Original Article

http://napas.org.ng

Investigation of Finite Sample Properties and Efficiency of Some Estimators for Panel Data Model with Normal and Non-Normal Error Structure

Torruam, Japheth Terande
Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria
torruam@gmail.com

Waheed Babatunde Yahya
Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria
wbyahya@unilorin.edu.ng

OPEN ACCESS
**Corresponding Authour:*
Torruam, Japheth Terande
Email:torruam@gmail.com

Specialty Section: *This article was submitted to Sciences section of NAPAS.*

Submitted date: 3rd August, 2023
Accepted date: 24th November, 2023
Published date:

Citation: Torruam,Japheth Terande and Waheed Babatunde Yahya (2023) Investigation of Finite Sample Properties and Efficiency of Some Estimators for Panel Data Model with Normal and Non-Normal Error Structure - ***Nigerian Annals of Pure and Applied*** **Sciences 6 (1) 218 - 232**

DOI:10.5281/zenodo.7338397

Publisher:

Email:

AccessCode

Abstract

The study investigates efficiency of some estimators for panel data model with non-normal error structure and varying sample sizes. It considers one-stage and two-stage error component models with three exogenous and one endogenous variable. The efficiency of four estimators of panel data model based on one-step and two-step error component models across varying finite samples were investigated under normal and non-normal error structures. The data set used for the panel linear model (PLM) and the general feasible generalized least squares (GFGLS) model for investigating efficiency of the four estimators in this study were simulated using R software. Three predictors were simulated from normal distributions at the various samples sizes and variances. The error structures were simulated from Gaussian distribution with mean 0 and variance 1 and Exponential distribution with lambda 1 in the plm library of the R software. The four estimators were utilized to estimate the fixed parameters that form the models and their efficiencies were assessed based on absolute bias, coefficient of multiple determination and root mean square error (RMSE) of parameter estimates. The results of the study indicated that the Within Ordinary Least Squares (WOLS) estimator is the most stable and most efficient estimator of panel data model parameters than the Pooling, Between (BTW) and the First Difference (FD) estimators with both one-stage and two-stage normal and non-normal error structures. It is evident from this study that the four estimators have increasing R^2_{adj} *and the FD estimator is the next most stable while both pooling and BTW are worse but pooling is more stable under varying samples sizes (dimension).*

Keywords: *Finite Sample Properties, Efficiency of Some Estimators, Panel Data Model, Non-Normal Error Structure*

Introduction

Panel Data are data in which we observe repeated cross-sections of the same individuals. They involve observations obtained from the same set of entities at several periods of time and in the same units. These units could be individuals, households, firms, regions or countries. It has the combination of both time-series and cross-sectional data characteristics (Garba et al., 2013). Generally, a panel data regression is different from a regular time-series or cross-sectional regression because of the double subscript on its variables, i.e.

where is the households, individuals, firms, countries, etc. and t is the time. The subscript, therefore, is the cross-sectional dimension whereas t is the time-series dimension. is a scalar, is and is the i^{th} observation of the k explanatory variables. Most of the panel data applications utilize a one-way error component model for the disturbances, withwhere is the unobservable individual-specific effect and is the remainder disturbance.

The increasing availability of data observed on cross-sections of units (i.e. households, firms, countries etc.) over time has given rise to a number of estimation approaches exploiting this double dimensionality to cope with some of the typical problems associated with economic data. A lot of research efforts have been invested by econometricians to investigate model specification for these estimation approaches, testing and tackling a number of issues arising from the particular statistical problems associated with such data but research efforts to study stability of parameter estimates from these estimation methods, especially with non-normal error structures cannot be exhausted. Results from several studies (See Maddala, 2008; Creel, 2011; Wooldridge, 2012) have shown that the use of classical ordinary least squares (OLS) estimator for modelling panel data is grossly inefficient due to violations of some basic assumptions. The critical assumptions of the classical linear regression model (CLRM) are that the error terms in the model are normally distributed, with constant variance, and there is no serial correlation. These assumptions are often rarely satisfied by real life panel data.

A number of research work on the methodologies and applications of panel data modelling have appeared in the literature (Kapetanios et al, 2023; Kapetanios et al, 2011; Chudik & Pesaran, 2015; Westerlund & Urbain 2015; Juodis 2022; Chudik & Pesaran, 2015; and Creel, 2011; Olofin et al., 2010). Similarly to this current study, most of the studies that discussed panel data modelling considered the violation of each of the classical assumptions separately.

In this study therefore, the efficiency (in terms of stability of parameter estimates) of estimators for four panel (Within (OLS), Pooling (OLS), Between (BTW), and First Difference (FD)) were investigated based on one-step and two-step error component models across varying finite samples under normal and non-normal error structures. The motivation behind emergence of new panel data modelling techniques is the idea that efficiency of existing panel model estimators might be affected by violation of normality assumption of the error structure, among others as well the various forms of panel data emanating in real life cases.

Research Methodology

This study considers one-stage and two-stage error component models with three exogenous and one endogenous variable. This is in line with the work of many authors including Garba et al. (2013) but differs in that it considers heteroskedastic and autocorrelated disturbances but non-normality of error structures across varying dimensions of panel data.

The general panel data model is given as follows:

(1)

where is considered to be the response for unit at time , is the individual specific intercept, vector contains regressors for unit at time , vector contains regression coefficients to be estimated and is the error component for unit at time , and t = 1, 2, . . .,5.

Specifically, we considered the panel data model that has two exogenous and one endogenous variable as shown below;

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + u_{it} \quad (2)$$

where $\alpha_i = \alpha + \varepsilon_i$. The individual specific intercept (α_i) captures the effects of those variables that are peculiar to the i^{th} individual and that are time invariant.
The model therefore becomes:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_i + u_{it} \quad (3)$$

where ε_i is the individual specific error component and u_{it} is the combined time series and cross section error component with variances σ_ε^2 and σ_u^2 respectively.

Suppose we let $w_{it} = \varepsilon_i + u_{it}$, then, model (3) becomes:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + w_{it} \quad (4)$$

General Feasible Generalized Least Squares (GFGLS)

General feasible generalized least squares (GFGLS) estimators are based on a two-step estimation process: First an OLS model is estimated, then its residuals $\hat{u}_{it}$ are used to estimate an error covariance matrix more general than the random effects one for use in a feasible-GLS analysis. Formally, the estimated error covariance matrix is $\hat{V} = I_n \otimes \hat{\Omega}$ with $\hat{\Omega} = \sum_{i=1}^{n} \frac{\hat{u}_{it}\hat{u}_{it}^T}{n}$ where is the pooled OLS residuals (Wooldridge, 2012).

This framework allows the error covariance structure inside every group of observations to be fully unrestricted and is therefore robust against any type of intragroup heteroskedasticity and serial correlation. This structure, by converse, is assumed identical across groups and thus general (FGLS) is inefficient under group-wise heteroskedasticity (Wooldridge, 2012).

Moreover, the number of variance parameters to be estimated with $N = n \times T$ data points is $T(T+1)/2$, which makes these estimators particularly suited for situations where $n >> T$. For example, considering labour or household income surveys, being problematic for "long" panels, where $\hat{V}$ tends to become singular and standard errors also become biased downwards. In a pooled time series context (effect="time"), symmetrically, this estimator is able to account for arbitrary cross-sectional correlation, provided that the latter is time-invariant (Greene, 2003). In this case serial correlation has to be assumed away and the estimator is consistent with respect to the time dimension, keeping n fixed.

The Four Estimators of Panel Data Model

i. Within Sample (OLS) Estimator: The estimator uses information that is not taken into account by the between estimator and is called within estimator as it uses only the variation within each cross-section unit. This is also known as the fixed effects or least squares dummy variables model, usually estimated by OLS on transformed data which gives consistent estimates for β. The data set is pre-multiplied by a matrix M_D, where $M_D = I_{nT} - D(D'D)^{-1}D$ and OLS is computed on the transformed data. The within estimator is $\hat{\beta}_w = [(M_D X)'(M_D X)]^{-1}(M_D X)'(M_D Y)$, this is further simplified to;

$$\hat{\beta}_w = (X' M_D X)^{-1} X' M_D y \quad (5)$$

i. Pooled Sample Estimator: This Estimator stacks the data over i and t into one long regression with nT observations, and estimates of the parameters are obtained by OLS using the model (Greene, 2008).

$$y = X'\beta + w \quad (6)$$

where y is an $nT \times 1$ column vector of response variables, X is an $nT \times k$ matrix of

regressors, β is a $(k+1) \times 1$ column vector of regression coefficients, w is an column vector of the combined error terms. The Pooled estimator is therefore given as follows

$$\hat{\beta}_{pooled} = (X'X)^{-1}X'y \quad (7)$$

iii. Between Sample Estimator (BTW): This regresses the group means of Y on the group means of $X's$ in a regression of n observations. It uses cross-sectional variation by averaging the observations over period t (Creel, 2011; Wooldridge, 2012). Explicitly, it converts all the observations into individual-specific averages and performs OLS on the transformed data.
Averaging over all t gives the following:

$$\bar{Y}_{i.} = \alpha + \beta_1\bar{X}_{1i.} + \beta_2\bar{X}_{2i.} + \beta_3\bar{X}_{3i.} + \bar{w}_{i.} \quad (8)$$

Where $\bar{Y}_{i.} = T^{-1}\sum_t Y_{it}$, $\bar{X}_{ji.} = T^{-1}\sum_t X_{jit}$ and $\bar{w}_{i.}$ $= T^{-1}\sum_t w_{it}$ $for\ i = 1,2,3,\cdots,n\ and$ $j = 1,2,3$

iv. First Difference Estimator (FD): This is the ordinary least squares estimation of the difference between the original model and its one-period-lagged model (Arellano, 2003; Baltagi, 2005). The FD model is given as follows:

$$\Delta Y_{it} = \beta_1\Delta X_{1it} + \beta_2\Delta X_{2it} + \beta_3\Delta X_{3it} + \Delta w_{it} \quad (9)$$

Where $\Delta Y_{it} = Y_{it} - Y_{i,\ t-1}; \Delta X_{1it}$ $= X_{1it} - X_{1i,\ t-1}; \Delta X_{2it}$ $= X_{2it} - X_{2i,\ t-1}; and\ \Delta w_{it} = w_{it} -$ $w_{1i,\ t-1}, for\ i = 1,2,\cdots,n\ and\ t = 2,3,\cdots,T.$

Simulation Scheme

The data set used for the panel linear model (PLM) and the general feasible generalized least squares (GFGLS) model for investigating efficiency of the four estimators in this study were simulated from the Gaussian (normal) and the exponential distributions for three time periods $(10, 25, and\ 50\ years)$ under five cross-sectional units (5, 10, 25 and 50), using R software for statistical computing and graphics. The response and three predictors were simulated from normal distributions for the four samples sizes with means 30, 40 and 50 and variances 5, 10 and 20, respectively. The predictors in the exponential models were simulated for the samples sizes with lambda values of 1/6, 1/4 and 2/5, respectively. The error structures were simulated from Gaussian distribution with mean 0 and variance 1 and Exponential distribution with lambda 1 in the plm library (Croissant and Millo, 2008) of the version 3.3.2 of the R software (R Core Team, 2015). Two panel models from each distribution were fitted with parameters fixed at

$\beta_0 = 20, \beta_1 = 3, \beta_2 = 2\ and\ \beta_3 = 6$ as:

$$Y_{it\ (Norm)} = 20 + 3X_{1it(Norm)} + 2X_{2it(Norm)} + 6X_{3it(Norm)} + u_{it(Norm)} + \varepsilon_{it(Norm)} \quad (10)$$

$$Y_{it(Exp)} = 20 + 3X_{1it(Exp)} + 2X_{2it(Exp)} + 6X_{3it(Exp)} + u_{it(Exp)} + \varepsilon_{it(Exp)} \quad (11)$$

Each of the combinations using equations 10 and 11 was iterated 1000 times and the assessments of the four estimators considered were based on the absolute bias, coefficient of multiple determination (R^2), adjusted coefficient of multiple determination $(R^2_{adj.})$ and RMSE of parameter estimates.

Table1: Scheme for Data Simulation from Gaussian Distribution

Dimension	Gaussian Distribution				
	X_1	X_2	X_3	Error (ε_{it})	$Error\ (u_{it})$
T10N5	$rnorm(50,30,5)$	$rnorm(50,4\ 010)$	$rnorm(50,50,20)$	$rnorm(50,0,1)$	$rep(rnorm(5,0,1),10)$
T10N10	$rnorm(100,30,5)$	$rnorm(100,4\ 010)$	$rnorm(100,50,20)$	$rnorm(100,0,1)$	$rep(rnorm(10,0,1),10)$
T10N25	$rnorm(250,30,5)$	$rnorm(250,4\ 010)$	$rnorm(250,50,20)$	$rnorm(250,0,1)$	$rep(rnorm(25,0,1),10)$
T10N50	$rnorm(500,30,5)$	$rnorm(500,4\ 010)$	$rnorm(500,50,20)$	$rnorm(500,0,1)$	$rep(rnorm(50,0,1),10)$
T25N5	$rnorm(125,30,5)$	$rnorm(125,4\ 010)$	$rnorm(125,50,20)$	$rnorm(125,0,1)$	$rep(rnorm(5,0,1),25)$
T25N10	$rnorm(250,30,5)$	$rnorm(250,4\ 010)$	$rnorm(250,50,20)$	$rnorm(250,0,1)$	$rep(rnorm(10,0,1),25)$
T25N25	$rnorm(6\ 2530,5)$	$rnorm(6\ 254\ 010)$	$rnorm(6\ 2550,20)$	$rnorm(6\ 250,1)$	$rep(rnorm(25,0,1),25)$
T25N50	$rnorm(1250,30,5)$	$rnorm(1250,4\ 010)$	$rnorm(1250,50,20)$	$rnorm(1250,0,1)$	$rep(rnorm(50,0,1),25)$
T50N5	$rnorm(250,30,5)$	$rnorm(250,4\ 010)$	$rnorm(250,50,20)$	$rnorm(250,0,1)$	$rep(rnorm(5,0,1),50)$
T50N10	$rnorm(500,30,5)$	$rnorm(500,4\ 010)$	$rnorm(500,50,20)$	$rnorm(500,0,1)$	$rep(rnorm(10,0,1),50)$
T50N25	$rnorm(1250,30,5)$	$rnorm(1250,4\ 010)$	$rnorm(1250,50,20)$	$rnorm(1250,0,1)$	$rep(rnorm(25,0,1),50)$
T50N50	$rnorm(2500,30,5)$	$rnorm(2500,4\ 010)$	$rnorm(2500,50,20)$	$rnorm(2500,0,1)$	$rep(rnorm(50,0,1),50)$

Note: rnorm is the function to simulate normal random sample in R software

Table2: Scheme for Data Simulation from Exponential Distribution

Dimension	Exponential Distribution				
	X_1	X_2	X_3	Error (ε_{it})	Error (u_{it})
T10N5	$rexp(50,5/30)$	$rexp(50,10/4\ 0$	$rexp(50,20/50)$	$rexp(50,1)$	$rep(rexp(5,1),10)$
T10N10	$rexp(100,5/30)$	$rexp(100,10/4\ 0$	$rexp(100,20/50)$	$rexp(100,1)$	$rep(rexp(10,1),10$
T10N25	$rexp(250,5/30)$	$rexp(250,10/4\ 0$	$rexp(250,20/50)$	$rexp(250,1)$	$rep(rexp(25,1),10$
T10N50	$rexp(500,5/30)$	$rexp(500,10/4\ 0$	$rexp(500,20/50)$	$rexp(500,1)$	$rep(rexp(50,1),10$
T25N5	$rexp(125,5/30)$	$rexp(125,10/4\ 0$	$rexp(125,20/50)$	$rexp(125,1)$	$rep(rexp(5,1),25)$
T25N10	$rexp(250,5/30)$	$rexp(250,10/4\ 0$	$rexp(250,20/50)$	$rexp(250,1)$	$rep(rexp(10,1),25$
T25N25	$rexp(6\ 255/30)$	$rexp(6\ 2510/4\ 0$	$rexp(6\ 2520/50)$	$rexp(6\ 251)$	$rep(rexp(25,1),25$
T25N50	$rexp(1250,5/30)$	$rexp(1250,10/4\ 0$	$rexp(1250,20/50)$	$rexp(1250,1)$	$rep(rexp(50,1),25$
T50N5	$rexp(250,5/30)$	$rexp(250,10/4\ 0$	$rexp(250,20/50)$	$rexp(250,1)$	$rep(rexp(5,1),50)$
T50N10	$rexp(500,5/30)$	$rexp(500,10/4\ 0$	$rexp(500,20/50)$	$rexp(500,1)$	$rep(rexp(10,1),50$
T50N25	$rexp(1250,5/30)$	$rexp(1250,10/4\ 0$	$rexp(1250,20/50)$	$rexp(1250,1)$	$rep(rexp(25,1),50$
T50N50	$rexp(2500,5/30)$	$rexp(2500,10/4\ 0$	$rexp(2500,20/50)$	$rexp(2500,1)$	$rep(rexp(50,1),50$

Note: **rexp** is the function to simulate exponential random sample in R software

Performance Measures

The absolute bias (B_{abs}) of parameters $\hat{\beta}_k$ was estimated over 1000 iterations is defined by

$$B_{abs}(\hat{\beta}_k) = \frac{1}{r}\sum_{j=1}^{1000} |\hat{\beta}_{kj} - \hat{\beta}_k| \qquad (12)$$

The adjusted coefficient of multiple determination (R^2_{adj}) over r iterations is defined by

$$R^2_{adj} = 1 - \frac{\frac{1}{r-k-1}\sum_j^{1000}(y_j - \hat{y}_j)^2}{\frac{1}{r-1}\sum_j^{1000}(y_j - \bar{y}_j)^2} \qquad (13)$$

The root mean square error (RMSE) over r iterations is defined as

$$RMSE(\hat{\beta}_k) = \sqrt{\frac{1}{r}\sum_{j=1}^{1000}\left(\hat{\beta}_k - \bar{\hat{\beta}}_k\right)^2} \qquad (14)$$

where $\hat{\beta}_k, k = 1,2,3$ indicates the k[th] parameter being estimated for j = 1, 2, 3, . . ., 1000 (number of iterations).

Results and Discussion

Table 3: Absolute Bias of Normal and Non-Normal Error PLM Estimated β_1

Dimension	Absolute Bias of Normal Error PLM Estimated β_1				Absolute Bias of Non-Normal Error PLM Estimated β_1			
	Within	Pooling	Between	First Diff.	Within	Pooling	Between	First Diff.
T10N5	0.0225	16.8115	31.8940	3.0158	2.9957	22.0263	22.6697	0.0046
T10N10	0.0159	16.8710	16.7422	3.0115	0.0151	18.9975	19.4813	2.9955
T10N25	0.0109	16.9580	16.4858	2.9958	0.0087	19.0308	19.1435	2.9985
T10N50	0.0081	16.9223	16.0995	2.9979	0.0061	18.9888	19.1043	2.9951
T25N5	0.0172	17.0931	33.5552	3.002	0.0119	18.9305	20.6377	2.9992
T25N10	0.0095	16.8847	18.2966	2.9959	0.0092	19.0269	19.5346	2.9987
T25N25	0.0066	16.9324	16.0447	3.0012	0.0054	19.0084	19.0369	3.0004
T25N50	0.0045	16.9409	16.1903	2.9994	0.0037	19.0235	19.0495	3.0006
T50N5	0.0102	16.9309	52.5558	2.9972	0.0084	19.059	26.6624	3.001
T50N10	0.0075	16.9583	23.7461	2.9992	0.0061	18.9649	19.7816	3.0003
T50N25	0.0046	17.0258	18.5438	2.9998	0.0035	19.002	18.742	3.0000
T50N50	0.0037	16.9762	16.0517	2.9999	0.0026	18.9855	18.9551	3.0002

Note: *Within is more efficient, follow by First Difference at normal and non-normal error GFGLS estimated* β_1

Table 3 shows that within estimator have the lower value at each sample size; this implies that it is more efficient, follow by First Difference at both normal and non-normal error PLM estimated β_1. The performance of within and the other estimators are as also shown graphically in figures 1 and 2 below for normal and non-normal error PLM estimated.

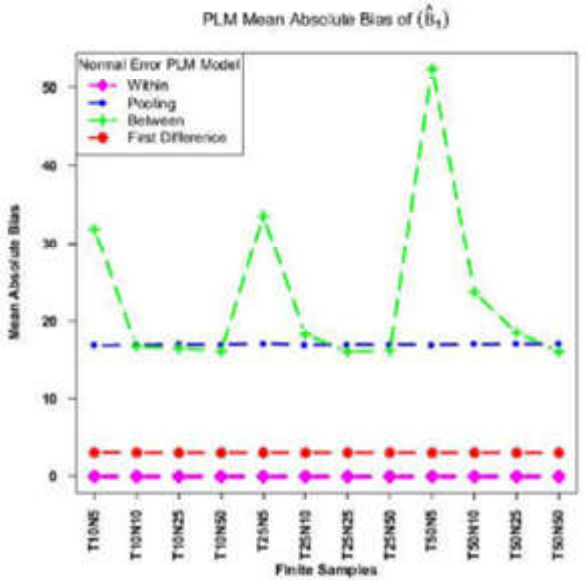

Fig. 1: Absolute Bias of Normal Error PLM Estimated (β_1)

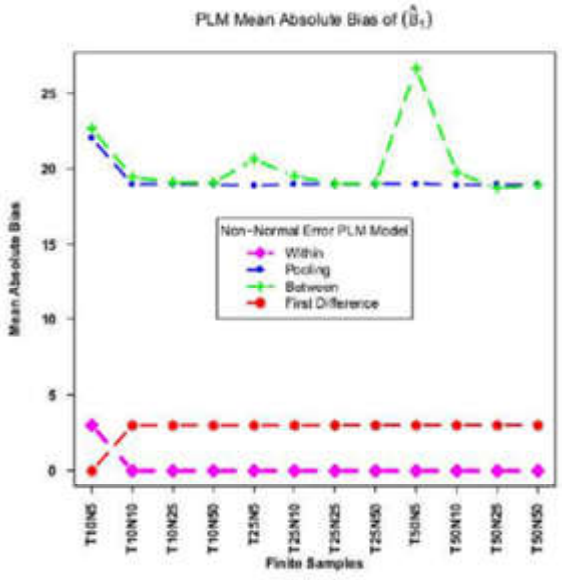

Fig. 2: Absolute Bias of Non-Normal Error PLM Estimated (β_1)

Table 4: Absolute Bias of Normal and Non-Normal Error GFGLS Estimated (β_1)

Dimension	Absolute Bias of Normal Error GFGLS Estimated β_1				Absolute Bias of Non-Normal Error GFGLS Estimated β_1			
	Within	Pooling	Between	First Diff.	Within	Pooling	Between	First Diff.
T10N5	0.0225	16.8115	39.5511	3.0158	0.0204	19.0263	39.5511	2.9954
T10N10	0.0159	16.871	40.463	3.0115	0.0151	18.9975	40.463	2.9955
T10N25	0.0109	16.958	39.192	2.9958	0.0087	19.0308	39.192	2.9985
T10N50	0.0081	16.9223	40.2633	2.9979	0.0061	18.9888	40.2633	2.9951
T25N5	0.0172	17.0931	39.7801	3.002	0.0119	18.9305	39.7801	2.9992
T25N10	0.0095	16.8847	40.1192	2.9959	0.0092	19.0269	40.1192	2.9987
T25N25	0.0066	16.9324	41.0657	3.0012	0.0054	19.0084	41.0657	3.0004
T25N50	0.0045	16.9409	40.0217	2.9994	0.0037	19.0235	40.0217	3.0006
T50N5	0.0102	16.9309	39.2341	2.9972	0.0084	19.0590	39.2341	3.0010
T50N10	0.0075	16.9583	40.1257	2.9992	0.0061	18.9649	40.1257	3.0003
T50N25	0.0046	17.0258	39.9748	2.9998	0.0035	19.0020	39.9748	3.0000
T50N50	0.0037	16.9762	40.2959	2.9999	0.0026	18.9855	40.2959	3.0002

Note: *Within is more efficient, follow by First Difference at normal and non-normal error GFGLS estimated* β_1

Table 4 shows that within estimator have the lower value at each sample size; this implies that it is more efficient, follow by First Difference at both normal and non-normal error GFGLS estimated β_1. The performanc e of within and the other estimators are as also shown graphically in figures 1 and 2 below for normal and non-normal error GFGLS estimated.

Fig.3:Absolute Bias of Normal Error GFGLS Estimated

Absolute Bias of Non-Normal Error GFGLS Estimated

Table 5: Absolute Bias of Normal and Non-Normal Error PLM Estimated

Dimension	Absolute Bias of Normal Error PLM Estimated $\boldsymbol{\beta_2}$				Absolute Bias of Non-Normal Error PLM Estimated $\boldsymbol{\beta_2}$			
	Within	Pooling	Between	FD	Within	Pooling	Between	FD
T10N5	0.0130	1.0016	1.0855	1.0024	0.0315	0.9978	1.0557	0.9952
T10N10	0.008	1.0019	1.0403	1.0021	0.0211	0.9994	0.9915	1.0001
T10N25	0.0049	1.0000	1.0006	0.9993	0.0138	0.9978	0.9901	0.9988
T10N50	0.0034	1.0010	1.0129	1.0000	0.0082	1.0000	0.9936	1.0013
T25N5	0.0072	0.9968	1.2027	1.0002	0.0176	1.0011	1.4609	0.9984
T25N10	0.0056	1.0018	1.0551	0.9998	0.0117	0.9992	0.9407	1.0008
T25N25	0.0034	1.0015	1.0448	1.0000	0.0084	1.0011	0.9944	1.0016
T25N50	0.0026	1.0011	1.0206	1.0002	0.0051	0.9988	0.9807	0.9998
T50N5	0.0048	0.9970	1.5646	0.9989	0.0133	0.9979	1.5835	0.9985
T50N10	0.0029	1.0024	1.0545	1.0014	0.0079	1.0014	0.9808	1.0007
T50N25	0.0025	1.0008	1.0205	1.0008	0.0049	1.0013	1.0275	1.0014
T50N50	0.0015	0.9999	1.0323	0.9992	0.0038	0.9999	0.9850	1.0000

Note: *Within is more efficient at normal and non-normal error PLM estimated*

Table 5 shows that within estimator have the lower value at each sample size; this implies that it is more efficient at both normal and non-normal error PLM estimated β_2. The performance of within and the other estimators are also as shown graphically in figures 1 and 2 below for normal and non-normal error PLM estimated.

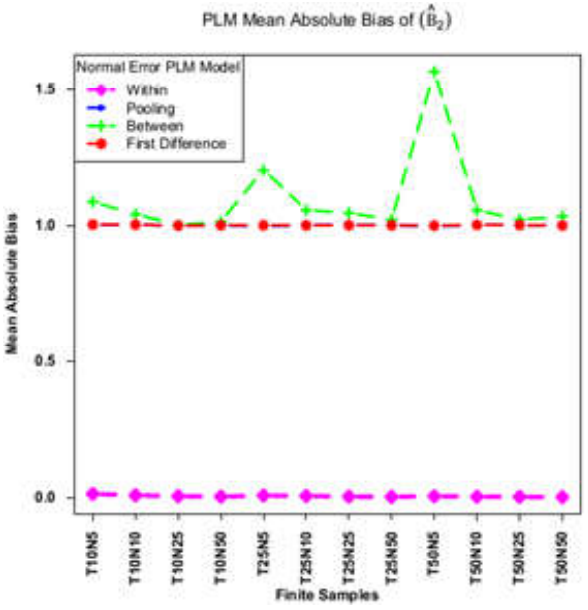

Fig.4:Absolute Bias of Normal Error PLM Estimated (β_2)

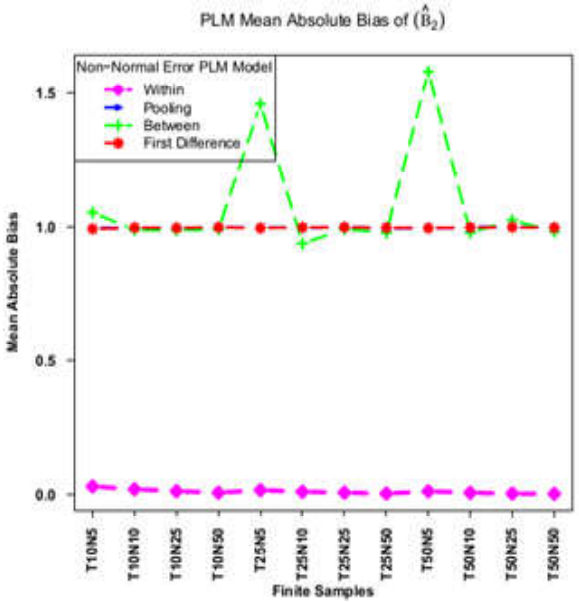

Fig.5:Absolute Bias of Non-Normal Error PLM Estimated (β_2)

Table 6: Absolute Bias of Normal and Non-Normal Error GFGLS Estimated β_2

Dimension	Absolute Bias of Normal Error GFGLS Estimated β_2				Absolute Bias of Non-Normal Error GFGLS Estimated β_2			
	Within	Pooling	Between	FD	Within	Pooling	Between	FD
T10N5	0.0130	1.0016	1.6939	1.0024	0.0315	0.9978	1.6939	0.9952
T10N10	0.008	1.0019	1.1898	1.0021	0.0211	0.9994	1.1898	1.0001
T10N25	0.0049	1.0000	1.0451	0.9993	0.0138	0.9978	1.0451	0.9988
T10N50	0.0034	1.0000	1.0086	1.0000	0.0082	1.0000	1.0086	1.0013
T25N5	0.0072	0.9968	2.7366	1.0002	0.0176	1.0011	2.7366	0.9984
T25N10	0.0056	1.0018	1.2320	0.9998	0.0117	0.9992	1.2320	1.0008
T25N25	0.0034	1.0015	0.9622	1.0000	0.0084	1.0011	0.9622	1.0016
T25N50	0.0026	1.0011	0.0048	1.0002	0.0051	0.9988	0.9688	0.9998
T50N5	0.0048	0.997	1.9179	0.9989	0.0133	0.9979	1.9179	0.9985
T50N10	0.0029	1.0024	1.112	1.0014	0.0079	1.0014	1.112	1.0007
T50N25	0.0025	1.0008	1.0298	1.0008	0.0049	1.0013	1.0298	1.0014
T50N50	0.0015	0.9999	0.9749	0.9992	0.0038	0.9999	0.9749	1.0000

Note: *Within is more efficient at normal and non-normal error GFGLS estimated* β_2

Table 6 shows that within estimator have the lower value at each sample size; this implies that it is more efficient at both normal and non-normal error GFGLS estimated . The performance of within and the other estimators are also as shown graphically in figures 1 and 2 below for normal and non-normal error GFGLS estimated

Fig.7:Absolute Bias of Normal Error GFGLS Estimated

Fig.8:Absolute Bias of Non-Normal Error GFGLS Estimated

Table 7: Absolute Bias of Normal and Non-Normal Error PLM Estimated

Dimension	Absolute Bias of Normal Error PLM Estimated β_3				Absolute Bias of Non-Normal Error PLM Estimated β_3			
	Within	Pooling	Between	FD	Within	Pooling	Between	FD
T10N5	0.0130	1.0016	1.0855	1.0024	0.0315	0.9978	1.0557	0.9952
T10N10	0.0080	1.0019	1.0403	1.0021	0.0211	0.9994	0.9915	1.0001
T10N25	0.0049	1.0000	1.0006	0.9993	0.0138	0.9978	0.9901	0.9988
T10N50	0.0034	1.0010	1.0129	1.0000	0.0082	1.0000	0.9936	1.0013
T25N5	0.0072	0.9968	1.2027	1.0002	0.0176	1.0011	1.4609	0.9984
T25N10	0.0056	1.0018	1.0551	0.9998	0.0117	0.9992	0.9407	1.0008
T25N25	0.0034	1.0015	1.0448	1.0000	0.0084	1.0011	0.9944	1.0016
T25N50	0.0026	1.0011	1.0206	1.0002	0.0051	0.9988	0.9807	0.9998
T50N5	0.0048	0.997	1.5646	0.9989	0.0133	0.9979	1.5835	0.9985
T50N10	0.0029	1.0024	1.0545	1.0014	0.0079	1.0014	0.9808	1.0007
T50N25	0.0025	1.0008	1.0205	1.0008	0.0049	1.0013	1.0275	1.0014
T50N50	0.0015	0.9999	1.0323	0.9992	0.0038	0.9999	0.985	1.0000

Note: *Within is more efficient at normal and non-normal error PLM estimated*

Table 7 shows that within estimator have the lower value at each sample size; this implies that it is more efficient at both normal and non-normal error PLM estimated . The performance of within and the other estimators are also as shown graphically in figures 1 and 2 below for normal and non-normal error PLM estimated .

Fig.9:Absolute Bias of Normal Error PLM Estimated

Fig.10:Absolute Bias of Non-Normal Error PLM Estimated

Table 8: Absolute Bias of Normal and Non-Normal Error GFGLS Estimated

Dimension	Absolute Bias of Normal Error GFGLS Estimated β_3				Absolute Bias of Non-Normal Error GFGLS Estimated β_3			
	Within	Pooling	Between	FD	Within	Pooling	Between	FD
T10N5	0.0130	1.0016	1.6939	1.0024	0.0315	0.9978	1.6939	0.9952
T10N10	0.0080	1.0019	1.1898	1.0021	0.0211	0.9994	1.1898	1.0001
T10N25	0.0049	1.0000	1.0451	0.9993	0.0138	0.9978	1.0451	0.9988
T10N50	0.0034	1.0010	1.0086	1.0000	0.0082	1.0000	1.0086	1.0013
T25N5	0.0072	0.9968	2.7366	1.0002	0.0176	1.0011	2.7366	0.9984
T25N10	0.0056	1.0018	1.232	0.9998	0.0117	0.9992	1.232	1.0008
T25N25	0.0034	1.0015	0.9622	1.0000	0.0084	1.0011	0.9622	1.0016
T25N50	0.0026	1.0011	0.9688	1.0002	0.0051	0.9988	0.9688	0.9998
T50N5	0.0048	0.997	1.9179	0.9989	0.0133	0.9979	1.9179	0.9985
T50N10	0.0029	1.0024	1.112	1.0014	0.0079	1.0014	1.112	1.0007
T50N25	0.0025	1.0008	1.0298	1.0008	0.0049	1.0013	1.0298	1.0014
T50N50	0.0015	0.9999	0.9749	0.9992	0.0038	0.9999	0.9749	1.0000

Note: Within is more efficient at normal and non-normal error GFGLS estimated

Table 8 shows that within estimator have the lower value at each sample size; this implies that it is more efficient at both normal and non-normal error GFGLS estimated . The performance of within and the other estimators are also as shown graphically in figures 1 and 2 below for normal and non-normal error PLM estimated ..

Fig.11:Absolute Bias of Normal Error GFGLS Estimated

Fig.12:Absolute Bias of Non-Normal Error GFGLS Estimated

Table 9: of PLM and GFGLS with Normal and Non-Normal Errors at Varying Sample Size

	Models	T10N5	T10N10	T10N25	T10N50	T25N5	T25N10	T25N25	T25N50	T50N5	T50N10	T50N25	T50N50
PLM with Normal Error	Within**	0.84	0.87	0.89	0.89	0.94	0.95	0.96	0.96	0.97	0.97	0.98	0.98
	Pooling***	0.92	0.96	0.98	0.99	0.97	0.98	0.99	1.00	0.98	0.99	1.00	1.00
	Between*	0.20	0.60	0.84	0.92	0.20	0.60	0.84	0.92	0.20	0.60	0.84	0.92
	First Diff.***	0.91	0.96	0.98	0.99	0.97	0.98	0.99	1.00	0.98	0.99	1.00	1.00
PLM with Exponential Error	Within**	0.84	0.87	0.89	0.89	0.93	0.95	0.95	0.96	0.97	0.97	0.98	0.98
	Pooling***	0.92	0.96	0.98	0.99	0.97	0.98	0.99	0.99	0.98	0.99	0.99	1.00
	Between*	0.20	0.59	0.83	0.90	0.20	0.58	0.81	0.89	0.20	0.56	0.78	0.86
	First Diff.***	0.91	0.95	0.98	0.99	0.97	0.98	0.99	1.00	0.98	0.99	1.00	1.00
GFGLS with Normal Error	Within**	0.84	0.87	0.89	0.89	0.94	0.95	0.96	0.96	0.97	0.97	0.98	0.98
	Pooling***	0.92	0.96	0.98	0.99	0.97	0.98	0.99	1.00	0.98	0.99	1.00	1.00
	Between*	0.20	0.60	0.84	0.92	0.20	0.60	0.84	0.92	0.20	0.60	0.84	0.92
	First Diff.***	0.91	0.96	0.98	0.99	0.97	0.98	0.99	1.00	0.98	0.99	1.00	1.00
GFGLS with Exponential Error	Within**	0.84	0.87	0.89	0.89	0.93	0.95	0.95	0.96	0.97	0.97	0.98	0.98
	Pooling***	0.92	0.96	0.98	0.99	0.97	0.98	0.99	0.99	0.98	0.99	0.99	1.00
	Between*	0.29	0.47	0.53	0.57	0.29	0.44	0.51	0.56	0.31	0.46	0.53	0.53
	First Diff.***	0.91	0.95	0.98	0.99	0.97	0.98	0.99	1.00	0.98	0.99	1.00	1.00

Note: $* : Least R^2_{adj}$; $** : Low R^2_{adj}$; $*** : Medium R^2_{adj}$; $**** : High\ R^2_{adj}$

Table 9 shows that across varying samples sizes (dimension), all the four estimators have increasing R^2_{adj}. This implies efficiency (predictor significance) of these estimators increases as panel data dimension increases. The ranks of the estimators in terms of R^2_{adj} is: Pooling > FD > Within > BTW. This is also graphically as shown below.

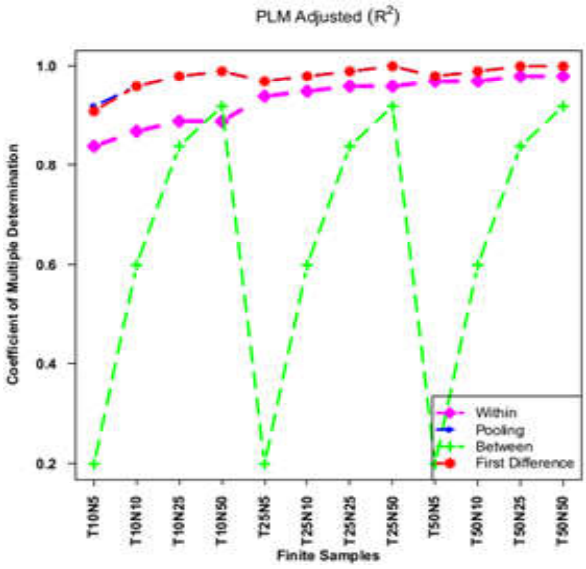

Fig.13: Adjusted R Squared from PLM Models with Normal Error

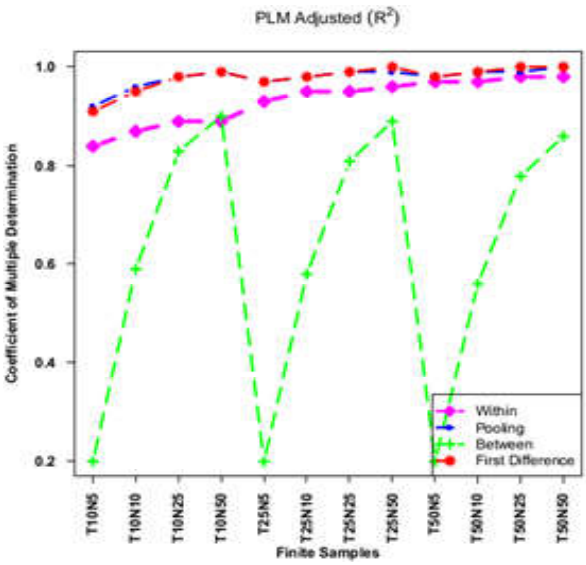

Fig.14: Adjusted R Squared from PLM Models with Non-Normal Errors

Note: where the estimations are favourably competitive, one overrides the other in the graphs.

Table 10: PLM and GFGLS RMSE of β_1

Dimension	RMSE of Normal Error PLM Estimated β_1				RMSE of Non-Normal Error PLM Estimated β_1			
	Within	Pooling	Between	FD	Within	Pooling	Between	FD
T10N5	0.0004	16.9298	57.7721	3.0172	0.0374	19.0509	24.1635	2.9973
T10N10	0.0283	16.9648	21.0548	3.0125	0.0265	19.0093	19.9751	2.9961
T10N25	0.0173	16.9984	18.7499	2.9961	0.0141	19.0351	19.2602	2.9989
T10N50	0.0141	16.9355	16.9007	2.9980	0.0100	18.9908	19.1743	2.9953
T25N5	0.0300	17.1598	60.5396	3.0022	0.0200	18.9466	30.1101	2.9994
T25N10	0.0173	16.9142	27.1096	2.996	0.0173	19.0343	21.1728	2.9988
T25N25	0.0100	16.9497	19.6911	3.0012	0.0100	19.0122	19.4119	3.0004
T25N50	0.0100	16.9479	18.008	2.9994	0.0000	19.0247	19.159	3.0006
T50N5	0.0173	16.962	106.0712	2.9973	0.0141	19.0717	46.8296	3.0011
T50N10	0.0141	16.978	38.9071	2.9992	0.0100	18.9702	22.7038	3.0003
T50N25	0.0100	17.0317	24.5577	2.9999	0.0000	19.0048	19.3038	3.0000
T50N50	0.0000	16.9803	19.5996	2.9999	0.0000	18.987	19.2141	3.0002

***Note:** within is more efficient, follow by first difference.*

Table 10 shows that in terms of stability of parameter estimate using absolute bias and rmse, the estimators rank as follows: Within > FD ≥ Pooling BTW.

Conclusion

Based on the foregoing investigation of efficiency and finite sample properties of the four panel data model estimators, it can be concluded that: The within estimator is the most stable and most efficient estimator of panel data model parameters with both one-stage and two-stage normal and non-normal error structures. The FD estimator is the next most stable while both pooling and BTW are worse but pooling is more stable. Across varying samples sizes (dimension), all the four estimators have increasing. This implies efficiency (predictor significance) of these estimators increases as panel data dimension increases. The ranks of the estimators in terms of is: Pooling > FD > Within > BTW. In terms of stability of parameter estimate using absolute bias and rmse, the estimators rank as follows: Within > FD Pooling BTW. This is in line with the result of Nwakuya & Biu, (2019) who examines the within-group and first difference fixed effect models using panel data set and found that in the within-group model, trade was the only independent variable that contributes significantly to GDP but in the first difference model both trade and population contributed significantly to GDP. The finding also reveals that within group model had a better fit with an R^2 of 0.77317 as compared to first difference model which reported R^2 of 0.75472.

References

Arellano, M. (2003). *Panel Data Econometrics*, Oxford University Press, Oxford.

Baltagi, B. H. (2005). *Econometrics analysis of panel data*, 3rd edition, John Wiley and Sons Ltd, England.

Chudik, A., and Pesaran, M. H. (2015), Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors, *Journal of Econometrics*, 188, 393–420.

Creel S, Christianson D, Winnie J. (2011). A survey of the effects of wolf predation risk onpregnancy rates and calf recruitment in elk. Ecological Applications 21: 2847–2853 http://dx.doi.org/10.1890/11-0768.1.

Garba, M-K. Oyejola B.A., and Yahya W.B. (2013). Investigations of Certain Estimators for Modeling Panel Data Under Violations of Some Basic Assumptions, *Journal of Mathematical Theory and Modeling*, 3(10), 47-54.

Greene W (2003). *Econometric Analysis*. 5th edition. Prentice Hall.

Greene, (2008). *Econometric analysis.* 6th ed., Upper Saddle River, N.J.: Prentice Hall.

Juodis, A. (2022), "A Regularization Approach to Common Correlated Effects Estimation," *Journal of Applied Econometrics*, 37,(1), 788–810.

Kapetanios, G., Serlenga, L., and Shin, Y., (2023), Testing for Correlation between the Regressors and Factor Loadings in Heterogeneous Panels with Interactive Effects, *Journal of Empirical Economics*, 64, 2611–2659.

Kapetanios, G., Pesaran, M. H., and Yamagata, T. (2011), "Panels with Nonstationary Multifactor Error Structures," Journal of Econometrics, 160,(11), 326–348.

Maddala, G.S. (2008). *Introduction to econometrics*, 3rd edition, John Wiley and Sons,Ltd, Chichester, UK.

Nwakuya M. T., Biu E. O., (2019). Comparative Study of Within-Group and First Difference Fixed Effects Models. *American Journal of Mathematics and Statistics,* 9(4), 177-181.

Olofin, S. O., Rebuttal, E. and Salisu, A. A. (2010), Testing for heteroscedasticity and serialcorrelation in a two-way error component model. Ph.D dissertation submitted to the Department of Economics, University of Ibadan, Nigeria.

Wooldridge, J. M. (2012), *Introductory Econometrics*: A Modern Approach, 5th edition, South-Western College.

Westerlund J, Urbain J-P (2015) Cross-sectional averages versus principal components. J Econom 185,(2),372–377.

Wooldridge, J. M. (2012). Introductory Econometrics: A Modern Approach, 5th edition, South- Western College.