



SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN



WORKING PAPER SERIES

**Small Sample Estimation in Dynamic
Panel Data Models**

by

Lorelie DA. Santos
and
Erniel B. Barrios

UPSS Working Paper No. 2010-02
January 2010

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

Small Sample Estimation in Dynamic Panel Data Models

Lorelie DA. Santos

*Assistant Professor
UP School of Statistics*

Erniel B. Barrios

*Professor
UP School of Statistics*

Abstract

The within-groups (WG) and first difference generalized method of moments (FD-GMM) estimators of a dynamic panel data (DPD) model are investigated. WG estimator performs best for long panels such as those with time dimension as large as 50. The advantage of FD-GMM estimator is observed on panels that are long and wide, say with time dimension at least 25 and cross-section dimension size of at least 30. We developed a parametric bootstrap version of WG and FD-GMM estimators for small samples. While WG and FDGMM estimators are asymptotically optimal, it is also true for small samples using the bootstrap method.

Keywords: *dynamic panel data model, within-groups estimator, first-difference generalized method of moments estimator, parametric bootstrap*

1 Introduction

Panel data combine cross-sectional and time series information, and because of the inherent nature of data involving a time series for each unit leads, a dynamic parameter is usually included modeling with such data. Dynamic panel data (DPD) model postulates the lagged dependent variable as explanatory variable. Just like in univariate time series analysis, modeling the dependency of the time series on its past value(s) gives valuable insights on the temporal behavior of the series. (Baltagi, 2005) noted that many economic relationships are dynamic in nature and the panel data allow the researcher to better understand the dynamics of adjustment exhibited by the units.

A good number of dynamic panel data estimators have been proposed and thoroughly characterized in the literature. The within-group (WG) estimator which provides consistent estimate for static models is among the early methods of estimation for DPD model introduced in the literature. In DPD models however, Nickell (1981) showed that the WG estimator of the coefficient of the lagged dependent variable parameter is downward biased and the bias only disappears as the number of time units grows large. Thus, the WG estimator is known to be biased whenever the time-dimension T is fixed even if the cross-section dimension N is large.

The inconsistency of the WG estimator leads to the development of DPD coefficient parameter estimators that are consistent for large N and fixed or large T , e.g., the use of instrumental variables (IV). For the IV estimators, (Anderson and Hsiao (1981) used either the dependent variable lagged two periods or its first-differences as instruments. Even the development of the generalized method of moments (GMM) estimators for DPD coefficient parameters is based on the IV approach. (Arellano and Bond, 1991) proposed GMM estimator

that uses all available lags at each period as instruments for the equations in first differences and this estimator is now known as the first-difference generalized method of moments estimator (FD-GMM). (Arellano and Bover, 1995) proposed the level GMM estimator which is based on the level of the model and uses lagged difference variables as instruments. (Blundell and Bond, 1998) further proposed the now called system GMM estimator which uses both the lags of the level and first difference as instruments.

The DPD model estimators have exhibited good asymptotic properties, see for example (Nickell, 1981), (Kiviet, 1995), and (Alvarez and Arellano, 2003). Some papers attempted to investigate the small sample properties of the said estimators, e.g., (Judson and Owen, 1999) and (Hayakawa, 2007). Alternative estimators that work well under small sample have to be developed since many panel data may have large number of units but with a relatively short time series realization. There are numerous studies on the properties of dynamic panel data model but are mostly focusing on data sets with large cross-section and small time dimensions. Other studies highlight datasets with sizeable cross-section dimensions and moderate size time dimensions.

We used intensive simulations to investigate both the small and large sample properties of two of the simplest and oldest DPD estimators, the within-groups and first-difference generalized method of moments estimators of the AR(1) DPD model. We also propose the use parametric bootstrap procedure to the WG and FD-GMM for the boundary scenario, i.e., when asymptotic optimality of WG and FD-GMM fail.

As (Houkannounon, 2008) pointed out, the application of bootstrap methodology in panel data analysis is currently in its embryonic stage. The bootstrap estimators proposed in this study can answer the possible limitations of the estimators by (Alvarez and Arellano, 2003). Over a short period, it is common for processes over time to be easily affected by random shocks. Thus, if long period data are used, it is very likely that structural change will manifest and the modeler will either incorporate the change into the model (more complicated), or analyze the series by shorter periods, leading to small sample data where the proposed estimator is applicable.

2 Dynamic Panel Data Model

Suppose the dynamic behavior of a time series for unit i (y_{it}) is characterized by the presence of a lagged dependent variable among the regressors, i.e.

$$y_{it} = \alpha y_{i,t-1} + \underline{x}'_{it} \underline{\beta} + \mu_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \quad 1$$

where α is a constant, \underline{x}'_{it} is $1 \times K$ vector of explanatory/exogenous variables, and $\underline{\beta}$ is $K \times 1$ vector of regression coefficients. μ_{it} follows a two-way error component model $\mu_{it} = \eta_i + \lambda_t + v_{it}$ where η_i and λ_t are the (unobserved) individual and time specific effects, which are assumed to stay constant for given i over t and for a given t over i , respectively; and v_{it} represents the unobserved random shocks over i and t . Hsiao (2003), Baltagi (2005) and Hsiao and Tahmiscioglu (2008) give detailed discussions on dynamic panel data models.

Consider the following AR(1) dynamic panel data model without exogenous variable

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \quad 2$$

where y_{it} is the dependent variable, α is the regression coefficient (parameter of interest) with $|\alpha| < 1$, η_i is the unobserved heterogeneity or individual effect which has mean 0 and variance $\sigma_\eta^2 (\geq 0)$ and v_{it} is unobserved disturbance with mean 0 and variance $\sigma_v^2 (> 0)$.

Alvarez and Arellano (2003) defined the WG estimator as

$$\hat{\alpha}_{WG} = \frac{\sum_{i=1}^N \underline{x}_i' \underline{Q}_T \underline{y}_i}{\sum_{i=1}^N \underline{x}_i' \underline{Q}_T \underline{x}_i} \quad 4$$

where $\underline{Q}_T = \underline{I}_T - \underline{t}_T \underline{t}_T' / T$, \underline{Q}_T is called the WG operator of order T . WG estimator may also be written in terms of the forward orthogonal deviations operator \underline{A} , an upper triangular-like matrix, with dimension $(T-1) \times T$ and with the following characteristics: $\underline{A}' \underline{A} = \underline{Q}_T$, $\underline{A} \underline{A}' = \underline{I}_{T-1}$ and $\underline{A} \underline{t}_T = 0$. Let $\underline{x}^* = \underline{A} \underline{x}$ and $\underline{y}^* = \underline{A} \underline{y}$, then WG estimator is given by

$$\hat{\alpha}_{WG} = \frac{\underline{x}^* \underline{y}^*}{\underline{x}^* \underline{x}^*} \quad 5$$

Alvarez and Arellano (2003) further analyzed an asymptotically efficient FD-GMM estimator given by

$$\hat{\alpha}_{FD-GMM} = \frac{\underline{x}^{*'} \underline{Z} (\underline{Z}' \underline{Z})^{-1} \underline{Z}' \underline{y}^*}{\underline{x}^{*'} \underline{Z} (\underline{Z}' \underline{Z})^{-1} \underline{Z}' \underline{x}^*} \quad 6$$

where $\underline{Z} = (\underline{Z}_1', \dots, \underline{Z}_N')$. A computationally useful alternative expression for $\hat{\alpha}_{FD-GMM}$ is:

$$\hat{\alpha}_{FD-GMM} = \frac{\sum_{t=1}^{T-1} \underline{x}_t^{*'} \underline{M}_t \underline{y}_t^*}{\sum_{t=1}^{T-1} \underline{x}_t^{*'} \underline{M}_t \underline{x}_t^*} \quad 7$$

where \underline{x}_t^* and \underline{y}_t^* are the $N \times 1$ vectors whose i^{th} elements are x_{it}^* and y_{it}^* , respectively, $\underline{M}_t = \underline{Z}_t (\underline{Z}_t' \underline{Z}_t)^{-1} \underline{Z}_t'$ and \underline{Z}_t is the $N \times t$ matrix whose i^{th} row is z_{it}' .

Alvarez and Arellano (2003) derived the asymptotic properties of several dynamic panel data estimators namely, within groups (WG), first-difference generalized method of moments (FD-GMM), limited information maximum likelihood (LIML), crude GMM and random effects ML estimators of the AR(1) parameter of a simple DPD model with random effects.

As observed by Alvarez and Arellano (2003), $\hat{\alpha}_{WG}$ is consistent for $T \rightarrow \infty$, regardless of N . However, when $N \rightarrow \infty$, the asymptotic distribution of $\hat{\alpha}_{WG}$ may contain an asymptotic bias term, depending on the relative rates of increase of T and N . When $0 < \lim(T/N) < \infty$, the asymptotic variance of WG estimator is the same as the of GMM estimator and they have similar (negative) asymptotic biases in which for WG has order

$(1 + \alpha)/T$. On the otherhand, $\hat{\alpha}_{FD-GMM}$ is a consistent estimator for α as both $N \rightarrow \infty$ and $T \rightarrow \infty$, provided that $(\log T)^2/N \rightarrow 0$. Also, the number of the FD-GMM orthogonality conditions $q = T(T-1)/2$ tends to infinity as $T \rightarrow \infty$. The F $\hat{\alpha}_{FD-GMM}$ is asymptotically normal, provided that $T \rightarrow \infty$ and $\lim(T/N) = 0$. When $0 < \lim(T/N) < \infty$, the asymptotic variance of FD-GMM estimator is the same as the WG estimator and they have similar expression for their (negative) asymptotic biases in which for FD-GMM has order $(1 + \alpha)/N$. When $T < N$, the asymptotic bias of GMM is smaller than the WG bias, when $T = N$, the two asymptotic biases are equal and when $N/T \rightarrow 0$, the asymptotic bias in the WG estimator disappears.

Kiviet (1995) reported from a simulation study that the variance of the WG estimators is usually much smaller than the variance of consistent GMM estimators. See Nickell (1981), Kiviet (1995), Judson and Owen (1999), Arellano and Bond (1991), Bond (2002), and Alvarez and Arellano (2003) for further details.

3 The Bootstrap Method

The bootstrap is a useful tool for estimation in finite samples. Bootstrap procedure entails the estimation of parameters in a model through resampling with a large number of replications, Mackinnon (2006). Efron (1979) developed the idea of bootstrap procedure known as a nonparametric method of resampling with replacement and it stems from older resampling methods such as the jackknife method and delta method. Originally, the bootstrap requires independent observations, i.e., \underline{x} consisting of n observations x_1, x_2, \dots, x_n , a random sample from the true distribution $F(x)$ generating the data.

For time series models, the sieve bootstrap and the block bootstrap are more popular. The sieve bootstrap starts by fitting the most adequate model and the behavior of the empirical distribution of the residuals is analyzed. The bootstrap errors ε_t^* are generated by resampling $T + k$ times from the empirical distribution $\tilde{F}(\tilde{\varepsilon}_t)$, $k > p$. In order to generate the b^{th} bootstrap sample $\underline{y}_b^* = [y_{1,b}^*, y_{2,b}^*, \dots, y_{T,b}^*]$, each element of \underline{y}_b^* is determined by the recursion $y_{t,b}^* = \sum_{i=0}^p \phi_i y_{t-i,b}^* + \varepsilon_{t,b}^*$, where the starting values for $y_{t,b}^*$ are set to zero and the first k generated values are thrown away, so that the needed b^{th} bootstrap sample $\underline{y}_b^* = [y_{1,b}^*, y_{2,b}^*, \dots, y_{T,b}^*]$ is obtained. See Bühlmann (2002) and Mackinnon (2006) for a more comprehensive discussion of bootstrap methods for time series models.

To define the bootstrap for dynamic panel data models, suppose y is defined as measurements from different cross-section units of the population over different time periods, so that the data can be represented by

$$\underline{y}_{PD} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1,T-1} & y_{1,T} \\ y_{21} & y_{22} & \cdots & y_{2,T-1} & y_{2,T} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{N-1,1} & y_{N-1,2} & \cdots & y_{N-1,T-1} & y_{N-1,T} \\ y_{N,1} & y_{N,2} & \cdots & y_{N,T-1} & y_{N,T} \end{bmatrix} = \begin{bmatrix} \underline{y}_{c1} \\ \underline{y}_{c2} \\ \vdots \\ \underline{y}_{cN-1} \\ \underline{y}_{cN} \end{bmatrix} = [\underline{y}_{t1} \quad \underline{y}_{t2} \quad \cdots \quad \underline{y}_{tT-1} \quad \underline{y}_{tT}] \quad 8$$

The b^{th} bootstrap sample is the $N \times T$ matrix \underline{y}_b^* , given by

$$\underline{y}_b^* = \begin{bmatrix} y_{11,b}^* & y_{12,b}^* & \cdots & y_{1,T-1,b}^* & y_{1,T,b}^* \\ y_{21,b}^* & y_{22,b}^* & \cdots & y_{2,T-1,b}^* & y_{2,T,b}^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{N-1,1,b}^* & y_{N-1,2,b}^* & \cdots & y_{N-1,T-1,b}^* & y_{N-1,T,b}^* \\ y_{N,1,b}^* & y_{N,2,b}^* & \cdots & y_{N,T-1,b}^* & y_{N,T,b}^* \end{bmatrix} = \begin{bmatrix} \underline{y}_{c1,b}^* \\ \underline{y}_{c2,b}^* \\ \vdots \\ \underline{y}_{cN-1,b}^* \\ \underline{y}_{cN,b}^* \end{bmatrix} = [\underline{y}_{t1,b}^* \quad \underline{y}_{t2,b}^* \quad \cdots \quad \underline{y}_{tT-1,b}^* \quad \underline{y}_{tT,b}^*] \quad 9$$

\underline{y}_b^* is generated by doing the AR-sieve bootstrap procedure for panel data. Hounkannounon (2008) identified five bootstrap methodologies for panel data namely: i.i.d. bootstrap, individual bootstrap, temporal bootstrap, block bootstrap and double resampling bootstrap. The i.i.d bootstrap refers to the bootstrap procedure defined by Efron (1979). Each of the $N \times T$ elements of the observed data matrix \underline{y}_{PD} is given $1/NT$ probability in the empirical distribution $\tilde{F}(y)$. The elements of the b^{th} bootstrap sample \underline{y}_b^* in 9 are obtained by resampling with replacement from the empirical distribution $\tilde{F}(y)$.

The rows of \underline{y}_{PD} in 8 are resampled with replacement in order to determine the b^{th} bootstrap sample \underline{y}_b^* of the form 9 in the individual bootstrap procedure. On the other hand, in the temporal bootstrap procedure, the columns of \underline{y}_{PD} in 8 are resampled with replacement in order to create the b^{th} bootstrap sample \underline{y}_b^* of the form 9. The resampling procedure for the block bootstrap is in the temporal dimension, so that the data matrix of the form 8 is used. The difference of block bootstrap from temporal bootstrap lies in the sampling of blocks of columns of \underline{y}_{PD} in 8 instead of single column/period in the temporal bootstrap case. Let $T = Kl$, where l is the length of a block and thus, there are K non-overlapping blocks. Block bootstrap resampling entails the construction of \underline{y}_b^* such as 9 with columns obtained by resampling with replacement the K non-overlapping blocks columns of \underline{y}_{PD} in 8.

4 Bootstrap Procedure for Dynamic Panel Data Estimators

While WG and FD-GMM estimators are established to possess optimal properties for large samples, there are doubts on their performance under small samples. Many panel data are usually formed from small samples of time points and/or panel units because of the structural change or random shocks that may occur in bigger datasets. For small samples, we propose to use the parametric bootstrap to WG and FD-GMM in mitigating the bias and inconsistency that these estimators are known to possess.

We consider the $AR(1)$ dynamic panel data model:

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it}, \quad \begin{matrix} i = 1, 2, \dots, N \\ t = 1, 2, \dots, T \end{matrix} \quad 10$$

where α is the parameter, η_i is the individual effect with mean zero and variance σ_η^2 , and disturbances v_{it} with mean zero and variance σ_v^2 . The bootstrap procedure below follows an ARSieve with the following steps:

Step 1: Given $\{y_{it}\}$ generated from the model in 10, we have N time-series data

$$\{y_{it}\} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} y_{10} & y_{11} & \cdots & y_{1T} \\ y_{20} & y_{21} & \cdots & y_{2T} \\ & & \ddots & \\ y_{N0} & y_{N1} & \cdots & y_{NT} \end{bmatrix}. \quad 11$$

For each cross-section unit i , we assume an $AR(1)$ model with slope α_i and intercept η_i , i.e., $y_{it} = \alpha_i y_{it-1} + \eta_i + v_{it}$, $t = 1, 2, \dots, T$. Using the method of least squares we obtain the estimators $\hat{\alpha}_i$ and $\hat{\eta}_i$ for the parameters.

Step 2: Determine $\bar{\alpha}$, the average of all $\hat{\alpha}_i$'s over all cross-section units, i.e.,

$$\bar{\alpha} = \frac{\sum_{i=1}^N \hat{\alpha}_i}{N}.$$

Step 3: For a fixed cross-section unit i , the predicted values $\hat{y}_{it} = \bar{\alpha} y_{it-1} + \hat{\eta}_i$, $t = 1, 2, \dots, T$, is computed and used to compute the

$$MSE = \hat{\sigma}_{v,i}^2 = \frac{\sum_{t=1}^T (y_{it} - \hat{y}_{it})^2}{(T-2)}. \text{ Simulate } B \text{ sets of } v_{it}^* \text{ from } N(0, \hat{\sigma}_{v,i}^2).$$

Step 4: Form one reconstructed panel data $\{y_{it}^*\}$ using $\bar{\alpha}$ in step 2, $\hat{\eta}_i$ in step 1 and one of the B sets of v_{it}^* in step 3. The reconstructed panel data $\{y_{it}^*\}$ is obtained using equation 10, i.e., $y_{it}^* = \bar{\alpha} y_{it-1}^* + \hat{\eta}_i + v_{it}^*$, $t = 1, 2, \dots, T$, where $y_{i0}^* = y_{i0}$ comes from $\{y_{it}\}$, the i th element of the first column of matrix 11.

Step 5: Repeat step 4 $B-1$ times, taking note that the used set of v_{it}^* should not be used again in subsequent reconstruction of the panel data. There will be B panel data $\{y_{it}^*\}$ available in the end.

Step 6: Compute for WG and FD-GMM estimators using equations 5 and 7 respectively, for each of the B panel data sets formed in steps 3 and 4.

Step 7: Do iid bootstrap from the B estimates in step 6.

When sample size is small, there is a tendency for estimators based on asymptotic optimality to become erratic. The AR sieve is used to reconstruct as many time series as possible that capture the same structure as the data. Resampling from each of the recreated

data and computation of WG and FD-GMM for each resamples can alleviate instability caused by small samples, inheriting the optimal properties of the bootstrap methods.

5 Simulation Study

In the simulation, we used AR(1) with individual random effects model given in equation 10, i.e., $y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$, $i = 1, 2, \dots, N$, $t = 1, 2, \dots, T$. We used an extensive Monte

Carlo design that aims to examine both the asymptotic and finite (small) sample properties of the two estimators of the parameter α . In studies where the asymptotic properties are examined, the cross-section dimension go as large as 500 (see Everaert, and Pozzi (2007)) or as small as 50 (see Alvarez and Arellano (2003)). We consider $N = 50$ corresponding to the large cross-section dimension scenario. On the other hand, $T = 50$ is the largest time dimension used in studies about asymptotic properties (see Alvarez and Arellano (2003)). Some studies use smaller T values such as $T = 10$ and $T = 25$ along with $N = 50$ and $N = 100$ to show the asymptotic properties of estimators, especially the GMM type, see Alvarez and Arellano (2003). We assume large time dimension here with $T = 50$. When smaller T is used such as, $T = 4$ and $T = 6$ (see Hayakawa (2007), Chigira and Yamamoto (2006)), even N is as large as 100, it still exhibit small sample properties of the estimator, specifically the GMM estimator. We consider as small time dimensions cases with $T = 3$ and $T = 5$. If the objective is to examine the finite sample properties of estimators especially the WG estimator, small to moderate sizes for both the cross-section and time dimensions are commonly used, e.g., $N = 20$; $N = 100$ and $T = 5$; $T = 10$; $T = 20$; $T = 30$, for details, see Judson and Owen (1999).

There are few studies where the time series and cross-section dimensions are both small such, e.g., $N, T = 10, 20$ (see Bun and Kiviet (2001)). Small cross-section dimensions considered are $N = 10$ and $N = 20$, and moderate cross-section dimensions are $N = 30$ and $N = 40$. We also consider as moderate time dimension cases where $T = 10$ and $T = 25$. Therefore, the values for T and N can capture the settings for both short and wide panel, typical of a micro-panel and long and narrow panel, which is a common set-up for macro-panels.

In panel data, the observations in a particular cross-section unit comprise a time series. Since, we employ a dynamic panel data model, the AR(1) coefficient parameter α , can be viewed as the common slope parameter for the N time series in an $N \times T$ panel data. Thus, given a time series with T observations, our choice of the values of the AR(1) coefficient parameter α ranges from an almost white noise series, where α is very small, e.g., when $\alpha = 0.1$ to an almost unit root series where α is very near to one, that is, when $\alpha = 0.9$.

The values for the variance of the individual effects accounts for both fixed effects when $\sigma_\eta^2 = 0$ and random effects, that is, $\sigma_\eta^2 \neq 0$. The variance of the random disturbance is

set at $\sigma_v^2 = 1$. Table 1 presents the different combination of parameter values for a total of 625 parameter combinations.

Table 1. Monte Carlo designs

$\frac{\sigma_\eta^2}{\sigma_v^2}$	N	T	α
0	10	3	0.1
0.2	20	5	0.2
0.8	30	10	0.5
1	40	25	0.8
1.25	50	50	0.9

First, we assume values of the ratio of the individual effects variance to the random disturbance variance, the possible values of the cross-section unit and time unit sizes that are considered small samples and varying values of the slope parameter. Then we generate N sets of 10,000 v_{it} 's where $v_{it} \sim N(0,1)$ and we generate N η_i 's where $\eta_i \sim N(0, \sigma_\eta^2)$, by choosing one value for the variance ratio $\sigma_\eta^2 / \sigma_v^2$ from Table 1. The generated v_{it} 's and the individual effect for the i^{th} cross-section η_i is used to compute the initial value y_{i0} for each

cross-section unit i , using equation 10, the initial value is $y_{i0} = \frac{\eta_i}{1-\alpha} + \omega_{i0}$, where

$\omega_{i0} = \sum_{j=0}^{\infty} \alpha^j v_{i(-j)}$. Also, we generate $N \times T$ v_{it} 's where $v_{it} \sim N(0,1)$. Then a value for α

is chosen from Table 1 and using the N η_i 's and $N \times T$ v_{it} 's, a set of dynamic panel data $\{y_{it}\}$, $i = 1, 2, \dots, N$; $t = 1, 2, \dots, T$ can be generated from equation 10, i.e., $y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$.

This will give rise to a data matrix of the form 11,

$$\underline{y}_{PD} = \{y_{it}\} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} y_{10} & y_{11} & \cdots & y_{1T} \\ y_{20} & y_{21} & \cdots & y_{2T} \\ & & \vdots & \\ y_{N0} & y_{N1} & \cdots & y_{NT} \end{bmatrix}.$$

The analysis for the asymptotic and finite sample properties of the WG and FD-GMM estimators was done using 100 replications, in this case, 100 panel data sets \underline{y}_{PD} for each of the 625 designs/parameter combinations. The WG and FD-GMM estimates for a total of 62,500 data sets using equations 5 and 7 respectively. The mean, median, quartiles of WG and FD-GMM estimates for each of the 625 sets containing 100 replicates of data are computed.

We compare the performance of the two estimators $\hat{\alpha}_{WG}$ and $\hat{\alpha}_{FD-GMM}$ using the sample medians and interquartile ranges. Note that the two estimators are both downward biased, so that comparisons will be more meaningful if resistant measures are used to assess the bias and efficiency. Hence, in assessing the finite sample properties of the two estimators, the median bias and median percent bias are used. On the other hand, efficiency is examined by looking at the dispersion using a more resistant measure like the interquartile range as compared to the standard deviation.

If we denote the WG and FD-GMM estimates from the r th replicate by $\hat{\alpha}_{WG,r}$ and $\hat{\alpha}_{FD-GMM,r}$, we get $\hat{\alpha}_{WG,1}, \hat{\alpha}_{WG,2}, \dots, \hat{\alpha}_{WG,100}$ and $\hat{\alpha}_{FD-GMM,1}, \hat{\alpha}_{FD-GMM,2}, \dots, \hat{\alpha}_{FD-GMM,100}$ respectively. Denote the sorted values of $\hat{\alpha}_{WG,r}$ by $\hat{\alpha}_{WG,(r)}$ and $\hat{\alpha}_{FD-GMM,r}$ by $\hat{\alpha}_{FD-GMM,(r)}$. The sample median for a particular design is given by $\text{Md}(\hat{\alpha}_{WG}) = \frac{1}{2} [\hat{\alpha}_{WG,(50)} + \hat{\alpha}_{WG,(51)}]$ for the WG estimator and $\text{Md}(\hat{\alpha}_{FD-GMM}) = \frac{1}{2} [\hat{\alpha}_{FD-GMM,(50)} + \hat{\alpha}_{FD-GMM,(51)}]$ for the FD-GMM estimator. The interquartile range for WG estimator is $\text{IQR}(\hat{\alpha}_{WG}) = \frac{1}{2} [\hat{\alpha}_{WG,(75)} + \hat{\alpha}_{WG,(76)}] - \frac{1}{2} [\hat{\alpha}_{WG,(25)} + \hat{\alpha}_{WG,(26)}]$ and the corresponding interquartile range for the FD-GMM estimator is equal to $\text{IQR}(\hat{\alpha}_{FD-GMM}) = \frac{1}{2} [\hat{\alpha}_{FD-GMM,(75)} + \hat{\alpha}_{FD-GMM,(76)}] - \frac{1}{2} [\hat{\alpha}_{FD-GMM,(25)} + \hat{\alpha}_{FD-GMM,(26)}]$.

Alvarez and Arellano (2003) computed the asymptotic approximations to the bias given by $(1 + \alpha)/T$ and $(1 + \alpha)/N$ for WG and FD-GMM estimators, respectively.

6 Results and Discussion

We report the Monte Carlo simulations on the WG and FD-GMM estimators for various combinations of values of N and T , and relatively wide range of values for α and $\sigma_{\eta}^2/\sigma_v^2$. The main focus of the analysis is on the bias of the WG and FD-GMM estimators as the number of cross-section dimension N and the number of time dimension T changes. The change in bias as the true value of the coefficient parameter α varies is also shown. The effect of the variance ratio between the individual effect and the random disturbance on the bias of the two estimators is explored.

6.1 Effect of the Sample Size on the Bias

The marginal effect of varying the cross-section dimension N and the time dimension T are presented separately. The joint effect of N and T is also presented as Alvarez and Arellano (2003) emphasized that the asymptotic bias of the DPD estimators depend on the relative rates of increase of N and T .

In theory, the cross-section dimension N has no effect on the bias of the WG estimator. This is confirmed in the simulation exercise, where the bias of WG estimator is relatively constant as N varies, given that T is fixed. On the other hand, the theory is that the FD-GMM estimator have a bias of order $1/N$, that is, the bias decreases as N becomes large. This pattern is not perfectly observed in the simulation. For instance, when the variance ratio $\sigma_\eta^2/\sigma_v^2 = 1$, only 10 out of the 25 cases have shown the pattern of decreasing bias of the FD-GMM estimates as the number of cross-section dimension increases. Also, when the variance ratio $\sigma_\eta^2/\sigma_v^2 = 0$, only 10 out of the 25 cases have shown the pattern of decrease in bias as N increases. A summary of the range of values of bias as percentage of the true parameter value for the FD-GMM estimator focusing on varying the cross-section dimension N is presented in Table 2. For a moderate time dimension $T = 10$, we could expect a FD-GMM estimate with a bias from 14% to 47% for $N = 30$, from 8% to 39% for $N = 40$ and from 6% to 38% of the true value of the coefficient even when $N = 50$.

Table 2. Percent Bias of $\hat{\alpha}_{FD-GMM}$

	$\sigma_\eta^2/\sigma_v^2 = 0$		$\sigma_\eta^2/\sigma_v^2 = 0.2$		$\sigma_\eta^2/\sigma_v^2 = 0.8$		$\sigma_\eta^2/\sigma_v^2 = 1$		$\sigma_\eta^2/\sigma_v^2 = 1.25$	
	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX
T=3										
N=10	66	114	75	130	77	183	73	189	84	162
N=20	41	138	33	95	62	103	28	101	48	95
N=30	14	46	30	81	15	81	38	82	28	88
N=40	9	70	12	72	10	83	33	126	9	68
N=50	15	21	15	75	11	74	7	85	20	88
T=5										
N=10	53	109	52	136	56	141	58	111	52	161
N=20	27	74	33	101	28	69	40	80	37	74
N=30	22	52	25	50	23	62	24	67	25	59
N=40	14	59	22	48	16	61	13	67	18	61
N=50	3	29	14	48	17	61	12	49	18	59
T=10										
N=10	28	124	30	127	34	95	36	97	32	101
N=20	18	83	21	44	19	55	27	39	26	71
N=30	14	32	15	27	16	34	18	47	19	43
N=40	8	39	10	29	11	32	14	32	11	34
N=50	8	26	11	38	8	28	10	30	6	31
T=25										
N=10	12	55	12	29	13	49	13	37	12	41
N=20	11	42	13	55	11	43	13	53	13	44
N=30	9	41	11	28	11	29	10	28	12	45
N=40	6	23	9	54	9	34	8	30	7	27
N=50	6	23	7	22	8	22	7	18	5	25
T=50										
N=10	5	18	6	18	5	16	5	25	5	31
N=20	5	24	6	24	5	18	6	20	6	20
N=30	5	23	5	27	6	22	5	21	5	22
N=40	5	17	6	27	6	24	6	22	6	20
N=50	4	23	6	21	6	20	6	23	5	21

The time dimension T affects the bias of the WG estimator. The WG estimator has bias of order $1/T$, that is, as T becomes larger, the WG bias becomes smaller. This is also observed in the simulation. As expected, as T increases from 3 to 50, the bias reduces tremendously within acceptable levels as T is nearing 50. This pattern is implied by the increase of the magnitude of the medians as T becomes larger.

The FD-GMM estimator is known to be affected by the cross-section dimension N , and the order of bias is $1/N$ and does not involve T . Alvarez and Arellano (2003) noted that consistency of the FD-GMM estimator requires $(\log T)^2/N \rightarrow 0$ where $N \rightarrow \infty$ and $T \rightarrow \infty$. This means that the bias of the FD-GMM estimator does not depend on N alone but also T . The FD-GMM estimates in the simulation exercise illustrates the decrease in bias as the time dimension increases, most specially for $\alpha \geq 0.5$. Also, from Table 2, a FD-GMM estimate is within 6% to 127% of the true value, when $T = 10$, within 6% to 55% of the true value when $T = 25$ and within 4% to 27% of the true value of the coefficient, implying the magnitude of FD-GMM bias decreases as T increases.

A summary of the range of values of bias as percentage of the true parameter value for the WG estimator with varying time dimension T is presented in Table 3. For a small cross-section dimension $N = 10$, one could expect a WG estimate with a bias from 27% to 132% for $T = 10$, from 10% to 54% for $T = 25$ and from 5% to 28% of the true value of the coefficient when $T = 50$. Even for large cross-section dimension, say $N = 50$, the WG estimates have similar range of percent bias as those estimates when $N = 10$. This indeed shows that there is notable decrease in the bias of WG estimates as T increases whatever the value of N .

Table 3. Percent Bias of $\hat{\alpha}_{WG}$

	$\sigma_{\eta}^2/\sigma_v^2 = 0$		$\sigma_{\eta}^2/\sigma_v^2 = 0.2$		$\sigma_{\eta}^2/\sigma_v^2 = 0.8$		$\sigma_{\eta}^2/\sigma_v^2 = 1$		$\sigma_{\eta}^2/\sigma_v^2 = 1.25$	
	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX
N=10										
T=3	81	533	76	569	82	552	82	514	81	561
T=5	55	254	51	305	53	276	56	260	58	313
T=10	28	130	29	132	27	120	29	120	30	119
T=25	11	54	10	26	10	54	11	39	11	41
T=50	5	19	5	19	5	16	5	25	5	28
N=20										
T=3	79	573	80	559	79	547	81	535	79	543
T=5	52	285	53	306	56	283	51	286	55	290
T=10	26	136	29	109	28	118	25	101	28	131
T=25	10	41	10	52	11	41	11	57	10	45
T=50	5	24	5	24	4	18	5	20	5	18
N=30										
T=3	78	558	82	557	81	551	82	547	82	545
T=5	53	273	53	272	51	298	53	289	55	267
T=10	29	124	28	121	29	115	30	126	28	113
T=25	11	55	10	40	11	42	10	42	12	49
T=50	5	22	5	26	5	23	5	22	5	23

		N=40									
T=3	83	554	90	557	84	541	80	543	80	551	
T=5	53	283	53	267	53	276	54	290	53	287	
T=10	28	133	27	122	28	124	27	130	27	132	
T=25	11	42	11	54	11	52	10	46	11	45	
T=50	5	19	5	25	5	24	5	22	5	22	
		N=50									
T=3	79	546	80	542	80	537	78	535	80	530	
T=5	53	259	52	266	52	283	54	282	54	265	
T=10	28	122	28	126	28	124	26	124	29	124	
T=25	11	45	11	45	11	46	10	43	11	51	
T=50	5	25	5	23	5	22	5	24	5	23	

It is interesting to note that the relationship between bias and the sample size represented by the pair (N, T) also takes into account the relative rates of increase of N with respect to T for the WG estimator and T with respect to N for the FD-GMM estimators. When we have a square panel, that is when the sample size is either, $(N = 10, T = 10)$ or $(N = 50, T = 50)$, the WG estimates and GMM estimates are almost the same. The similarity of the WG estimates to the GMM estimates is also seen for an almost square panel, such as $(N = 20, T = 25)$, $(N = 30, T = 25)$, and $(N = 40, T = 50)$. This confirms the theory of Alvarez and Arellano (2003) that the asymptotic bias of WG and FD-GMM are the same when $N = T$, we confirmed here to be also true for moderate samples and even small samples. When $T = 50$, regardless of the size of N , the value of WG estimates are similar to the value of FD-GMM estimates. This is attributed to the fact that in the previously stated scenario, the time-series dimension is always greater than or equal to the cross-section dimension, in this case the workable formula for FD-GMM whenever $t > N$ is almost identical to the WG formula.

6.2 Effect of Parameter Values on the Bias

The first-order DPD model considered in this study has three parameters, but we focused only on the coefficient of the lagged dependent variable (α). The two other parameters are the variances of the one-way random effects error component, namely the variance of the individual effect σ_{η}^2 and the variance of the random disturbance σ_v^2 . Instead of analyzing the effect of σ_{η}^2 and σ_v^2 separately, we focus on the variance ratio $\sigma_{\eta}^2/\sigma_v^2$.

Both the WG and FD-GMM estimators are downward biased, that is, the estimates are smaller than the true value of the coefficient parameter α . As shown by Nickell (1981) for WG estimator, the bias decreases as α increases. This is true for both the WG and FD-GMM estimators as illustrated in the study. Also, one may think that the percent bias will increase as we decrease the value of α , since the smaller the value of the denominator the larger the fraction becomes. This is confirmed in the results of simulation, the FD-GMM bias decreases as T increases provided that α is large.

The exact distribution of WG estimator is said to be invariant to both the variance of the individual effect σ_{η}^2 and the variance of the random disturbance σ_v^2 , while the

distribution of the FD-GMM estimator is invariant only to the variance ratio σ_η^2/σ_v^2 , see Alvarez and Arellano (2003). In the simulation exercise, varying the variance ratio does not show sizeable changes on the bias, when the sample size (N, T) and the value of the parameter coefficient α are fixed.

6.3 Other Asymptotic and Finite Sample Properties of WG and FD-GMM Estimators

We compare the median of the estimators to the approximate bias values computed by Alvarez and Arellano (2003). Percent different between the computed bias and the approximate bias are reported in Table 4. The following asymptotic properties: (a) when $T < N$, the asymptotic bias of GMM is smaller than the WG bias, (b) when $T = N$, the expression for the two asymptotic biases are equal and (c) when $N/T \rightarrow 0$, the asymptotic bias in the WG estimator disappears, derived by Alvarez and Arellano (2003) still hold for smaller samples considered in this study. The findings of Kiviet (1995) that the WG is more efficient than GMM is also supported by the simulation study. Since the bias of WG estimator does not depend on N , the values are similar, that is, the number of WG estimates with percent difference values less than 5% are almost the same for different values of N . On the other hand, percent difference of FD-GMM estimate increases with N . The asymptotic approximation of Alvarez and Arellano (2003) performs well for $N \geq 40$, since 72.8% of the MC medians of FD-GMM estimates are within 20% of the approximated value.

Table 4. Percent Difference of Bias Estimates From Asymptotic Biases

%Difference	WG	FD-GMM
		N=10
<=20	70 (56%)	35 (28%)
<=10	58 (46.4%)	13 (10.4%)
<5	39 (31.2%)	6 (4.8%)
		N=20
<=20	71 (56.8%)	56 (44.8%)
<=10	66 (52.8%)	46 (36.8%)
<5	43 (34.4%)	26 (20.8%)
		N=30
<=20	76 (60.8%)	81 (64.8%)
<=10	61 (48.8%)	59 (47.2%)
<5	46 (36.8%)	40 (32%)
		N=40
<=20	75 (60%)	91 (72.8%)
<=10	64 (51.2%)	72 (57.6%)
<5	46 (36.8%)	52 (41.6%)
		N=50
<=20	73 (58.4%)	96 (76.8%)
<=10	61 (48.8%)	76 (60.8%)
<5	49 (39.2%)	57 (45.6%)

The bias of FD-GMM is smaller than the bias of WG estimates, that is 71 (56.8%) out of 125 cases follow this pattern. Note that 60% of the cases have the set-up $T < N$, and the

other 40% have the set-up $T > N$. It is interesting to note that only when the variance of the individual effect σ_{η}^2 equal to zero, we see that majority, that is, 20 out of the 25 cases considered have bias of FD-GMM less than the bias of WG. On the other hand, when $\sigma_{\eta}^2 > 0$, half of the cases have bias of FD-GMM smaller than bias of WG and the other half have bias of WG less than or equal to bias of FD-GMM. The cases where bias of WG is less than or equal to the bias of FD-GMM estimates have moderate to large T , but when $T = 10$, the bias of WG is smaller and closer to the FD-GMM bias only when α is at least 0.5. Specifically, bias of WG is close to bias of FD-GMM for square panels where $\alpha \geq 0.5$ and for $T \geq 25$.

We also analyzed moderately-sized cross-section dimension, i.e., $N = 30$. This allows for 80% of the 125 cases to have $T < N$ and the other 20% are designs where $T > N$. We expect that more percentage of FD-GMM estimates have smaller bias than WG estimates as compared to where the cross-dimension size is small. There are 90 (72%) of the 125 cases where bias of FD-GMM is smaller than the bias of WG. The other 35 (28%) cases have either large time-dimensions, that is $T = 50$ or larger value for the coefficient parameter, that is $\alpha \geq 0.8$. This is intuitively true, since when $T = 50$, $N < T$ and the bias of WG is expected to be less than the bias of FD-GMM. For moderately-sized cross-section dimension, about 73% of the FD-GMM estimates have smaller bias than their WG counterpart and the other 27% have designs where the time dimension is large or the value of the coefficient of parameter is close to one.

Table 5. Comparison Of Bias of WG and FD-GMM Estimates

Cases	Number of cases with Bias(FD-GMM)<Bias(WG)	Set-up not satisfying the inequality
N=10, $\sigma_{\eta}^2=0$	15(60%)	moderate to large T, α close to 1
N=10, $\sigma_{\eta}^2=0.2$	11(44%)	moderate T, α close to 1
N=10, $\sigma_{\eta}^2=0.8$	12(48%)	moderate T, α close to 1
N=10, $\sigma_{\eta}^2=1.0$	14(56%)	moderate to large T, α close to 1
N=10, $\sigma_{\eta}^2=1.25$	14(56%)	moderate to large T, α close to 1
Total	66(52.8%)	
N=20, $\sigma_{\eta}^2=0$	20(80%)	moderate T or α close to 1
N=20, $\sigma_{\eta}^2=0.2$	14(56%)	moderate to large T, α close to 1
N=20, $\sigma_{\eta}^2=0.8$	13(52%)	moderate to large T, α close to 1
N=20, $\sigma_{\eta}^2=1.0$	12(48%)	moderate to large T, α close to 1
N=20, $\sigma_{\eta}^2=1.25$	12(48%)	moderate to large T, α close to 1
Total	71(56.8%)	
N=30, $\sigma_{\eta}^2=0$	22(88%)	T=50. $\alpha =0.1,0.8,0.9$
N=30, $\sigma_{\eta}^2=0.2$	18(72%)	T=25,50, α not 0.5
N=30, $\sigma_{\eta}^2=0.8$	17(68%)	T=50 or α close to 1
N=30, $\sigma_{\eta}^2=1.0$	17(68%)	T=50 or α close to 1
N=30, $\sigma_{\eta}^2=1.25$	16(64%)	T=50 or α close to 1
Total	90(72%)	
N=40, $\sigma_{\eta}^2=0$	21(84%)	almost the same bias for T=50

N=40, $\sigma_\eta^2=0.2$	18(72%)	T=25 or 50 or α close to 1
N=40, $\sigma_\eta^2=0.8$	18(72%)	T=25 or 50 or α close to 1
N=40, $v=1.0$	17(68%)	T=25 or 50 or α close to 1
N=40, $\sigma_\eta^2=1.25$	17(68%)	T=25 or 50 or α close to 1
Total	91(72.8%)	
N=50, $\sigma_\eta^2=0$	25(100%)	
N=50, $\sigma_\eta^2=0.2$	23(92%)	T=50 and α close to 1
N=50, $\sigma_\eta^2=0.8$	21(84%)	T=50 or α close to 1
N=50, $\sigma_\eta^2=1.0$	19(76%)	
N=50, $v=1.25$	16(64%)	T=50 or α close to 1
Total	104(83.2%)	

Some 80% of the 125 cases have designs where $T < N$ and 20% of the cases come from square panel, that is, $N = T = 50$. We expect that most of the FD-GMM estimates have smaller bias than the WG estimates and when square panels are considered the biases are the same. There are 104 (83.2%) of the 125 cases considered have FD-GMM bias smaller than their WG counterpart. The other 21 (16.8%) cases have designs where the time dimension is large, that is, $T = 50$ and the coefficient parameter is close to one, that is, $\alpha = 0.8$ or 0.9 .

Table 6. Comparison of Variability of WG and FD-GMM Estimates

Cases	Interquartile Range (WG) < Interquartile Range (GMM)
N=10, $\sigma_\eta^2=0$	21(84%)
N=10, $\sigma_\eta^2=0.2$	18(72%)
N=10, $\sigma_\eta^2=0.8$	20(80%)
N=10, $\sigma_\eta^2=1.0$	20(80%)
N=10, $v=1.25$	18(72%)
Total	97(77.6%)
N=20, $v=0$	22(88%)
N=20, $\sigma_\eta^2=0.2$	22(88%)
N=20, $\sigma_\eta^2=0.8$	23(92%)
N=20, $\sigma_\eta^2=1.0$	23(92%)
N=20, $\sigma_\eta^2=1.25$	21(84%)
Total	111(88.8%)
N=30, $v=0$	17(68%)
N=30, $\sigma_\eta^2=0.2$	20(80%)
N=30, $v=0.8$	24(96%)
N=30, $\sigma_\eta^2=1.0$	22(88%)
N=30, $v=1.25$	22(88%)
Total	105(84%)
N=40, $v=0$	22(88%)
N=40, $v=0.2$	22(88%)
N=40, $\sigma_\eta^2=0.8$	20(80%)

N=40, $\sigma_\eta^2=1.0$	24(96%)
N=40, $\sigma_\eta^2=1.25$	24(96%)
Total	112(89.6%)
N=50, $\sigma_\eta^2=0$	24(96%)
N=50, $v=0.2$	23(92%)
N=50, $v=0.8$	23(92%)
N=50, $\sigma_\eta^2=1.0$	22(88%)
N=50, $v=1.25$	21(84%)
Total	113(90.4%)

Table 6 summarizes comparison of the variability of WG and FD-GMM estimates as measured by the interquartile range (IQR). The WG estimates generally have less variability than FD-GMM estimates, for the moderate size cross-section dimension. There are 105 (84%) cases where IQR of WG is less than IQR of FD-GMM. The strongest evidence that the WG estimator has smaller variability than FD-GMM is seen in the summary of Table 6, particularly, 113 (90.4%) of the 125 cases have IQR of WG smaller than IQR of FD-GMM.

6.4 Comparison of Bootstrapped DPD and Conventional DPD Estimators

To be able to assess the benefits from bootstrapping the WG and GMM estimators, we identified settings in the simulation scenarios where they yield the largest bias. We report in Tables 7 and 8 the top 25 of estimates with the largest bias for both WG and GMM, together with their respective design specifications.

Table 7. 25 Largest Bias of WG Estimates and their Design Specifications

Rank	Bias	α	N	T	σ_η^2
1	0.749	0.9	40	3	0
2	0.746	0.9	10	3	0.8
3	0.743	0.9	40	3	0.8
4	0.741	0.9	10	3	1
5	0.739	0.9	30	3	0.2
6	0.739	0.9	30	3	1.25
7	0.737	0.9	20	3	1
8	0.735	0.9	50	3	1.25
9	0.733	0.9	30	3	1
10	0.731	0.9	30	3	0.8
11	0.731	0.9	50	3	0.2
12	0.730	0.8	50	3	1.25
13	0.728	0.8	40	3	0.2
14	0.726	0.9	40	3	1
15	0.725	0.9	10	3	0
16	0.725	0.9	10	3	1.25
17	0.724	0.9	50	3	0.8
18	0.721	0.9	40	3	0.2

19	0.720	0.9	20	3	0.2
20	0.720	0.9	40	3	1.25
21	0.719	0.9	20	3	1.25
22	0.719	0.9	20	3	0.8
23	0.719	0.8	30	3	1
24	0.716	0.8	10	3	0
25	0.715	0.8	20	3	1

Table 8. 25 Largest Bias of FD-GMM Estimates and their Design Specifications

Rank	bias	α	N	T	σ_η^2
1	0.951	0.9	50	3	1
2	0.922	0.9	10	3	1
3	0.882	0.9	20	3	1.25
4	0.871	0.9	40	3	0.8
5	0.858	0.9	20	3	1
6	0.857	0.9	40	3	1
7	0.854	0.9	20	3	0.8
8	0.820	0.9	50	3	1.25
9	0.799	0.8	10	3	1
10	0.784	0.9	30	3	1.25
11	0.765	0.9	10	3	0.8
12	0.756	0.8	10	3	1.25
13	0.747	0.9	20	3	0.2
14	0.741	0.9	40	3	1.25
15	0.739	0.9	30	3	0.2
16	0.738	0.8	20	3	1
17	0.722	0.9	30	3	0.8
18	0.703	0.8	20	3	1.25
19	0.690	0.8	20	3	0.8
20	0.684	0.9	50	3	0.2
21	0.678	0.8	50	3	1
22	0.669	0.8	30	3	1
23	0.662	0.9	10	3	1.25
24	0.661	0.8	10	3	0.2
25	0.660	0.8	10	3	0.8

The order of bias of the WG estimator is $1/T$ and therefore, as T increases the bias decreases. The order of bias of the GMM estimator is $1/N$ and thus, as N increases, the bias

decreases. In terms of the bias, the worst WG and GMM estimates came from designs with very small time dimension. This is congruent to the theoretical properties of the WG estimator, but at first quite surprising for GMM estimator. When $T = 3$, FD-GMM estimator uses only one instrument and thus equivalent to the IV estimator, less appealing than the GMM estimators. Moreover, the FD-GMM estimator when $T = 3$, does not show a decrease in bias as N becomes larger.

It is known that the bias of WG estimator increases with the coefficient parameter α , see Nickell (1981) and Kiviet (1995). In this simulation exercise, all the WG and GMM worst estimates came from designs with large α , see Table 9. As the autoregression component of the model becomes nearly nonstationary, both WG and FD-GMM estimates can suffer tremendously.

Table 9. Monte Carlo Designs Yielding Worst Estimates (33 Parameter Combinations).

Design	α	N	T	σ_{η}^2	σ_v^2
1	0.8	10	3	0	1
2	0.8	10	3	0.2	1
3	0.8	10	3	0.8	1
4	0.8	10	3	1	1
5	0.8	10	3	1.25	1
6	0.8	20	3	0.8	1
7	0.8	20	3	1	1
8	0.8	20	3	1.25	1
9	0.8	30	3	1	1
10	0.8	40	3	0.2	1
11	0.8	50	3	1	1
12	0.8	50	3	1.25	1
13	0.9	10	3	0	1
14	0.9	10	3	0.8	1
15	0.9	10	3	1	1
16	0.9	10	3	1.25	1
17	0.9	20	3	0.2	1
18	0.9	20	3	0.8	1
19	0.9	20	3	1	1
20	0.9	20	3	1.25	1
21	0.9	30	3	0.2	1
22	0.9	30	3	0.8	1
23	0.9	30	3	1	1
24	0.9	30	3	1.25	1
25	0.9	40	3	0	1
26	0.9	40	3	0.2	1
27	0.9	40	3	0.8	1

28	0.9	40	3	1	1
29	0.9	40	3	1.25	1
30	0.9	50	3	0.2	1
31	0.9	50	3	0.8	1
32	0.9	50	3	1	1
33	0.9	50	3	1.25	1

The variances of the individual effect for the worst GMM estimates are mostly large, in fact 15 out of the 25 designs considered have $\sigma_{\eta}^2 \geq 1$ and 21 out of the 25 cases considered have $\sigma_{\eta}^2 \geq 0.8$. The variances of the individual effects of the 25 worst WG estimates are similar.

When we combine the designs for the 25 largest bias of WG and 25 largest bias of GMM, they have 17 common designs, thus 33 designs in Table 9 represent the worst estimate (largest bias) both from WG and GMM.

The bootstrap methodology is used for these 33 designs. The bias, standard deviation and the root mean square error for the original estimators WG and FD-GMM together with the bootstrap estimators WGb and FD-GMM b are presented in Tables 10-A to 10-E.

Table 10-A. Estimates for $\alpha = 0.8$, $N = 10, 20$

Design		Bias	Std. Dev.	MSE
1	WG	0.716	0.221	0.561
	WGb	0.751	0.117	0.578
	FD-GMM	0.596	0.662	0.793
	FD-GMMb	0.701	0.137	0.510
2	WG	0.705	0.215	0.543
	WGb	1.264	0.259	1.665
	FD-GMM	0.661	0.692	0.916
	FD-GMMb	1.226	1.004	2.511
3	WG	0.678	0.200	0.500
	WGb	1.196	0.162	1.457
	FD-GMM	0.660	0.905	1.255
	FD-GMMb	1.291	0.699	2.156
4 ^{*wg,gmm}	WG	0.708	0.189	0.537
	WGb	0.489	0.063	0.243
	FD-GMM	0.799	1.242	2.181
	FD-GMMb	0.518	0.261	0.337
5 ^{*wg,gmm}	WG	0.713	0.215	0.555
	WGb	0.706	0.019	0.498
	FD-GMM	0.756	0.744	1.125
	FD-GMMb	0.705	0.085	0.504
6	WG	0.684	0.167	0.496

	WGb	0.965	0.036	0.933
	FD-GMM	0.690	1.062	1.604
	FD-GMMb	0.966	0.253	0.997
7	WG	0.715	0.162	0.537
	WGb	0.889	0.061	0.794
	FD-GMM	0.738	1.334	2.324
	FD-GMMb	0.865	0.264	0.819
8	WG	0.674	0.127	0.470
	WGb	0.990	0.052	0.982
	FD-GMM	0.703	0.864	1.241
	FD-GMMb	0.977	0.223	1.004

Table 10-B Estimates for $\alpha = 0.8$, $N = 30, 40, 50$

		Bias	Std. Dev.	MSE
9	WG	0.719	0.127	0.533
	WGb	0.964	0.051	0.932
	FD-GMM	0.669	0.555	0.756
	FD-GMMb	0.936	0.164	0.902
10	WG	0.728	0.096	0.539
	WGb	0.768	0.076	0.595
	FD-GMM	0.448	0.765	0.786
	FD-GMMb	0.829	0.463	0.901
11 ^{*wg.gmm}	WG	0.698	0.09	0.495
	WGb	0.437	0.006	0.191
	FD-GMM	0.678	0.688	0.933
	FD-GMMb	0.433	0.021	0.188
12	WG	0.73	0.091	0.541
	WGb	1.014	0.016	1.029
	FD-GMM	0.502	0.867	1.004
	FD-GMMb	1.007	0.080	1.021

Table 10-C. Estimates for $\alpha = 0.9$, $N = 10, 20$

		Bias	Std. Dev.	MSE
13	WG	0.725	0.254	0.590
	WGb	1.468	0.243	2.215
	FD-GMM	0.614	1.119	1.629
	FD-GMMb	1.230	0.443	1.710
14 ^{*wg.gmm}	WG	0.746	0.249	0.619
	WGb	0.618	0.018	0.382
	FD-GMM	0.765	0.785	1.201
	FD-GMMb	0.614	0.029	0.377
15 ^{*wg.gmm}	WG	0.741	0.217	0.596
	WGb	0.216	0.058	0.050
	FD-GMM	0.922	0.950	1.753
	FD-GMMb	0.203	0.086	0.049
16	WG	0.725	0.254	0.590
	WGb	1.088	0.026	1.184
	FD-GMM	0.662	0.787	1.058
	FD-GMMb	1.075	0.168	1.183
17 ^{*wg.gmm}	WG	0.720	0.167	0.546
	WGb	0.396	0.022	0.158
	FD-GMM	0.747	0.760	1.136
	FD-GMMb	0.393	0.072	0.160
18	WG	0.719	0.157	0.542

	WGb	0.864	0.055	0.750
	FD-GMM	0.854	0.738	1.274
	FD-GMMb	0.869	0.204	0.797
19	WG	0.737	0.172	0.573
	WGb	1.150	0.030	1.324
	FD-GMM	0.858	0.832	1.428
	FD-GMMb	1.144	0.069	1.314
20	WG	0.719	0.163	0.544
	WGb	1.316	0.026	1.733
	FD-GMM	0.882	0.808	1.431
	FD-GMMb	1.314	0.057	1.729

Table 10-D. Estimates for $\alpha = 0.9$, $N = 30, 40$

		Bias	Std. Dev.	MSE
21	WG	0.739	0.120	0.561
	WGb	-0.873	0.000	0.762
	FD-GMM	0.596	0.875	1.121
	FD-GMMb	-0.873	0.001	0.762
22	WG	0.731	0.128	0.551
	WGb	1.145	0.010	1.311
	FD-GMM	0.722	0.710	1.025
	FD-GMMb	1.141	0.024	1.301
23	WG	0.733	0.132	0.555
	WGb	0.907	0.024	0.823
	FD-GMM	0.607	0.665	0.811
	FD-GMMb	0.894	0.090	0.808
24 ^{*gmm}	WG	0.738	0.121	0.559
	WGb	0.753	0.013	0.567
	FD-GMM	0.784	0.647	1.033
	FD-GMMb	0.747	0.026	0.559
25	WG	0.749	0.111	0.573
	WGb	0.953	0.074	0.914
	FD-GMM	0.317	0.603	0.464
	FD-GMMb	0.897	0.205	0.847
26	WG	0.721	0.116	0.533
	WGb	0.825	0.023	0.682
	FD-GMM	0.650	0.868	1.176
	FD-GMMb	0.820	0.177	0.704
27	WG	0.743	0.109	0.564
	WGb	1.428	0.017	2.040
	FD-GMM	0.871	0.828	1.444
	FD-GMMb	1.420	0.034	2.019
28 ^{*gmm}	WG	0.726	0.115	0.540
	WGb	0.785	0.019	0.616
	FD-GMM	0.857	0.652	1.160
	FD-GMMb	0.790	0.117	0.638
29 ^{*wg.gmm}	WG	0.719	0.113	0.530
	WGb	0.614	0.014	0.378
	FD-GMM	0.741	0.762	1.130
	FD-GMMb	0.609	0.044	0.372

Table 10-E. Estimates for $\alpha = 0.9$, $N = 50$

		Bias	Std. Dev.	MSE
30	WG	0.731	0.096	0.544
	WGb	1.208	0.025	1.460
	FD-GMM	0.687	0.691	0.949
	FD-GMMb	1.189	0.084	1.420
31	WG	0.724	0.103	0.535
	WGb	0.776	0.016	0.603
	FD-GMM	0.651	0.709	0.926
	FD-GMMb	0.765	0.120	0.600
32 ^{*wg.gmm}	WG	0.708	0.104	0.512
	WGb	0.698	0.021	0.487
	FD-GMM	0.951	2.010	4.945
	FD-GMMb	0.686	0.111	0.483
33	WG	0.735	0.094	0.549
	WGb	0.936	0.016	0.875
	FD-GMM	0.819	0.755	1.241
	FD-GMMb	0.940	0.109	0.896

When the true value of the coefficient α is equal to 0.8, the bootstrap estimators WGb and FD-GMMb work well under the smallest sample case scenario, i.e., when $N = 10$ and $T = 3$ where the variance of the individual effect is at least one. Also, for the largest N and smallest T combination, i.e., when $N = 50$ and $T = 3$ and the variance of the individual effect is equal to one, both bootstrap estimators perform better than the conventional estimators. The bootstrap estimators for $\alpha = 0.9$ are also better than the conventional estimators for the smallest sample case and when the variance of the individual effect is as large as 0.8 and 1.0. However, when $\sigma_{\eta}^2 = 1.25$, the bootstrapped estimators perform as badly as the conventional estimators. Still under small sample case, $N = 20$ and $T = 3$, and time the variance of the individual effect is small but nonzero, i.e., $\sigma_{\eta}^2 = 0.2$, we obtain the most improved bootstrap estimates for both WG and FD-GMM.

Given moderate size cross-section dimension ($N = 30$ and $N = 40$), the bootstrap FD-GMM is better than the conventional FD-GMM for cases where the variance of the individual effect is as large as 1 or 1.25. Also, the bootstrap WG has smaller bias than the conventional WG when $N = 40$ and $\sigma_{\eta}^2 = 1.25$.

Both the bootstrap WG and bootstrap FD-GMM are better estimators than their conventional counterparts for the largest N and smallest T case where the variance of the individual effect is equal to one.

7 Conclusions

In estimating a dynamic panel data (DPD) model using WG and FD-GMM estimators, the cross-section dimension N has no effect on the bias of the WG estimator, but the bias of FD-GMM decreases as N increases in some cases. The bias of WG estimator decreases as T increases quite consistently in most cases, however, the bias of FD-GMM decreases as T increases. In square panels such as panels with dimensions, e.g., $N = 10 = T$, $N = 50 = T$,

the WG and FD-GMM estimates are similar. However, when $N = 10 = T$, the two estimators are similar only when $\alpha \geq 0.5$. Also, the WG estimates are similar to FD-GMM estimates for almost square panels such as panels, e.g., $(N = 20, T = 25)$, $(N = 30, T = 25)$, and $(N = 40, T = 50)$. When $T = 50$ regardless of the value of N , WG and FD-GMM estimates are similar.

WG and FD-GMM estimators are both downward biased, the bias increases with α . However, the percent bias, that is, the bias as a percentage of the true value of α decreases as α increases. Varying the variance ratio (variance of the individual effect divided by the variance of the random disturbance) does not show sizeable changes on the bias. The bias differ by at most 20% from the approximate bias provided by Alvarez and Arellano (2003) for WG when $T = 10$ and $\alpha \geq 0.5$ or when $T \geq 25$. For the FD-GMM, at most 20% bias difference from the approximate large sample bias happen when $(N = 10, 20, \alpha \geq 0.2$ and $T = 10)$ or $(N \geq 30, \alpha \geq 0.2$ and $T = 10, 20)$.

WG estimates are less variable than FD-GMM estimates based on the interquartile range. The WG estimator is best to use when the time-series dimension is as large as 50 and the cross-section dimension as low as 10 for models where the ratio of the variance of the individual effect to the variance of the random disturbance to be less than one and regardless of the true value of the coefficient parameter. Also, when the $T = 50$, $N = 20$ and $\sigma_{\eta}^2 / \sigma_v^2 \geq 1$, the WG estimates are good. These conditions will assure that the percent bias of the WG estimate will not exceed 20%. The FD-GMM estimator has bias not exceeding 20% of the true value of the parameter coefficient for small and moderate samples in terms of N when: (1) $N = 10$ and $T = 50$, for $\sigma_{\eta}^2 / \sigma_v^2 < 1$, (2) $N = 20$ and $T = 50$, for $\sigma_{\eta}^2 / \sigma_v^2 \geq 1$. The importance of having large N to reduce the bias and percent bias is evident for designs where the time-series dimension is as large as 25.

The bootstrap estimators for both WG and FD-GMM, labelled as WGb and FD-GMMb respectively, work well for the smallest sample size, i.e., $N = 10$ and $T = 3$ and the extreme sample size set-up where $N = 50$ and $T = 3$, provided that the ratio of the variance of the individual effect to the variance of the random disturbance is equal to one.

References:

- Alvarez, J. and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71:1121-1159.
- Anderson, T. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of American Statistical Association* 76(375):598-606.
- Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: monte carlo evidence and an application to employment equations. *Review of Economic Studies* 58:277-297.
- Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-component models. *Journal of Econometrics* 68:29-45.
- Baltagi, B. (2005). *Econometric Analysis of Panel Data*, 3rd ed.. New York: John Wiley and Sons.

- Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87:115-143.
- Bond, S. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal* 1(2):141-162.
- Buhlman, P. (2002). Bootstrap for time series. *Statistical Science* 1(1):52-72.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* 14:1171-1179.
- Chigira, H. and Yamamoto, T. (2006). A bias-corrected estimation for dynamic panel models in small samples. *Hi-Stat Discussion Paper Series No.177*. Institute of Economic Research, Hitotsubashi University.
- Efron B. (1979). Bootstrap method: another look at the Jackknife. *The Annals of Statistics* 7(1): 1-26.
- Everaert, G. and Pozzi, L. (2007). [Bootstrap based bias correction for dynamic panels](#). *Journal of Economic Dynamics and Control* 31:1160-1184.
- Hayakawa, K.(2007). Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Letters* 95:32-38.
- Hounkannounon, B. (2008). Bootstrap for panel data. Available at SSRN: <http://ssrn.com/abstract=1309902>
- Hsiao, C. (2003). *Analysis of Panel Data*, 2nd ed. Cambridge: Cambridge University Press.
- Hsiao, C. and Tahmiscioglu, A. (2008). Estimation of dynamic panel data models models with both individual and time-specific effects. *Journal of Statistical Planning and Inference* 138: 2698-2721.
- Islam, N. (2000). Small sample performance of dynamic panel data estimators: a monte carlo study on the basis of growth data. In: Baltagi, B., ed. *Advances in Econometrics* (Vol. 15): *Nonstationary Panels, Panel Cointegration and Dynamic Panels*. New York: Elsevier Science.
- Judson, R. and Owen, A. (1999). Estimating dynamic panel data models: a practical guide for macroeconomist. *Economics Letters* 65(1):9-15.
- Kiviet, J. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68:53-78.
- Kiviet, J. and Bun, M. (2001). The accuracy of inference in small samples of dynamic panel data models: simulation evidence and empirical results. Tinbergen Institute Discussion Paper TI 2001-006/4, Faculty of Economics and Econometrics, University of Amsterdam and Tinbergen Institute.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17:1217-1241.
- MacKinnon, J. (2002). Bootstrap inference in econometrics. *The Canadian Journal of Economics* 35:615-645.
- MacKinnon, J. (2006). Bootstrap methods in econometrics. *The Economic Record* 82(Special Issue):S2-S18.
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica* 49:1417-1425.