}_{\text{WG}|\text{CSI}}+\eta\hat{\rho}_{\text{F,WG}}+o_p\left(T^{-1}\right),$$

where $\hat{\rho}_{\text{WG}|\text{CSI}}$ is the WG estimator when the observations are cross-sectionally independent, $\eta = m_\delta^2\sigma_\theta^2\left(\sigma^2+m_\delta^2\sigma_\theta^2\right)^{-1}$, $m_\delta^2=(NK)^{-1}\sum_{i=1}^{N}\sum_{s=1}^{K}\delta_{si}^2$, $\sigma_\theta^2=(KT)^{-1}\sum_{s=1}^{K}\sum_{t=1}^{T}\theta_{st}^2$, and $\sigma^2 = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}^2$. η has the interpretation of the degree of cross-sectional dependence. When $\eta=1$, the observations are maximally dependent and when $\eta=0$ they are independent. Since $\hat{\rho}_{\text{F,WG}}$ does not depend on N, the $N-$asymptotic bias for $\hat{\rho}_{\text{WG}|\text{CSD}}$ is seen to depend on the inconsistency of $\hat{\rho}_{\text{F,WG}}$. As shown by Phillips and Sul (2007),

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

as $N\to\infty$. The strategy of the proof follows the strategy taken by Phillips and Sul (2007).

For the inconsistency for WG estimator, see Proposition 3 in Phillips and Sul (2007). Here we take $K=1$ case and then provide inconsistency for RMA estimator under cross-sectional dependence. For RMA, define

$$\text{plim}_{N\to\infty}\hat{\rho}_{\text{RMA}|\text{CSD}}=-(1-\rho)\,\text{plim}_{N\to\infty}\frac{C_{NT}^C}{D_{NT}^C},$$

where

$$C_{NT}^C\equiv-(1-\rho)\textstyle\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\bar{y}_{it-1}\right)\bar{z}_{it-1}+\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\bar{y}_{it-1}\right)u_{it},\quad D_{NT}^C\equiv\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1}-\bar{y}_{it-1})^2.$$

Further decompose $B(\rho, T)$ into

$$\begin{aligned} B(\rho, T) &= \text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{RMA}|\text{CSI}} - \rho\right) \\ &= \frac{B_{CT}}{B_{DT}} = \frac{-(1-\rho)\,\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)\bar{z}_{it}}{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it}-\bar{z}_{it}\right)^2}. \end{aligned}$$

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

$$y_{it} = \mu_i + z_{it},\ \ z_{it} = \rho z_{it-1} + u_{it}, \quad u_{it} = \delta_i\theta_t + \epsilon_{it},$$

with

$$z_{it} = \delta_i\sum_{j=0}^{\infty}\rho^j\theta_{t-j} + \sum_{j=0}^{\infty}\rho^j\epsilon_{it-j} = \delta_i F_t + x_{it}.$$

Since $y_{it} - \bar{y}_{it} = z_{it} - \bar{z}_{it}$, we have

$$\begin{aligned} &\text{plim}_{N\to\infty}\frac{1}{N}C_{NT}^{C} \\ &= -(1-\rho)\,\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\bar{y}_{it-1}\right)\bar{z}_{it-1} \\ &= -(1-\rho)\,\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(\delta_i F_{t-1}+x_{it-1}\right)\left(\delta_i\bar{F}_{t-1}+\bar{x}_{it-1}\right) - \sum_{t=1}^{T}\left(\delta_i\bar{F}_{t-1}+\bar{x}_{it-1}\right)^2\right] \\ &= -(1-\rho)\,\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left\{\left[\sum_{t=1}^{T}x_{it-1}\bar{x}_{it-1} - \sum_{t=1}^{T}\bar{x}_{it-1}^2\right] + \delta_i^2\left[\sum_{t=1}^{T}F_{t-1}\bar{F}_{t-1} - \sum_{t=1}^{T}\bar{F}_{t-1}^2\right]\right\} \\ &= B_{CT} + m_\delta^2\left[\sum_{t=1}^{T}F_{t-1}\bar{F}_{t-1} - \sum_{t=1}^{T}\bar{F}_{t-1}^2\right] \end{aligned}$$

because all probability limits of the cross product terms are zero. That is, $\text{E}\delta_i x_{it-1} = \text{E}\delta_i\bar{x}_{it-1} = 0$.

Similarly, we have

$$\begin{aligned} \text{plim}_{N\to\infty}\frac{1}{N}D_{NT}^{C} &= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(z_{it-1}-\bar{z}_{it-1}\right)^2 \\ &= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it-1}-\bar{x}_{it-1}\right)^2 + \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\sum_{t=1}^{T}\left(F_{t-1}-\bar{F}_{t-1}\right)^2 \\ &= B_{DT} + m_\delta^2\sum_{t=1}^{T}\left(F_{t-1}-\bar{F}_{t-1}\right)^2. \end{aligned}$$

Combining the two results gives

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

Let $T\to\infty$ to have

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

Denote

$$\Delta_T = \sigma_\theta^2 T^{-1} \left[\frac{\sigma_\theta^2}{1-\rho^2} + O_p\left(T^{-1/2}\right) \right]^{-1} = \left(1-\rho^2\right) T^{-1} + O_p\left(T^{-3/2}\right)$$

since

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

It follows that

$$\frac{\sigma_\theta^2}{\sum_{t=1}^T \left(F_{t-1} - \bar{F}_{t-1}\right)^2} = \Delta_T \text{ as } T \to \infty.$$

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

$$\begin{aligned}
B_{CT}\Delta_T &= \frac{\sigma^2}{1-\rho^2} \left(\rho \log T + B_{1T} - (1-\rho) B_{2T}\right) \left[\left(1-\rho^2\right) T^{-1} + O_p\left(T^{-3/2}\right)\right] \\
&= \sigma^2 \frac{\rho \log T}{T} + O\left(T^{-1}\right), \\
B_{DT}\Delta_T &= \frac{\sigma^2}{1-\rho^2} \left(T - \log T + \frac{2}{1-\rho} B_{1T} - B_{2T}\right) \left[\left(1-\rho^2\right) T^{-1} + O_p\left(T^{-3/2}\right)\right] \\
&= \sigma^2 + O\left(\frac{\log T}{T}\right),
\end{aligned}$$

so that we have

$$\begin{aligned}
\frac{\text{plim}_{N\to\infty} \frac{1}{N} C_{NT}^C}{\text{plim}_{N\to\infty} \frac{1}{N} D_{NT}^C} &= \frac{B_{CT}\Delta_T + \sigma_\theta^2 m_\delta^2 \left(\hat{\rho}_{\text{F,RMA}} - \rho\right)}{B_{DT}\Delta_T + \sigma_\theta^2 m_\delta^2} \\
&= \frac{\sigma^2 \frac{\rho \log T}{T} + \sigma_\theta^2 m_\delta^2 \left(\hat{\rho}_{\text{F,RMA}} - \rho\right)}{\sigma^2 + \sigma_\theta^2 m_\delta^2} + O_p\left(T^{-3/2}\right) \\
&= \eta \, \text{plim}_{N\to\infty} \left(\hat{\rho}_{\text{RMA|CSI}} - \rho\right) + (1-\eta) \left(\hat{\rho}_{\text{F,RMA}} - \rho\right) + o_p\left(T^{-1}\right).
\end{aligned}$$

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

			Bias						MSE ×100					
T_0	N	ρ	AB1	AB2	AB3	WG	HK	RMA	AB1	AB2	AB3	WG	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

$$y_{it} = \mu_i + z_{it},$$
$$z_{it} = \rho z_{it-1} + \epsilon_{it},$$

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 WG and RMA in Panel AR(2) case

				Bias				Variance × 100				MSE × 100			
				$\hat{\rho}^p_{\text{RMA}}$		$\hat{\rho}_2$		$\hat{\rho}^p_{\text{RMA}}$		$\hat{\rho}_2$		$\hat{\rho}^p_{\text{RMA}}$		$\hat{\rho}_2$	
T_0	N	ρ	ρ_2	WG	RMA	WG	RMA	WG	RMA	WG	RMA	WG	RMA	WG	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

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

where $\rho = \rho_1 + \rho_2$, $\epsilon_{it} \overset{iid}{\sim} N(0,1)$, and $\mu_i \overset{iid}{\sim} N\left(0, \sigma^2_\mu\right)$.

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

			Bias		Variance × 100		MSE × 100		Rej. of t-test	
			AR(1) case							
T_0	N	ρ	WG	RMA	WG	RMA	WG	RMA	WG	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	WG	RMA	WG	RMA	WG	RMA	WG	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 for AR(2) case is

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

where $\rho = \rho_1 + \rho_2$, $\epsilon_{it} \overset{iid}{\sim} N(0,1)$, $\mu_i \overset{iid}{\sim} N(0,1)$, $q_{it} \overset{iid}{\sim} N(0,1)$ 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	WG	HK	RMA	AB1	AB2	AB3	WG	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

$$y_{it} = a_i + \rho y_{it-1} + u_{it},$$
$$u_{it} = \delta_i F_t + \epsilon_{it},$$

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