

# Bias correction methods for dynamic panel data models with fixed effects

Mohamed Reda Abonazel \*

Department of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, Cairo University, Giza, Egypt

\*Corresponding author E-mail: [mabonazel@hotmail.com](mailto:mabonazel@hotmail.com)

## Abstract

This paper considers the estimation methods for dynamic panel data (DPD) models with fixed effects, which suggested in econometric literature, such as least squares (LS) and generalized method of moments (GMM). These methods obtain biased estimators for DPD models. The LS estimator is inconsistent when the time dimension (T) is short regardless of the cross-sectional dimension (N). Although consistent estimates can be obtained by GMM procedures, the inconsistent LS estimator has a relatively low variance and hence can lead to an estimator with lower root mean square error after the bias is removed. Therefore, we discuss in this paper the different methods to correct the bias of LS and GMM estimations. The analytical expressions for the asymptotic biases of the LS and GMM estimators have been presented for large N and finite T. Finally, we display new estimators that presented by Youssef and Abonazel [40] as more efficient estimators than the conventional estimators.

**Keywords:** Bias-Corrected Estimators; First-Order Autoregressive Panel Model; Generalized Method of Moments Estimators; Kantorovich Inequality; Least Squares Dummy Variable Estimators.

## 1. Introduction

In econometrics literature, the panel data refers to the pooling of observations on a cross-section of households, countries, firms, etc. over several time periods. Panel data is now widely used to estimate dynamic econometric models.<sup>1</sup> Its advantage over cross-section data in this context is obvious: we cannot estimate dynamic models from observations at a single point in time, and it is rare for single cross section surveys to provide sufficient information about earlier time periods for dynamic relationships to be investigated. Its advantages over aggregate time-series data include the possibility that underlying microeconomic dynamics may be obscured by aggregation biases, and the scope that panel data offers to investigate heterogeneity in adjustment dynamics between different types of individuals, household, or firms.

In time series regression models it is common practice to deal with these by including in the specification lagged values of the covariate, the dependent variable, or both. The inclusion of lags of the dependent variable seems to provide an adequate characterization of many economic dynamic adjustment processes. However, in panel data analysis with a small number of time periods, there often appear to be inference problems, such as small sample bias in coefficient estimation and hypothesis testing. Therefore, there were many estimation methods for DPD models.

In DPD models, the least squares methods lead to inconsistent estimates for the parameters when T is short regardless of N. This inconsistency stems from the fact that the disturbance terms are correlated with the lagged endogenous variable. Moreover, under large N fixed T Asymptotic, Nickell [36] showed that the standard maximum likelihood (ML) estimator suffers from an incidental

parameter problem leading to inconsistency. In order to avoid this problem, the literature has focused on GMM estimation applied to first differences, such as Anderson and Hsiao [7] and Arellano and Bond [8]. However, the standard GMM estimator obtained after first differencing has been found to suffer from substantial finite sample bias, especially when the instruments are weak and the number of moments is large relative to the cross section sample size. See Alonso-Borrego and Arellano [4].

This low precision of GMM is also evident in more general contexts. To improve the finite sample properties of GMM estimators, a number of alternative estimators have been suggested, such as level and system GMM estimators, which presented by Arellano and Bover [9]. These estimators based on use many instruments variables to improve the efficiency of GMM estimator. However, these estimators still biased and need for further improvement. Recently, Youssef and Abonazel [40] proposed a new approach to improve the efficiency of GMM estimators. Furthermore, they presented new GMM estimators. These estimators are more efficient than the conventional GMM estimators.

From above review, we can conclude that all the estimators for DPD models are biased and need to improvement. Therefore, the main objective of this paper is discussing the different methods to correct the bias, and also displaying other methods that improve the efficiency of the estimation in DPD models.

This paper is organized as follows. Section 2 provides the model and reviews the LS estimators. Section 3 presents the asymptotic bias of the LS estimator which has been examined by Nickell [36]. In Section 4, we will review the papers proposing the bias-corrected LS estimators, such as Kiviet [27], Hansen [24], and Bun and Carree [19]. While in Section 5, we will display the bias of GMM estimators which has been examined by Hayakawa [25]. Section 6 provides bias-corrected GMM estimators. In Section 7, we will present the new efficient GMM estimators which have

<sup>1</sup> See, e.g., Bond [16], Baltagi [10], and Hsiao [26].



introduced by Youssef and Abonazel [40]. Finally, Section 8 offers the concluding remarks.

## 2. The model and the least squares estimation

Anderson and Hsiao [6] presented the first contribution to estimate the DPD models. They used LS method as an initial estimation for these models. They considered the first-order autoregressive, AR (1), panel data model with P additional time-varying regressors:

$$y_{it} = \phi y_{i,t-1} + x_{it}\beta + u_{it}, i = 1, \dots, N; t = 1, \dots, T, \tag{1}$$

where  $y_{it}$  and  $x_{it}$  denote the (scalar) dependent variable and the  $1 \times P$  vector of exogenous variables corresponding to cross sectional unit  $i$  in period  $t$ ,  $\phi$  is scalar, where  $|\phi| < 1$ , and  $\beta$  represent corresponding  $P \times 1$  parameters, and  $u_{it}$  denotes the overall disturbance term consisting of individual effects  $\mu_i$  and an innovation  $\varepsilon_{it}$ , i.e.  $u_{it} = \mu_i + \varepsilon_{it}$ . Stacking the observations over time, we get

$$y_i = \phi y_{i,-1} + X_i\beta + e_T\mu_i + \varepsilon_i,$$

Where

$$y_i = (y_{i1}, \dots, y_{iT})', y_{i,-1} = (y_{i0}, \dots, y_{i,T-1})', X_i = (x_{i1}, \dots, x_{iT})', e_T = (1, \dots, 1)', \varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'. \tag{2}$$

Under the following assumptions:

A1: The error terms are distributed as  $\varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2)$ .

A2: The error terms are orthogonal to the exogenous variables, i.e.  $E(x'_{it}\varepsilon_{it}) = 0$ .

A3: The exogenous variables might be correlated with the individual effect, i.e.  $E(x'_{it}\mu_i) \neq 0$ .

A4: The error terms are uncorrelated with the lagged endogenous variable, i.e.  $E(y_{i,t-1}\varepsilon_{it}) = 0$ .

Stacking the observations once again across individuals, we get

$$y = \begin{matrix} NT \times 1 \\ NT \times (P+1) \end{matrix} \begin{matrix} \delta \\ (I_N \otimes e_T) \mu \end{matrix} + \begin{matrix} \varepsilon \\ \varepsilon \end{matrix}$$

Where  $y = (y_1, \dots, y_N)'$ ,  $\mu = (\mu_1, \dots, \mu_N)'$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$ , and  $I_N$  is identity matrix with dimension  $N$ . While

$$\delta = (\phi, \beta)'; Z = (y_{-1}, X); y_{-1} = (y_{1,-1}, \dots, y_{N,-1})'; X = (X_1, \dots, X_N)'. \tag{3}$$

The within group estimator of  $\delta$ , which is called the covariance estimator by Anderson and Hsiao [6], can simply be expressed as

$$\hat{\delta}_{lsdv} = \begin{pmatrix} \hat{\phi}_{lsdv} \\ \hat{\beta}_{lsdv} \end{pmatrix} = (Z' A Z)^{-1} Z' A y, \tag{4}$$

Where the  $NT \times NT$  within group transformation matrix  $A$  is defined as

$$A = I_N \otimes \left( I_T - \frac{e_T e_T'}{T} \right). \tag{5}$$

Anderson and Hsiao [6] showed that this estimator is inconsistent for fixed  $T$  due to presence of individual effects in both the disturbances  $\varepsilon$  and the regressors  $y_{-1}$ .

## 3. The asymptotic bias for LSDV estimator

Nickell [36] derived the asymptotic bias of least squares dummy variable (LSDV) estimator of this model in (1) under the assumptions A1: A4. He focused on the case of random start-up (no conditioning on  $y_{i0}$ ). For the DPD model with no exogenous regressor variables he obtained:

$$\text{Plim}_{N \rightarrow \infty}(\hat{\phi}_{lsdv} - \phi) = -\frac{1+\phi}{T-1} \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right) \times \left[ 1 - \frac{2\phi}{(1-\phi)(T-1)} \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right) \right]^{-1} \tag{6}$$

Where  $\hat{\phi}_{lsdv}$  is the LSDV estimator as given in (4), this formula clearly showed that the inconsistency is  $O(T^{-1})$  and negative for positive  $\phi$  (if  $T > 1$ ); Moreover, it does not depend on  $\sigma_\varepsilon^2$ . Even for small  $N$  it has been found to approximate the true bias, as assessed from Monte Carlo studies, pretty close, except for large values of  $\phi$ . Similar results have been found by Sevestre and Trognon [37]. They considered the situation where the individual effects were random and examined the consequences of various assumptions regarding the initial observations. They did not consider just the LSDV estimator, but the class of  $\lambda$ -type estimators, see Maddala [32], which included LSDV and OLS as special cases.

For the DPD model with exogenous regressor variables, Nickell [36] used partitioned regression techniques to express the LSDV estimation errors of  $\phi$  and  $\beta$  in (1) and he obtained:

$$\hat{\phi}_{lsdv} - \phi = (\tilde{y}'_{-1} M \tilde{y}_{-1})^{-1} \tilde{y}'_{-1} M \tilde{\varepsilon},$$

$$\hat{\beta}_{lsdv} - \beta = -(\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{y}_{-1} (\hat{\phi}_{lsdv} - \phi) + (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{\varepsilon}$$

Where  $\hat{\phi}_{lsdv}$  and  $\hat{\beta}_{lsdv}$  are the LSDV estimators as given in (4) while  $\tilde{y} = Ay$ ,  $\tilde{y}_{-1} = Ay_{-1}$ ,  $\tilde{X} = AX$ ,  $\tilde{\varepsilon} = A\varepsilon$ , and  $M = I_{NT} - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ , where  $A$  is defined in (5). Hence, the inconsistency reads

$$\text{Plim}_{N \rightarrow \infty}(\hat{\phi}_{lsdv} - \phi) = \left( \text{Plim}_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}'_{-1} M \tilde{y}_{-1} \right)^{-1} \text{Plim}_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}'_{-1} M \tilde{\varepsilon}, \tag{7}$$

$$\text{Plim}_{N \rightarrow \infty}(\hat{\beta}_{lsdv} - \beta) = -\text{Plim}_{N \rightarrow \infty} (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{y}_{-1} \text{Plim}_{N \rightarrow \infty}(\hat{\phi}_{lsdv} - \phi), \tag{8}$$

From which it is seen that the inconsistency critically depends on  $\text{Plim}_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}'_{-1} M \tilde{\varepsilon}$ . Because of the assumed strict exogeneity of  $X$  this term can be written as  $\text{Plim}_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}'_{-1} M \tilde{\varepsilon} = \text{Plim}_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}'_{-1} \tilde{\varepsilon}$ . Nickell [36] calculated this term as

$$\text{Plim}_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}'_{-1} \tilde{\varepsilon} = \frac{-\sigma_\varepsilon^2}{T(1-\phi)} \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right). \tag{9}$$

Substituting (9) into (7), we have

$$\text{Plim}_{N \rightarrow \infty}(\hat{\phi}_{lsdv} - \phi) = \left( \text{Plim}_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}'_{-1} M \tilde{y}_{-1} \right)^{-1} \frac{-\sigma_\varepsilon^2}{T(1-\phi)} \times \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right). \tag{10}$$

Formulas (8) and (10) are not very helpful in providing a clearcut insight into the asymptotic bias, and may even be very inaccurate as far as the actual magnitude of the bias of the LSDV estimator in small samples is concerned. He gives an indication on how a more accurate approximation might be obtained. A comparable suggestion to approximate the bias to  $O(N^{-1}T^{-1})$  is put forward by Beggs and Nerlove [11], but they did not pursue this line of approach. Moreover, their suggestion seems only applicable for the approximation of the bias in  $\hat{\phi}_{lsdv}$ , and not for the complete coefficient vector  $\hat{\delta}_{lsdv}$ .

## 4. Bias-corrected LSDV estimators

Although consistent estimates can be obtained by IV or GMM procedures, the inconsistent LSDV estimator has a relatively low variance and hence can lead to an estimator with lower root mean square error after the bias is removed. So, we interested in correcting the bias for LSDV estimator to take advantage of this low

variance. In this section, the articles which proposed the bias-corrected estimators of LSDV will be reviewed, such as Kiviet [27], Hansen [24], and Bun and Carree [19].

#### 4.1. Kiviet estimator

Kiviet [27] proposed a direct bias correction method, by deriving an approximating formula for the bias of the full vector of LSDV coefficient estimates in the DPD model in (1) with exogenous regressors. As he showed its magnitude can be evaluated and exploited easily.

The basic idea, of Kiviet's procedure to correction the bias, is the approximation of the unknown bias by a two-step procedure. In the first step empirical estimates are derived, while in the second step the bias is derived which leads to a correction of the biased estimator. The motivation for the direct correction lies in the well-known fact, that the LSDV estimator is biased but has a variance much smaller compared to IV estimators, like the Anderson-Hsiao [7] estimator.<sup>2</sup> Kiviet [27] derived the expected bias approximation for the LSDV estimator  $\hat{\delta}_{lsdv}$  of (4):

$$E(\hat{\delta}_{lsdv} - \delta) = -\sigma_{\varepsilon}^2(\bar{D})^{-1} \left( \frac{N}{T} (e_T' C e_T) [2q - \bar{Z}' A \bar{Z} (\bar{D})^{-1} q] + \text{tr}\{\bar{Z}' (I_N \otimes A_T C A_T) \bar{Z} (\bar{D})^{-1}\} q + \bar{Z}' (I_N \otimes A_T C A_T) \bar{Z} (\bar{D})^{-1} q + \sigma_{\varepsilon}^2 N q' (\bar{D})^{-1} q \times \left[ -\frac{N}{T} (e_T' C e_T) \text{tr}\{C' A_T C\} + 2 \text{tr}\{C' A_T C A_T C\} q \right] + O\left(N^{-1} T^{-\frac{3}{2}}\right) \right)$$

Where  $\bar{D} = \bar{Z}' A \bar{Z} + \sigma_{\varepsilon}^2 N \text{tr}\{C' A_T C\} q q'$ , and  $A \bar{Z} = E(AZ)$ , where  $Z$ ,  $A$ , and  $e_T$  are given in (3), (5), and (2) respectively, and  $A_T = [I_T - (e_T e_T' / T)]$ , and  $q$  is the  $(P+1) \times 1$  unitvector:  $q = (1, 0, \dots, 0)'$ , and  $C$  is  $T \times T$  matrix:

$$C = \begin{pmatrix} 0 & 0 & \dots & \dots & 0 & 0 \\ 1 & 0 & & & & 0 \\ \lambda & 1 & 0 & & & \cdot \\ \lambda^2 & \lambda & 1 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \lambda^{T-2} & \cdot & \cdot & \cdot & \lambda & 1 & 0 \end{pmatrix}$$

Where  $\lambda$  is the estimate of  $\phi$  which obtained from the first step of Kiviet's procedure. Kiviet suggested use of Anderson-Hsiao estimator as a consistent first step estimator.

#### 4.2. Hansen estimator

Hansen [24] suggested an alternative bias correction method, for the model in (1), based on the estimator which proposed by Kiviet [27]. The basic idea is to approximate the unknown bias by making use of the first step biased estimator. As the starting point the biased estimators  $\hat{\phi}_{lsdv}$  and  $\hat{\beta}_{lsdv}$  are obtained. The asymptotic bias expression is then approximated by making use of first round regression results.

He simplified the asymptotic bias formulas of  $\hat{\phi}_{lsdv}$  and  $\hat{\beta}_{lsdv}$  which are given in (10) and (8) respectively. By inserting  $M = I_{NT} - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'$  into (10), and by using  $\tilde{\varepsilon}_{-1} = \tilde{y}_{-1} - \bar{X}\hat{\beta}_{lsdv}$ , and  $\hat{\beta}_{lsdv} = (\bar{X}'\bar{X})^{-1}\bar{X}'\tilde{y}_{-1}$ , he obtained:

$$Plim_{N \rightarrow \infty}(\hat{\phi}_{lsdv} - \phi) = \left( Plim_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}_{-1}' (I_{NT} - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}') \tilde{y}_{-1} \right)^{-1} \times \frac{-\sigma_{\varepsilon}^2}{T(1-\phi)} \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right) = \left( Plim_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}_{-1}' (\tilde{y}_{-1} - \right)$$

$$\bar{X}\hat{\beta}_{lsdv} \right)^{-1} \frac{-\sigma_{\varepsilon}^2}{T(1-\phi)} \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right) = \left( Plim_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}_{-1}' \tilde{\varepsilon}_{-1} \right)^{-1} \frac{-\sigma_{\varepsilon}^2}{T(1-\phi)} \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right) \quad (11)$$

Also, he proved that the term

$$\left( Plim_{N \rightarrow \infty} \frac{1}{NT} \tilde{y}_{-1}' \tilde{\varepsilon}_{-1} \right)^{-1},$$

Is approximated by

$$\frac{NT}{\tilde{\varepsilon}_{-1}' \tilde{\varepsilon}_{-1}} \quad (12)$$

And

$$\frac{-\sigma_{\varepsilon}^2}{T(1-\phi)} \left( 1 - \frac{(1-\phi^T)}{T(1-\phi)} \right),$$

Is approximated by

$$\frac{-\hat{\sigma}_{\varepsilon}^2}{T(1-\hat{\phi}_{lsdv})} \left( 1 - \frac{(1-\hat{\phi}_{lsdv}^T)}{T(1-\hat{\phi}_{lsdv})} \right) \quad (13)$$

By using (12) and (13) in (11), he gets on the approximated bias on the basis:

$$\hat{B} = \frac{NT}{\tilde{\varepsilon}_{-1}' \tilde{\varepsilon}_{-1}} \times \frac{-\hat{\sigma}_{\varepsilon}^2}{T(1-\hat{\phi}_{lsdv})} \left( 1 - \frac{(1-\hat{\phi}_{lsdv}^T)}{T(1-\hat{\phi}_{lsdv})} \right).$$

Hansen [24] suggested a bias-corrected estimator,  $\hat{\phi}_{bc,H}$  (where the subscript *bc* means "bias-corrected" and subscript *H* refers to Hansen), by minimizing the quadratic difference between the unknown bias ( $\hat{\phi}_{lsdv} - \phi$ ) and the approximated bias on the basis of the first step estimation ( $\hat{B}$ ):

$$\hat{\phi}_{bc,H}: \text{Min}_{\phi} (B - \hat{B})^2 = \text{Min}_{\phi} \left[ \left( \hat{\phi}_{lsdv} - \phi \right) - \frac{NT}{\tilde{\varepsilon}_{-1}' \tilde{\varepsilon}_{-1}} \times \frac{-\hat{\sigma}_{\varepsilon}^2}{T(1-\hat{\phi}_{lsdv})} \left( 1 - \frac{(1-\hat{\phi}_{lsdv}^T)}{T(1-\hat{\phi}_{lsdv})} \right) \right]^2$$

The problem has to be solved iteratively. Because the unknown parameter  $\phi$  is expected to be in a rather narrow interval,  $-1 \leq \phi \leq 1$ , a grid-search is applied.

While the asymptotic bias of the vector  $\hat{\beta}_{lsdv}$  of the remaining explanatory variables is given by:

$$Plim_{N \rightarrow \infty}(\hat{\beta}_{lsdv} - \beta) = -Plim_{N \rightarrow \infty}(\bar{X}'\bar{X})^{-1}\bar{X}'\tilde{y}_{-1} Plim_{N \rightarrow \infty}(\hat{\phi}_{lsdv} - \phi) = -Plim_{N \rightarrow \infty} \hat{\beta}_{lsdv} Plim_{N \rightarrow \infty}(\hat{\phi}_{lsdv} - \phi)$$

By making use of the bias-corrected parameter of the lagged endogenous variable  $\hat{\phi}_{bc,H}$  the bias-corrected estimator for the exogenous variables  $\hat{\beta}_{bc,H}$  is estimated making again use of the first step regression results:

$$\hat{\beta}_{bc,H} = \hat{\beta}_{lsdv} + \hat{\beta}_{lsdv}(\hat{\phi}_{lsdv} - \hat{\phi}_{bc,H})$$

Behr [12] studied the behavior of several DPD estimators (LSDV, GMM) and compared between Kiviet and Hansen bias-corrected estimators by using Monte Carlo simulation. He showed that the bias of bias-corrected estimator which proposed by Hansen [24] less than bias-corrected estimator which proposed by Kiviet [27]. But the superiority of Hansen estimator decreases with growing numbers of individuals.

#### 4.3. Bun-carree estimator

As is well-known, the LSDV estimator is not consistent for large  $N$  and finite  $T$  in DPD models. Bun and Kiviet [21] and Bruno [18]

<sup>2</sup> Anderson and Hsiao [7] suggested two IV estimators, first estimator based on lagged levels as instruments. While the second estimator uses the lagged differences as instruments.

derived the infeasible bias approximations of this estimator. The bias approximations can be estimated using an initial consistent estimator such as Anderson-Hsiao or GMM estimator. All their bias correction methods depend on initial consistent estimates. While Bun and Carree [19] proposed an alternative bias correction method to the bias that directly uses LSDV estimator, obviating the need to resort to initial consistent estimates.

Bun and Carree [19] considered the model in (1), and they formulated the expressions for the case of a balanced panel to correct the bias of the inconsistent LSDV estimator. They reintroduced the LSDV estimator in (4) as follows:

$$\begin{aligned} \hat{\delta}_{lsdv} &= \begin{pmatrix} \hat{\phi}_{lsdv} \\ \hat{\beta}_{lsdv} \end{pmatrix} = (Z'AZ)^{-1}Z' Ay \\ &= \begin{pmatrix} \hat{\sigma}_{y-1}^2 & \hat{\Sigma}'_{xy-1} \\ \hat{\Sigma}_{xy-1} & \hat{\Sigma}_{xx} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\sigma}_{y-1} y \\ \hat{\Sigma}_{xy} \end{pmatrix} \end{aligned} \tag{14}$$

Where  $Z$  and  $A$  are defined in (3) and (5) respectively, while  $\hat{\Sigma}_{xy}$ ,  $\hat{\Sigma}_{xx}$ ,  $\hat{\sigma}_{y-1} y$ , and  $\hat{\Sigma}_{xy-1}$  are the sample analogs of:

$$\begin{aligned} \Sigma_{xy} &= Plim_{N \rightarrow \infty} \frac{1}{N(T-1)} X' Ay; \Sigma_{xx} = Plim_{N \rightarrow \infty} \frac{1}{N(T-1)} X' AX; \\ \hat{\sigma}_{y-1} y &= Plim_{N \rightarrow \infty} \frac{1}{N(T-1)} y'_{-1} Ay; \Sigma_{xy-1} = Plim_{N \rightarrow \infty} \frac{1}{N(T-1)} X' Ay_{-1}. \end{aligned}$$

The inconsistency of the LSDV estimator as

$$\delta^* = Plim_{N \rightarrow \infty} (\hat{\delta}_{lsdv} - \delta).$$

Bun and Carree [19] introduced the (asymptotic) squared multiple correlation coefficient of the regression of  $\hat{y}_{i,t-1}$  on  $\hat{x}_{i,t}$  through  $\hat{x}_{pit}$  as

$$\rho_{xy-1}^2 = \frac{\Sigma'_{xy-1} \Sigma_{xx}^{-1} \Sigma_{xy-1}}{\sigma_{y-1}^2}; \xi = (\xi_1, \dots, \xi_p)' = \Sigma_{xx}^{-1} \Sigma_{xy-1},$$

As the corresponding vector of regression coefficients. The inconsistency  $\delta^* = (\phi^*, \beta^*)'$  is now conveniently expressed as

$$\phi^* = \frac{-\sigma_{\varepsilon}^2 h(\phi, T)}{(1 - \rho_{xy-1}^2) \sigma_{y-1}^2}, \text{ and } \beta_p^* = -\xi_p \phi^*; p = 1, \dots, P, \tag{15}$$

Where

$$h(\phi, T) = \frac{(T-1) - T\phi + \phi^T}{T(T-1)(1-\phi)^2}. \tag{16}$$

Note that the denominator  $(1 - \rho_{xy-1}^2) \sigma_{y-1}^2$  in the first expression of (15) is the conditional variance of  $\hat{y}_{-1}$  given  $\hat{X}$ .

From the first expression in (15), it is clear that the LSDV estimator  $\hat{\phi}_{lsdv}$  is downward-biased. The extent of the (asymptotic) bias depends on five parameters:  $\phi$ ,  $T$ ,  $\sigma_{\varepsilon}^2$ ,  $\sigma_{y-1}^2$ , and  $\rho_{xy-1}^2$ . The bias of the LSDV estimator will be especially severe when (a) the value of  $\phi$  is close to 1; (b) the number of time periods,  $T$ , is low; (c) the ratio of variances,  $\sigma_{\varepsilon}^2 / \sigma_{y-1}^2$ , is high; or (d) the lagged endogenous variable and the exogenous variable are highly correlated, either positively or negatively. The second expression in (15) shows that the inconsistency of  $\hat{\beta}_{lsdv}$  is proportional to that of  $\hat{\phi}_{lsdv}$ . The bias of the LSDV estimator  $\hat{\beta}_{lsdv}$  can be either positive or negative, depending on the sign of the (asymptotic) covariances between  $\hat{y}_{-1}$  and  $\hat{X}$ .

The principle of correction the bias can be explained straightforwardly using (15). First, assume that the values for  $\sigma_{\varepsilon}^2, \rho_{xy-1}^2, \sigma_{y-1}^2$ , and  $\xi$  are known. Then use as a bias-corrected estimator,  $\hat{\phi}_{bc}$  (where the subscript  $bc$  means ‘‘bias-corrected’’; the fact that  $bc$  also are the initials of the authors’ surnames), that value of  $\phi$  for which

$$\hat{\phi}_{lsdv} = \phi - \frac{\sigma_{\varepsilon}^2 h(\phi, T)}{(1 - \rho_{xy-1}^2) \sigma_{y-1}^2}. \tag{17}$$

This estimator can then be inserted into the second expression in (15) to find the bias-corrected estimator

$$\hat{\beta}_{bc} = \hat{\beta}_{lsdv} + \xi (\hat{\phi}_{lsdv} - \hat{\phi}_{bc}).$$

The function  $h(\phi, T)$  as defined in (16) plays an important role in this nonlinear bias correction procedure. This function is always positive and monotonically increasing for  $\phi \geq -1$ , a condition that usually can be safely assumed to hold in applications.

In practice, the values for  $\sigma_{\varepsilon}^2, \rho_{xy-1}^2, \sigma_{y-1}^2$ , and  $\xi$  are unknown. The values of the latter three variables can be estimated consistently using their sample analogs (14), as follows:  $\hat{\rho}_{xy-1}^2 = \hat{\Sigma}'_{xy-1} \hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{xy-1} / \hat{\sigma}_{y-1}^2$ ,  $\hat{\sigma}_{y-1}^2$ , and  $\hat{\xi} = (\hat{\xi}_1, \dots, \hat{\xi}_p) = \hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{xy-1}$ . However, the LSDV estimator of  $\sigma_{\varepsilon}^2$  is inconsistent, and the variance of the error term can be consistently estimated only when the LSDV estimators for  $\phi$  and  $\beta$  have been bias-corrected. Bun and Carree [19] proposed three solutions to this problem that lead to the same bias-corrected estimates and all these solutions depend on the iterative procedure. In this paper, we display one of these solutions. This solution based on use of an iterative procedure for (17): substitute the LSDV estimate for  $\sigma_{\varepsilon}^2$  in (17) to achieve one-step estimates for  $\phi$  and  $\beta$ . These estimates are used to compute the one-step estimate for  $\sigma_{\varepsilon}^2$ . This one-step estimate is again substituted in (17) to achieve two-step estimates for  $\phi$  and  $\beta$  and so on until convergence.

Bun and Carree [19] used Monte Carlo simulation to compare the performance of their bias-corrected estimator (17) with the original LSDV estimator (4), the bias-corrected estimator which introduced by Kiviet [27], and the GMM estimator of Arellano and Bond [8]. Their simulation results on various designs showed that, based on root mean squared error criterion, their bias-corrected estimator performs well when  $T$  is small and  $N$  is large. Also they showed that bias-corrected LSDV estimators (Kiviet and Bun-Carree estimators) perform well against GMM estimators. In cases where both  $T$  and  $N$  are small, the limiting distributions for the estimators may have little to say about the actual distribution (especially when  $\phi$  is close to unity). However, given the strong (relative) performance of Bun-Carree estimator in the Monte Carlo exercises in cases where  $T$  is as small as 2 or 3, this estimator appears suitable for research efforts with samples with large numbers of individuals/firms and a (very) small number of time periods.

Bun and Carree [20] studied the extended case on bias-corrected estimator for the fixed-effects dynamic panel data model which has been presented by them in Bun and Carree [19]. They derived the inconsistency of the LSDV estimator for finite  $T$  and  $N$  large in case of both time-series and cross section heteroscedasticity and they showed how to implement it in bias correction procedures.

Lokshin [31] compared the performance of three proposed estimators for DPD models (bias-corrected LSDV which introduced by Bun and Carree, ML, and MD) along with GMM by using Monte Carlo. He showed that ML and Bun-Carree estimators have the smallest bias and are good alternatives for the GMM especially when  $T$  is small ( $T \leq 6$ ).

## 5. The bias in GMM estimators

Since the work of Arellano and Bond [8], the GMM technique has been widely used in the estimation of dynamic panel data models. However, subsequent examinations of the finite sample performance of the GMM estimator showed that it is substantially biased. One source of the bias, first discovered by Nelson and Startz [34], [35], is weak instruments. Staiger and Stock [38] showed that the instrumental variables estimator would be inconsistent under weak instrument asymptotics. We call this the ‘‘weak instruments problem’’. The other source of bias is the relative number of instruments to sample size. Especially, in linear simultane-

ous equation models, Kunitomo [29], Morimune [33], and Bekker [13] showed that the two stage least squares (2SLS) estimator is inconsistent as the number of the instruments tends to infinity. Hahn and Hausman [23] showed that the finite sample bias of 2SLS estimator is monotonically increasing in the number of instruments. One important finding of the papers listed above is that the magnitude of the bias is proportional to the relative size of the number of instruments to the sample size. We label this the “many instruments problem”.

These considerations are still binding in the estimation of DPD models. It is well known that in the first differencing models the bias is sizable when the parameter concerning the lagged dependent variable is close to unity (Alonso-Borrego and Arellano [4]). Blundell and Bond [14] showed that this is due to weak instruments, and to overcome the “weak instruments problem” they proposed the system GMM estimator. They first showed that the level GMM estimators by Arellano and Bover [9] are free from weak instruments when the parameter concerning the lagged variable is close to unity and then combined the moment conditions which are used in the first differencing (by Arellano and Bond [8]) and the level GMM estimators to improve the efficiency of the estimator. The system GMM estimator is becoming widely used in empirical analyses. Empirical applications include the estimation of production functions (Blundell and Bond [15]) and empirical growth models (Bond et al. [17]), among others.

In this section, we will display the finite sample properties of the GMM estimators. Specifically, we focus in our study on three GMM estimators that were provided by Arellano and Bond [8], Arellano and Bover [9], and Blundell and Bond [14].

### 5.1. The AR (1) panel model and GMM estimators

The AR (1) panel data model without exogenous variables can be written as

$$y_{it} = \phi y_{i,t-1} + \mu_i + \varepsilon_{it}, i = 1, \dots, N; t = 2, \dots, T, \quad (18)$$

Under the assumptions:

- i)  $\varepsilon_{it}$  are i.i.d across time and individuals and independent of  $\mu_i$  and  $y_{i1}$  with  $E(\varepsilon_{it}) = 0, Var(\varepsilon_{it}) = \sigma_\varepsilon^2$ .
- ii)  $\mu_i$  are i.i.d across individuals with  $E(\mu_i) = 0, Var(\mu_i) = \sigma_\mu^2$ .
- iii) The initial observations satisfy  $y_{i1} = \frac{\mu_i}{1-\phi} + w_{i1}$  for  $i = 1, \dots, N$ , where  $w_{i1} = \sum_{j=0}^{\infty} \phi^j \varepsilon_{i,1-j}$  and independent of  $\mu_i$ .

Assumptions (i) and (ii) are the same as in Blundell and Bond [14], while assumption (iii) has been developed by Alvarez and Arellano [5]. Stacking equation (18) over time, we obtain

$$y_i = \phi y_{i,-1} + u_i, \text{ for } i = 1, \dots, N \quad (19)$$

Where

$$y_i = \begin{pmatrix} y_{i3} \\ \vdots \\ y_{iT} \end{pmatrix}, y_{i,-1} = \begin{pmatrix} y_{i2} \\ \vdots \\ y_{i,T-1} \end{pmatrix}, u_i = \begin{pmatrix} u_{i3} \\ \vdots \\ u_{iT} \end{pmatrix}; \text{ With } u_{it} = \mu_i + \varepsilon_{it}.$$

Given these assumptions, we get three types of GMM estimators. These include first-difference GMM (DIF) estimator, level GMM (LEV) estimator, and system GMM (SYS) estimator.

#### 5.1.1. First-difference GMM estimator

The individual effect ( $\mu_i$ ) in (18) causes a severe correlation between the lagged endogenous variable ( $y_{i,-1}$ ) and the error term ( $u_i$ ). In order to eliminate the individual effect, Arellano and Bond [8] used the first differences of the model in (19):

$$\Delta y_i = \phi \Delta y_{i,-1} + \Delta u_i,$$

Where

$$\Delta y_i = \begin{pmatrix} y_{i3} - y_{i2} \\ \vdots \\ y_{iT} - y_{i,T-1} \end{pmatrix}, \Delta y_{i,-1} = \begin{pmatrix} y_{i2} - y_{i1} \\ \vdots \\ y_{i,T-1} - y_{i,T-2} \end{pmatrix}, \Delta u_i = \begin{pmatrix} u_{i3} - u_{i2} \\ \vdots \\ u_{iT} - u_{i,T-1} \end{pmatrix},$$

And then they showed that

$$E(H_i^{D'} \Delta u_i) = 0, \quad (20)$$

Where

$$H_i^D = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{i,T-2} \end{pmatrix}. \quad (21)$$

Using (20) as the orthogonal conditions in the GMM, Arellano and Bond [8] constructed the one-step DIF (DIF1) estimator for  $\phi$ , which is given by:

$$\hat{\phi}^D = (\Delta y'_{-1} H^D W^D H^{D'} \Delta y_{-1})^{-1} \Delta y'_{-1} H^D W^D H^{D'} \Delta y, \quad (22)$$

Where

$$\Delta y_{-1} = \begin{pmatrix} \Delta y_{1,-1} \\ \vdots \\ \Delta y_{N,-1} \end{pmatrix}, \Delta y = \begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_N \end{pmatrix}, H^D = \begin{pmatrix} H_1^D \\ \vdots \\ H_N^D \end{pmatrix},$$

And they used the following matrix as an initial weighting matrix:

$$W^D = \left( \frac{1}{N} \sum_{i=1}^N H_i^{D'} D H_i^D \right)^{-1},$$

Where  $D$  is a  $(T-2) \times (T-2)$  first-difference operator matrix

$$D = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & \dots & 0 & -1 & 2 \end{pmatrix}$$

To get the two-step DIF (DIF2) estimator, the moment conditions are weighted by

$$W_{(2)}^D = \left( \frac{1}{N} \sum_{i=1}^N H_i^{D'} \Delta \hat{u}_i \Delta \hat{u}_i' H_i^D \right)^{-1},$$

Where  $\Delta \hat{u}_i$  are the fitted residuals from DIF1 in (22).

Blundell and Bond [14] showed that when  $\phi$  is close to unity and/or  $\sigma_\mu^2 / \sigma_\varepsilon^2$  increases the instruments matrix (21) becomes invalid. This means that the DIF estimator has the weak instruments problem.

#### 5.1.2. Level GMM estimator

Arellano and Bover [9] suggested use of the instrumental variables to eliminate the individual effect from the DPD model, while, as mentioned above, Arellano and Bond [8] used the first differences of the DPD model to eliminate it. Explicitly, Arellano and Bover [9] considered the level model of (19) and then they used the following matrix as instrumental variables:

$$H_i^L = \begin{pmatrix} \Delta y_{i2} & 0 & \dots & 0 \\ 0 & \Delta y_{i3} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \Delta y_{i,T-1} \end{pmatrix}$$

Which not contains individual effect and satisfies the orthogonal conditions

$$E(H_i^{L'} u_i) = 0. \tag{23}$$

Using (23), Arellano and Bover's [9] one-step LEV (LEV1) estimator is calculated as follows:

$$\hat{\phi}^L = (y'_{-1} H^L W^L H^{L'} y_{-1})^{-1} y'_{-1} H^L W^L H^{L'} y. \tag{24}$$

Where

$$y_{-1} = \begin{pmatrix} y_{1,-1} \\ \vdots \\ y_{N,-1} \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, H^L = \begin{pmatrix} H_1^L \\ \vdots \\ H_N^L \end{pmatrix} \tag{25}$$

And the y used the following matrix as an initial weighting matrix:

$$W^L = \left( \frac{1}{N} \sum_{i=1}^N H_i^{L'} H_i^L \right)^{-1}.$$

To get the two-step LEV (LEV2) estimator, similarly as in DIF2 estimator, the moment conditions are weighted by

$$W_{(2)}^L = \left( \frac{1}{N} \sum_{i=1}^N H_i^{L'} \hat{u}_i^L H_i^L \right)^{-1},$$

Where  $\hat{u}_i$  are the fitted residuals from LEV1 estimator in (24).

### 5.1.3. System GMM estimator

Arellano and Bover [9] and Blundell and Bond [14] proposed a system GMM estimator in which the moment conditions of DIF and LEV are used jointly to avoid weak instruments and improve the efficiency of the estimator. The moment conditions used in constructing the system GMM estimator are given by

$$E(H_i^{S'} u_i^S) = 0. \tag{26}$$

Where  $u_i^S = (\Delta u_i', u_i')'$  and  $H_i^S$  is a  $2(T-2) \times (T+1)(T-2)/2$  block diagonal matrix given by

$$H_i^S = \begin{pmatrix} H_i^D & 0 \\ 0 & H_i^L \end{pmatrix}$$

Using (26), the one-step SYS (SYS1) estimator is calculated as follows:

$$\hat{\phi}^S = (y_{-1}^{S'} H^S W_G^S H^{S'} y_{-1}^S)^{-1} y_{-1}^{S'} H^S W_G^S H^{S'} y^S, \tag{27}$$

Where

$$y_{-1}^S = \begin{pmatrix} \Delta y_{1,-1} \\ y_{1,-1} \\ \vdots \\ \Delta y_{N,-1} \\ y_{N,-1} \end{pmatrix}, y^S = \begin{pmatrix} \Delta y_1 \\ y_1 \\ \vdots \\ \Delta y_N \\ y_N \end{pmatrix}, H^S = \begin{pmatrix} H_1^S \\ \vdots \\ H_N^S \end{pmatrix}.$$

And they used the following matrix as an initial weighting matrix:

$$W_G^S = \left( \frac{1}{N} \sum_{i=1}^N H_i^{S'} G H_i^S \right)^{-1}; \text{ With } G = \begin{pmatrix} D & 0 \\ 0 & I_{T-2} \end{pmatrix}. \tag{28}$$

To get the two-step SYS (SYS2) estimator, the moment conditions are weighted by:

$$W_{G(2)}^S = \left( \frac{1}{N} \sum_{i=1}^N H_i^{S'} \hat{u}_i^S \hat{u}_i^{S'} H_i^S \right)^{-1},$$

Where  $\hat{u}_i^S$  are the fitted residuals from SYS1 estimator in (27). Also, they used the identity matrix ( $I_{2T-4}$ ) instead of  $G$  in (28), in this case the weighting matrix is:

$$W_I^S = \left( \frac{1}{N} \sum_{i=1}^N H_i^{S'} H_i^S \right)^{-1},$$

In their first step of two-step system GMM estimator, which yields the simple system GMM estimator. This is certainly not optimal either, but is easy and could perhaps suit well as first step in a two-step procedure. To get the two-step SYS estimator, the moment conditions are weighted by

$$W_{I(2)}^S = \left( \frac{1}{N} \sum_{i=1}^N H_i^{S'} \hat{u}_i^S \hat{u}_i^{S'} H_i^S \right)^{-1},$$

Where  $\hat{u}_i^S$  are the fitted residuals from one-step SYS estimator with use  $W_I^S$  as the initial weighting matrix.

Note that  $\hat{\phi}^S$  in (27) can be expressed as:

$$\hat{\phi}^S = \hat{\gamma} \hat{\phi}^D + (1 - \hat{\gamma}) \hat{\phi}^L,$$

$$\text{Where } \hat{\gamma} = \frac{\Delta y_{-1}' H^D (H^{D'} H^D)^{-1} H^{D'} \Delta y_{-1}}{\Delta y_{-1}' H^D (H^{D'} H^D)^{-1} H^{D'} \Delta y_{-1} + y_{-1}' H^L (H^{L'} H^L)^{-1} H^{L'} y_{-1}}.$$

It is worth mentioning that Abonazel [2] provided R-code to calculate DIF, LEV, and SYS estimates. Moreover, this code has been designed to enable the researcher to make a simulation study in DPD models, such as the simulation study in Youssef et al. [42].<sup>3</sup>

### 5.2. Small sample bias properties of GMM estimators

Hayakawa [25] considered the model in (18) under the assumptions (i) to (iii), and he derived the second order bias for DIF, LEV, and SYS estimators. He considered the general one-step GMM estimator based on the moment condition  $E[f(H_i, \phi_0)] = E[f_i(\phi_0)] = 0$ . An inefficient one-step GMM estimator is defined as

$$\hat{\phi} = \text{arg min}_{\phi} \left[ \frac{1}{N} \sum_{i=1}^N f_i(\phi) \right]' W_N \left[ \frac{1}{N} \sum_{i=1}^N f_i(\phi) \right],$$

Where  $W_N$  is a positive semidefinite weighting matrix which satisfies  $\text{plim}_{N \rightarrow \infty} W_N = W$ , and  $W_i = W(H_i)$  are symmetric and positive definite matrices which do not depend on parameter  $\phi$ .

Generally, an estimator of  $\phi$ ,  $\hat{\phi}$ , based on a sample of size  $N$  allows for an expansion of the form:

$$\sqrt{N}(\hat{\phi} - \phi) = \phi^{(1)} + \frac{1}{\sqrt{N}} \phi^{(2)} + O\left(\frac{1}{N}\right),$$

Where  $\phi^{(1)}$  and  $\phi^{(2)}$  are  $O(1)$ . Typically,  $\phi^{(1)}$  has a zero mean and converges in distribution to a normal distribution. By taking an expectation and ignoring the  $O\left(\frac{1}{N}\right)$  term, then the approximate mean of  $\sqrt{N}(\hat{\phi} - \phi)$  is  $\frac{1}{\sqrt{N}} E(\phi^{(2)})$ . Therefore, the second order bias of  $\hat{\phi}$  is  $\frac{1}{\sqrt{N}} E(\phi^{(2)})$ .

To display the formulas of the second order bias for DIF, LEV, and SYS estimators, Hayakawa [25] defined the following notations:

<sup>3</sup> For information about a Monte Carlo simulation study, see Abonazel [3]. He provided the main steps (a simple guide) for making the Monte Carlo simulation study using R language. Moreover, He also provided some empirical examples in econometrics.

$$\pi_d = \begin{bmatrix} \frac{-\sigma_\varepsilon^2}{(1+\phi)(R+D)} & \frac{\sigma_\varepsilon^2 R(1-\phi)}{F(1+\phi)} & \frac{-\sigma_\varepsilon^2(1-\phi)[R+D(1+\phi)]}{F(1+\phi)} \end{bmatrix}' =$$

$$[\pi_{d1} \quad \pi_{d2} \quad \pi_{d3}]'$$

$$\pi_l = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}' = [\pi_{l1} \quad \pi_{l2}]'$$

Where

$$R = \frac{\sigma_\mu^2}{(1-\phi)^2}, D = \frac{\sigma_\varepsilon^2}{1-\phi^2}, \sigma_y^2 = R + D, F = \left[ \frac{\sigma_\varepsilon^2}{1+\phi} \right] \left[ 2R + \frac{\sigma_\varepsilon^2}{1-\phi} \right].$$

And

$$\alpha_d = \sigma_y^2(\pi_{d1}^2 + \pi_{d2}^2 + \pi_{d3}^2) + 2\pi_{d2}\pi_{d3}(R + \phi D), \alpha_l = \frac{\sigma_\varepsilon^2}{1+\phi}, \gamma = \frac{\alpha_d}{\alpha_d + \alpha_l}.$$

The second order biases of  $\hat{\phi}^D$ ,  $\hat{\phi}^L$ , and  $\hat{\phi}^S$  are given by

$$N \cdot \text{Bias}(\hat{\phi}^D) = B_1^D + B_2^D + B_3^D + B_4^D,$$

$$N \cdot \text{Bias}(\hat{\phi}^L) = B_1^L + B_2^L + B_3^L + B_4^L,$$

$$N \cdot \text{Bias}(\hat{\phi}^S) = \gamma(B_1^D + B_2^D) + (1-\gamma)(B_1^L + B_2^L) + \gamma^2(B_3^D + B_4^D) + (1-\gamma)^2(B_3^L + B_4^L) - \frac{2}{(\alpha_d + \alpha_l)^2} \psi_3 + \frac{1}{(\alpha_d + \alpha_l)^2} \psi_4.$$

Where

$$B_1^D = 0, B_2^D = -\frac{\sigma_\varepsilon^2}{\alpha_d} \left[ 1 + \frac{2\sigma_y^4}{F} - \frac{2(R+\phi D)^2}{F} \right],$$

$$B_3^D = \frac{2\sigma_\varepsilon^2}{\alpha_d^2} \left[ \sigma_y^2 \{ \pi_{d1}^2 + \pi_{d2}^2 + \pi_{d3}^2 + \pi_{d1}\pi_{d2}(\phi - 2) \} - \pi_{d1}\pi_{d3} \{ R(2 - \phi) - \phi D(2\phi - 3) \} + 2\pi_{d2}\pi_{d3}(R + \phi D) \right],$$

$$B_4^D = -\frac{2\sigma_\varepsilon^2}{\alpha_d^2} \left[ \pi_{d1}\pi_{d3}^2(R + \phi D) + \pi_{d1}\pi_{d2}\pi_{d3}\sigma_y^2 \right],$$

$$B_1^L = 0, B_2^L = \frac{2\sigma_\mu^2}{\alpha_l(1-\phi)}, B_3^L = -\frac{\sigma_\varepsilon^2}{\alpha_l^2} \left[ \frac{\sigma_\mu^2}{1-\phi} + \frac{\sigma_\varepsilon^2(2\phi-1)}{2(1+\phi)} \right],$$

$$B_4^L = \frac{\sigma_\varepsilon^4(\phi-1)}{4\alpha_l^2(1+\phi)},$$

$\psi_3 =$

$$\pi_d' \begin{bmatrix} \frac{\sigma_\varepsilon^2 \sigma_\mu^2 (1-3\phi)}{(1-\phi)^2(1+\phi)} - \frac{\sigma_\varepsilon^4(2\phi-1)}{1-\phi^2} & \frac{\sigma_\varepsilon^2 \sigma_\mu^2 (1+3\phi-2\phi^2)}{(1-\phi)^2(1+\phi)} - \frac{\phi \sigma_\varepsilon^4(1+4\phi-2\phi^2)}{1-\phi^2} \\ \frac{-\sigma_\varepsilon^2 \sigma_\mu^2}{1+\phi} - \frac{\sigma_\varepsilon^4}{1+\phi} & \frac{\sigma_\varepsilon^2 \sigma_\mu^2 (1-3\phi)}{(1-\phi)^2(1+\phi)} - \frac{\phi \sigma_\varepsilon^4(2\phi-1)}{1-\phi^2} \\ \frac{-\sigma_\varepsilon^2 \sigma_\mu^2}{1+\phi} - \frac{\sigma_\varepsilon^4}{1+\phi} & \frac{\sigma_\varepsilon^2 \sigma_\mu^2 (1-3\phi)}{(1-\phi)^2(1+\phi)} - \frac{\phi \sigma_\varepsilon^4(2\phi-1)}{1-\phi^2} \end{bmatrix} \pi_l,$$

$$\psi_4 = \pi_d' \begin{bmatrix} \frac{2\pi_{l1}\sigma_\varepsilon^4}{1+\phi} - \frac{2\pi_{d1}\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi^2} & \frac{2\pi_{l2}\phi\sigma_\varepsilon^4(\phi-2)}{1+\phi} - \frac{2\pi_{d1}\phi\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi^2} \\ \frac{-2\pi_{d2}\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi^2} & \frac{2\pi_{l2}\sigma_\varepsilon^4}{1+\phi} - \frac{2\pi_{d2}\phi\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi^2} - \frac{\pi_{d3}\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi} \\ \frac{2\pi_{d3}\phi\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi^2} & \frac{2\pi_{l2}\sigma_\varepsilon^4}{1+\phi} - \frac{\pi_{d2}\phi\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi} - \frac{2\pi_{d3}\sigma_\varepsilon^2\sigma_\mu^2}{1-\phi^2} \end{bmatrix} \pi_l.$$

He confirmed how well the second order biases explain the actual biases by comparing theoretical values with simulation values. Since the biases of all estimators are characterized by  $N$ ,  $\phi$  and  $\sigma_\mu^2/\sigma_\varepsilon^2$ , he calculated the theoretical values of the biases for the cases  $\sigma_\mu^2/\sigma_\varepsilon^2 = 0.25, 1, 4$  with  $\phi = 0.1, \dots, 0.9$  and  $N = 50$ . He found that the theoretical and simulation values are close when  $\phi \leq 0.5$ .

Also, he found that the bias of the SYS estimator is a weighted sum of the biases in opposite directions of the DIF and LEV estimators. In addition, he found that the role of the weight is also important since it adjusts the difference of the magnitudes of the biases. And he provided theoretical evidence why the SYS estimator

has smaller bias. When  $\sigma_\mu^2$  and  $\sigma_\varepsilon^2$  are of almost the same value is an important reason why the system estimator has small bias. In the case when  $\sigma_\mu^2/\sigma_\varepsilon^2 = 4$ , the biases of the entire GMM estimator are sizable.

## 6. Bias-corrected GMM estimators

Chigira and Yamamoto [22] proposed a bias-corrected estimator based on reduce the bias in GMM estimator. Let  $T$  be an even integer, they defined the bias-corrected estimator ( $\hat{\phi}_{CY}^{BC}$ ) as follows:

$$\hat{\phi}_{CY}^{BC} = 2\tilde{\phi} - \frac{1}{2}(\tilde{\phi}_I + \tilde{\phi}_{II}), \quad (29)$$

Where  $\tilde{\phi}$  is a GMM estimator for the whole period, while  $\tilde{\phi}_I$  and  $\tilde{\phi}_{II}$  are GMM estimators based on a sample of the first period ( $t = 3, \dots, T/2 + 1$ ) and of the second period ( $t = T/2 + 2, \dots, T$ ), respectively. Next, we define the asymptotic bias of  $\tilde{\phi}$ ,  $ABias(\tilde{\phi})$ , as  $E(\tilde{\phi} - \phi)$ , ignoring the terms whose orders are lower than  $O\left(\frac{1}{NT}\right)$ .

They showed that the bias-corrected estimator in (29) has no asymptotic bias, i.e.  $ABias(\hat{\phi}_{CY}^{BC}) = o\left(\frac{1}{NT}\right)$  if satisfies the following condition:

$$ABias(\tilde{\phi}) = \frac{c}{NT} + o\left(\frac{1}{NT}\right), \quad (30)$$

Where  $c$  is a finite-valued constant independent of  $T$  and  $N$ . Note that, when  $T$  is odd integer, say  $T = 2m + 1$ , the bias-corrected estimator defined as:

$$\hat{\phi}_{CY}^{BC} = 2\tilde{\phi} - \frac{m}{2m+1}\tilde{\phi}_I - \frac{m+1}{2m+1}\tilde{\phi}_{II}.$$

Where  $\tilde{\phi}$  is a GMM estimator for the whole period, while  $\tilde{\phi}_I$  and  $\tilde{\phi}_{II}$  are GMM estimators based on a sample of the first period ( $t = 3, \dots, m + 1$ ) and of the second period ( $t = m + 2, \dots, T$ ), respectively.

The  $\hat{\phi}_{CY}^{BC}$  estimator in (29) has no asymptotic bias as long as (30) holds. Unfortunately, the asymptotic biases of DIF, LEV, and SYS GMM estimators that were displayed in above are not given by (30). Explicitly, these estimators have, as shown by Bun and Kiviet [21], a bias of the order  $1/N$  and do not satisfy (30). Chigira and Yamamoto [22] solved this problem by providing a GMM estimator whose bias is given by (30).

## 7. Alternative GMM estimators

In this section, we review the new GMM estimators which presented by Youssef and Abonazel [40] as alternative estimators for the conventional GMM estimators. They proposed a new approach to improve the efficiency of the conventional GMM estimators. Their approach (Youssef-Abonazel's approach) is based on the optimal (or at least suboptimal) weighting matrix<sup>4</sup> of GMM estimation, then use of these matrices as new weighting matrices in GMM estimation, and then we get new GMM estimators. The new GMM estimators are more efficient than the conventional GMM estimators.

In level GMM estimation, Youssef et al. [41] showed that  $W^L$  is an optimal weighting matrix only in the case of  $\sigma_\mu^2 = 0$ , i.e., no individual effects case, and they presented an optimal weighting matrix for LEV estimator, in the general case, as:

$$W^{OL} = \left( \frac{1}{N} \sum_{i=1}^N H_i' J_{T-2} H_i \right)^{-1}; \text{ With } J_{T-2} = I_{T-2} + \rho l_{T-2} l_{T-2}',$$

<sup>4</sup> Youssef et al. [41] studied the conventional GMM weighting matrices that using in LEV and SYS GMM estimators, and also they proposed other more efficient weighting matrices.

Where  $\rho = \sigma_{\mu}^2 / \sigma_{\varepsilon}^2$  and  $l_{T-2}$  is a  $(T - 2) \times 1$  vector of ones.

Note that the use of the weighting matrix  $W^{OL}$  can be described as inducing cross-sectional heterogeneity through  $\rho$ , and also can be explained as partially adopting a procedure of generalized least squares to the level estimation. So, using  $W^{OL}$ , instead of  $W^L$ , certainly improve the efficiency of LEV estimator. So, using  $W^{OL}$ , we can obtain the optimal level GMM (OLEV) estimator:

$$\hat{\phi}^{OL} = (y'_{-1} H^L W^{OL} H^L y_{-1})^{-1} y'_{-1} H^L W^{OL} H^L y.$$

In system GMM estimation, Youssef and Abonazel [40] used  $J_{T-2}$  in the weighting matrix to improve the efficiency for SYS estimator as follows:

$$W^{SS} = \left( \frac{1}{N} \sum_{i=1}^N H_i^S G_j H_i^S \right)^{-1}; \text{ With } G_j = \begin{pmatrix} D & 0 \\ 0 & J_{T-2} \end{pmatrix}.$$

So, they presented the suboptimal system GMM (SSYS) estimator<sup>5</sup> depending on the suboptimal weighting matrix ( $W^{SS}$ ):

$$\hat{\phi}^{SS} = (y_{-1}^S H^S W^{SS} H^S y_{-1}^S)^{-1} y_{-1}^S H^S W^{SS} H^S y^S.$$

Since an asymptotically efficient estimator can be obtained through the two-step procedure in the standard GMM estimation. In the first step, an initial positive semidefinite weighting matrix is used to obtain consistent estimates of the parameters. Given these, a weighting matrix can be constructed and used for asymptotically efficient two-step estimates. It is well known<sup>6</sup> that the two-step estimated standard errors have a small-sample downward bias in dynamic panel data setting, and one-step estimates with robust standard errors are often preferred. Therefrom Youssef and Abonazel [40] suggested using the three-step procedure in GMM estimation; this can obtain by replacing the residuals from the two-step estimation into new weighting matrix to achieve more efficiency for all GMM estimators.<sup>7</sup>

Note that if  $\rho = 0$ , we get  $J_{T-2} = I_{T-2}$  then  $W^{OL} = W^L$  and  $W^{SS} = W_G^S$ . Therefrom Youssef and Abonazel's estimators (OLEV and SSYS) are equivalent to the conventional GMM (LEV and SYS) estimators.

Youssef and Abonazel [40] used the Kantorovich inequality (KI)<sup>8</sup> to study the efficiency gain for the new estimators against the conventional estimators. They find that the new estimators are more efficient than the conventional estimators. Moreover, the potential efficiency gain for the new estimators becomes large when the variance of individual effects ( $\sigma_{\mu}^2$ ) increases compared with the variance of the errors ( $\sigma_{\varepsilon}^2$ ). In other words, the advantages from OLEV and SSYS estimators are increasing when  $\rho$  is increasing. Moreover, They make Monte Carlo simulation study to illustrate the moderate and large samples performance of LEV, SYS, OLEV, and SSYS estimators in different situations of  $\phi, \rho$ , and  $T$ . The simulation study confirms the KI conclusion, i.e., the simulation study indicates that the OLEV and SSYS estimators are more efficient than LEV and SYS estimators, respectively.<sup>9</sup> Since, the bias and RMSE of SSYS are smaller, in most situations, than the bias and RMSE of OLEV especially when  $\rho > 1$ . Consequently, they concluded that the SSYS estimator will provide useful parameter estimates for the practitioner.

<sup>5</sup> Kiviet [28] proposed a similar estimator using an optimal weighting matrix based on the particular values of  $\phi, \sigma_{\mu}^2$ , and  $\sigma_{\varepsilon}^2$ .

<sup>6</sup> See, e.g., Arellano and Bond [8] and Windmeijer [39].

<sup>7</sup> For more details about the three-step procedure in GMM estimation, See Youssef and Abonazel [40].

<sup>8</sup> Youssef and Abonazel [40] calculated the IK upper bounds of SYS and SSYS estimators. These bounds have been derived by Liu and Neudecker [30].

<sup>9</sup> See Tables 1 to 5 in Youssef and Abonazel [40].

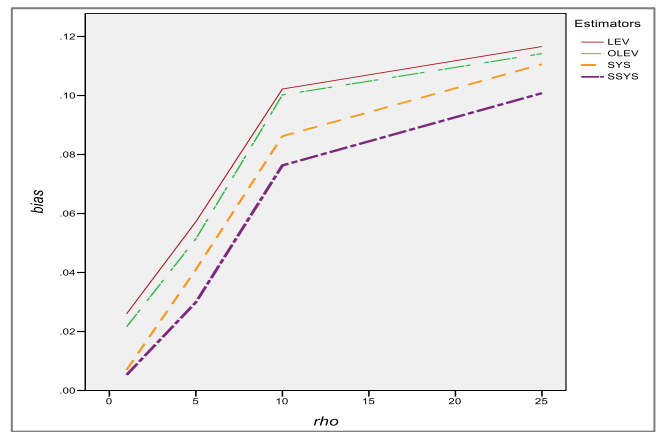


Fig. 1: Bias of the conventional and new GMM estimators when  $\phi = .85, T = 8$ , and  $N = 600$ .

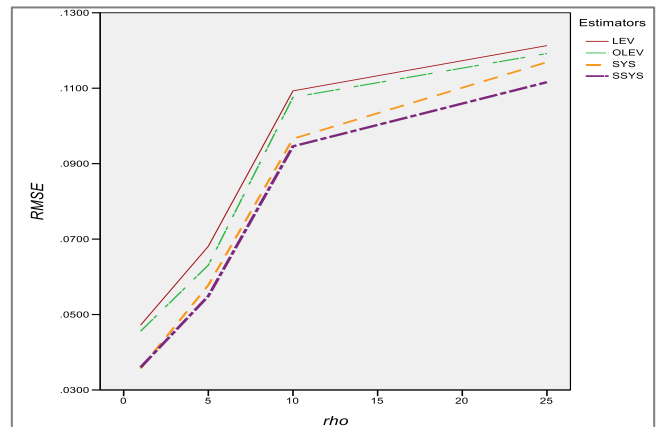


Fig. 2: RMSE of the conventional and new GMM estimators when  $\phi = .85, T = 8$ , and  $N = 600$ .

For further clarification, the important simulation results for Youssef and Abonazel [40] were summarized in this paper in two Figures. Fig. 1 presents the bias values of LEV, SYS, OLEV, and SSYS estimators when  $\phi = .85, T = 8$ , and  $N = 600$ . While Fig. 2 presents RMSE values of the same estimators in the same values of  $\phi, T$ , and  $N$ . Figs. 1 and 2 show that Youssef-Abonazel's approach improves the bias and RMSE of GMM estimators together, not only RMSE. Moreover, SSYS estimator is better than other GMM estimators, especially when  $\rho$  is increasing ( $\rho > 1$ ) even if  $T$  is small and  $\phi$  close to one.

## 8. Conclusions

In this paper, we discuss the estimation methods for DPD models with fixed effects which suggested in econometric literature, and we focus on LS and GMM methods. All these methods obtain biased estimators for DPD models. Therefore, we discuss the different methods to correct the bias of LS and GMM estimations. And we present the analytical expressions for the asymptotic biases of the LS and GMM estimators for large  $N$  and finite  $T$ . Finally, we discuss the properties of Youssef and Abonazel's [40] estimators.

From this review, we can divide all improving methods of the estimation in DPD models into two approaches. The first approach includes all the bias correction methods (all methods above except Youssef-Abonazel's approach). In this approach, we improve the estimation by remove (or at least reduce) the bias from the estimates, without any improvement in the efficiency of it. While the second approach includes the methods that improve the efficiency of the estimation, without any direct improvement in bias, as in Youssef-Abonazel's approach. However, the bias of estimates from the second approach is also very small. Therefrom, we conclude that use of the second approach is more efficient than the first approach.



## References

- [1] Abonazel, M. R. (2014). Some estimation methods for dynamic panel data models. PhD thesis. Institute of Statistical Studies and Research. Cairo University.
- [2] Abonazel, M. R. (2015). R-Codes to Calculate GMM Estimations for Dynamic Panel Data Models. Working paper, No. 70627. University Library of Munich, Germany.
- [3] Abonazel, M. R. (2015). How to Create a Monte Carlo Simulation Study using R: with Applications on Econometric Models. *Annual Conference on Statistics, Computer Sciences and Operations Research*. Institute of Statistical Studies and Research. Cairo University.
- [4] Alonso-Borrego, C., Arellano, M. (1999). Symmetrically normalized instrumental variable estimation using panel data. *Journal of Business and Economic Statistics* 17:36–49.
- [5] Alvarez, J., Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71:1121–1159. <http://dx.doi.org/10.1111/1468-0262.00441>.
- [6] Anderson, T., Hsiao, C. (1981). Estimation of dynamics models with error components. *Journal of the American Statistical Association* 76:598–606. <http://dx.doi.org/10.1080/01621459.1981.10477691>.
- [7] Anderson, T., Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18:47–82. [http://dx.doi.org/10.1016/0304-4076\(82\)90095-1](http://dx.doi.org/10.1016/0304-4076(82)90095-1).
- [8] Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58:277–98. <http://dx.doi.org/10.2307/2297968>.
- [9] Arellano, M., Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68:29–51. [http://dx.doi.org/10.1016/0304-4076\(94\)01642-D](http://dx.doi.org/10.1016/0304-4076(94)01642-D).
- [10] Baltagi, B. H. (2013). *Econometric Analysis of Panel Data*. 5th ed. Chichester: John Wiley and Sons.
- [11] Beggs, J., Nerlove, M. (1988). Biases in dynamic models with fixed effects. *Economics Letters* 26:29–31. [http://dx.doi.org/10.1016/0165-1765\(88\)90046-8](http://dx.doi.org/10.1016/0165-1765(88)90046-8).
- [12] Behr, A. (2003). A Comparison of dynamic panel data estimators: Monte Carlo evidence and an application to the investment function. Working paper. Research Centre of Deutsche Bundesbank, Economic Studies with number 2003, 05.
- [13] Bekker, P.A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62:657–681. <http://dx.doi.org/10.2307/2951662>.
- [14] Blundell, R., Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87:115–143. [http://dx.doi.org/10.1016/S0304-4076\(98\)00009-8](http://dx.doi.org/10.1016/S0304-4076(98)00009-8).
- [15] Blundell, R., Bond, S. (2000). GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews* 19:321–340. <http://dx.doi.org/10.1080/07474930008800475>.
- [16] Bond, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal* 1:141–162. <http://dx.doi.org/10.1007/s10258-002-0009-9>.
- [17] Bond, S., Hoeffler, A., Temple, J. (2001). GMM estimation of empirical growth models. Working Paper, No. 2001-W21. University of Oxford, Nuffield College, Economics Group.
- [18] Bruno, G.S. (2005). Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models. *Economics Letters* 87:361–366. <http://dx.doi.org/10.1016/j.econlet.2005.01.005>.
- [19] Bun, M., Carree, M. (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23:200–210. <http://dx.doi.org/10.1198/073500104000000532>.
- [20] Bun, M., Carree, M. (2006). Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters* 92:220–227. <http://dx.doi.org/10.1016/j.econlet.2006.02.008>.
- [21] Bun, M., Kiviet, J. (2003). On the diminishing returns of higher-order terms in asymptotic expansions of bias. *Economics Letters* 79:145–152. [http://dx.doi.org/10.1016/S0165-1765\(02\)00299-9](http://dx.doi.org/10.1016/S0165-1765(02)00299-9).
- [22] Chigira, H., Yamamoto, T. (2006). A bias-corrected estimation for dynamic panel models in small samples. Hi-Stat Discussion Paper, No. 177, Hitotsubashi University.
- [23] Hahn, J., Hausman, J. (2002). Notes on bias in estimators for simultaneous equation models. *Economics Letters* 75:237–241. [http://dx.doi.org/10.1016/S0165-1765\(01\)00602-4](http://dx.doi.org/10.1016/S0165-1765(01)00602-4).
- [24] Hansen, G. (2001). A bias-corrected least squares estimator of dynamic panel models. *Allgemeines Statistisches Archiv* 85:127–140. <http://dx.doi.org/10.1007/s101820100054>.
- [25] Hayakawa, K. (2007). Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Letters* 95:32–38. <http://dx.doi.org/10.1016/j.econlet.2006.09.011>.
- [26] Hsiao, C. (2014). *Analysis of Panel Data*. 3rd ed. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139839327>.
- [27] Kiviet, J.F. (1995). On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68:53–78. [http://dx.doi.org/10.1016/0304-4076\(94\)01643-E](http://dx.doi.org/10.1016/0304-4076(94)01643-E).
- [28] Kiviet, J. F. (2007). On the optimal weighting matrix for the GMM system estimator in dynamic panel data models, Discussion Paper, No. 2007/08. University of Amsterdam.
- [29] Kunitomo, N. (1980). Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association* 75:693–700. <http://dx.doi.org/10.1080/01621459.1980.10477535>.
- [30] Liu, S., Neudecker, H. (1997). Kantorovich inequalities and efficiency comparisons for several classes of estimators in linear models. *Statistica Neerlandica* 51:345–355. <http://dx.doi.org/10.1111/1467-9574.00058>.
- [31] Lokshin, B. (2008). Monte Carlo comparison of alternative estimators for dynamic panel data models. *Applied Economics Letters* 15:15–18. <http://dx.doi.org/10.1080/13504850600706545>.
- [32] Maddala, G.S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica* 39:341–358. <http://dx.doi.org/10.2307/1913349>.
- [33] Morimune, K. (1983). Approximate distributions of  $k$ -class estimators when the degree of overidentification is large compared with the sample size. *Econometrica* 51:821–841. <http://dx.doi.org/10.2307/1912160>.
- [34] Nelson, C., Startz, R. (1990). Some further results on the exact small sample properties of the instrumental variables estimator. *Econometrica* 58:967–976. <http://dx.doi.org/10.2307/2938359>.
- [35] Nelson, C., Startz, R. (1990). The distribution of the instrumental variable estimator and its  $t$  ratio when the instrument is a poor one. *Journal of Business* 63:125–140. <http://dx.doi.org/10.1086/296497>.
- [36] Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica* 49:1417–1426. <http://dx.doi.org/10.2307/1911408>.
- [37] Sevestre, P., Trognon, A. (1985). A note on autoregressive error component models. *Journal of Econometrics* 28:231–245. [http://dx.doi.org/10.1016/0304-4076\(85\)90122-8](http://dx.doi.org/10.1016/0304-4076(85)90122-8).
- [38] Staiger, D., Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica* 65:557–586. <http://dx.doi.org/10.2307/2171753>.
- [39] Windmeijer, F. (2005). A Finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126:25–51. <http://dx.doi.org/10.1016/j.jeconom.2004.02.005>.
- [40] Youssef, A. H., Abonazel, M. R. (2017). Alternative GMM estimators for first-order autoregressive panel model: an improving efficiency approach. *Communications in Statistics-Simulation and Computation* 46:3112-3128. <http://dx.doi.org/10.1080/03610918.2015.1073307>.
- [41] Youssef, A., El-sheikh, A., Abonazel, M. (2014). Improving the efficiency of GMM estimators for dynamic panel models. *Far East Journal of Theoretical Statistics* 47:171–189.
- [42] Youssef, A., El-sheikh, A., Abonazel, M. (2014). New GMM estimators for dynamic panel data models. *International Journal of Innovative Research in Science, Engineering and Technology* 3:16414–16425.