Volume : 4 | Issue : 10 | October 2015 • ISSN No 2277 - 8179

Research Paper

Statistics

KEYWORDS : finite sample; consistent estimators; instrumental variables; GMM ;autocorrelation; autoregressive and robust.

Robust Regression Models of Dynamic Panel Data

Okeke,Joseph Uchenna — Faculty of Pure and Applied Sciences, Federal University Wukari, Katsina-Ala RoadWukari, Nigeria

Nwabueze,Joy Chioma — Department of Statistics, Michael Okpara University, Umudike, Nigeria.

Okeke, Evelyn Nkiruka — Faculty of Pure and Applied Sciences, Federal University Wukari, Katsina-Ala RoadWukari, Nigeria

ABSTRACT

This paper reviewed the finite sample performance of the consistent estimators of the dynamic panel data models with fixed effects using the Minimum Root Mean Squares Error property of the estimates of the autoregressive term, γ (i.e. RMSE(γ)) and the Akaike Information Criteria(AIC) but specifically draws comparison among the robust estimators with specific interest in their behaviours, in the light of varying degrees of the autocorrelation coefficient(ρ) and the autoregressive coefficient(γ) , The Anderson-Hsiao (AH) Instrumental Variable method; Arellano-Bond (AB) GMM method and Blundel-Bond (BB) system GMM estimator are robust in relation to the conventional Least Squares (LS) estimators . The experiment is of Monte-Carlo type performed on some data sets simulated for time periods(T) and number of individuals(N) with each data set replicated three hundred and fifty (350) times. It is suspected that the degree as well as the structure of autocorrelation may affect 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 degree of autocorrelation. The studies showed that: the magnitude of γ and the degree of autocorrelation(ρ) do affect the quality of results obtained from the estimators therefore lending credence to results from estimation of dynamic panel data models which is still evolving. The BB performed better than the other IV-GMM estimation methods considered in terms of the RMSE(γ) for any values of ρ at γ≥0.8 ; for any values of γ at ρ≥0.8 and also in terms of RMSE and AIC for γ≥ 0.6.

INTRODUCTION

Let us consider the general panel data model as in [1] :

$$Y_{it} = X_{it}^{1}\beta + \alpha_i + \varepsilon_{it} \qquad (1)$$

Which explains the relationship between Y_{it}, an (NT X I) vector of dependent variables, X_{it}, an (NT X (K + 1)) matrix of independent variables, $_i$, an (NT X 1) vector of individual effects and ε_{it} , an (NT x I) vector of error terms. More clearly, i denotes the individuals or groups under study while t is the time period studied. The individual effect term $_i$ and ε_{it} make up the error terms of the model such that

$$V_{it} = {}_i + \varepsilon_{it}$$

If $_i$ or ε_{it} or both are autoregressive then the error term is autoregressive in the account of [2]. (1) above can be rewritten as

$$Y_{it} = X_{it}^{1}\beta + V_{it}. \qquad (2)$$

In some empirical econometrics, the static model in (2) has been replaced with the distributed lag model denoted by

$$Y_{it} = \beta_o X_{it} + \beta_1 X_{it-1} + \dots + \beta_s X_{i\ t-s} + V_{it} \qquad (3)$$

To determine the minimum number of lags required, different lag schemes were considered as in [3]. The introduction of a lagged dependent variable to substitute for the lags will yield the dynamic panel data model:

$$Y_{it} = X_{it}^{1}\beta_1 + \gamma Y_{i\,t-1} + V_{it} \qquad (4)$$

Where

The dynamic panel data model proposed under different lag schemes are as in [3] and [4]. Basically, the violation of the assumptions of absence of autocorrelation and lack of multicollinearity when applying OLS or Generalized Least Squares in estimating model have some consequences such as parameter estimate having statistical bias, variance of the parameter estimate β being large when ε's are positively auto-correlated and the values of X being also positively auto-correlated, underestimation of the standard error of the random error term, OLS estimates are imprecise as shown by t- ratio obtained as Which is overestimated as in the accounts of [2,3 and 4].

[5] compared AH, AB and OLS. and concluded that estimators relying on further lagged values of the dependent variables as instrument performed worse than those that do not relying on them.

[6] conducted a Monte-Carlo analysis and concluded that the least squares dummy variable corrected for bias (LSDVC) and AH have smaller bias than AB, BB and LSDV. [6] conducted this investigation for an unbalanced panel and finally stated that the AB bias for is always negative.

Most estimators are special cases of GMM Estimators. GMM is a large sample Estimator. Two reasons for the popularity of GMM Estimator are: (I) It provides useful framework for testing, comparing and Evaluating many Estimators; (II) It provides a simple alternative to other Estimators, especially when it is difficult to write down the maximum likelihood function.

This paper proceeds as follows: section 2 reviews the theoretical framework of the estimation methods for our simulation study and 3 specifies the model while sections 4 and 5 presents the results with brief discussion and conclusions respectively.

Materials and Methods

OLS

The pooled OLS, when applied, estimates the dynamic panel data model as

$$Y_{it} = X_{it}^{1}\beta + \gamma\, Y_{i-1} + V_{it} \quad \text{where} \quad V_{it} = \alpha_i + \varepsilon_{it} \qquad (5)$$

This estimation is carried out on the basis of equation in levels.

Let $y = [y_{11}, y_{21}, \dots, y_{N1}, \dots, y_{1T}\, y_{2T}, \dots, y_{NT}]^1$

$y_{-1} = [y_{10}, y_{20}, \dots, y_{N0}, \dots, y_{1T-1}\, y_{2T-1}, \dots, y_{N\,T-1}]^1$

$X = [X_{11}, X_{21}, \dots, y_{N1}, \dots, X_{1T}, X_{2T}, \dots, X_{NT}]^1$

Volume : 4 | Issue : 10 | October 2015 • ISSN No 2277 - 8179

Also, let us define $W = [X\ y_{-1}]$ an (N x k) matrix of endogenous and strictly exogenous explanatory variables. The OLS estimator of the parameter vector

$[\beta\ \gamma]^1 = \delta$ is given by [4] as $\delta = (w^1w)^{-1}\ w'y$ (6)

Under the assumption of homoscedasticity the standard error of the estimate is obtained from var $(\delta) = S^2\ (w^1w)^{-1}$ with $S^2 = \varepsilon I\ \varepsilon\ /\ (NT - 2)$,Where $\varepsilon = (y - w\ \delta)$

Instrumental Variable (IV) Estimator

The instrumental variable (IV) is a proxy for Y_{t-1} that is highly correlated with y_{t-1} but is uncorrelated with the random error term ε_{it}. [7]

The Procedure involves

(i) regressing Y_t on lagged values of X's only to obtain

$\hat{Y}_t = \hat{a}_0 + \hat{a}_1\ X_{t-1} + \hat{a}_2\ X_{t-2} + \dots\ \hat{a}_f\ X_{t-f}$

The number of lags depends on the improvement in fit as additional values of X's are introduced.

(ii) Lagging $\hat{Y}_t$ by one period to obtain $\hat{Y}_{t-1}$ which is used to replaceY_{t-1} in the original model. Then apply OLS to the model:

$$Y_{it} = X'_{it}\beta + \gamma\ \hat{Y}_{it-1} + \alpha_i + \varepsilon_{it}. \quad (7)$$

Anderson and Hsiao (instrumental variable), AH Estimator

They started by differencing the model to eliminate individual effect term α_I to yield:

$(Y_{it}-Y_{it-1}) = (X_{it} - X_{it-1})'\ \beta + r\ (Y_{it-1} - Y_{it-2}) + (\varepsilon_{1t} - \varepsilon_{it-1})$ so that

$(Y_{it}-Y_{it-1}) = \tilde{Y}_{it,}\ (Y_{it-1} - Y_{it-2}) = \tilde{Y}_{it-1}$ and $(X_t-X_{it-1}) = \tilde{X}_{it}$

Then, letting $(\tilde{Y}_{it})_1 = [\tilde{Y}_{12}, \tilde{Y}_{22},\dots, \tilde{Y}_{N2}, \tilde{Y}_{13}, \tilde{Y}_{23},\dots, \tilde{Y}_{N3},\dots, \tilde{Y}_{1T}, \tilde{Y}_{2T},\dots, \tilde{Y}_{NT}]^1$

$(\tilde{Y}_{it-1})_1 = [\tilde{Y}_{11}, \tilde{Y}_{21}, \dots, \tilde{Y}_{N1}, \tilde{Y}_{12}, \tilde{Y}_{22}, \dots, \tilde{Y}_{N2},\dots, \tilde{Y}_{1T-1}, \tilde{Y}_{2T-1},\dots, \tilde{Y}_{NT-1}]^1$

$(\tilde{\tilde{X}}_{it})_1 = [\tilde{\tilde{X}}_{12}, \tilde{\tilde{X}}_{22},\dots, \tilde{\tilde{X}}_{N2}, \tilde{\tilde{X}}_{13}, \tilde{\tilde{X}}_{23},\dots, \tilde{\tilde{X}}_{N3},\dots, \tilde{\tilde{X}}_{1T}, \tilde{\tilde{X}}_{2T},\dots, \tilde{\tilde{X}}_{NT}]^1$

$(\tilde{Y}_{1,\ t-2})_1 = [\tilde{Y}_{10}, \tilde{Y}_{20},\dots, \tilde{Y}_{No}, \tilde{Y}_{11}, \tilde{Y}_{21}, \dots, \tilde{Y}_{N1}, \dots, \tilde{Y}_{1T-2}, \tilde{Y}_{2T-2},\dots, \tilde{Y}_{NT-2}]^1$

Therefore, $W1 = [\tilde{Y}_{it-1}\ \tilde{X}_{it,}]$ and $Z_1 = [\ \tilde{Y}_{1t-2}\ \tilde{X}_{it}]$

Such that

AH (1) estimator of δ is given by

$$\delta = (Z_1^1\ W_1)^{-1}\ Z^1_i(\tilde{Y}_{it})_1 \quad (8)$$

Under the assumption of homoscedasticity the asymptotic standard error of this estimate are obtained from Var $(\delta) = S_1^1\ (Z_1^1\ W_1)^{-1}\ Z_1^1\ Z_1(W_1^1\ Z_1)^{-1}$,where $S_1^2 = \varepsilon^1{}_1\ \varepsilon_1\ /\ [N(T-1) - 2]$ and $\varepsilon_1 = (\tilde{Y}_{it})_1 - W_1\delta$

The general heteroscedasticity standard error are obtained from var $(\delta) = (Z_1^1W_1)^{-1}Z^1$ diag $\varepsilon_1^1\varepsilon_1\ (W_1^1Z_1)$

One common thing about the IV-GMM estimators is that their properties holds for N large but can be severely biased and imprecise in panel data with small number of cross-sectional units in the account of [4]. [6,8] employed the measurement of the bias and root mean squares error (RMSE) to determine the performance of the IV-GMM estimators.

Arellano and Bond (AB) Estimation AGMMI

Arellano and Bond adopted and extended the ideas of Anderson –Hsiao estimators by proposing as instrument, all further lagged values of Y_{it} that qualifies as instrument in view of serial uncorrelatedness of ε_{it}. The number of instruments differs depending on the time period for which the equation is considered [1].

Arellano used the Generalized Method of Moment (GMM) framework to adopt a multi-equation approach and used all the qualified instrument for each equation.

The block diagonal matrix of instruments that would result is denoted by Z_i.

Where

$$Z_i = \left(\begin{array}{cccccccccc|c} y_{i0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots 0 & x_{i1} - x_{i0} \\ 0 & y_{i0} & y_{i1} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots 0 & x_{i2} - x_{i1} \\ 0 & 0 & 0 & y_{i0} & y_{i1} & y_{i2} & 0 & 0 & 0 & \cdots 0 & x_{i3} - x_{i2} \\ m & m & m & m & m & m & m & m & m & \cdots m & m \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{i0} & y_{i1} & y_{i2} & \cdots\ y_{iT-2} & x_{iT-1} - x_{iT-2} \end{array}\right)$$

Let $\tilde{\varepsilon}$ denote the N(T-1) x1 vector of differenced error term $(\varepsilon_{it}-\varepsilon_{it-1})$ and Z denote the N(T-1) x m stacked matrix of the instrument matrices Z_i's, where m is the number of columns in the instrument matrix Z_i.

Then the GMM estimator is given by

$$\delta = \arg \operatorname{Min}\ (\tilde{\varepsilon}^1\ Z)\ A_N\ (Z^1\tilde{\varepsilon}) \quad (9)$$

Where

A_N is an appropriate weighting matrix. If denote the stacked vector $(y_{it}-y_{it-1})$ and denote the stacked matrix $[(y_{it-1} - y_{it-2})\ (X_{it} - X_{it-1})]$, then the formular for the GMM estimator from the minimization is given by

$$\boldsymbol{\delta = (\widetilde{W}^1 Z A_N Z^1 \widetilde{W})^{-1} \widetilde{W}^1 Z A_N Z^1 \widetilde{Y}} \quad \textbf{(10)}$$

The one step GMM estimator (AGMM1) is obtained by setting

$A_N = N^{-1}\sum_{i=1}^{N}(Z_i^{'}HZ_i)^{-1}$

Where H is an arbitrary square matrix of dimension (T-2). The standard error for AGMM1 estimates under the assumption of homoscedasticity are obtained from

$\operatorname{Var}(\delta) = (W'ZA_NZ'W)^{-1}(W'ZA_N\widehat{v_{NT}}\ A_N\ Z'W)\ (W'ZA_N\ Z'W)^{-1}$

Where

$\widehat{v_{NT}}$ is estimated using residuals form AGMMI

2.5. Blundel and Bond(BB) Method

The BB method proposed the system GMM estimator which solves the included equations of the system simultaneously using first differenced instruments for the equation in levels and instruments in levels for the first-differenced equation [1]. It involves the use of statistical package specially designed to solve this complex process.

2.6. Model Specification

The simple dynamic panel data model with fixed effects(i.e. α_i , being time invariant but varies between groups) has full specification given as:

$Y_{it} = X^1_{it}\beta + \gamma Y_{it-1} + \alpha_i + \varepsilon_{it}$ (11)

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

$\varepsilon_{it} = \rho\varepsilon_{it-1} + e_t \quad {}_{i=1,2,\ldots,N}$ And $e_t \sim N(0,1)$ (12)

The exogenous variable x_{it} with the specification

$X_{it} = \rho X_{it-1} + e_t$, $e \sim N(O,1)$, $/\rho/<1$ (13)

Was generated as in [6] and [9] by specifying ρ from 0.1 to 0.9 under 21 different specifications and a life data on the GDP of thirty(30) randomly selected countries observed for eight years. X_{i0} was generated using the procedure spelt out by [10] while X_{it} was finally obtained from the specification model for it in (2).

For $\alpha_i \sim$ iid (o, σ_α^2), $\sigma_\alpha = \sigma_e(1-\gamma)$, we normalize $\sigma_e{}^2$ to unity. Then for N individuals we generated N random numbers using excel packages specifying: mean 0 and variance σ_α^2. Finally we standardize the result to obtain the specifications. The start up values Y_{i0} are independent of all e_{it} for t>0 in line with the procedure spelt out by [10] : Where $n_o = \frac{\beta}{1-r}$, $n_1 = \frac{1}{1-r}$, and $n_2 = \pm\sqrt{\frac{1}{1-r^2}}$

Likewise is X_{i0} . Stata 10.0, Excel 2003/2007 and Minitab statistical packages were used in this paper.

We considered the mean of parameter estimates in three hundred and fifty replications and the standard errors of parameter estimates as the yardsticks for parameter estimates. In other words, for $\widehat{\beta_\iota}$, the ith estimator of the true parameter value β, we have: Bias($\widehat{\beta_\iota}$) = ($\widehat{\beta_\iota} - \beta$), Bias($\hat{\beta}$) = $\frac{1}{n}\sum_{i=1}^{n}(\widehat{\beta_\iota} - \beta)$, MSE($\hat{\beta}$) = $\frac{1}{n}\sum_{i=1}^{n}[VAR(\hat{\beta}_i)$ + Bias$(\widehat{\beta_\iota})^2$].Then taking the root we got RMSE($\hat{\beta}$). Specifically we used RMSE(γ) Where the mean values of the parameter estimates is obtained as $\bar{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\beta_\iota}$, n is the number of replications. The Mean Square Error(MSE) of the model given as MSE = $\frac{1}{n-k}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$ is computed. 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 number of regressors, n is number of observations while RSS is the residual sum of squares [1].

To determine the effect of the autoregressive coefficient (γ) of the lagged dependent variable Y_{it-1}, on the estimation, we, first, fixed the value of ρ for generating the values of the exogenous variable and estimated the model at alternating values of γ from 0.1 to 0.9.

Volume : 4 | Issue : 10 | October 2015 • ISSN No 2277 - 8179

3. Results

In this section the average values of the computed RMSE(γ) for all the models specifications and the GDP life data are presented in table 1 .

Table 1: Average performance of AB, BB, AH and Least Squares Estimators at specifications of γ and ρ and GDP life data

MODEL	OLS	LSDV	AH(1)	AB	BB	2SLS	3SLS	SUR
SPEC.	0.230165	0.170495	2.490042	0.260304	0.240405	0.230863	0.250034	0.250034
GDP	0.095184	0.534696	1.149652	0.145602	0.084083	0.095184	0.094868	0.094868

Discussion and Conclusions

The Root Mean Square Error of the autoregressive coefficient, γ, RMSE(γ), at γ≥0.8 for any values of ρ and at ρ ≥0.8 for any values of γ, yields a minimum average RMSE(γ) =0.240405 for BB, 0.260304 and 2.490042 for AH(1) and AB respectively. Analysis of variance test showed that for alternating values of γ and ρ, there exist significant differences in the values of the RMSE(y) of the estimators considered. The AB has minimum RMSE(γ)=0.008515 at γ=0.3 and ρ=0.1 while BB has minimum RMSE(γ)=0.03265 at γ=0.9 and ρ=0.1. The BB has a minimum RMSE(γ)=0.084083 in the life data.

From the analysis conducted the following conclusions were reached:

(1) The magnitude of γ alone can affect the performance of the Estimators.

(2) The magnitude of ρ alone can affect the performance of the Estimators.

(3) The BB performs better than the AB and AH in terms of RMSE(γ) as in [6] at γ ≥ 0.8 and in terms of RMSE and AIC for γ ≥ 0.6.

REFERENCE

[1] Okeke, J.U. et al.2015. Individual Effects on First Order Autoregressive Panel Data Model. International Journal of Basic and Applied Research. 19(). 235-241. [2] Greene, W.H.2003. Econometric Analysis. 4th Eds. Engle-wood cliffs. Prentice Hall. . [3] Koutsoyiannis,A.2001. Theory of Econometrics. 2nd Eds. New York. Palmgrove Publishers Ltd. [4] Johnston,J & D'inardo,J.1997.Econometric Methods. 4th Eds. New York. McGraw-Hill. [5] Islam, N. 1998. 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. [6] Bruno, G.S.F.2004. Estimation, Inference and Monte Carlo Analysis in Dynamic Panel Data Models with a Small Number of Individuals, Instituto di Economia Politica, Milan: Universitá Bocconi,. [7] Gujarati, D.N. 2003. Basic of Econometrics. 4th Eds. New York .Tata McGraw-Hill, [8] Kiviet, J.F. 1999. On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models. Journal of Econometrics. 68. 53-78. [9] Kiviet, J.F.1985. Judging Contending Estimation by simulation: Tournaments in Dynamic Panel Data Models. Tinbergen Institute Discussion paper. University of Amsterdam. [10] McLeod, A.I. & Hipel, K.W.1978. Simulation Procedure for Box-Jenkins Models. Water Resources Research, 14. 969-975.