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*

[b] *Department 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 (α_i) at alternating values of the autocorrelation coefficient(ρ) and the autoregressive coefficient(γ) 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(γ) and MSE(γ) at $\gamma = 0.2$ and at the two specifications of α_i . The specification of α_i does not affect the ranked performance of the IV-GMM estimators. Also, the estimators recorded lower values at the uncorrelated specification of α_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

$$Y_{it} = X_{it}^1 \beta + \alpha_i + \varepsilon_{it} \tag{1}$$

which explains the relationship between Y_{it}, an (NT X I) vector of dependent variable, X_{it}, an (NT X (K + 1)) matrix of independent variables, α_i, an (NT X 1) vector of individual effects and ε_{it} , an (NT x 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

$$Y_{it} = \beta_o X_{it} + \beta_1 X_{it-1} + \ldots + \beta_s X_{i\ t-s} + V_{it} \tag{2}$$

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

$$Y_{it} = X_{i\ t}^{1} \beta_1 + \gamma Y_{i\ t-1} + V_{i\ t} \tag{3}$$

Where

$$Y_{it-1} = \sum_{i=1}^{n} \sum_{j=1}^{s} \beta_j X_{it-j} + V_{it-j}$$

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$\rightarrow \infty$ and T$\rightarrow \infty$ as in [5,4]. LSDV yield consistent estimates only as T$\rightarrow\infty$, the estimates will still be inconsistent as N$\rightarrow \infty$ 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 (γ)=0.8, AB bias for γ 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:

$$Y_{it} = X^1_{it}\beta + \gamma Y_{it-1} + \alpha_i + \varepsilon_{it} \qquad\qquad (4)$$

$X_{it} = \rho X_{it-1} + e_t,$ e~N(O,1), $/\rho/$<1, Y_{i0}= $\eta_0 + \eta_1\alpha_i + \eta_2\varepsilon_{i0}$, $X_{i0} = \lambda_0 + \lambda_1\varepsilon_{i0}$

Where $\varepsilon_{it} = \rho\varepsilon_{it-1} + e_t$ $_{i=1,2,...,N}$ And $e_t \sim N(0,1)$

The exogenous variable x_{it} with the specification

$X_{it} = \rho X_{it-1} + e_t,$ e~N(O,1), $/\rho/$<1

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

Where $n_o = \frac{\beta}{1-r}$, $n_1 = \frac{1}{1-r}$, and $n_2 = \pm\sqrt{\dfrac{1}{1-r^2}}$

The model has two error components: the individual effect term α_i and the random error component ε_{it}.

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

(1) $\alpha_i \sim \text{iid} (o, \sigma_\alpha^2), \qquad \sigma_\alpha = \sigma_e(1 - \gamma)$

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

For $\alpha_i \sim \text{iid} (o, \sigma_\alpha^2)$, $\sigma_\alpha = \sigma_e(1 - \gamma)$, we normalize $\sigma_e{}^2$ to unity. Then for N individuals we generate N random numbers using excel packages specifying: mean 0 and variance σ_α^2. 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 θ 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 α_i, an (NX1) vector.

To achieve the aims of this study, we vary the sizes of gamma (γ) 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 $\widehat{\beta}_i$, the ith estimate of the true parameter value, β, we have: Bias $(\widehat{\beta}_i) = (\widehat{\beta}_i - \beta)$;

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

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

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

is used to determine the combined effects of the estimator of β and γ in the model.

The Akaike Information Criterion(AIC) is given as

$AIC = e^{\frac{2k}{n}}\frac{RSS}{n}$ 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 α_i specified as $\alpha_i \sim$ iid (o, σ_α^2)

Bias(β)	Estimate	SE(β)	SE(γ)	MSe(β)	MSE(γ)	MSE	RMSE	AIC	Bias(γ)
0.1872	OLS (nc)	0.051608	0.041216	0.03771	0.00238	1.7856	1.3363	0.5897	-0.02624
0.5279	LSDV (nc)	0.028716	0.02390	0.279504	0.12124	0.3815	0.6173	-1.0547	-0.34720
0.7977	AH(1)	0.064312	0.16317	0.06408	0.21257	0.6535	0.8084	-0.413	-0.43122
0.8766	AB	0.01724	0.011624	0.76868	0.2354	2.7124	1.6469	1.0102	-0.48503
0.7696	BB	0.03598	0.03297	0.59354	0.17157	2.4975	1.5803	0.9263	-0.41297
0.1872	2SLS	0.051610	0.041216	0.03771	0.00238	1.7856	1.3363	0.5897	-0.02624
0.1872	3SLS	0.051362	0.0410097	0.037685	0.00237	1.7856	1.3363	1.5890	-0.2624
0.1872	SUR	0.051362	0.0410097	0.037685	0.00237	1.7856	1.3363	1.5890	-0.2624

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

Bias(β)	Estimators	SE(β)	SE(γ)	MSe(β)	MSE(γ)	MSE	RMSE	AIC	Bias(γ)
0.5878	OLS (nc)	0.02969	0.02473	0.34639	0.20609	0.3350	0.5788	0.3517	-0.4533
0.6301	LSDV (nc)	0.03146	0.02606	0.39802	0.19112	0.3361	0.5797	0.3208	-0.4364
0.8415	AH (1)	0.03756	0.14130	0.70953	0.217102	0.6028	0.7764	0.6202	-0.4440
0.8660	AB	0.01545	0.00640	0.075020	0.2251	0.2716	0.5211	0.2932	-0.4744
0.7587	BB	0.01494	0.01040	0.57585	0.20039	0.2900	0.5385	0.3124	-0.4354
0.5878	2SLS	0.02969	0.02473	0.34639	0.20609	0.3350	0.5788	0.3517	-0.4533
0.5878	3SLS	0.02954	0.02460	0.34638	0.20609	0.3350	0.5760	0.3517	-0.4533
0.5878	SUR	0.02954	0.02460	0.34638	0.20609	0.3350	0.5760	0.3517	-0.4533

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 γ 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 $\gamma = 0.2, for\ \alpha_i$ specified as $\alpha_i = \theta \bar{X}_{it} + e_{it}$ the BB has the least bias for γ and MSE(γ) while the AB has the minimum AIC among all the estimators studied.
2.   At $\gamma = 0.2, for\ \alpha_i$ specified as $\alpha_i \sim iid(0, \sigma_\alpha^2)$ the BB has the least bias for γ and MSE(γ) 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(γ) and MSE(γ) at $\gamma = 0.2$ at the two specifications .
4. The specification of α_i does not affect the ranked performance of the estimators

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

Acknowledgements

References

[1] K.M. Carlson.. "The Lag from Money to Price Review." *Federal Reserve Bank of St. Louis,* Vol. 1, P. 4. 1980.

[2] P.R. Krugman and M. Obstfeld. *International Economics:Theory and Practice*. 3rd edition. New York: Harper Collins. 1994. Pp 100-160.

[3] A. Koutsoyiannis. *Theory of Econometrics*. 2nd edition. New York: Palmgrove Publishers Ltd. 2001. Pp. 294- 327.

[4] W.H. Greene. *Econometric Analysis*. 4th edition. Engle-wood cliffs: Prentice Hall. 2003. Pp 230-310.

[5] G.S.F Bruno. (2004). Estimation, Inference and Monte Carlo Analysis in Dynamic Panel Data Models with a Small Number of Individuals". Instituto di Economia Politica, Universitá Bocconi, Milan.

[6] J.F. Kiviet. "Judging Contending Estimation by simulation: Tournaments in Dynamic Panel Data Models." Tinbergen Institute Discussion paper, University of Amsterdam, Netherlands. 2005.

[7] D.N. Gujarati. *Basic of Econometrics.* 4th edition. New York: Tata McGraw-Hill. 2003. Pp. 636-719.

[8] J. Johnston and J. D'inardo. *Econometric Methods*. 4th Edition . New York: McGraw-Hill. 1997.Pp. 210-270

[9] N. Islam. "Small Sample Performance of Dynamic Panel Data Estimators: A Monte Carlo Study on the Basis of Growth Data." Seminar paper at Harvard University and Emory University. 1998.

[10] J.J. Spitzer. "Small Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Auto-correlated Errors." *Journal of American Statistical Associaton.* Vol.74, Pp. 126-138. 1979.

[11] A.I. McLeod and K.W. Hipel. "Simulation Procedure for Box-Jenkins Models." *Water Resources Research*, Vol. 14, Pp. 969-975. 1978.