Individual Effects on First Order Autoregressive Panel Data Model

Okeke, Joseph Uchenna a*, Nwabueze, Joy Chioma b, Okeke, Evelyn Nkiruka c

a,c Department of Mathematics and Statistics, Federal University Wukari, Nigeria
bDepartment of Statistics, Michael Okpara University, Umudike, Nigeria

a uche70ng@yahoo.com
b teeubueze@yahoo.co.uk
c evelyn70ng@yahoo.co

Abstract

A Monte-Carlo study is carried out on some data sets generated to investigate the stable first order autoregressive panel data model at two different specifications of the individual effect terms ($\alpha_i$) at alternating values of the autocorrelation coefficient($\rho$) and the autoregressive coefficient($\gamma$) from 0.1 through 0.5 to 0.9 at varying time periods(T) and number of individuals(N) with each data set replicated 300 times. Anderson-Hsiao (AH) Instrumental Variable method, Arellano-Bond (AB) GMM method and Blundell-Bond (BB) system GMM estimator are studied using the minimum Mean Squares Error (MSE) property and the Akaike Information Criteria(AIC) and their results placed alongside conventional Ordinary Least Squares (OLS) , Least Squares Dummy Variable (LSDV), the inefficient Two Stage Least Squares (2SLS), Three Stage Least Squares (3SLS) and the Seemingly Unrelated Regression (SUR) methods. The result shows that the BB performed better than the AH and AB in terms of bias($\gamma$) and MSE($\gamma$) at $\gamma = 0.2$ and at the two specifications of $\alpha_i$. The specification of $\alpha_i$ does not affect the ranked performance of the IV-GMM estimators. Also, the estimators recorded lower values at the uncorrelated specification of $\alpha_i$ than the correlated.

Keywords: Individual effects; Autoregressive; Panel data; Instrumental variable; estimators; Specifications.

* Corresponding author. Tel.: +23408036026806
E-mail address: uche70ng@yahoo.com.
1. Introduction

The panel data model explains the relation that may exist in a process that has time and space dimensions where heterogeneity across space (units) is an integral part and often the central point of the analysis. In economic relations, the dependence of \( Y \) on \( X \) is rarely spontaneous or instantaneous in the accounts of [1] but very often requires some time lag in the accounts of [2]. Thereby making the panel data model denoted by

\[
Y_{it} = X_{it}^{'} \beta + \alpha_i + \epsilon_{it} \tag{1}
\]

which explains the relationship between \( Y_{it} \), an \((NT \times 1)\) vector of dependent variable, \( X_{it} \), an \((NT \times (K + 1))\) matrix of independent variables, \( \alpha_i \), an \((NT \times 1)\) vector of individual effects and \( \epsilon_{it} \), an \((NT \times I)\) vector of error terms, \( i \) denotes the individuals or groups under study while \( t \) is the time period studied, to be replaced with the distributed lag model denoted by

\[
Y_{it} = \beta_o X_{it} + \beta_1 X_{it-1} + \ldots + \beta_s X_{it-s} + V_{it} \tag{2}
\]

where

\( V_{it} \) is a component error term that accounts for individual effects or spatial differences in \( x \) as well as the random error term.

Acknowledging that lagged variables are one way for taking into account the duration in the adjustment process of economic behaviour and perhaps the most efficient way for rendering them dynamic as in [3], the result is the dynamic panel data model

\[
Y_{it} = X_{it}^{'} \beta_1 + \gamma Y_{it-1} + V_{it} \tag{3}
\]

Where

\[
Y_{it-1} = \sum_{j=1}^{n} \sum_{j=1}^{m} \beta_j X_{it-j} + V_{it-j}
\]

The dynamic panel data model proposed under different lag schemes ranged from the works of Koyck, Cagan, Nerlove, Solow and Gorgenson as in [3,4].

Various estimation methods for the parameters of the dynamic panel data models includes: the conventional ones such as the: Ordinary Least Squares (OLS), Least Squares Dummy Variables(LSDV), Two Stage Least Squares (2SLS), Three Stage Least Squares (3SLS) and Seemingly Unrelated Regression (SUR) methods. The Instrumental Variable (IV) method and the Generalized Methods of Moments (GMM) such as the: Anderson-Hsiao (AH), Arellano-Bond (AB) and Blundell and Bond (BB) methods.

The OLS and LSDV estimators are applied to equation in level form and all the T cross-sections can be used in actual estimation. Both yields inconsistent estimates for finite T according to Nickel as in [5,6]. AH(1), AB,
2SLS and 3SLS are all applied to the model in first differenced form and results to a loss of at least one cross-section in the estimation. BB is applied with the first difference instruments for the equations in level and instruments in level for the first differenced equation. AH, AB and BB yields consistent estimates of the coefficients as \( N \to \infty \) and \( T \to \infty \) as in [5,4]. LSDV yield consistent estimates only as \( T \to \infty \), the estimates will still be inconsistent as \( N \to \infty \) in the accounts of [7, 8,5]. [9] estimated the growth convergence equation obtained from summer-Huston (1988) data set and concluded that the estimators not using lagged dependent variables as instruments performed better than the ones that do for dynamic panel data model. [5] studied an unbalanced panel and concluded that the least squares dummy variable corrected for bias (LSDVC) estimator and AH have smaller bias than AB, BB and LSDV and that for the coefficient of \( Y_{t-1} \) at \( \gamma = 0.8 \), AB bias for \( \gamma \) is always negative. The authors in [3,4] believed that the degree as well as the structure of autocorrelation affects the efficiency of estimators making it necessary for a study to be conducted to investigate and determine the sensitivity of the various estimators under study to the structures of autocorrelation of the error terms.

This paper studies the effects of two different specifications of the individual effect term on some properties of the estimators of the dynamic panel data model estimators as an extension to [9,5].

The rest of the paper proceeds as follows: section 2 describes the specification of the model to be estimated, the data used for the analysis and the estimators. The estimation results of our simulation study is presented in section 3 while discussion and conclusions are presented in sections 4 and 5 respectively.

2. The Design of the Sampling Experiment

This study is focused on the simple dynamic panel data model which is the stable first order autoregressive panel data model with exogenous regressors, endogenous regressors, unobserved individual effect term and random disturbance term with full specification given as:

\[
Y_{it} = X_{it}^{\prime}\beta + \gamma Y_{i,t-1} + \alpha_i + \varepsilon_{it} \quad (4)
\]

\[
X_{it} = \rho X_{i,t-1} + \varepsilon_t, \quad e \sim N(0,1), \quad \rho / < 1, \quad Y_{i0} = \eta_0 + \eta_1\alpha_i + \eta_2\varepsilon_{i0}, \quad X_{i0} = \lambda_0 + \lambda_1\varepsilon_{i0}
\]

Where \( \varepsilon_{it} = \rho \varepsilon_{i,t-1} + \varepsilon_t \), \( i = 1,2,...,N \) \quad And \quad \( \varepsilon_t \sim N(0,1) \)

The exogenous variable \( x_i \) with the specification

\[
X_{it} = \rho X_{i,t-1} + \varepsilon_t, \quad e \sim N(0,1), \quad / \rho / < 1
\]

Was generated as in [5] and [6] by specifying \( \rho \) from 0.1 to 0.9. \( X_{i0} \) was generated using the [11] procedure while \( X_{i0} \) was finally obtained from the specification model for it in (4).

Where \( n_0 = \frac{\beta}{1-\rho}, \quad n_1 = \frac{1}{1-\rho^2} \), and \( n_2 = \pm \sqrt{\frac{1}{1-\rho^2}} \)

281
The model has two error components: the individual effect term $\alpha_i$ and the random error component $\epsilon_{it}$.

The individual effect term has two (2) specifications for the experiment given by

(1) $\alpha_i \sim iid (0, \sigma^2_{\alpha})$, $\sigma_{\alpha} = \sigma_{\epsilon} (1 - \gamma)$

(2) $\alpha_i = \theta \bar{X}_{it} + e_{it}$, $e_{it} \sim N(0,1), / \theta_1 / < 1$

For $\alpha_i \sim iid (0, \sigma^2_{\alpha})$, $\sigma_{\alpha} = \sigma_{\epsilon} (1 - \gamma)$, we normalize $\sigma_{\epsilon}^2$ to unity. Then for N individuals we generate N random numbers using excel packages specifying: mean 0 and variance $\sigma^2_{\alpha}$. Finally, we standardize the result to obtain the specification.

For $\alpha_i = \theta \bar{X}_{it} + e_{it}$, $e_{it} \sim N(0,1)$, we fix $\theta$ at 0.8 as in [10]. Using the average $X_{it}$ for each individual ($\bar{X}_i$) and the standardized error term generated for each individual, we generate the individual effect term $\alpha_i$, an (NX1) vector.

To achieve the aims of this study, we vary the sizes of gamma ($\gamma$) from 0.1 through 0.5 to 0.9 inclusively. We also studied the effects of various mix of time period and cross-sections on the estimators by varying the periods of time as 10 and 20 and; individuals as 5, 20 and 30 . The experiments were replicated 300 times, as recommended by [4], at each specification of the individual effect term(s).

These properties of a good estimator as: minimum bias, minimum variance and minimum root mean square error(RMSE) are deployed as criteria for determining the performance of the estimators in this study. We also considered the Akaike Information Criterion(AIC) for the model selection. We considered the mean of parameter estimates in three hundred replications and standard errors of parameter estimates – as yardsticks for measuring parameter estimates. In other words, for $\bar{\beta}_i$, the ith estimate of the true parameter value, $\beta$, we have:

Bias ($\bar{\beta}_i$) = ($\bar{\beta}_i - \beta$)

Bias ($\bar{\beta}$) = $\frac{1}{n} \sum_{i=1}^{n} (\bar{\beta}_i - \beta)$

MSE ($\bar{\beta}$) = $\frac{1}{n} \sum_{i=1}^{n} [VAR(\bar{\beta}_i) + Bias(\bar{\beta}_i)]^2$. Where the mean values of the parameter estimates is obtained as $\bar{\beta} = \frac{1}{n} \sum_{i=1}^{n} \bar{\beta}_i$, where $n$ is the number of replications. The Mean Square Error(MSE) of the model given as

$MSE = \frac{1}{n-k} \sum_{i=1}^{n} (\bar{Y}_i - \bar{Y_i})^2$

is used to determine the combined effects of the estimator of $\beta$ and $\gamma$ in the model.

The Akaike Information Criterion(AIC) is given as

$AIC = \frac{2k}{n} RSS$ or $\ln AIC = \frac{2k}{n} + ln(\frac{RSS}{n})$
where k is the number of regressors, n is number of observations while RSS is the residual sum of squares.

3. Estimation Results for our Simulation Study

Table 1 and Table 2 presents the results.

**Table 1.** Estimators of model 1 at true parameter values of $\beta = 0.8$, $\gamma = 0.2$ and at $\rho = 0.8$, $N=20$ and $T=10$ and $\alpha_i$ specified as $\alpha_i \sim \text{iid} (0, \sigma^2_\alpha)$

<table>
<thead>
<tr>
<th>Bias($\beta$)</th>
<th>Estimators</th>
<th>SE($\beta$)</th>
<th>SE($\gamma$)</th>
<th>MSe($\beta$)</th>
<th>MSE($\gamma$)</th>
<th>MSe</th>
<th>RMSe</th>
<th>AIC</th>
<th>Bias($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1872</td>
<td>OLS (nc)</td>
<td>0.051608</td>
<td>0.041216</td>
<td>0.03771</td>
<td>0.00238</td>
<td>1.7856</td>
<td>1.3363</td>
<td>0.5897</td>
<td>-0.02624</td>
</tr>
<tr>
<td>0.5279</td>
<td>LSDV (nc)</td>
<td>0.028716</td>
<td>0.02390</td>
<td>0.279504</td>
<td>0.12124</td>
<td>0.3815</td>
<td>0.6173</td>
<td>-1.0547</td>
<td>-0.34720</td>
</tr>
<tr>
<td>0.7977</td>
<td>AH(1)</td>
<td>0.064312</td>
<td>0.16317</td>
<td>0.06408</td>
<td>0.21257</td>
<td>0.6535</td>
<td>0.8084</td>
<td>-0.413</td>
<td>-0.43122</td>
</tr>
<tr>
<td>0.8766</td>
<td>AB</td>
<td>0.01724</td>
<td>0.011624</td>
<td>0.76868</td>
<td>0.2354</td>
<td>2.7124</td>
<td>1.6469</td>
<td>1.0102</td>
<td>-0.48503</td>
</tr>
<tr>
<td>0.7696</td>
<td>BB</td>
<td>0.03598</td>
<td>0.03297</td>
<td>0.59354</td>
<td>0.17157</td>
<td>2.4975</td>
<td>1.5803</td>
<td>0.9263</td>
<td>-0.41297</td>
</tr>
<tr>
<td>0.1872</td>
<td>2SLS</td>
<td>0.051610</td>
<td>0.041216</td>
<td>0.03771</td>
<td>0.00238</td>
<td>1.7856</td>
<td>1.3363</td>
<td>0.5897</td>
<td>-0.02624</td>
</tr>
<tr>
<td>0.1872</td>
<td>3SLS</td>
<td>0.051362</td>
<td>0.0410097</td>
<td>0.037685</td>
<td>0.00237</td>
<td>1.7856</td>
<td>1.3363</td>
<td>1.5890</td>
<td>-0.2624</td>
</tr>
<tr>
<td>0.1872</td>
<td>SUR</td>
<td>0.051362</td>
<td>0.0410097</td>
<td>0.037685</td>
<td>0.00237</td>
<td>1.7856</td>
<td>1.3363</td>
<td>1.5890</td>
<td>-0.2624</td>
</tr>
</tbody>
</table>

**Table 2.** Estimators of model 2 at the true parameter value $\beta = 0.8$, $\gamma = 0.2$ and at $\rho = 0.8$ for $N=20$, $T=10$ and $\alpha_i$ specified as $\alpha_i = \theta \bar{X}_{it} + e_{it}$

<table>
<thead>
<tr>
<th>Bias($\beta$)</th>
<th>Estimators</th>
<th>SE($\beta$)</th>
<th>SE($\gamma$)</th>
<th>MSe($\beta$)</th>
<th>MSE($\gamma$)</th>
<th>MSe</th>
<th>RMSe</th>
<th>AIC</th>
<th>Bias($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5878</td>
<td>OLS (nc)</td>
<td>0.02969</td>
<td>0.02473</td>
<td>0.34639</td>
<td>0.20609</td>
<td>0.3350</td>
<td>0.5788</td>
<td>0.3517</td>
<td>-0.4533</td>
</tr>
<tr>
<td>0.6301</td>
<td>LSDV (nc)</td>
<td>0.03146</td>
<td>0.02606</td>
<td>0.39802</td>
<td>0.19112</td>
<td>0.3361</td>
<td>0.5797</td>
<td>0.3208</td>
<td>-0.4364</td>
</tr>
<tr>
<td>0.8415</td>
<td>AH(1)</td>
<td>0.03756</td>
<td>0.14130</td>
<td>0.70953</td>
<td>0.217102</td>
<td>0.6028</td>
<td>0.7764</td>
<td>0.6202</td>
<td>-0.4440</td>
</tr>
<tr>
<td>0.8660</td>
<td>AB</td>
<td>0.01545</td>
<td>0.00640</td>
<td>0.075020</td>
<td>0.22251</td>
<td>0.2716</td>
<td>0.5211</td>
<td>0.2932</td>
<td>-0.4744</td>
</tr>
<tr>
<td>0.7587</td>
<td>BB</td>
<td>0.01494</td>
<td>0.01040</td>
<td>0.57585</td>
<td>0.20039</td>
<td>0.2900</td>
<td>0.5385</td>
<td>0.3124</td>
<td>-0.4354</td>
</tr>
<tr>
<td>0.5878</td>
<td>2SLS</td>
<td>0.02969</td>
<td>0.02473</td>
<td>0.34639</td>
<td>0.20609</td>
<td>0.3350</td>
<td>0.5788</td>
<td>0.3517</td>
<td>-0.4533</td>
</tr>
<tr>
<td>0.5878</td>
<td>3SLS</td>
<td>0.02954</td>
<td>0.02460</td>
<td>0.34638</td>
<td>0.20609</td>
<td>0.3350</td>
<td>0.5760</td>
<td>0.3517</td>
<td>-0.4533</td>
</tr>
<tr>
<td>0.5878</td>
<td>SUR</td>
<td>0.02954</td>
<td>0.02460</td>
<td>0.34638</td>
<td>0.20609</td>
<td>0.3350</td>
<td>0.5760</td>
<td>0.3517</td>
<td>-0.4533</td>
</tr>
</tbody>
</table>

4. Discussion

The study of the unobserved individual effects distinguished Econometrics as a field of study from Pure Mathematics and Economics. However not much attention has been given to the effects of its specifications in estimation of panel data model rather attention has always been focused on the specification of the autoregressive term of the dynamic panel data model. From tables 1 and 2 respectively, model 1 represents the uncorrelated effect terms while model 2 represents the correlated effect terms (correlated with the exogenous explanatory variable). The bias of $\gamma$ is positive in model 2 but negative in model 1. Model 1 recorded lower
values relative to model 2. In both specifications, the performances of the consistent estimators produced the same performance by rank.

5. Conclusions

1. At $\gamma = 0.2$, for $\alpha_t$ specified as $\alpha_t = \theta \tilde{X}_{it} + e_{it}$ the BB has the least bias for $\gamma$ and MSE($\gamma$) while the AB has the minimum AIC among all the estimators studied.
2. At $\gamma = 0.2$, for $\alpha_t$ specified as $\alpha_t \sim iid(0, \sigma_{\alpha}^2)$ the BB has the least bias for $\gamma$ and MSE($\gamma$) while the AH has the minimum AIC among the IV-GMM estimators studied.
3. The BB performed better than the AH and AB in terms of bias($\gamma$) and MSE($\gamma$) at $\gamma = 0.2$ at the two specifications.
4. The specification of $\alpha_t$ does not affect the ranked performance of the estimators.

The estimators recorded lower values at the uncorrelated specification than the correlated.

Acknowledgements

I sincerely appreciate the contributions of all the authors whose work formed the bedrock of this work as contained in the references since their materials provided thought provoking insight into this research.

References