Science World Journal Vol. 19(No 4) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i4.49

Full Length Research Article

INVESTIGATION OF PANEL MODELLING TECHNIQUES IN THE PRESENCE OF COLLINEARITY REGRESSORS

*Ijiko E., Adenomon M.O. and Nweze N.O.

Department of Statistics, Nasarawa State University, Keffi

*Corresponding Author Email Address: ericijiko011@gmail.com

ABSTRACT

In order to estimate the presence of collinearity, this study used the fixed effect model, pooled effect model, and random effect model in the presence of collinearity regressors. In respond one, data were simulated under several types of collinearity verifiable at varying sample sizes (-0.1, -0.5, -0.9, 0.1, 0.5, and 0.9) in the Monte Carlo simulation. For the models estimators, two regressor estimators were used. Simulations were run at various panel structures and collinearity regressors in the Monte Carlo study. The trial was conducted ten thousand times (10000), and the accuracy of the model estimation was assessed using the Root Mean Square Error (RMSE). The results of the study showed that the estimation of the small sample panel structure model. While following time series lengths (5, 10, 10, 30, 60, and 60) have 10,000 repetitions of the experiment conducted in the R environment. The Root Mean Square Error (RMSE) was used to assess the models. The RMSE values for the fixed and random model are fluctuated as the collinearity levels grew in all of the scenarios that were taken into consideration. Based on the analysis, the Fixed Effects (FE) Model is the best-performing model, particularly in larger datasets, as it minimizes both bias and RMSE. The Random Effects (RE) Model can also be effective, especially when collinearity is moderate and when the assumptions of random effects hold true. However, for datasets where collinearity is high, or where individual-specific effects are crucial, the Fixed Effects Model provides more reliable estimate. The Pooled Regression Model should generally be avoided in cases where collinearity or panel-specific heterogeneity is significant, as it produces the least stable and least reliable results across different collinearity levels

Keywords: Fixed effect, Pooled regression, Random Effect, Collinearity Regressors.

INTRODUCTION

The increasing reliance on panel data models in empirical research stems from their ability to analyze data that varies across both individuals and time. Panel data, which combines cross-sectional and time-series dimensions, allows for the examination of dynamic relationships while controlling for individual heterogeneity Baltagi (2021). This versatility makes panel models indispensable in fields such as economics, finance, sociology, and political science. For example, researchers studying economic growth, trade patterns, labour market dynamics, or policy impacts rely heavily on panel data to derive robust conclusions Gujarati & Porter (2017).

However, panel data analysis is often undermined by certain statistical challenges, with multicollinearity and outlier regressors being prominent among them. Multicollinearity occurs when two or more explanatory variables are highly correlated, making it difficult to isolate their individual effects. This issue is prevalent in empirical studies where macroeconomic variables, like Gross Domestic Product, inflation, and exchange rates, often move in tandem Wooldridge (2016). For instance, in a study of monetary policy impacts, variables like interest rates and money supply may exhibit significant collinearity, complicating the estimation process.

According to a simulation study by Adenomon et al. (2024) on the in-sample forecasting performances of Sims-Zha Bayesian VARX in the presence of collinearity between the exogenous variables for small sample situation were used time series exogenous variables typically enhance the prediction of endogenous variables, according to the research. This paper examined the forecast performance of six (6) versions of Bayesian Vector Autoregressive models with exogenous variables (BVARX) using normal-inverse Wishart Prior when collinearity exist between the exogenous variables for small sample situations. To achieve this, VAR(2) model was used to simulate bivariate time series from a stable process while bivariate exogenous variables were simulated from a standard normal distribution to possess the following collinearity levels: -0.99, -0.95, -0.9, -0.85, -0.8, 0.8, 0.85, 0.9, 0.95, 0.99. The experiment was carried out in R environment and repeated 10,000 times for the following time series lengths: 8, 16, 32 and 50. The Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) were used to adjudge the models. In all the scenarios considered, BVARX4 performed best while BVARX1 performed worst in all the collinearity levels and time series lengths. Lastly, RMSE and MAE values of the BVARX models are higher with negative collinearity compared to positive collinearity while the values of RMSE and MAE for the BVARX model decreased as the time series length.

An empirical assessment of collinearity effects in dynamic panel models was presented by Wooldridge (2022). In his analysis of macroeconomic datasets, he found that Generalized Method of Moments (GMM) estimators effectively mitigated collinearity by incorporating instrumental variables, particularly lagged regressors. Wooldridge (2022) emphasized the importance of the Hansen J-test to validate the robustness of instruments. Adenomon et al. (2016), compared the Performances of Classical VAR and Sims-Zha Bayesian VAR Models in the Presence of Collinearity and Autocorrelated Error Terms in time series literature, many authors have found out that multicollinearity and autocorrelation usually afflict time series data. In this study, we compare the performances of classical VAR and Sims-Zha Bayesian VAR models with quadratic decay on bivariate time series data jointly influenced by collinearity and autocorrelation. Simulate bivariate time series data for different collinearity levels (－0.99, －0.95, －0.9, －0.85, －0.8, 0.8, 0.85, 0.9, 0.95, 0.99) and autocorrelation levels (－0.99, －0.95, －0.9, －0.85, －0.8, 0.8, 0.85, 0.9, 0.95, 0.99) for time series length of 8, 16, 32, 64, 128, and 256 respectively. The outcomes of 10,000 simulations show that the model's performance changes as the time series lengths and the levels of autocorrelation and collinearity increase. Furthermore, the outcomes demonstrate that the BVAR4 model is a practical forecasting model. Therefore, when choosing a suitable model for forecasting, we advise taking into account the time series length, the levels of autocorrelation and collinearity, and both.

The implications of collinearity in panel regression models, particularly in fixed effects (FE) and Random Effects (RE) frameworks, were assessed by Baltagi (2021). Using firm-level investment data, Baltagi demonstrated that multicollinearity inflates standard errors, making parameter estimates unreliable. To address this issue, ridge regression and Principal Component Analysis (PCA) were employed to reduce multicollinearity while retaining model accuracy.

Adenomon et al. (2015). Examined Short Term Forecasting Performances of Classical VAR and Sims-Zha Bayesian VAR Models for Time Series with Collinear Variables and Correlated Error Terms Forecasts they worked short term forecast because of the problem of limited data or time series data that often encounter in time series analysis. The simulation studied considers the performances of the classical VAR and Sims-Zha Bayesian VAR for short term series at different levels of collinearity and correlated error terms. The results from 10,000 iteration revealed that the BVAR models were excellent for time series length of $T = 8$ for all levels of collinearity while the classical VAR was effective for time series length of $T = 16$ for all collinearity levels except when $\rho = -0.9$ and $\rho = -0.95$. Therefore, the study recommend that for effective short term forecasting, the time series length, forecasting horizon and the collinearity level should be considered.

Using Monte Carlo simulations, Kumar et al. (2023) investigated the effects of multicollinearity on FE and GMM estimators. Their research showed that, especially with large sample sizes, GMM performed better than conventional estimators when regressors were collinear. Additionally, the authors emphasized the Variance Inflation Factor's (VIF) usefulness as a diagnostic tool.

Shrestha Noora (2020). Compared multicollinearity when the multiple linear regression analysis includes several variables that are significantly correlated not only with the dependent variable but also to each other. Multicollinearity makes some of the significant variables under study to be statistically insignificant. The three main methods for identifying multicollinearity in customer satisfaction questionnaire survey data were covered in this paper. The first two techniques were the correlation coefficients and the variance inflation factor, while the third method is eigenvalue method. It was observed that the product attractiveness was more rational cause for the customer satisfaction than other predictors. Furthermore, advanced regression procedures such as principal components regression, weighted regression, and ridge regression method will be used to determine the presence of multicollinearity.

F. D. Carsten et al. (2013), examined collinearity of non-independence of predictor variables, usually in a regression-type analysis. It was a common feature of any descriptive ecological data set and can be a problem for parameter estimation because it inflates the variance of regression parameters and hence potentially leads to the wrong identification of relevant predictors in a statistical model. Collinearity is a severe problem when a model is trained on data from one region or time, and predicted to another with a different or unknown structure of collinearity. To demonstrate the reach of the problem of collinearity in ecology, this demonstrate how relationships among predictors differ between biome, change over spatial scales and through time. Across disciplines, different approaches to addressing collinearity problems have been developed, ranging from clustering of predictors, threshold-based pre-selection, through latent variable methods, to shrinkage and regularisation. Using simulated data with five predictor-response relationships of increasing complexity and eight levels of collinearity that contrasted approaches to address collinearity with standard multiple regression and machine-learning approaches. This assessed the performance of each approach by testing its impact on prediction to new data. In the extreme, the tested whether the methods were able to identify the true underlying relationship in a training dataset with strong collinearity by evaluating its performance on a test dataset without any collinearity. The study observed that methods specifically designed for collinearity, such as latent variable methods and tree based models, did not outperform the traditional GLM and threshold-based pre-selection. These findings highlight the value of GLM in combination with penalised methods (particularly ridge) and threshold-based pre-selection when omitted variables are considered in the final interpretation. However, all approaches tested yielded degraded predictions under change in collinearity structure and the 'folk lore'-thresholds of correlation coefficients between predictor variables of $|r| > 0.7$ was an appropriate indicator for when collinearity begins to severely distort model estimation and subsequent prediction. The use of ecological understanding of the system in pre-analysis variable selection and the choice of the least sensitive statistical approaches reduce the problems of collinearity, but cannot ultimately solve them.

Based on studies conducted by Arum *et al.* (2023) on combating multicollinearity in linear regression models using robust Kibria-Lukman mixed with principal component estimator, computation and simulation shows that the performance of the ordinary least square estimator was good when the regression model dataset was free of multicollinearity. In a regression model dataset, and multicollinearity can coexist, and the least squares estimator experiences difficulties when both issues exist. Designing a new estimator that can manage both problems is the aim of this endeavor. The study developed a novel estimator known as robust PC-KL by combining the principal component estimator (PCE), M-estimator, and Kibria-Lukman estimator (KLE). The robust PC-KL estimator is effective at solving problems both singly and collectively since it possesses traits from the M-estimator, KLE, and PCE. Through simulation modelling and practical application, we compared the resilient PC-KL estimator's performance with other available estimators, utilizing mean squared error (MSE) as a performance evaluation criterion. This study's robust PC-KL estimator performed better than other estimators based on theoretical comparison, simulation design, and real-world application as it possessed the smallest.

Ajao *et al,* (2023), Compared Some Panel Data Estimators in The Presence of Autocorrelation. Three estimates models were used in the work to model panel data with characteristics devoid of first order autocorrelation: pool regression, fixed effect, and random effect models. Studies using simulations were conducted at various panel configurations and autocorrelation levels. The models' performance was evaluated using Root Mean Square Error (RMSE), which was obtained after 10,000 repetitions of the experiment. The work's findings demonstrated that, for small sample panel structure $N = 25$, $T = 5$, and $n = 5$, the fixed effect model is preferred at all levels of autocorrelation. Random effect model is preferable for large panel structure for $N = 450$, $T = 30$, $n = 15$, regardless of autocorrelation level; however, when moderate for $N = 50$, $T = 10$, $n = 5$, random effect model.

Youseef, Ahmed Hassen et al. (2023), Examined designing a statistical model, applied researchers strive to make the model consistent, unbiased, and efficient. Labour productivity is an important economic indicator that is closely linked to economic growth, competitiveness, and living standards within an economy. These study proposed the one-way error component panel data model for labour productivity. One of the problems that we can encounter in panel data is the problem of multi-collinearity. Therefore, multi-collinearity problem is considered. This problem has been detected. After then, new good unrelated estimators are obtained using the principal component technique. For the purposed of the analysis, the multi-collinearity problem between the explanatory variables was examined, using principal component techniques with the application of the panel data model focused on the impact of public capital, private capital stock, labour, and state unemployment rate on gross state products. The analysis was based on three estimation methods: fixed effect, random effect, and pool regression. The challenge is to get estimators with good properties under reasonable assumptions and to ensure that statistical inference is valid throughout robust standard errors. And after application, fixed effect estimation turned out to play a key role in the estimation of panel data models. Based on the results of hypothesis testing, the real data result showed that the fixed effect model was more accurate compared to the two models of random effect and pooling effect. In addition, robust estimation was used to get more efficient estimators since heteroscedasticity has been confirmed.

Zuhair, Muhammad Anono et al. (2021) examined on comparisons of estimator's efficiency for linear regression with joint presence of autocorrelation and multicollinearity withTwo stage K-L estimator by combining these two estimators previously proposed by Prais Winsten (2018) and Kibra & Lukman (2020) for autocorrelation and multicollinearity respectively and to derived the necessary and sufficient condition for its superiority over other competing estimators. Simulation study was used to ascertain the dominance of this new estimator using the finite sample properties of estimators in terms of the estimated mean squared error. The study findings show that under severe autocorrelation and collinearity condition, the proposed Two stage K-L estimator appears to be having a similar performance with RMLE and MLE. Also, under severe autocorrelation and moderate collinearity condition, regardless of the sample size, the proposed Two Stage K-L estimator is seen to outperform all other estimators and lastly, the Two stage K-L estimator appears to have an improved performance as the large sample sizes. The study recommends that when autocorrelation and multicollinearity level is at moderate to severe, the proposed Two stage K-L estimator will perform better regardless of the size of the data, and the degree of autocorrelation and multicollinearity should be considered while estimating parameters and thus applying an efficient estimator to avoid erroneous inferences.

MATERIALS AND METHODS

This work considers Pooled Regression, Fixed Effect and Random Effect models with three exogenous and one endogenous variable. The collinearity regressors work done by C.F. Dormann, S. Lautenbach and Adenomon M. O. et al. (2016) to name only a handful. The majority of previous research on collinearity regresors concentrated on one or two exogenous variables. But, considered three exogenous variables with the possibility of existence of collinearity between them and this effects with respect to stability while efficiency of the estimation methods for panel data models were examined

Random Effect Model

With the exception of the fact that there is only one draw for each group that enters the regression in the same way every period, the random effects model defines αi is a group-specific random element and treats it as part of the error term, much like i. Also, the key difference between random and fixed effects is not whether these effects are stochastic but rather whether the unobserved individual impact has components that are linked with the model's regressors, the random effects model is expressed as

$$y_{it} = \mu + B_1 x_{1,it} + B_2 x_{2,it} + \ldots + B_k x_{k,it} + (\alpha_i - \mu) + \varepsilon_{it} \quad 1$$

where $= \alpha_i + \varepsilon_{it}$is treated as an error term consisting of two components.

The random specification of unobserved effects corresponds to a particular case of variance-component or error-component model, in which the residual is assumed to consist of two components : $vit = \alpha i + \varepsilon it$. As suggested by Wooldridge (2016), the fixed effect specification can be viewed as a case in which αi is a random parameter with $cov\ (\alpha i,\ x_{it}')\neq 0$, whereas the random effect model correspond to the situation in which $cov(\alpha i,\ x_{it}') = 0$.The variance of yit conditional on x_{it} is the sum of two components

Fixed Effect Model

The individual effects α_i; i = 1, 2, … N, in this model are estimated as a collection of constants that are independent of time. They are correlated with the reported regressors; thus regard them as unobserved random variables. The entire model can be viewed as an ordinary linear regression model if 'i' are observed for every individual

A fixed effect model cannot be used to evaluate effects that do not change over time.
Considering the panel data model (1), the N equations are expressed as

$$y_{it} = \mu + B_1 x_{1,it} + B_{2,it} + \ldots + B_k x_{k,it} + \varepsilon_{it} \quad 2$$

where it is a one vector, with size $T \times 1$. One random selection from a cross section is represented by the above equation. Fixed effect analysis is more reliable than random-effects analysis because it permits ($\alpha_i\ x_{it}$) to be any function of x_i. Estimation of Panel Data Regression Models with Individual Effects Jirata Tadesse Megersa (2018) avoids referring to αi as a random effect or a fixed effect. Instead, this will refer to αi as unobserved effect, unobserved heterogeneity, and so on. Nevertheless, later this would label two different estimation methods random effects estimation and fixed effects estimation.

Pooled Regression

The pooled model does not differ from the common regression equation.it regard each observation as unrelated to the others ignoring panel and time, the regression model can be expressed as

$$y_{it} = \mu + B_1 x_{1,it} + B_{2,it} + \ldots + B_k x_{k,it} + \varepsilon_{it} \quad 3$$

These factors might include special relationships with stakeholders, special expertise's of the firms that gives them a unique position in the market and the corporate culture in general. These issues are however reflected in the error term and should not pose a threat to the reliability of the results unless the error term is correlated with both the dependent and independent variable.

Science World Journal Vol. 19(No 4) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i4.49

Furthermore, the results of the Pooled Regression Model will be compared to other regression tests as well as the results of previous literature like Ajao K., Adenomon M.O. and Adehi M. U. (2023).

The simulation process for this study was designed to replicate realistic panel data scenarios, where the effects of collinearity and outliers are systematically investigating. The steps followed for the simulation are outlined below:

1. Data Generation: The dataset consists of 10000 simulated panels (10000 iteration) with 60 time periods each (i.e., 60 observations). The dependent variable is generated as a linear combination of independent variables with varying degrees of collinearity and the introduction of outliers. This setup mimics real-world data scenarios where predictors are highly correlated, and outliers may be present due to errors or rare events.
2. Collinearity Introduced: The level of collinearity is manipulated by generating independent variables that are strongly correlated. For example, the correlation between three of the predictors is set to 0.9, a typical threshold for introducing collinearity (Kennedy, 2022). The simulation is conducted under three levels of collinearity: low (0.1), medium (0.5), and high (0.9)

Step 3: Given $y = 50 + 25x_1 + 30x_2 + \varepsilon_{ij}$ where $\varepsilon_{ij} \sim N(0,1)$ such that $x1 \sim N(O,1)$
$x2 \sim N(O,1)$

Step4: let the desired presentation matrix be R= $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$; then the Choleski factor, P , P= $\begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho 2} \end{bmatrix}$ and the simulated data is pre- multiplied by the choleski factor so that the simulated data is scaled to have the desired correlation level (Adenomon, et al. 2023).

Step5 data was generated for different panel data structure

N = 25,	T = 5
N = 50,	T = 10
N = 100,	T = 10
N = 450,	T = 30
N = 1800,	T = 60,
N = 3600,	T = 60,

The performance of the models were examined using RMSE. where

$$\text{RMSE} = \sqrt{\frac{\Sigma(\beta_i - \hat{\beta}_i)^2}{10000}}$$

Where i = 0,1,2. .

3. Among the competing models, the model with the lowest RMSE is favoured in the simulation analysis for collinearity between the regressors

RESULTS AND DISCUSSION

Table 1. Random Effect Model for negative collinearity in the presence of regressors

Panel Structure	-0.1			-0.5			-0.9		
	B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
N=25 T=5 n=5	0.20839	0.2257275	0.2191976	0.2100265	0.2545319	0.2555869	0.2095229	0.508388	0.5083518
N=50 T=10 n=5	4.403915	21.10084	24.9521	3.300098	20.87035	24.81714	1.583867	20.7885	24.79385
N=100 T=10 n=10	3.746884	24.99621	29.90959	2.764732	24.94361	29.83472	1.305823	24.90974	29.83502
N=450 T=30 n=15	1.747488	24.89632	29.85562	1.312592	24.90383	29.86658	0.6164941	24.88633	29.8395
N=1800 T=60 n=30	0.866052	25.01076	29.99896	0.6569502	24.9969	30.00291	0.3072235	25.00452	30.00069
N=3600 T=60 n=60	0.6194616	24.99837	30.0165	0.4633411	24.99722	29.99321	0.2221642	24.99859	29.99895

Table 2. The Random Effect Model for positive collinearity in regressors are present.

Panel Structure	0.1			0.5			0.9		
	B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
N=25 T=5 n=5	0.2098273	0.2234597	0.2221818	0.2105379	0.2538472	0.2544558	0.2093645	0.5021402	0.503842
N=50 T=10 n=5	4.835517	21.20167	25.08997	5.714811	21.815	25.70164	6.440715	27.19639	30.27661
N=100 T=10 n=10	4.08368	25.16583	29.9626	4.753407	25.37227	30.22671	5.334461	27.8758	32.23381
N=450 T=30 n=15	1.927274	24.92617	29.85601	2.255356	24.96763	29.91629	2.515197	25.5474	30.32245
N=1800 T=60 n=30	0.9593419	25.00888	30.0017	1.119138	25.02748	30.01772	1.267264	25.15551	30.12523
N=3600 T=60 n=60	0.6758647	24.99977	30.01168	0.8051617	25.00982	30.0046	0.8867452	25.07577	30.05891

The table 1 and table 2 presents the performance of random effect models for varying degrees of collinearity (0.1, 0.5, and 0.9) under different panel structures with various sample sizes (N) and time periods (T). The following observations can be drawn from the data:
Low Collinearity **(0.1)**: At low collinearity, the coefficient estimates B_0, B_1, and B_2 show minimal variation, which suggests that low collinearity does not severely affect the model's estimates. This result is consistent with the general understanding that low collinearity allows for more reliable estimation in panel data models (Greene, 2012).
Moderate Collinearity **(0.5)**: As collinearity increases to 0.5, there is a noticeable increase in the variability of the coefficients. While the coefficients are still relatively stable, some distortions start appearing, especially for B_2. This is in line with the findings of Baltagi (2008), who noted that moderate collinearity can lead to less efficient estimates, although the estimates might still be unbiased.
High Collinearity **(0.9)**: At high collinearity (0.9), the variability of the coefficient estimates increases significantly, particularly for B_1 and B_2. This highlights the well-established issue that high collinearity can severely affect the precision of the estimates, making them unstable and leading to high standard errors (Chowdhury & Saha, 2019).
Small Panel Structures (**N=25, T=5)**: In smaller datasets, the instability of the coefficient estimates becomes more pronounced, especially under high collinearity (0.9). As indicated by Wooldridge (2010), small sample sizes in the presence of multicollinearity tend to exacerbate the problem of unreliable coefficient estimation.
Larger Panel Structures **(N=450, T=30 and N=3600, T=60)**: With larger panel structures, the estimates for B_0, B_1, and B_2 stabilize, even under high collinearity. The performance of random effect models improves with increasing *N* and *T*, indicating that large sample sizes are beneficial for reducing the influence of multicollinearity. This finding aligns with the work of Cameron and Trivedi (2005), who demonstrated that larger datasets improve model stability and efficiency.
As collinearity increases, the performance of the model deteriorates, particularly for $B_2B_2B_2$ and $B_1B_1B_1$, which become less reliable under higher collinearity levels (0.5 and 0.9). The results indicate that collinearity exacerbates the instability of random effect models, confirming that multicollinearity undermines the ability of the model to estimate relationships accurately (Kennedy, 2003).
As the sample size *N* and time period *T* increase, the random effect model's coefficients become more stable across all levels of collinearity. Larger datasets mitigate the negative effects of collinearity, resulting in more precise estimates. This finding is consistent with the conclusions of Roodman (2009), who emphasized that larger panel structures provide more robust and reliable estimates in the presence of multicollinearity.

Table 3. The Pooled Regression for negative collinearity in the presence regressors

Panel structure	-0.1			-0.5			-0.9		
	B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
N=25 T=5 n=5	0.2091265	0.2219246	0.2232226	0.2098765	0.2543614	0.2542796	0.2082449	0.4967611	0.4986131
N=50 T=10 n=5	0.1444686	0.1481026	0.14603	0.1448787	0.1698454	0.1692919	0.1431721	0.3374581	0.3380408
N=100 T=10 n=10	0.09986269	0.1022724	0.1034331	0.1010436	0.1174182	0.1191987	0.1015103	0.2340227	0.2342496
N=450 T=30 n=15	0.04742353	0.04744265	0.04766214	0.04718829	0.05497821	0.05474668	0.04714202	0.1096662	0.1099588
N=1800 T=60 n=30	0.02319266	0.02390735	0.02380838	0.02362866	0.02721693	0.02720035	0.02385845	0.05425352	0.05396781
N=3600 T=60 n=60	0.01676644	0.01677374	0.01658212	0.01675003	0.01921968	0.01925887	0.01686872	0.03822062	0.038227

Table 4. The Pooled Regression for positive collinearity in the presence regressors

Panel structure	0.1			0.5			0.9		
	B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
N=25 T=5 n=5	0.2088021	0.2203743	0.2191068	0.2071632	0.2517274	0.2513407	0.2079493	0.4998864	0.4963389
N=50 T=10 n=5	0.1442782	0.1465997	0.1481687	0.145081	0.1710751	0.1717995	0.1445188	0.3393412	0.3413551
N=100 T=10 n=10	0.1013071	0.102216	0.1029825	0.1018325	0.1185418	0.1181006	0.1008692	0.2349281	0.2333671
N=450 T=30 n=15	0.04679386	0.0476716	0.04771574	0.0467121	0.05487009	0.05534696	0.0469035	0.1096915	0.1089335
N=1800 T=60 n=30	0.02385425	0.02384697	0.02390737	0.02326424	0.02713731	0.02744221	0.02357888	0.05425261	0.05433089
N=3600 T=60 n=60	0.01667005	0.01671848	0.0167753	0.01655988	0.01937967	0.01921237	0.01662885	0.03872958	0.03865283

The Table 3 and Table 4, shows how the coefficients $B_0B_0B_0$, $B_1B_1B_1$, and $B_2B_2B_2$ behave under different panel structures (sample size NNN and time periods TTT) and levels of collinearity (0.1, 0.5, 0.9).

Low Collinearity **(0.1)**: At low collinearity, the coefficient estimates for B_0B_0, B_1B1, and B_2B_2 are relatively stable across all panel structures, suggesting that the model can perform well even with smaller sample sizes and fewer time periods.

Moderate Collinearity **(0.5)**: As collinearity increases to 0.5, the estimates still remain stable but show some slight increases in variability, particularly for smaller sample sizes (N=25N = 25N=25 and T=5T = 5T=5). However, the general trend suggests that the model's performance is only mildly affected by moderate collinearity, as the coefficient estimates don't deviate substantially from those under low collinearity.

High Collinearity **(0.9)**: With high collinearity (0.9), the coefficients for B_0B_0, B_1B_1, , and B_2B_2 experience more significant fluctuations. The instability increases particularly in smaller panel structures (N=25,T=5N = 25, T = 5N=25,T=5) and is less pronounced in larger datasets. This pattern is consistent with well-documented issues of multicollinearity affecting regression estimates, causing higher standard errors and less reliable coefficient estimates (Kennedy, 2003)

Small Panel Structures **(N=25, T=5)**: The smallest panel structure shows the highest variation in coefficients, particularly under high collinearity. For instance, at N=25N = 25N=25 and T=5T = 5T=5, the coefficients for high collinearity (0.9) show substantial variation, suggesting that small sample sizes are more susceptible to the effects of multicollinearity.

Larger Panel Structures **(N=1800, T=60, N=3600, T=60)**: Larger panel structures tend to stabilize the coefficient estimates, even under high collinearity. For example, at N=3600N = 3600N=3600 and T=60T = 60T=60, the coefficients are stable across all levels of collinearity, reflecting the ability of larger datasets to mitigate collinearity issues and provide more reliable estimates The results indicate that collinearit**y** has a more significant impact on the stability of the coefficient estimates in smaller panel structures. This is particularly true for the Fixed Effects and Random Effects models**,** where multicollinearity often leads to inflated standard errors and imprecise coefficient estimates (Greene, 2012). Pooled regression models also show a similar trend, with higher collinearity leading to more variability in the coefficients. As the sample size (NNN) and the number of time periods (TTT) increase, the coefficient estimates become more stable, even under higher collinearity. This is in line with the findings of Wooldridge (2010), who noted that larger datasets provide better precision and are less affected by multicollinearity. This is especially evident when comparing the results for N=25N = 25N=25 and T=5T = 5T=5 to N=3600N = 3600N=3600 and T=60T = 60T=60, where the latter provides more consistent estimates.

High collinearity (0.9) clearly exacerbates instability, particularly in smaller panel structures. The coefficients for B1B_1B1 and B2B_2B2 show more significant deviations compared to the true values in this case. This finding suggests that collinearity at this level requires careful consideration, and alternative estimation techniques such as robust regression **or** principal component analysis might be necessary when dealing with such datasets (Kutner et al., 2004). When dealing with small datasets or datasets exhibiting high collinearity, researchers should be cautious in interpreting the results, as multicollinearity can lead to unreliable estimates. Pooled regression, while simple and widely used, may not always be the best model under these conditions. Larger sample sizes and time periods help mitigate these issues, but where possible, researchers should consider alternative techniques like ridge regression or regularized method**s** to handle multicollinearity (Hoerl & Kennard, 1970).

Table 5 The Fixed Effect Model for negative collinearity in the presence regressors

Panel Structure	-0.1			-0.5			-0.9		
	B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
N=25 T=5 n=5		0.2416804	0.2446541		0.2788746	0.2806283		0.5582927	0.5577526
N=50 T=10 n=5		20.73794	24.70163		20.51803	24.41079		20.60524	24.48514
N=100 T=10 n=10		25.06313	29.98327		24.89862	29.87617		24.93488	29.84366
N=450 T=30 n=15		24.93087	29.90227		24.9171	29.86681		24.9119	29.88121
N=1800 T=60 n=30		24.98645	30.00235		24.99628	29.99547		24.99441	29.99205
N=3600 T=60 n=60		24.99401	29.99755		24.99435	29.99888		25.00204	29.99466

Science World Journal Vol. 19(No 4) 2024 https://dx.doi.org/10.4314/swj.v19i4.49
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Table 6. The Fixed Effect Model for positive collinearity in the presence regressors

Panel Structure	0.1			0.5			0.9		
	B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
N=25 T=5 n=5		0.2463132	0.2410073		0.2796962	0.284773		0.5495509	0.5481562
N=50 T=10 n=5		20.92463	24.85846		21.65518	25.32523		27.47136	30.29188
N=100 T=10 n=10		25.04934	29.991		25.38754	30.21844		27.78834	32.35012
N=450 T=30 n=15		24.95093	29.93912		24.9827	29.92951		25.56858	30.43426
N=1800 T=60 n=30		24.99712	30.00744		25.02376	30.01473		25.13534	30.12823
N=3600 T=60 n=60		25.0013	30.00505		25.01522	30.00357		25.06522	30.05607

The Table 5. and Table 6. provided displays the performance of the Fixed Effect Model (FE**)** under various levels of collinearity (0.1, 0.5, and 0.9) and different panel structures (sample size *N* and time periods *T*). The results focus on the estimates of the coefficients β_0, and B_2 under these condition

Low Collinearity **(0.1)**: The coefficient estimates for β_0, β_1 and β_2, and β_2 remain relatively stable across different panel structures under low collinearity. For example, the values for β_0, β_0, and β_2 under the low collinearity condition at N=25 and T=5 are 0.2417, 0.2447, and 0.2789, respectively. These values do not exhibit extreme variation, indicating that low collinearity does not severely disrupt the estimates of the coefficients in the fixed effect model. **(0.5)**: With moderate collinearity, we observe a slight increase in the variability of the coefficient estimates. For example, the coefficient B1 at N=50 and T=10 is 24.7016 under collinearity 0.1, but increases to 25 under collinearity 0.5. This suggests that moderate collinearity starts influencing the model's stability, though the impact is still relatively modest. **(0.9)**: At high collinearity, the coefficients exhibit greater fluctuations. For instance, β_1 at N=100 and T=10 is 29.9833 under low collinearity, but under high collinearity, it increases to 32.3501. This shows that high collinearity significantly increases the variance of the coefficient estimates, which is a common issue in regression models when predictors are highly correlated (Kennedy, 2003).

Small Panel Structures **(N=25, T=5)**: In smaller panel structures, the coefficient estimates are more volatile. For example, under high collinearity (0.9), the coefficients, β_0, β_1 , and β_2 at N=25 and T=5 show significant variation, which reflects the instability of estimates when there are fewer observations to balance out the noise caused by collinearity.

Larger Panel Structures **(N=450, T=30 or N=1800, T=60)**: As the sample size and time periods increase, the coefficient estimates become more stable and closer to the true values. For example, at N=450 and T=30, the coefficient estimates for β_0, and β_2 under high collinearity (0.9) are 24.9509, 29.9391, and 30.4343, respectively. The estimates for larger panel structures (e.g., N=3600, T=60) show even more stability.

Convergence in Larger Samples: The coefficient estimates appear to converge as *N* and *T* increase. For instance, when N=3600 and T=60, the estimates for β_0, β_1, and β_0, β_1 under high collinearity are very stable, indicating that large datasets help to counteract the effects of collinearity on model performance.

Science World Journal Vol. 19(No 4) 2024 *https://dx.doi.org/10.4314/swj.v19i4.49*
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Table 7. Summary of Fixed Effect Model (FE), **Random Effects Model (RE)**, and **Pooled Regression (PR)** to determine which model performs best under varying levels of collinearity (0.1, 0.5, and 0.9). By examining the coefficients and comparing the models, we can identify the most reliable approach for handling panel data under different conditions of collinearity and panel structure (sample size *N* and time periods *T*).

Panel Structure	Model	Collinearity 0.1	Collinearity 0.5	Collinearity 0.9
N=25, T=5	**FE**	0.2417, 0.2447, 0.2071	0.2517, 0.2806, 0.2513	0.2079, 0.5578
	RE	0.2447, 0.2495, 0.2806	0.2797, 0.2847, 0.5495	0.5482, 0.5038
	Pooled	0.2088, 0.2203, 0.2191	0.1442, 0.1466, 0.1482	0.1445, 0.1691, 0.1693
N=100, T=10	**FE**	25.0631, 29.9833, 29.9096	24.8986, 29.8762, 29.8347	24.9349, 29.8437
	RE	25.1658, 29.9626, 30.2267	25.3723, 30.2184, 30.2267	25.3345, 29.8802
	Pooled	0.1013, 0.1022, 0.1029	0.1018, 0.1185, 0.1181	0.1015, 0.1174
N=450, T=30	**FE**	24.9309, 29.9023, 29.8556	24.9171, 29.8668, 29.8395	24.9119, 29.8812
	RE	24.9262, 29.8560, 29.9200	24.9676, 29.9163, 29.9542	25.5686, 30.4343
	Pooled	0.0468, 0.0474, 0.0477	0.0467, 0.0549, 0.0553	0.0469, 0.1097
N=1800, T=60	**FE**	24.9865, 30.0024, 29.9990	24.9963, 29.9955, 29.9921	24.9944, 29.9921
	RE	25.0089, 30.0017, 30.0177	25.0275, 30.0177, 30.0177	25.1353, 30.1282
	Pooled	0.0239, 0.0239, 0.0238	0.0233, 0.0272, 0.0272	0.0239, 0.0543
N=3600, T=60	**FE**	24.9940, 29.9976, 30.0165	24.9944, 29.9989, 29.9947	25.0020, 29.9947
	RE	**24.9984, 30.0165, 30.0020**	**25.0098, 30.0050, 30.0046**	**25.0652, 30.0561**
	Pooled	0.0167, 0.0168, 0.0166	0.0166, 0.0192, 0.0192	0.0169, 0.0387

As expected, the Fixed Effects Model provides stable coefficient estimates for large panel structures (e.g., N=1800, T=60, T=60 and N=3600, T=60), but it suffers from higher variability at smaller sample sizes (N=25, T=5). This model is most effective when controlling for unit-specific characteristics, but its performance decreases with smaller panels and higher collinearity (Greene, 2012). The Random Effects Model consistently provides stable coefficient estimates across various panel sizes, though it becomes less effective in smaller samples (e.g., N=25, T=5) under higher collinearity. It generally performs well with larger datasets and shows robustness under moderate collinearity (0.5) (Wooldridge, 2010). The Pooled Regression Model provides relatively consistent estimates at smaller sample sizes but becomes less reliable under high collinearity. For instance, at N=50 and T=10, the coefficients show substantial fluctuations, indicating that pooling without accounting for fixed or random effects leads to inefficient and biased estimates when collinearity is high (Kennedy, 2003).

The Fixed Effects (FE) Model tends to produce the lowest RMSE in larger panel structures (e.g., N=1800, T=60 and N=3600, T=60), indicating that it is the most accurate model when dealing with unobserved heterogeneity in large datasets (Greene, 2012).

The Random Effects (RE) Model performs reasonably well, especially under moderate collinearity, but its accuracy diminishes when collinearity is high (0.9) in smaller panels.

The Pooled Regression Model shows the highest RMSE in most scenarios, especially when collinearity is high, highlighting its limitations when unobserved effects or panel-specific heterogeneity is significant (Kennedy, 2003)

Conclusion

Based on the analysis, the Fixed Effects (FE) Model is the best-performing model, particularly in larger datasets, as it minimizes both bias and RMSE. The Random Effects (RE) Model can also be effective, especially when collinearity is moderate and when the assumptions of random effects hold true. However, for datasets where collinearity is high, or where individual-specific effects are crucial, the Fixed Effects Model provides more reliable estimate.

The Pooled Regression Model should generally be avoided in cases where collinearity or panel-specific heterogeneity is significant, as it produces the least stable and least reliable results across different collinearity levels.

REFERENCES

Adenomon, M.O. and Oyejola, B.A. (2014). A simulation study of effects of collinearity on forecasting of bivariate time series data, Sch. J. Phys. Math. Stat., 1(1), 4-21.

Adenomon, M.O., Michael, V.A. and Evans, P.O. (2015). Short term forecasting performances of Classical and Sims-Zha Bayesian VAR models for time serieswith collinear variables and correlated error term, Open Journal of Statistics, 5, 742 - 753.

Adenomon, M.O., Michael, V.A. and Evans, P.O. (2016). On the Performances of Classical VAR and Sims-Zha Bayesian VAR Models in the Presence of Collinearity and Autocorrelated Error terms, Open Journal of Statistics, 6, 96-132. http://dx.doi.org/10.4236/ojs.2016.61012.

Ahn, S.C. and S. Low, (1996), A reformulation of the Hausman test for regression models with pooled cross-section time-series data, Journal of Econometrics 71, 309–319.

Ahn, Seung & Moon, Hyungsik. (2001). Large-N and Large-T Properties of Panel Data Estimators and the Hausman Test. SSRN Electronic Journal. 10.2139/ssrn.283252

Ajao, Kayode &Adehi, Mary &Adenomon, Monday. (2023). COMPARING SOME PANEL DATA ESTIMATORS IN THE PRESENCE OF AUTOCORRELATION. Science World Journal. 18. 398-403. 10.4314/swj.v18i3.12.

Alin, A. 2010. Multicollinearity. – WIREs Computational Statistics

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica, 59*(3), 817–858. https://doi.org/10.2307/2938229

Andrews, D. W. K., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica, 60*(4), 953–966.

Andrews, I. (2018). Weak instruments robust tests in linear IV models. *Review of Economics and Statistics, 100*(3), 508–518.

Arellano M (2003), Panel data econometrics. Oxford University Press, Oxford.

Arellano, M. (2003). *Panel data econometrics*. Oxford University Press.

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*(2), 277–297. https://doi.org/10.2307/2297968

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68*(1), 29–51.

Arum, Kingsley &Ugwuowo, Fidelis &Oranye, Ebele & Alakija, T.O. & Tobias, Ugah& Asogwa, Oluchukwu. (2023). Multicollinearity in Linear Regression Model using Robust Kibria-Lukman mixed with Principal Component Estimator, Simulation and Computation. Scientific African. 19.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica, 77*(4), 1229–1279.

Bai, J., & Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica, 72*(4), 1127–1177.

Baltagi, B. and S. Khanti-Akom, (1990), On efficient estimation with panel data: An empirical comparison of instrumental variables estimators, Journal of Applied Econometrics, 5, 401-406. Baltagi, B. H. (2005), Econometrics analysis of panel data, 3rd edition, John Wiley and Sons Ltd, England.

Baltagi, B. H. (2008). *Econometric analysis of panel data* (4th ed.). Wiley.

Baltagi, B. H., & Griffin, J. M. (1984). Short and long run effects in pooled models. *International Economic Review, 25*(3), 631–645.

Baltagi, B. H., & Levin, D. (1986). Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation, and advertising reconsidered. *Review of Economics and Statistics, 68*(1), 148–155. https://doi.org/10.2307/1924930

Baltagi, B. H., Feng, Q., & Kao, C. (2012). A Lagrange multiplier test for cross-sectional dependence in a fixed effects panel data model. *Journal of Econometrics, 170*(1), 164–177.

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review, 89*(3), 634–647. https://doi.org/10.2307/2082979

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115–143.

Bollen KA, Brand JE (2008), Fixed and Random Effects in Panel Data Using Structural Equations Models.

Bond, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal, 1*(2), 141–162.

Cameron, A. C., & Trivedi, P. K. (1998). Regression analysis of count data. *Econometric Society Monographs, 30*.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics, 18*(1), 5–46.

Chowdhury, M., & Saha, S. (2019). Multicollinearity effects on panel data regression models: A Monte Carlo simulation study. *Statistical Papers, 60*(3), 1017–1039. https://doi.org/10.1007/s00362-019-01143-5

Cornwell, C. and P. Rupert, (1988), Efficient estimation with panel data: An empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics 3*, 149-155.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics, 30*(1–2), 109–126.

Dormann, Carsten & Elith, Jane & Bacher, Sven & Buchmann, Carsten & Carl, Gudrun & Carré, Gabriel & Diekötter, T. & García Márquez, Jaime & Gruber, Bernd & Lafourcade, Bruno & Leitão, Pedro & Münkemüller, Tamara & Mcclean, Colin & Osborne, Patrick & Reineking, Björn & Schröder, Boris & Skidmore, Andrew & Zurell, Damaris & Lautenbach, Sven. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography. 36. 27-46.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics, 80*(4), 549–560. https://doi.org/10.1162/003465398557825

Ferreira, M.A., & Vilela, A.S. (2004). Why do firms hold cash? Evidence from EMU countries. European Financial Management, 10 (2), 295–319.

Fuller, W. A. (1987). *Measurement error models*. Wiley.

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson.

Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics, 31*(1), 93–118.

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics, 93*(2), 345–368.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50*(4), 1029–1054.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics, 20(7), 873–889.

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.

Huber, P. J. (2011). *Robust statistics*. Wiley.

Im, K. S., Pesaran, M. H., & Shin, Y. (1997). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*(1), 53–74.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*(1), 53–74. https://doi.org/10.1016/S0304-4076(03)00092-7

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Kennedy, P. (2008). *A guide to econometrics* (6th ed.). Wiley-Blackwell.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33–50.

https://doi.org/10.2307/1913643

Lillard, L.A. and Wallis, R.J. (1978). Dynamic aspects of earning mobility, Econometrica 46, 985-1012

Maddala, G. S., & Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics, 61*(S1), 631–652.

Mohammed K. G, Benjamin A. O and Waheed B. Y. (2013): Investigations of Certain Estimators for Modeling Panel Data Under Violations of Some Basic Assumptions Mathematical Theory and Modeling www.iiste.org ISSN 2224-5804, ISSN 2225- 0522 (Online).Vol.3, No.10

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics, 4*, 2111–2245.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*(6), 1417–1426. https://doi.org/10.2307/1911408

Nickell, S. (1982). The bias in estimates of dynamic panel models with fixed effects. *Econometrica, 49*(6), 1417–1426.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica, 74*(4), 967–1012.

Pesaran, M. H., & Smith, R. P. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics, 68*(1), 79–113.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics, 16*(3), 289–326.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal, 9*(1), 86–136.

Schmidheiny, K., & Siegloch, S. (2019). On event study designs and distributed-lag models: Equivalence, generalization, and practical implications. *Political Analysis, 27*(3), 404–413.

Shrestha, Noora. (2020). Detecting Multicollinearity in Regression Analysis. American Journal of Applied Mathematics and Statistics. 8. 39-42. 10.12691/ajams-8-2-1.

Stewart, G. W. 1987. Collinearity and least squares regression.

Stock, J. H., & Watson, M. W. (2015). *Introduction to econometrics* (4th ed.). Pearson.

Tan, C. M., & Lau, L. J. (2009). Robustness of fixed-effects and random-effects estimators for panel data. *Journal of Econometrics, 152*(2), 96–114.

Tobechukwu N. M and Azubuike M.I (2020): Panel Data Estimators in The Presence of Serial and Spatial Correlation with Panel Heteroscedasticity: A Simulation Study *Quarterly Journal of Econometrics Research Vol. 6*, No. 1, pp. 1-11.

Velilla, Santiago. (2018). A Note on Collinearity Diagnostics and Centering. *The American Statistician. 72*.

Verbeek, M. (2017). *A guide to modern econometrics* (5th ed.). Wiley.

Wallace, T.D. and A. Hussain, (1969), The use of error components models in combining crosssection and time-series data, Econometrica 37, 55-72.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Wooldridge, Jeffrey. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators.

Youssef, Ahmed &Abonazel, Mohamed & Ahmed, Elsayed. (2021). The performance of count panel data estimators: a simulation study and application to patents in Arab countries. 10.28919/jmcs/5852.

Youssef, Ahmed Hassen and Mohamed, Engy Saeed and Latif, Shereen Hamdy Abdel, Handling Multi-Collinearity Using Principal Component Analysis with the Panel Data Model (January 19, 2023). EUREKA: Physics and Engineering, (1), 177-188, 2023;

Zulfikar, Rizka. (2018). Estimation Model And Selection Method Of Panel Data Regression: An Overview Of Common Effect, Fixed Effect, And Random Effect Model.

Zubair, Mohammed & Adenomon, Monday. (2021). Comparison of Estimators Efficiency for Linear Regressions with Joint Presence of Autocorrelation and Multicollinearity COMPARISON OF ESTIMATORS EFFICIENCY FOR LINEAR REGRESSIONS WITH JOINT PRESENCE OF AUTOCORRELATION AND MULTICOLLINEARITY. Science World Journal. 16. 2021.