Mean Group Estimation in Presence of Weakly Cross-Correlated Estimators*

Alexander Chudik
Federal Reserve Bank of Dallas

M. Hashem Pesaran
University of Southern California, USA, and Trinity College, Cambridge, UK

14 November 2018

Abstract

This paper extends the mean group (MG) estimator for random coefficient panel data models by allowing the underlying individual estimators to be weakly cross correlated. Weak cross-sectional dependence of the individual estimators can arise, for example, in panels with spatially correlated errors. We establish that the MG estimator is asymptotically correctly centered, and its asymptotic covariance matrix can be consistently estimated. The random coefficient specification allows for correct inference even when nothing is known about the weak cross-sectional dependence of the errors. This is in contrast to the well known homogenous case, where cross-sectional dependence of errors results in incorrect inference unless the nature of the cross-sectional error dependence is known and can be taken into account. Evidence on small sample performance of the MG estimators are provided using Monte Carlo experiments with both strictly and weakly exogenous regressors and cross-sectionally correlated innovations.

Keywords: Mean Group Estimator, Cross Sectional Dependence, Spatial Models, Panel Data
JEL Classification: C12, C13, C23

*We are grateful to Natalia Bailey and Ron Smith for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas.

1 Introduction

Averaging estimates of individual cross-section units in a panel to estimate a common mean is a well established idea explored in a variety of contexts. See Hsiao and Pesaran (2008) for an overview.[1] Pesaran and Smith (1995) refer to this estimator as the Mean Group (MG) estimator and show that it is consistent in the context of estimating long-run relationships in dynamic panel data models. In this paper, we consider the asymptotic properties of the MG estimator in a more general (linear or nonlinear) random coefficient panel data setup with N cross section units and T time periods, where the estimators of individual cross-section units are allowed to be weakly cross correlated.

We establish the asymptotic distribution of the MG estimator in a random coefficient (linear or non-linear) panel data setup under fairly general assumptions on the underlying individual estimators. We assume variances of the individual estimators exist and the estimators are weakly cross correlated (to be made precise in Section 2 below). We distinguish between cases when the individual estimators are unbiased (that could arise in the case of linear panels with strictly exogenous regressors), and when they are biased which could arise in the case of non-linear panels or linear panel data models with weakly exogenous regressors. Under these conditions, we show that the asymptotic distribution of the MG estimator is asymptotically normal and correctly centered as both panel dimensions N and $T \to \infty$, jointly. The required relative expansion rates of N and T depend on whether the underlying estimators are biased, and if so the order of their bias. No restrictions are imposed on the relative expansion rates of N and T when the estimators are unbiased. When the bias is of order T^{-1} then it is required that $\sqrt{N}/T \to 0$. If the order of the bias is reduced to T^{-2}, by use of half-panel Jackknife estimators, for example, then the validity of the Jackknifed-MG estimators only requires $N/T^3 \to 0$, as N and $T \to \infty$, jointly.

A consistent estimator of the asymptotic variance is also proposed under some additional regularity conditions, without needing to specify the nature and sources of the weak cross-sectional correlation. As discussed in Pesaran and Tosetti (2011), the asymptotic variance of MG or pooled estimators depends on the pattern of weak cross-sectional correlation of errors in the case of panels with homogenous slopes, and consistent estimation of their asymptotic covariance matrix does require some knowledge of the cross-sectional correlation of the errors (which is often difficult to

[1] See, for example, Baltagi (2013) and Chapter 28 of Pesaran (2015).

know in practice). In contrast, the MG estimator investigated in this paper for panels with heterogeneous slopes, *does not* require any such knowledge, and its proposed covariance matrix being non-parametric is robust to error serial correlation and heteroskedasticity.

The general assumptions for the MG estimation in this paper encompass a number of interesting models. We discuss panel regressions with strictly exogenous regressors and spatially correlated errors as an important although not the only illustrative example that falls within our framework.

Small sample properties of the MG estimator are investigated by Monte Carlo experiments. We find the small sample performance of the MG estimator satisfactory in the strictly exogenous case. In the case of weakly exogenous regressors, larger values of T are required for valid inference, but the size distortions can be substantially reduced with bias correction methods, such as the half-panel jackknife procedure.

The remainder of the paper is organized as follows. Section 2 sets out the assumptions, and provides the main theoretical results. Section 3 discusses examples. Section 4 presents MC evidence, and Section 5 concludes the paper. Proofs are provided in an appendix.

Notations. $K_0, K_1, ...,$ and K denote generic finite constants that do not depend on the sample size, (N, T). These constants can take different values at different instances. $\|\mathbf{A}\| = \sqrt{\varrho\left(\mathbf{A}'\mathbf{A}\right)}$ is the spectral norm of $\mathbf{A}$, $\varrho\left(\mathbf{A}\right) \equiv \left|\lambda_{\max}\left(\mathbf{A}\right)\right|$ is the spectral radius of $\mathbf{A}$, and $\lambda_{\max}\left(\mathbf{A}\right)$ is the largest eigenvalue (in absolute value) of $\mathbf{A}$.[2] All vectors are column vectors.

2 The Mean Group Estimator

Let $\hat{\boldsymbol{\theta}}_{iT}$ be an estimator of the $k \times 1$ vector of unknown parameters $\boldsymbol{\theta}_i$ for unit i in the panel, estimated using T observations. Suppose that $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{iT} - \boldsymbol{\theta}_{i0}\right)$ is asymptotically distributed with zero mean and a finite variance, where $\boldsymbol{\theta}_{i0}$ is the true value of $\boldsymbol{\theta}_i$. We do not require asymptotic normality. Let $\hat{\underline{\boldsymbol{\theta}}}_T = \left(\hat{\boldsymbol{\theta}}'_{1T}, \hat{\boldsymbol{\theta}}'_{2T}, ..., \hat{\boldsymbol{\theta}}'_{NT}\right)'$ and $\underline{\boldsymbol{\theta}}_0 = \left(\boldsymbol{\theta}'_{1,0}, \boldsymbol{\theta}'_{2,0}, ..., \boldsymbol{\theta}'_{N,0}\right)'$ be the $Nk \times 1$ vectors that stack the estimators and the true values of the parameters for all the N units. We assume there exists some T_0 such that for all $T > T_0$, the finite sample variance of $\sqrt{T}\hat{\underline{\boldsymbol{\theta}}}_T$, denoted by $Var\left(\sqrt{T}\hat{\underline{\boldsymbol{\theta}}}_T\right)$, exists. We distinguish between two possibilities for the bias of $\hat{\boldsymbol{\theta}}_{iT}$, and we allow $\hat{\boldsymbol{\theta}}_{iT}$, for $i = 1, 2, ..., N$, to be weakly cross correlated.

[2] Note that if $\mathbf{x}$ is a vector, then $\|\mathbf{x}\| = \sqrt{\varrho\left(\mathbf{x}'\mathbf{x}\right)} = \sqrt{\mathbf{x}'\mathbf{x}}$ corresponds to the Euclidean length of vector $\mathbf{x}$.

Assumption 1 ***(Bias of estimators)***

(a) ***No bias:*** *$E\left(\hat{\boldsymbol{\theta}}_{iT}\right)=0$, for each $i=1,2,...,N$.*

(b) ***Small-T bias:*** *$E\left(\hat{\boldsymbol{\theta}}_{iT}\right)=\boldsymbol{\theta}_{i0}+K_i/T+O\left(T^{-3/2}\right)$, for each $i=1,2,...,N$, where $\sup_i|K_i|<\infty$.*

Assumption 2 ***(Cross correlation of estimators)*** *There exists T_0 such that for all $T>T_0$, the $Nk\times Nk$ symmetric covariance matrix of $\sqrt{T}\hat{\boldsymbol{\theta}}_T$, denoted by $\mathbf{V}_{NT}$, exists and $\lambda_{\max}\left(\mathbf{V}_{NT}\right)<K<\infty$.*

Assumption 3 ***(Random coefficient model)*** *For each $i=1,2,...,N$, $\boldsymbol{\theta}_{i0}=\boldsymbol{\theta}_0+\boldsymbol{\eta}_i$, where $\boldsymbol{\eta}_i\sim IID\left(\mathbf{0}_k,\boldsymbol{\Omega}_\eta\right)$ and $\boldsymbol{\Omega}_\eta$ is a $k\times k$ symmetric positive definite matrix. In addition, there exists some $\epsilon>0$ such that $E\left\|\boldsymbol{\eta}_i\right\|^{2+\epsilon}<K<\infty$.*

Assumption 1.b allows for bias of order T^{-1}, which is important for the case of panels with weakly exogenous regressors. This assumption could be generalized to allow for a different upper bound on bias in terms of T and/or N, which would alter the relative rates of N and T required for our main results. One possibility is the bias of order $O\left(T^{-2}\right)$, which we discuss below. Assumption 2 does not require the existence of higher moments, besides variances. In addition, Assumption 2 allows for weak cross-sectional correlation of $\left\{\hat{\boldsymbol{\theta}}_{iT}, \text{ for } i=1,2,...,N\right\}$, which is important, for example, in panel data models with spatial dependence. This assumption can be relaxed so long as $T^{-1}\lambda_{\max}\left(\mathbf{V}_{NT}\right)\rightarrow 0$, which allows for a stronger cross-correlations of $\hat{\boldsymbol{\theta}}_{iT}$, $i=1,2,...,N$. Assumption 3 is the standard random coefficient assumption. $\{\boldsymbol{\theta}_{i0}, \text{ for } i=1,2,...,N\}$ are distributed independently around the common mean $\boldsymbol{\theta}_0$, which is the target of estimation and inference in this paper. This assumption can be relaxed by assuming $\boldsymbol{\eta}_i\sim IID\left(\mathbf{0}_k,N^{-\delta}\boldsymbol{\Omega}_\eta\right)$, for some $0\leq\delta\leq 1/2$, with some implications on the relative expansion rates of N and T.

The following theorem establishes asymptotic normality of the MG estimator under the above assumptions.

Theorem 1 *(Asymptotic distribution of the MG estimator) Suppose that $\hat{\boldsymbol{\theta}}_{iT}$ is an estimator of $\boldsymbol{\theta}_{i0}$, for $i=1,2,...,N$, where $E\left(\boldsymbol{\theta}_{i0}\right)=\boldsymbol{\theta}_0$, for all i, and Assumptions 1.a, 2, and 3 hold. Then the* *mean group estimator of $\boldsymbol{\theta}_0$, defined by*

$$\hat{\boldsymbol{\theta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{iT}, \tag{1}$$

is asymptotically distributed as

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}_0\right) \overset{a}{\sim} N\left(\mathbf{0}_k, \boldsymbol{\Omega}_\eta\right), \tag{2}$$

for N and $T \to \infty$, jointly. If Assumption 1.b holds in place of 1.a, then (2) holds so long as $\sqrt{N}/T \to 0$, as N and $T \to \infty$, jointly.

Inference based on (2) requires a consistent estimator of $\boldsymbol{\Omega}_\eta$. The standard estimator of $\boldsymbol{\Omega}_\eta$ is given by

$$\hat{\boldsymbol{\Omega}}_\eta = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{iT} - \hat{\boldsymbol{\theta}}_{MG}\right)\left(\hat{\boldsymbol{\theta}}_{iT} - \hat{\boldsymbol{\theta}}_{MG}\right)'. \tag{3}$$

The following theorem establishes the conditions under which $\hat{\boldsymbol{\Omega}}_\eta$ continues to be a consistent estimator of $\boldsymbol{\Omega}_\eta$ even if the individual estimators are cross-correlated.

Theorem 2 *(Consistency of $\hat{\boldsymbol{\Omega}}_\eta$) Consider $\hat{\boldsymbol{\Omega}}_\eta$ given by (3) and suppose Assumptions 1.a, 2, and 3 hold. Then $\hat{\boldsymbol{\Omega}}_\eta \to_p \boldsymbol{\Omega}_\eta$, as N and $T \to \infty$, at any rate. If Assumption 1.b holds in place of 1.a, then $\hat{\boldsymbol{\Omega}}_\eta \to_p \boldsymbol{\Omega}_\eta$ so long $\sqrt{N}/T \to 0$, as N and $T \to \infty$, jointly.*

There is no restriction on the relative expansion rates of N and T when the underlying estimators are unbiased (Assumption 1.a). The requirement $\sqrt{N}/T \to 0$ in Theorems 1 and 2 is a consequence of the $O\left(T^{-1}\right)$ bias allowed by Assumption 1.b. If the unit-specific estimators are biased, but with a bias of smaller order than the one postulated in Assumption 1.b, then a less stringent requirement on the relative expansion rates of N and T would be required for the asymptotic distribution of the MG estimator to be correctly centered. One example of an estimator with a bias of lower order are the jackknife-bias-corrected estimators discussed in Dhaene and Jochmans (2015) and Chudik, Pesaran, and Yang (2018), which we will also examine in the Monte Carlo section below. But first, we provide some example from the literature.

3 Applications

There are a number of models in panel data literature that meet our high level assumptions. Examples include heterogeneous version of the non-linear panel data models discussed in Dhaene and Jochmans (2015), heterogeneous spatial autoregressive models, and global VARs (Pesaran, Schuermann, and Weiner (2004)). Here we shall illustrate the applicability of our assumptions to linear panel regressions with spatially correlated errors.

Consider the following panel regression

$$y_{it} = \boldsymbol{\theta}_{i0}'\mathbf{x}_{it} + u_{it}, \tag{4}$$

for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$, where $\boldsymbol{\theta}_{i0}$ satisfies the random coefficient Assumption 3. It is convenient to define $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$ and $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{iT})'$, so that (4) can be written as $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta}_{i0} + \mathbf{u}_i$, for $i = 1, 2, ..., N$. Suppose that $\mathbf{x}_{it}$ is strictly exogenous and $E\left(u_{ij} \left| \mathbf{X}_i \right.\right) = 0$ for all i and j. In addition, suppose that

$$E\left(\mathbf{u}_i\mathbf{u}_j' \left| \mathbf{X}_1, ..., \mathbf{X}_N \right.\right) = \sigma_{ij}\mathbf{I}_T, \tag{5}$$

and

$$\sup_i \sum_{j=1}^{N} \left|\sigma_{ij}\right| < K < \infty. \tag{6}$$

We consider the least squares (LS) estimator

$$\hat{\boldsymbol{\theta}}_{iT} = \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{y}_i. \tag{7}$$

Under the above conditions, $\hat{\boldsymbol{\theta}}_{iT}$ is unbiased, namely $E\left(\hat{\boldsymbol{\theta}}_{iT}\right) = \boldsymbol{\theta}_{i0}$. Hence, Assumptions 1.a holds. Also

$$cov\left(\sqrt{T}\hat{\boldsymbol{\theta}}_{iT}, \sqrt{T}\hat{\boldsymbol{\theta}}_{jT}\right) = \sigma_{ij}\left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\mathbf{X}_j}{T}\right)\left(\frac{\mathbf{X}_j'\mathbf{X}_j}{T}\right)^{-1} = \sigma_{ij}\mathbf{P}_i'\mathbf{P}_j,$$

where $\mathbf{P}_i' = \left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'}{\sqrt{T}}$, for $i = 1, 2, ..., N$. Now using (7) note that $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) = \mathbf{P}'\mathbf{u}$, where

$\hat{\boldsymbol{\theta}}_T = \left(\hat{\boldsymbol{\theta}}'_{1T}, \hat{\boldsymbol{\theta}}'_{2T}, ..., \hat{\boldsymbol{\theta}}'_{NT}\right)'$, $\boldsymbol{\theta}_0 = \left(\boldsymbol{\theta}'_{1,0}, \boldsymbol{\theta}'_{2,0}, ..., \boldsymbol{\theta}'_{N,0}\right)'$, $\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2, ..., \mathbf{u}'_N)'$, and

$$\mathbf{P}' = \begin{pmatrix} \mathbf{P}'_1 & \mathbf{0}_{k\times T} & \cdots & \mathbf{0}_{k\times T} \\ \mathbf{0}_{k\times T} & \mathbf{P}'_2 & & \mathbf{0}_{k\times T} \\ \vdots & & \ddots & \\ \mathbf{0}_{k\times T} & \mathbf{0}_{k\times T} & & \mathbf{P}'_N \end{pmatrix}.$$

Also $E(\mathbf{u}\mathbf{u}') = \boldsymbol{\Sigma} \otimes \mathbf{I}_T$, where $\boldsymbol{\Sigma} = (\sigma_{ij})$ is $N \times N$ covariance matrix with elements σ_{ij}, and

$$\mathbf{V}_{NT} = Var\left[\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right)\right] = \mathbf{P}'\left(\boldsymbol{\Sigma} \otimes \mathbf{I}_T\right)\mathbf{P}.$$

Taking the spectral norm, and using the triangle inequality, we have $\|\mathbf{V}_{NT}\| \leq \|\mathbf{P}'\| \, \|\mathbf{P}\| \, \|\boldsymbol{\Sigma} \otimes \mathbf{I}_T\|$. But $\|\mathbf{P}'\| = \|\mathbf{P}\| = \lambda^{1/2}_{\max}(\mathbf{P}'\mathbf{P})$ and $\lambda^{1/2}_{\max}(\mathbf{P}'\mathbf{P}) \leq \sup_i \lambda^{1/2}_{\max}(\mathbf{P}'_i\mathbf{P}_i)$. In addition, $\|\boldsymbol{\Sigma} \otimes \mathbf{I}_T\| = \lambda^{1/2}_{\max}(\boldsymbol{\Sigma})$. Hence

$$\|\mathbf{V}_{NT}\| \leq \sup_i \lambda_{\max}\left(\mathbf{P}'_i\mathbf{P}_i\right) \lambda^{1/2}_{\max}(\boldsymbol{\Sigma}).$$

Note that $\mathbf{P}'_i\mathbf{P}_i = (\mathbf{X}'_i\mathbf{X}_i/T)^{-1}$, and $\lambda_{\max}(\mathbf{P}'_i\mathbf{P}_i) = [\lambda_{\min}(\mathbf{X}'_i\mathbf{X}_i/T)]^{-1}$. Consequently, Assumption 2 holds if $\lambda_{\max}(\boldsymbol{\Sigma}) < K$, and $\inf_i \lambda_{\min}\left(\frac{\mathbf{X}'_i\mathbf{X}_i}{T}\right) > 0$. The latter is the standard condition for identification of $\{\boldsymbol{\theta}_{i0}$, for $i = 1, 2, ..., N\}$. Condition $\lambda_{\max}(\boldsymbol{\Sigma}) < K$ is satisfied if $\{u_{it}$, for $i = 1, 2, ..., N\}$ is weakly cross-correlated. An example of weakly cross-sectionally correlated errors is the spatially autoregressive SAR(1) model: (See, for example, Anselin (1988)),

$$u_{it} = \rho \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}, \tag{8}$$

or (in vector notations) $\mathbf{u}_t = (\mathbf{I}_N - \rho\mathbf{W})^{-1} \boldsymbol{\varepsilon}_t$. In this example, we obtain

$$E\left(\mathbf{u}_t\mathbf{u}'_t\right) = (\sigma_{ij}) = (\mathbf{I}_N - \rho\mathbf{W})^{-1} \boldsymbol{\Lambda} \left(\mathbf{I}_N - \rho\mathbf{W}'\right)^{-1}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with elements $\sigma^2_i = Var(\varepsilon_{it})$ on the diagonal. Condition (6) is satisfied if $\sup_i Var(\varepsilon_{it}) < K$, $|\rho| < \max(1/\|\mathbf{W}\|_1, 1/\|\mathbf{W}\|_\infty)$, and $\|\mathbf{W}\|_1 < K$, $\|\mathbf{W}\|_\infty < K$.

4 Monte Carlo evidence

4.1 Data generating process

The dependent variable and the regressor are generated as:

$$y_{it} = \alpha_i + \theta_i x_{it} + e_{it}, \tag{9}$$

$$x_{it} = \alpha_{i1} + \kappa_i y_{i,t-1} + \alpha_{i2} f_t + v_{it}, \tag{10}$$

for $i = 1, 2, ..., N$ and $t = -49, ..., 0, 1, ..., T$, with the starting values $y_{i,-50} = 0$, where $\theta_i \sim IIDN(1, 0.25)$. The first 50 observations are discarded to minimize the effects of the initial values on the outcomes. Coefficients κ_i capture feedback from $y_{i,t-1}$ to x_{it}, and we consider two options: $\kappa_i = 0$ for all i (strictly exogenous regressors), and $\kappa_i \sim IIDU(0.1, 0.3)$ (weakly exogenous regressors). In addition to feedbacks, $\kappa_i y_{i,t-1}$, the regressors, x_{it}, contain the strong common factor, f_t, and the weakly cross-sectionally correlated idiosyncratic components, v_{it}, generated as AR(1) processes:

$$f_t = \rho_f f_{t-1} + \sqrt{1-\rho_f^2} v_{ft},\ v_{it} = \rho_{vi} v_{i,t-1} + \sqrt{1-\rho_{vi}^2} \xi_{it},\ t = -49, ..., 0, 1, ..., T, \tag{11}$$

$$v_{ft}, \xi_{it} \sim IIDN(0,1),\ \rho_f = 0.5,\ f_{-50} = 0,\ \rho_{vi} \sim IIDU(0, 0.8),\ v_{i,-50} = 0. \tag{12}$$

The errors e_{it} and ξ_{it} are generated as spatial autoregressive process:

$$e_{it} = \delta_e \sum_{j=1}^{N} w_{ij} e_{jt} + \varepsilon_{it}, \text{ and } \xi_{it} = \delta_\xi \sum_{j=1}^{N} w_{ij} \xi_{jt} + \zeta_{it}, \tag{13}$$

where $\varepsilon_{it} \sim IIDN(0, \sigma_i^2)$, $\sigma_i^2 \sim IIDU(0.5, 1.5)$, and $\zeta_{it} \sim IIDN(0,1)$. The spatial coefficients, δ_e and δ_ξ, capture the degree of cross correlations of the errors, and w_{ij} are the elements of the $N \times N$ spatial weights matrix $\mathbf{W} = (w_{ij})$. We follow Kelejian and Prucha (2007) and assume units are set out on a rectangular grid at locations (s, r), for $r = 1, 2, ..., m_1$ and $s = 1, 2, ..., m_2$ such that $N = m_1 m_2$.[3] $\mathbf{W}$ is a rook type matrix, where two units are neighbors if their Euclidean distance is less than or equal to one. The weights matrix is normalized such that rows sum to one. We set $\delta_e = \delta_\xi = 0.6$. The fixed effects (α_i) and the loadings $(\alpha_{i1}, \alpha_{i2})$ do not change across replications

[3] We consider $(m_1, m_2) \in \{(5,4), (6,5), (10,5), (10,10), (40,25), (75,40)\}$, for $N = 20$, 30, 50, 100, 1000, and 3000, respectively.

and are generated as $\alpha_i \sim IIDN\left(1,1\right)$ and $\left(\alpha_{i1},\alpha_{i2}\right)' \sim IIDN\left(0.5\boldsymbol{\tau}_2, 0.5\mathbf{I}_2\right)$, for $i=1,2,...,N$, where $\boldsymbol{\tau}_2=(1,1)'$, and $\mathbf{I}_2$ is 2×2 identity matrix. We conduct $R=2,000$ replications per each MC experiment and consider sample size combinations $N=20,30,50,100,1000,3000$, and $T=10,20,30,50,100,1000$.

4.2 Findings

Table 1 reports findings for the bias, Root Mean Square Error (RMSE), size (5% nominal level) and power of the MG and jackknifed-MD estimators.[4] The nominal size of the tests is set to 5 percent, and power is computed under the alternatives $E\left(\theta_i\right)=0.9$ for all i. The bias of the MG estimator is very small when regressors are strictly exogenous ($\kappa_i=0$), reported in Part A of Table 1. Size of the tests based on the MG estimator is close to the nominal level of 5% for all values of N and T, except for the few of the smallest sample sizes considered, where the tests based on the MG estimator do not achieve correct size (for $T=10$). But reported size distortions are very mild. Jackknife bias correction is clearly not required in the experiments with strictly exogenous regressors, and we see that the jackknifed-MG estimator has somewhat larger RMSE and consequently lower power.

The MG estimator exhibits downward bias in the weakly exogenous case, reported in Part B of Table 1. This bias declines in T, as expected. As a consequence, we see significant size distortions for sample sizes where the ratio $\sqrt{N}/T$ is not sufficiently small, whereas the size remains close to the nominal level for experiments where T is not too small as compared to N. Jackknife bias correction does help in reducing the bias of the MG estimator, and the size of the tests based on the jackknifed-MG estimator is close to the 5% nominal level for all sample sizes considered, except for $T=10$ where small size distortions are reported. Overall, the simulations accord well with the asymptotic results established in the paper.

5 Conclusion

This paper has established asymptotic results for the MG estimator when the underlying individual estimators are weakly cross-sectionally correlated, and possibly biased. Our high-level setup allows

[4] The MG estimator is computed using the simple average of $\hat{\theta}_i$. Jackknifed-MG is based on averaging the estimators $\tilde{\theta}_i=2\hat{\theta}_i-0.5\left(\hat{\theta}_{ai}+\hat{\theta}_{bi}\right)$, where $\hat{\theta}_i$ is the estimator using the full sample of T periods, and $\hat{\theta}_{ai}$, $\hat{\theta}_{bi}$ are the estimators using the first and the second halves of the sample, respectively. When T is odd one of the observations from the start or the end of the sample is dropped.

for a number of linear or non-linear panel data specifications. We conclude that in large panels MG estimator could be useful, especially when regressors are strictly exogenous. In the case of panels with weakly exogenous regressors we suggest a simple Jackknifed-MG estimator. The random coefficient specification allows for correct inference even when nothing is known about the weak cross-sectional dependence of the errors. This is in contrast to the well known homogenous case, where cross-sectional dependence of errors results in incorrect inference unless the nature of the cross-sectional error dependence is known and can be taken into account.

Table 1: Monte Carlo Findings for the Performance of MG and Jackknifed-MG Estimators

	Bias (x100)						RMSE (x100)						Size(x100)						Power (x100)					
$N \setminus T$	10	20	30	50	100	1000	10	20	30	50	100	1000	10	20	30	50	100	1000	10	20	30	50	100	1000
Part A:	**Experiments with strictly exogenous regressors, $\kappa_i = 0$ for $i = 1, 2, ..., N$**																							
	MG																							
20	0.04	-0.06	-0.15	-0.09	-0.10	-0.02	14.80	12.69	12.13	11.67	11.33	11.10	9.3	8.1	7.3	6.8	6.7	6.5	16.95	15.05	16.30	15.60	16.00	16.45
30	0.08	-0.21	0.03	-0.07	0.01	-0.07	12.15	10.32	9.71	9.41	9.19	9.01	8.5	6.7	5.8	5.4	5.3	5.0	20.00	19.25	20.20	19.75	20.50	20.35
50	-0.23	-0.14	-0.06	-0.15	-0.13	-0.14	9.31	8.23	7.85	7.51	7.42	7.23	7.8	7.4	7.1	7.1	6.0	5.7	24.30	27.65	28.30	27.70	29.55	29.50
100	-0.06	0.04	-0.03	0.00	0.02	-0.04	6.74	5.87	5.51	5.30	5.25	5.09	8.3	7.3	6.2	5.9	6.1	5.7	40.10	46.00	48.20	49.85	51.30	51.10
1000	0.02	0.00	0.01	-0.01	0.00	0.01	2.03	1.85	1.74	1.69	1.67	1.61	6.5	7.2	6.3	6.3	6.5	5.7	99.95	100.00	100.00	100.00	100.00	100.00
3000	0.01	-0.02	-0.01	-0.01	0.01	0.00	1.23	1.07	1.02	0.97	0.97	0.94	7.7	7.6	6.8	5.7	6.4	6.4	100.00	100.00	100.00	100.00	100.00	100.00
	Jackknifed-MG																							
20	-0.06	-0.04	-0.17	-0.09	-0.11	-0.02	16.53	12.86	12.19	11.70	11.34	11.10	8.7	8.3	7.2	7.1	6.8	6.6	16.05	14.60	16.00	15.25	16.05	16.35
30	-0.01	-0.27	0.07	-0.09	-0.01	-0.07	13.36	10.62	9.85	9.45	9.21	9.01	7.8	7.0	5.9	5.2	5.3	5.0	17.35	18.45	20.25	19.90	20.80	20.35
50	-0.09	-0.18	-0.08	-0.14	-0.13	-0.14	10.34	8.35	7.91	7.53	7.43	7.23	7.5	7.8	7.0	7.1	6.1	5.7	21.60	27.55	27.85	27.85	29.35	29.45
100	-0.03	0.04	-0.04	0.02	0.02	-0.04	7.52	5.98	5.58	5.30	5.25	5.09	8.2	7.3	6.4	5.8	6.2	5.7	34.95	45.15	47.20	49.95	51.10	51.15
1000	-0.02	0.01	0.01	-0.01	0.00	0.01	2.28	1.89	1.75	1.70	1.67	1.61	7.0	7.8	6.6	6.0	6.5	5.7	99.65	100.00	100.00	100.00	100.00	100.00
3000	0.00	-0.02	-0.01	-0.01	0.01	0.00	1.35	1.09	1.03	0.97	0.97	0.94	7.5	7.5	7.1	5.8	6.7	6.4	100.00	100.00	100.00	100.00	100.00	100.00
Part B:	**Experiments with weakly exogenous regressors, $\kappa_i \sim IIDU\,(0.1, 0.3)$**																							
	MG																							
20	-2.42	-1.36	-0.91	-0.52	-0.33	-0.07	14.53	12.85	12.01	11.64	11.35	11.07	9.2	8.0	6.5	7.0	6.3	6.4	13.10	15.10	15.40	15.60	16.05	16.25
30	-2.37	-1.51	-0.75	-0.55	-0.30	-0.12	12.05	10.25	9.60	9.35	9.18	9.01	9.1	7.4	5.5	5.6	5.4	5.0	14.05	16.20	18.35	18.90	19.85	20.55
50	-2.54	-1.31	-0.92	-0.61	-0.34	-0.17	9.54	8.24	7.81	7.58	7.44	7.23	9.2	7.5	7.0	6.5	6.6	5.9	18.10	24.00	25.50	26.80	28.70	29.00
100	-2.57	-1.20	-0.87	-0.58	-0.27	-0.07	7.05	5.89	5.48	5.34	5.25	5.08	10.4	7.3	6.0	6.4	5.8	6.0	27.35	37.75	41.75	46.00	48.95	50.50
1000	-2.44	-1.24	-0.84	-0.53	-0.25	-0.03	3.20	2.18	1.94	1.77	1.68	1.61	27.1	12.8	9.6	7.6	6.4	5.8	96.95	100.00	100.00	100.00	100.00	100.00
3000	-2.42	-1.24	-0.83	-0.49	-0.24	-0.03	2.72	1.62	1.31	1.10	1.00	0.95	59.9	25.8	15.3	10.0	7.4	6.5	100.00	100.00	100.00	100.00	100.00	100.00
	Jackknifed-MG																							
20	-0.20	-0.18	-0.07	0.02	-0.08	-0.04	16.32	13.04	12.14	11.65	11.37	11.08	9.4	7.8	7.1	7.0	6.1	6.4	14.10	16.35	16.85	16.75	16.65	16.30
30	-0.36	-0.39	0.06	-0.05	-0.06	-0.09	13.31	10.40	9.77	9.40	9.19	9.01	8.2	6.6	5.7	5.4	5.2	5.0	15.60	17.90	19.90	20.05	20.55	20.70
50	-0.40	-0.13	-0.13	-0.11	-0.09	-0.14	10.25	8.31	7.90	7.59	7.45	7.23	7.8	7.1	6.7	6.6	6.6	5.9	20.50	27.50	27.55	28.95	29.75	28.95
100	-0.43	-0.04	-0.02	-0.09	-0.02	-0.04	7.41	5.95	5.52	5.34	5.26	5.09	8.8	7.2	6.3	6.1	5.8	6.0	30.90	44.25	48.25	49.10	50.40	50.85
1000	-0.33	-0.05	-0.02	-0.04	0.01	0.00	2.36	1.84	1.77	1.70	1.66	1.61	8.2	6.2	6.3	6.5	6.2	5.9	99.15	100.00	100.00	100.00	100.00	100.00
3000	-0.28	-0.06	-0.01	0.01	0.01	0.00	1.41	1.06	1.02	0.99	0.97	0.95	9.2	7.2	6.9	6.5	6.7	6.4	100.00	100.00	100.00	100.00	100.00	100.00

Notes: The DGP is given by $y_{it} = \alpha_i + \theta_i x_{it} + e_{it}$, and $x_{it} = \alpha_{i1} + \kappa_i y_{i,t-1} + \alpha_{i2} f_t + v_{it}$, where f_t and v_{it} are AR(1) processes generated according to $f_t = \rho_f f_{t-1} + \sqrt{1-\rho_f^2} v_{ft}$, and $v_{it} = \rho_{vi} v_{i,t-1} + \sqrt{1-\rho_{vi}^2} \xi_{it}$, see (11)-(12). The fixed effects (α_i) and loadings (α_{i1}, α_{i2}) do not change across replications and are generated as $\alpha_i \sim IIDN\,(1,1)$, $(\alpha_{i1}, \alpha_{i2})' \sim IIDN\,(0.5\boldsymbol{\tau}_2, 0.5\mathbf{I}_2)$. Innovations e_{it} and ξ_{it} are generated according to the spatial autoregressive processes in (13) with the spatial AR coefficients $\delta_e = \delta_\xi = 0.6$.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands. doi: 10.1007/978-94-015-7799-1.

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data* (5th ed.). John Wiley & Sons Ltd. ISBN: 978-1-118-67232-7.

Chudik, A., M. H. Pesaran, and J.-C. Yang (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics 33*(6), 816–836. doi: 10.1002/jae.2623.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*(3), 991–1030. doi: 10.1093/restud/rdv007.

Hsiao, C. and M. H. Pesaran (2008). Random coefficient models. In *The Econometrics of Panel Data*, Chapter 6, pp. 185–213. Springer. ISBN: 978-3-540-75892-1.

Kelejian, H. H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics 140*(1), 131–154. doi: 10.1016/j.jeconom.2006.09.005.

Pesaran, M. H. (2015). *Time series and panel data econometrics*. Oxford University Press. ISBN: 9780198736912.

Pesaran, M. H., T. Schuermann, and S. Weiner (2004). Modelling Regional Interdependencies using a Global Error-Correcting Macroeconometric Model. *Journal of Business and Economics Statistics 22*, 129–162. doi:10.1198/073500104000000019.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics 68*(1), 79–113. doi: 10.1016/0304-4076(94)01644-F.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*(2), 182–202. doi: 10.1016/j.jeconom.2010.12.003.

A Proofs

Proof of Theorem 1. Let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \boldsymbol{\eta}_2', ..., \boldsymbol{\eta}_N')'$ and note that

$$\hat{\boldsymbol{\theta}}_{MG} = \frac{1}{N}\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)\left(\hat{\boldsymbol{\theta}}_T - \underline{\boldsymbol{\theta}}_0\right) + \frac{1}{N}\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)\underline{\boldsymbol{\theta}}_0, \tag{A.1}$$

where $\boldsymbol{\tau}_N$ is $N \times 1$ vector of ones, and $\mathbf{I}_k$ is a $k \times k$ identity matrix. Also

$$\underline{\boldsymbol{\theta}}_0 = \left(\boldsymbol{\tau}_N' \otimes \boldsymbol{\theta}_0\right) + \boldsymbol{\eta}. \tag{A.2}$$

Using (A.1)-(A.2) in (1) yields

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}_0\right) = \frac{1}{\sqrt{N}}\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)\boldsymbol{\eta} + \frac{1}{\sqrt{NT}}\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \underline{\boldsymbol{\theta}}_0\right).$$

Note that $\underline{\boldsymbol{\theta}}_0$ is $Nk \times 1$ and $\boldsymbol{\theta}_0$ is $k \times 1$. Let

$$\boldsymbol{\zeta}_{NT} = \frac{1}{\sqrt{NT}}\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \underline{\boldsymbol{\theta}}_0\right), \tag{A.3}$$

and note that it can also be written equivalently as

$$\boldsymbol{\zeta}_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{iT} - \boldsymbol{\theta}_{i0}\right). \tag{A.4}$$

We now establish that $\boldsymbol{\zeta}_{NT} \to_p \mathbf{0}$, by showing that $E\left(\boldsymbol{\zeta}_{NT}\right) \to \mathbf{0}$ and $\left\|Var\left(\boldsymbol{\zeta}_{NT}\right)\right\| \to 0$. Consider first the mean of $\boldsymbol{\zeta}_{NT}$. Under Assumption 1.a, $E\left(\boldsymbol{\zeta}_{NT}\right) = \mathbf{0}$, regardless of the relative rate of N and T. Under Assumption 1.b, we have

$$E\left(\boldsymbol{\zeta}_{NT}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{K_i}{T} + O\left(T^{-3/2}\right)\right] = \frac{\sqrt{N}}{T}\frac{1}{N}\sum_{i=1}^{N}K_i + O\left(\frac{\sqrt{N}}{T^{3/2}}\right),$$

where note that $N^{-1}\sum_{i=1}^{N}|K_i| \leq \sup_i |K_i| < \infty$. Hence, under Assumption 1.b, $E\left(\boldsymbol{\zeta}_{NT}\right) \to 0$, as N and $T \to \infty$ such that $\sqrt{N}/T \to 0$.

Consider now (using (A.3))

$$Var\left(\boldsymbol{\zeta}_{NT}\right) = \frac{1}{NT}\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)Var\left(\sqrt{T}\hat{\boldsymbol{\theta}}_T\right)\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)',$$

and, using spectral norm, note that $\left\|Var\left(\boldsymbol{\zeta}_{NT}\right)\right\| \leq \frac{1}{NT}\left\|\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right\|^2\left\|\mathbf{V}_{NT}\right\|$, where

$$\mathbf{V}_{NT} = Var\left(\sqrt{T}\hat{\boldsymbol{\theta}}_T\right). \tag{A.5}$$

But $\left\|\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right\|^2 = \lambda_{\max}\left(\boldsymbol{\tau}_N'\boldsymbol{\tau}_N \otimes \mathbf{I}_k\right) = N$, and $\left\|Var\left(\boldsymbol{\zeta}_{NT}\right)\right\| \leq T^{-1}\left\|\mathbf{V}_{NT}\right\|$. Hence, $\left\|Var\left(\boldsymbol{\zeta}_{NT}\right)\right\| \to 0$, as N and $T \to \infty$ if $T^{-1}\left\|\mathbf{V}_{NT}\right\| \to 0$, which is clearly satisfied under Assumption 2.[5] Result (2) now follows by noting that under Assumption 3, $N^{-1/2}\left(\boldsymbol{\tau}_N' \otimes \mathbf{I}_k\right)\boldsymbol{\eta} \overset{a}{\sim} N\left(\mathbf{0}_k, \boldsymbol{\Omega}_\eta\right)$. This completes the proof. ■

Proof of Theorem 2. Note that

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{iT} - \hat{\boldsymbol{\theta}}_{MG} &= \left(\hat{\boldsymbol{\theta}}_{iT} - \boldsymbol{\theta}_{i0}\right) + \left(\boldsymbol{\theta}_{i0} - \boldsymbol{\theta}_0\right) - \left(\hat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}_0\right) \\
&= \boldsymbol{\eta}_i + \boldsymbol{\xi}_{iT} - \mathbf{v}_{NT},
\end{aligned}$$

where $\boldsymbol{\xi}_{iT} = \hat{\boldsymbol{\theta}}_{iT} - \boldsymbol{\theta}_{i0}$, $\mathbf{v}_{NT} = \hat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}_0$, and we used Assumption 3. Hence

$$\begin{aligned}
\left(\hat{\boldsymbol{\theta}}_{iT} - \hat{\boldsymbol{\theta}}_{MG}\right)\left(\hat{\boldsymbol{\theta}}_{iT} - \hat{\boldsymbol{\theta}}_{MG}\right)' &= \boldsymbol{\eta}_i\boldsymbol{\eta}_i' + \boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}' + \mathbf{v}_{NT}\mathbf{v}_{NT}' \\
&\quad + \boldsymbol{\eta}_i\boldsymbol{\xi}_{iT}' - \boldsymbol{\eta}_i\mathbf{v}_{NT}' + \boldsymbol{\xi}_{iT}\boldsymbol{\eta}_i' \\
&\quad - \boldsymbol{\xi}_{iT}\mathbf{v}_{NT}' - \mathbf{v}_{NT}\boldsymbol{\eta}_i' - \mathbf{v}_{NT}\boldsymbol{\xi}_{iT}'.
\end{aligned} \tag{A.6}$$

Assuming $E\left\|\boldsymbol{\eta}_i\right\|^{2+\epsilon} < K$ for some $\epsilon > 0$, we have $N^{-1}\sum_{i=1}^N \boldsymbol{\eta}_i\boldsymbol{\eta}_i' \to_p \boldsymbol{\Omega}_\eta$, as $N \to \infty$ (regardless of T). We establish convergence of the cross section averages of the remaining terms on the right side of (A.6) next. Proof of Theorem 1 established

$$E\left\|\sqrt{N}\mathbf{v}_{NT}\right\| < K, \tag{A.7}$$

as N and $T \to \infty$ at any rate under Assumption 1.a, and as N and $T \to \infty$ such that $\sqrt{N}/T \to 0$ under Assumption 1.b. In addition, Assumption 2 implies

$$\boldsymbol{\xi}_{iT} = \hat{\boldsymbol{\theta}}_{iT} - \boldsymbol{\theta}_{i0} = O_p\left(T^{-1/2}\right). \tag{A.8}$$

Using (A.8) and Assumption 3 , we have $N^{-1}\sum_{i=1}^N \boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}' = O_p\left(T^{-1/2}\right)$, and $N^{-1}\sum_{i=1}^N \boldsymbol{\eta}_i\boldsymbol{\xi}_{iT}' = O_p\left(T^{-1/2}\right)$. Using (A.7), (A.8) and Assumption 3 , we obtain $N^{-1}\sum_{i=1}^N \mathbf{v}_{NT}\mathbf{v}_{NT}' = \mathbf{v}_{NT}\mathbf{v}_{NT}' \to_p \mathbf{0}$, $N^{-1}\sum_{i=1}^N \boldsymbol{\eta}_i\mathbf{v}_{NT}' = \left(N^{-1}\sum_{i=1}^N \boldsymbol{\eta}_i\right)\mathbf{v}_{NT}' \to_p \mathbf{0}$, and $N^{-1}\sum_{i=1}^N \boldsymbol{\xi}_{iT}\mathbf{v}_{NT}' = \left(N^{-1}\sum_{i=1}^N \boldsymbol{\xi}_{iT}\right)\mathbf{v}_{NT}' \to_p \mathbf{0}$, as $N, T \to \infty$ such that $\sqrt{N}/T \to 0$. Hence $\hat{\boldsymbol{\Omega}}_\eta \to_p \boldsymbol{\Omega}_\eta$, as $N, T \to \infty$ such that $\sqrt{N}/T \to 0$. This completes the proof. ■

[5] Note that $\left\|\mathbf{V}_{NT}\right\| = \lambda_{\max}^{1/2}\left(\mathbf{V}_{NT}^2\right) = \lambda_{\max}\left(\mathbf{V}_{NT}\right)$.