International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I6P107

Volume 72 Issue 6, 50-58, June 2024

Original Article

Comparative Analysis of Panel Data Regression Models on Nigeria Money Deposit Bank Dataset

Oludele Awodele[1], Afolashade Ayankoya[2], Emmanuel Ogu[3], Rotimi Olugbohungbe[4]

[1]*Computer Science Department, Babcock University, Ogun State, Nigeria.*
[2] *Information Technology Department, Babcock University, Ogun State, Nigeria.*

[4]*Corresponding Author : rotimiolugbohungbe@gmail.com*

Received: 22 April 2024 Revised: 26 May 2024 Accepted: 06 June 2024 Published: 15 June 2024

Abstract - *Panel data regression models have gained significant attention in empirical research due to their ability to capture both cross-sectional and time-series variations. This study conducts a comparative analysis of panel data regression models using a dataset from Nigeria's Money Deposit Banks. The research focuses on examining the performance of these models in estimating the relationship between key financial indicators and bank profitability. The dataset spans from 2001 to 2020, providing a comprehensive view of the banks' financial status using the Return on Asset (RoA). The study employs two-panel data regression models: Fixed Effects and Random Effects models. The models are compared based on their goodness-of-fit metrics' values, as measured by the Adjusted R-squared, F-statistic, Log-likelihood and AIC. The study also considers the significance and direction of the coefficients of the independent variables. Preliminary results suggest that the Fixed Effects model outperforms the Random effects model with 0.71, 9.66, 97.15, and -148.3 values for Adjusted R-squared, F-statistic, Log-likelihood and AIC metrics, respectively.*

Keywords - *Panel data regression models, Bank profitability, Nigeria money deposit banks, Comparative analysis, Fixed effects model.*

1. Introduction

Panel data regression models have become vital tools in empirical research, notably in economics and finance, since they can handle both cross-sectional and time-series changes at the same time [1]. In banking research, these models provide strong analytical frameworks for evaluating numerous aspects of bank performance, such as profitability, risk management, and regulatory compliance [2]. With Nigeria's banking system rapidly evolving in the face of dynamic market circumstances and regulatory changes, stakeholders must grasp the factors that influence Money Deposit Bank (MDB) performance [3].

Existing research on the Nigerian banking industry frequently uses traditional time series or cross-sectional analysis, which may not completely capture the dynamic interactions between variables. Studies using panel data for the Nigerian banking industry may focus on certain models without thoroughly assessing the performance of various models. This study fills a knowledge gap by comparing two-panel data regression models used on a panel dataset of Nigerian money deposit institutions. The study assesses the merits and shortcomings of two distinct models for describing the links between major variables impacting Nigeria's banking system.

Despite the proliferation of research using panel data regression in money deposit banks, detailed comparison analyses are required to determine the best appropriate modelling method between the Random effects and Fixed effects models in the Nigerian Money Deposit Banks setting. This study is an extract from a PhD thesis, which validates the intention of the researcher in the adoption of panel data regression models. Therefore, the data preprocessing algorithm analysis is not within the scope of this paper.

The purpose of this study is to compare the empirical outputs of the Random effects and Fixed effects models using the Hausman test to determine their certainty. To achieve this goal, the panel dataset from ten Nigerian money deposit banks was studied using a dataset spanning 2001 to 2020, offering a full assessment of the institutions' financial situation.

2. Panel Data Description

Repeated data collection from the same population necessitates a cross-sectional examination of the influence of factor variables on outcome variables. Panel data analysis arose as a result of the advancement of econometric modelling approaches, sophisticated statistical methodologies, and computer data processing technologies [4]. The first uses of this sort of data may be found in longitudinal research on societal problems [5]. The increased interest in understanding macroeconomic events, as well as the great availability of data for specific samples, have significantly contributed to the continued use of panel data analysis in the study of macroeconomic indicator dynamics. Assessing an organization's performance, particularly in the banking sector, is critical to shareholder decision-making because shareholders base their decisions to invest available cash in the assets of the firm on performance criteria [6].

The return on equity is reflected in the assessment of corporate performance based on the degree of shareholder equity [7]. Return on Equity (ROE) varies from one organization to organization; it varies from one financial record to another within the same organization, and to be able to determine such dynamics require the use of panel data to assess over time the effects of determinant factors on return on equity.

2.1. Conceptual Model of Panel Data

The conceptual model of panel data describes the methodology flow used in panel data analysis. Figure 1 shows a diagrammatic representation of the existing general framework used for panel data analysis, which was suggested and used in [8].

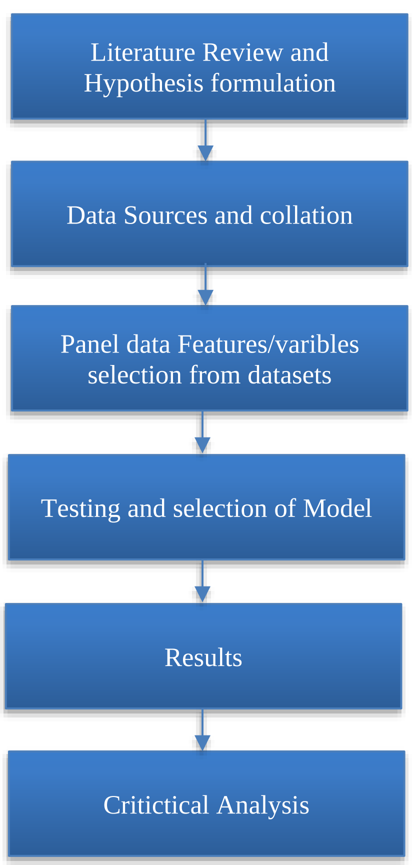

Fig. 1 Panel data analysis conceptual framework [8]

2.2. Groundwork for Panel Data

The panel data are sometimes referred to as pooled data or longitudinal data. A panel dataset is a collection of cross-section data Y_{nt} (n = 1, ..., N and t = 1, ..., T) produced from statistical observation of characteristics typical of a group of N persons done on a regular basis during a defined time interval T [9]. A panel dataset has a variation in observations for the same persons throughout time, resulting in the recording of N.T observations. According to this representation, statistical observations exhibit a variety of individual traits, which contributes to an increase in the variability of observations and accuracy of estimation.

2.3. Characteristics of Panel Data

Panel data, according to [10], is the result of many recordings of the same persons in a specified sample during a specific time period. Even if the criterion for random selection is quite stringent in the observed sample, eventually, correlations between markers describing people over time can be formed.

A. If panel data is analyzed by the size of the sample, then;
   i. Balanced panel data: This is obtained when an individual data point is observed over equal periods of time
   ii. Unbalanced panel data: This is obtained when individual data points are observed over different periods of time.
B. If panel data is analyzed by the selection methods of individual data points, then;
   i. Continuous panel data: The chosen individuals for the sample do not vary during the recording of observations.
   ii. Rotative panel data: When a population is observed over a set amount of time, they may be removed from the sample and replaced with other people for whom fresh observations will be collected.

The features of panel data are represented using the function in Equation (1)

$$Y_{nt}(n = 1, \dots, N \text{ and } t = 1, \dots, T) \quad (1)$$

For statistical observation using panel data, it defines three perspectives: person N, time T, and variable Y. Y_{nt} is the observed variable Y for individual N at instant T, according to these notations. Panel data sets, as indicated in equation 1, are distinguished by their double-dimensional representation, temporal and transverse, giving them a major advantage over other forms of data.

The temporal dimension tracks an individual's progress across time in relation to the factors under consideration. This dimension specifies the statistical recording of data from each observed statistical unit as a time series. On this dimension, the breakdown of total variability in each recorded observation should take into account the number of research periods. The entire variance in this example may be broken down as shown in equations (2 – 7);

$$Total\ variance = Intertemporal\ variance + Intratemporal\ variance \quad (2)$$

Which is represented mathematically in equation 2.6.

$$\sum_{n=1}^{N}\sum_{t=1}^{T}(y_{nt} - y_{..})^2 = N\ \sum_{t=1}^{T}(y_{.t} - y_{..})^2 + \sum_{n=1}^{N}\sum_{t=1}^{T}(y_{nt} - y_{.t})^2 \quad (3)$$

Transversal dimension examines the variation of characteristics from one individual to another regardless of the time t during which data have been collected, and total variance can be decomposed.

$$Total\ variance = Inter - individual\ variance + Intra - individual\ variance \quad (4)$$

Which is represented mathematically in Equation (5)

$$\sum_{n=1}^{N}\sum_{t=1}^{T}(y_{nt} - y_{..})^2 = N\ \sum_{n=1}^{N}(y_{n.} - y_{..})^2 + \sum_{n=1}^{N}\sum_{t=1}^{T}(y_{nt} - y_{n.})^2 \quad (5)$$

The whole variance of recorded observations can be dissected by active combining of the two dimensions as shown in Equation (6),

$$Total\ variance = Inter - individual\ variance + Inter - temporal\ variance + Intra - individual - temporal\ variance \quad (6)$$

The mathematical form of Equation (6) is shown in Equation (7).

$$\sum_{n=1}^{N}\sum_{t=1}^{T}(y_{nt} - y_{..})^2 = N\ \sum_{t=1}^{T}(y_{.t} - y_{..})^2 + T\ \sum_{n=1}^{N}(y_{n.} - y_{..})^2 + \sum_{n=1}^{N}\sum_{t=1}^{T}(y_{nt} - y_{n.} - y_{.t} - y_{..})^2 \quad (7)$$

The main difference between the last breakdown and the first two is that it takes into account both intra-temporal and intra-individual variances. The fundamental advantage of investigating people's behaviour from the perspective of the person and the temporal dimensions is the breakdown technique of total variance, as in the previous model.

2.4. Panel Analysis Models

To carry out panel data analysis, a general model given in [8] is usually the starting model. This model is adjusted to suit specific panel data analysis tasks. This model is frequently built using a set of data collected for N individuals over a T-year period. Given this data, the following generic model for the analysis of a resultative variable (Y) by determinant factors (X_k) may be written as suggested in Equation (8) [8];

$$y_{nt} = b_{0nt} + \sum_{k=1}^{K} b_{knt} + x_{knt} + w_{nt} \quad (8)$$

n = 1, ..., N and t = 1, ..., T, where y_{nt} represents dependant variable values, x_{knt} represents independent variable values K, b_{0nt} is a constant, and w_{nt} is the error component.

b_{0nt} and b_{knt} coefficients, k = 1, ..., K vary over time and across people. Because individuals' behaviour towards dependent variables of the researched sample may change over time, there may be a lack of recorded data homogeneity in the examined sample. It is difficult to estimate the model using standard approaches since the number of coefficients (NT(K + 1)) is more than the total number of data (NT). Contrasts between coefficients should be employed in this scenario by defining two classical models: fixed effects models (individual or temporal) and compounded error models (random effects).

2.4.1. Random Effects Model

In fixed effects models, it is believed that the influence of the investigated factor variables (x_{knt}) on the dependent variable (y_{nt}) is the same for all people across the whole studied time ($b_{knt} = b_k$). In this scenario, the constant b_{0nt} can be decomposed as given in Equation (9);

$$b_{0nt} = b_0 + a_n + d_t \quad (9)$$

Where, b_{0nt} is the constant of the regression model, b_0 , a constant

a_n indicates unobservable differences between individuals and d_t temporal differences that may appear in individuals.

The regression model is given in Equation (10).

$$y_{nt} = b_0 + a_n + d_t + \sum_{k=1}^{K} b_k\, x_{knt} + w_{nt} \quad (10)$$

In estimating fixed effects model parameters, take into account individual and temporal specificity by incorporating unique effects, also known as fixed effects, in individuals and periods that indicate coefficients to be calculated. Two banks with the same observable properties should have the same values for the resultative variables in the case of a model for a certain period.

2.4.2. Fixed Effect Model

Generally, the fixed effect models are given, as shown in Equation (11). The unpredictable nature of specific effects distinguishes composite effect models from fixed effect models.

$$y_{nt} = b_{0nt} + \sum_{k=1}^{K} b_k\, x_{knt} + \varepsilon_{nt} \quad (11)$$

Where,

$$\varepsilon_{nt} = U_n + V_t + w_{nt} \quad (12)$$

If the individual (U_n) and temporal (V_t) effects are random, with a zero mean and variance, σ_u^2 and σ_v^2. The model is easily decomposed, with the error factor consisting of three elements: a component that does not exhibit autocorrelation (w_{nt}) either individually or temporally, a component as an individual-specific effect (U_n), and a component as a temporal-specific effect (V_t), which are not correlated with each other or with each other.

These features give a condition mean value y_{nt} as shown in Equation (13).

$$E(y_{nt}/x_{1nt}, \dots, x_{Knt}) = b_0 + \sum_{k=1}^{K} b_k x_{knt} \quad (13)$$

Individual effects in the random effect model represent unobservable human characteristics and are uncorrelated with dependent observable variables. F and Hausman tests are used to choose one of two types of models (with fixed or random effects).

2.5. Hausman Tests

The Hausman test is also referred to as a model misspecification test. The Hausman test can help to select between a fixed effects model and a random effects model in panel data analysis (data analysis over time) [11]. The null hypothesis states that the chosen model has random effects, whereas the alternative hypothesis states that the model has fixed effects. The tests essentially try to check if there is a relationship between the unique mistakes and the regressors in the model. The null hypothesis states that no association exists between the two. The interpretation of a Hausman test result is simple: if the p-value is tiny (less than 0.05), reject the null hypothesis.

2.6. Reviewed Works

In carrying out panel data analysis, [14] investigated the impact of bank-specific and macroeconomic major drivers on the profitability of Islamic retail banks in Bahrain. It utilized explanatory research with secondary financial data and utilized panel data from six Islamic retail banks from 2013 to 2019. The two primary profitability measurements employed in this study are Return on Assets (ROA) and Return on Equity (ROE). The two statistical models used in this study are the random effect regression model and the fixed effect regression model.

According to the random effect regression model, bank size is strongly favourably connected to ROA, but operational efficiency and GDP growth have significant and negative relationships with ROA. The fixed effect regression model reveals that credit risk, operating efficiency, and GDP growth rate have negative significant effects on banks' ROE. The study did not predict the future outcome values for ROE and ROA, and the basis for selecting a regression model was not established in the study.

In assessing the performance of the banks in Bangladesh through the efficiency of the banks, Uddin and Nezum [15] used a Z score methodology, where the Z score was also used as the dependent variable. The fraction of total operating expenditures and net operating income of banks is calculated. Other indices are regressed on the Z score. From 2015 to 2017, the panel data estimation approach is employed. In that respect, two models were developed, with model-1 including 6(six) explanatory factors and model-2 containing an extra independent variable in addition to the previous six variables. Two effect models (fixed effect and random effect) are also calculated in the panel data estimation way. The random effect is shown to be appropriate for both models. According to the random effect, [15] discovered that three factors had a negative influence on the Z score in model-1, whereas three variables had a favourable impact. In model-2, three indices have a negative influence on the Z score, whereas four indices have a favourable effect. Surprisingly, the efficiency ratio has a significantly favourable influence on the GDP growth rate in both models. Panel data analysis was used in [16] to examine the condition and dynamics of the financial performance of firms listed on the Bucharest Stock Exchange in relation to determinant variables. The study presented a theoretical foundation as well as an applied panel data analysis of two case studies using fixed and random-effects models (investigating the influence of ROE of the previous period on ROE for the current period). The findings of the Hausman test are used to choose which of the two types of models to use. The outcome of the study in [16] justifies panel analysis of return on equity for monitoring the performance measures of financial organizations.

Other than assessing the performance of banks, panel data analysis was used to discover the profitability factors of the banking sector in [12], where the impact of the determinants of profitability on the commercial banks in Asian countries was the objective study in [12]. To evaluate the data, the panel data research approach was employed as an estimating tool. Data was also examined using the ordinary least squares (OLS) regression model. The Breusch-Pagan Lagrange Multiplier (LM) Test was used to determine whether the models were adequate. Bank-specific and microeconomic parameters varied virtually identically for private banks in Bangladesh and India.

E-views econometric software was used to analyze all models and tests. The study discovers that the Return on Asset (ROA) from the banks' individual factors, Bank Size (BS), and Debt to Asset Ratio (DAR), are positive and substantial. The Deposit to Asset Ratio (DTAR) and Loan to Deposit Ratio (LDR) for banks are shown to be negative and substantial. The Equity to Asset (EAR) and Debt to Equity (DER) ratios have neither positive nor negative influence. But, the study did not apply predictive analytics for future determination. Also, panel data analysis has been used to determine the dividends and performance of firms in India.

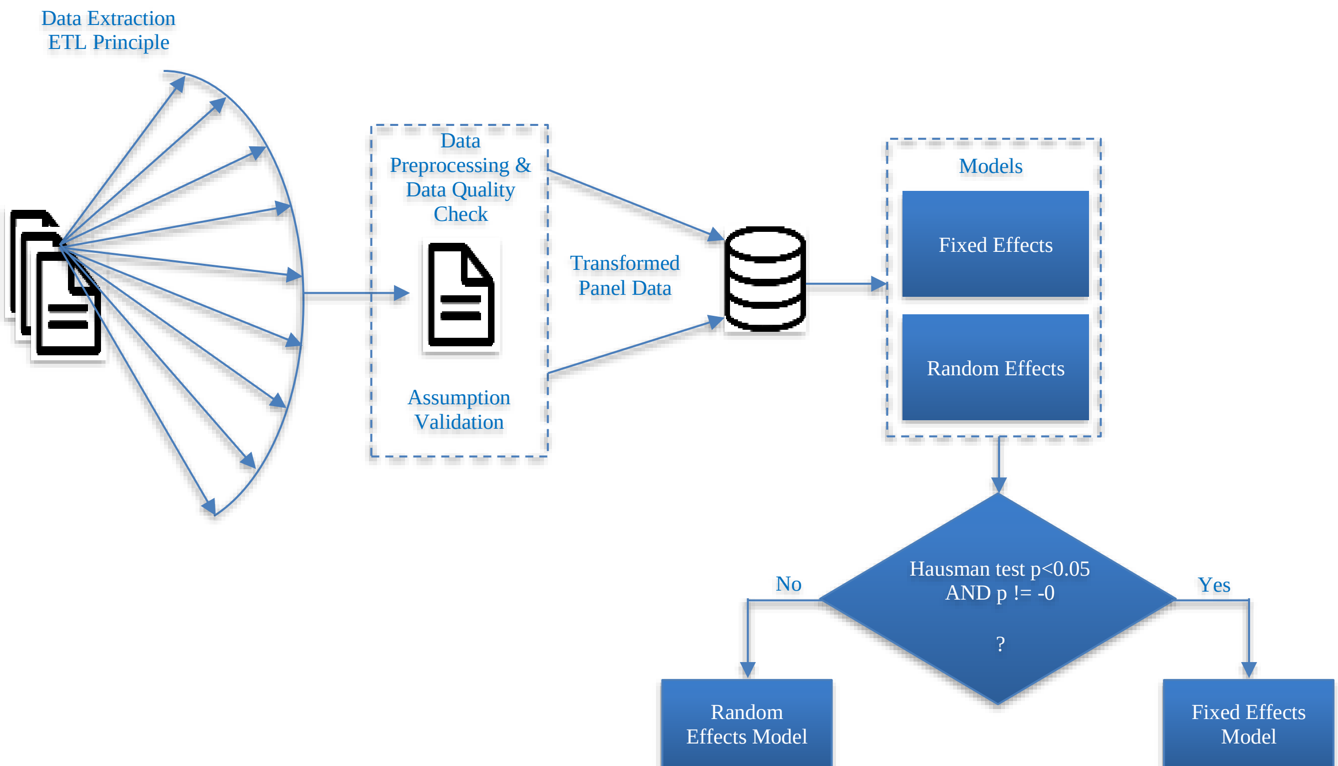

Fig. 2 The general process model

3. Methodology

Figure 2 shows the general flow of the developed model. The sourced data was obtained from different sources. The data was cleaned and preprocessed, and the quality check was done to ascertain that the features obtained were complete with no irregularities in the dataset.

The dataset was transformed to panel data having 19 features from 10 Nigerian banks (Union, Sterling, Wema, Access, GT, Zenith, Fidelity, First, FCMB, and UBA banks) over a period of 22 years. The panel data was analyzed using two-panel data analysis models: the random effects and the fixed effects models. The best-performing model was selected based on the Hausman Test, with p values tending more to zero (0).

4. Result and Discussion

The focus is to examine how the Fixed effects and the Random effects models performed in the face of the panel dataset. Although the predicting and predicted factors were not of major concern, the performance of the two models. The Return on Asset (RoA) was used as the performance indicator for the panel dataset.

Figure 3 shows the output of the fixed effects model on the panel dataset. Figure 4 shows the output of the random effects model on the panel data. In developing the two regression models, the least squares dummy variable method was used so as to factor in the variance in the unobserved variables. In doing this, the bank column containing eight (8) variables was converted to 8 dummy variables, and one of these dummy variables was left out so as to avoid perfect Multicollinearity between the 8 dummy variables. The regression model's intercept will hold the value of the coefficient for the left-out dummy variable.

The result of the regression analysis for the fixed effects model showed values for the effects of endogenous and exogenous elements on bank performance (ROA). The coefficient value for the Debt Burden Ratio (DBR) is statistically significant at $p < 0.05$ with a p-value = 0.025. Also, the bank-specific effects FCMB, FirstBank, GTB, and Wema are statistically significant at a $p < 0.01$ with p-values = 0.000 each.

We may choose to qualify Board Oversight Function (BOF) to be statistically significant at $p < 0.05$. This can be attributed to the unobserved behaviour of the variable within the model. The remaining predictors are all statistically insignificant at p-value < 0.05, as indicated in the $p > |t|$ column.

The outcome of the random effects model, as depicted in Figure 3, only indicates that only Debt Burden Ratio (DBR) coefficient is statistically significant at $p < 0.05$ with the p-value = 0.022.

```
================================================================================
============================== Fixed Effects Model ==============================
                            OLS Regression Results
==============================================================================
Dep. Variable:                    roa   R-squared:                       0.669
Model:                            OLS   Adj. R-squared:                  0.600
Method:                 Least Squares   F-statistic:                     9.663
Date:                Mon, 20 Mar 2023   Prob (F-statistic):           1.35e-16
Time:                        02:00:01   Log-Likelihood:                 97.156
No. Observations:                 128   AIC:                            -148.3
Df Residuals:                     105   BIC:                            -82.72
Df Model:                          22
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.3810      0.262     -1.454      0.149     -0.901       0.139
ldr            0.0788      0.091      0.863      0.390     -0.102       0.260
dbr            0.1980      0.087      2.282      0.025      0.026       0.370
cr            -0.0493      0.128     -0.384      0.702     -0.303       0.205
ocf            0.2702      0.149      1.817      0.072     -0.025       0.565
bg             0.1633      0.086      1.896      0.061     -0.008       0.334
car            0.1952      0.173      1.129      0.262     -0.148       0.538
ecs            0.0557      0.155      0.359      0.720     -0.251       0.363
crm           -0.1241      0.164     -0.758      0.450     -0.449       0.201
ls             0.1206      0.094      1.285      0.202     -0.065       0.307
cas            0.0611      0.088      0.694      0.489     -0.114       0.236
aq             0.0139      0.159      0.088      0.930     -0.301       0.329
bof            0.1920      0.100      1.915      0.058     -0.007       0.391
ta             0.0944      0.097      0.971      0.334     -0.098       0.287
infl          -0.0404      0.068     -0.598      0.551     -0.174       0.094
gdp           -0.0960      0.057     -1.684      0.095     -0.209       0.017
Acess          0.0802      0.051      1.566      0.120     -0.021       0.182
FCMB           0.4842      0.061      7.935      0.000      0.363       0.605
Fidelity       0.1760      0.054      3.280      0.001      0.070       0.282
FirstBank      0.2729      0.066      4.118      0.000      0.142       0.404
GTB            0.1811      0.050      3.610      0.000      0.082       0.281
UBA            0.1108      0.061      1.818      0.072     -0.010       0.232
Wema           0.4407      0.057      7.736      0.000      0.328       0.554
==============================================================================
Omnibus:                        5.295   Durbin-Watson:                   1.241
Prob(Omnibus):                  0.071   Jarque-Bera (JB):                4.891
Skew:                           0.376   Prob(JB):                       0.0867
Kurtosis:                       3.593   Cond. No.                         62.4
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
sigma2_epsilon = 0.01579114947328158
sigma2_pooled = 0.03569319800343661
sigma2_u = 0.019902048530155026
theta = 0.7826357564864186
```

Fig. 3 The outcome of the fixed effects model

```
================================================================================
================================== Random Effects Model ==================================
                            OLS Regression Results
==============================================================================
Dep. Variable:                    roa   R-squared:                       0.280
Model:                            OLS   Adj. R-squared:                  0.184
Method:                 Least Squares   F-statistic:                     2.909
Date:                Mon, 20 Mar 2023   Prob (F-statistic):           0.000662
Time:                        02:04:42   Log-Likelihood:                 91.524
No. Observations:                 128   AIC:                            -151.0
Df Residuals:                     112   BIC:                            -105.4
Df Model:                          15
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1031      0.253     -0.408      0.684     -0.605       0.398
ldr            0.0753      0.090      0.835      0.406     -0.103       0.254
dbr            0.2031      0.087      2.329      0.022      0.030       0.376
cr            -0.0301      0.129     -0.233      0.816     -0.286       0.226
ocf            0.2580      0.150      1.719      0.088     -0.039       0.556
bg             0.1586      0.087      1.823      0.071     -0.014       0.331
car            0.1437      0.173      0.830      0.408     -0.199       0.486
ecs            0.0630      0.156      0.403      0.688     -0.247       0.373
crm           -0.1344      0.163     -0.822      0.413     -0.458       0.189
ls             0.1153      0.093      1.239      0.218     -0.069       0.300
cas            0.0479      0.089      0.541      0.589     -0.127       0.223
aq             0.0051      0.160      0.032      0.974     -0.312       0.322
bof            0.1724      0.098      1.765      0.080     -0.021       0.366
ta             0.0666      0.097      0.684      0.495     -0.126       0.260
infl          -0.0316      0.068     -0.465      0.642     -0.166       0.103
gdp           -0.1002      0.057     -1.743      0.084     -0.214       0.014
==============================================================================
Omnibus:                       14.159   Durbin-Watson:                   1.068
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               16.836
Skew:                           0.677   Prob(JB):                     0.000221
Kurtosis:                       4.151   Cond. No.                         13.4
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Fig. 4 The outcome of the random effects model

This becomes a challenge in deciding on the model to choose. Therefore, there is a need to choose a better model between the Random effects and Fixed effects models that actually model the choice of panel data accurately. Therefore, the study adopted the Hausman Test to help make a decision on a better model through the hypothesis testing as shown;

H_0: The Random Effects model is the preferred model for panel data analysis
H_1: The fixed effects Model is the preferred model.

In determining the most appropriate estimating approach, the study carried out a Hausman test to decide whether to use the Fixed Effect (FE) or Random Effect (RE) Model. The outcome of the Hausman test is shown in Figure 5. Judging by the p-value of 4.4310667577614615e-17, which is very close to 0.00, we reject the null hypothesis (random effect) and accept the use of the fixed effect analysis as there is a correlation between the unique errors and the regressors in the model.

```
Hausman Test

chi-Squared: 24.452667918156116
degrees of freedom: 16
p-Value: 4.4310667577614615e-17
```

Fig. 5 Hausman result

Table 1. Co-efficient values for the predicting variables

Ratio	Full Name	Co-efficient
Intercept		-0.3810
LDR	Loan Deposit Ratio	0.0788
DBR	Debt Burden Ratio	0.1980
CR	Current Ratio	-0.0493
OCF	Operating Cashflow	0.2702
BG	Bank Growth	0.1633
CAR	Capital Adequacy Ratio	0.1952
ECS	Equity Capital Structure	0.0557
CRM	Credit Risk Management	0.1244
LS	Loan Structure	0.1206
CAS	Cash Asset Structure	0.0611
AQ	Accrual Quality	0.0139
TA	Total Asset	0.0944
BOF	Board Oversight Function	0.1920
INFL	Inflation	-0.0404
GDP	Gross Domestic Product	-0.0960

The result of the Fixed effects model is thoroughly explained in this section. Table 1 shows the summary of the co-efficient values for the exogenous or predicting variable.

In Table 1, the coefficient values of the exogenous variables have a direct influence on the targeted variable, ROA. The positive values indicate that as the values increase, the effect on ROA also increases, unlike the negative value coefficients, which cause a reduction effect on ROA.

4.1. Goodness-of-fit Evaluation Metrics

Aside the coefficients and the p values, the goodness-of-fit of the Fixed Effects Model (FEM) needs to be analyzed. The goodness-of-fit measures how well the fixed effects model fits the given collection of panel data. This can be placed on the check by first looking at the following;

The Adjusted R-Squared value, which yielded 0.600 from the FEM results in Fig. 3, where;

R-Squared = 0.669
Adj. R-Squared = 0.600

After accounting for the degrees of freedom lost owing to the existence of regression variables, the adjusted R-squared reflects the proportion of the variation in the response variable ROA that the model was able to explain. The adjusted R-squared of 0.600 (or around 60%) indicates a good fit despite the unobserved variables.

The F-statistics for a fixed effects regression model analysis determines if all model coefficients are jointly significant and, hence, whether the FE model has a good goodness-of-fit. It can be observed that the F-test value of 9.663 is significant at $p<.001$, showing that the model's goodness-of-fit is correct. This is evident in the FEM result shown in Figure 4, where;

F-statistic = 9.663
Prob (F-statistic) = 1.35e-16

The Meaningless values (Log-likelihood and Akaike Information Criterion (AIC)) are the values used for measuring a model fit that may be used to compare different types of models (or variations on the same model). Higher values indicate a better model fit. The log-likelihood and AIC are accessible from the FEM results in Figure 3, where,

Log-likelihood = 97.156

4.2. Generated Panel Data Regression Equation

The coefficient of the variables obtained from this study can be fitted into the panel data model given in equation 8.

$$Y_{it} = \beta_0 + \beta_1 x1_{it} + \beta_2 x2_{it} + \beta_3 x3_{it} + \cdots + \beta_n xn_{it} * \varepsilon_{it} \quad (14)$$

Where Y_{it} is the dependent variable,
$x1 \dots . xn$ represent the number of predicting variables,
$\beta_0 \dots . . \beta_n$ are the coefficients of the predicting variables, and *it* is the change in the group being observed over time t.
By substituting these variables with the variables used in this study, equation 15 was obtained

$$ROA_{it} = \beta_0 + \beta_1\, ldr_{it} + \beta_2\, dbr_{it} + \beta_3\, ocf_{it} + \dots + \beta_n\, gdp_{it} * \varepsilon_{it} \quad (15)$$

Then, substitute the actual coefficient values from Table 1 to obtain equation 16. Equation 9 becomes the panel data regression model.

$$ROA_{it} = -0.3810 + (0.0788)ldr_{it} + (0.1980)dbr_{it} + (-0.0493)ocf_{it} + \dots + (-0.0960)gdp_{it} * \varepsilon_{it} \quad (16)$$

4.3. Fixed Effects Panel Data Model Benchmark Evaluation

The outcome of the panel data fixed effects model was benchmarked with other models by fitting the models on the panel data. The benchmark was measured using the goodness-of-fit metrics (Adjusted R-squared, F-statistic, Log-likelihood and AIC). The outcome, as evident in Table 2, shows that the panel data fixed effects model has the best good-of-fit values. The competing models are the pooled OLS panel data model used in [11] and the Fixed effects panel data model used in [13].

Table 2. Benchmarking the Fixed Effects model against Pooled OLS and Random Effects panel data Models

Models	Adj. R-squared (%)	F-statistic (%)	Log-likelihood	AIC
Random Effects	0.184	2.909	91.524	-151.0
Fixed Effects	0.600	0.033	97.156	-148.3

5. Summary

In summary, the above results imply that all the variables jointly affect ROA performance. Also, the value of the adjusted coefficient of multiple determination of 0.600 means that the combined changes in the exogenous and endogenous variables would cause 60% changes in ROA while the remaining changes of 40% per cent are caused by other factors which are not within the coverage of the model. These variables are referred to as the unobserved variables. The adjusted R-squared is a modified version of the R-squared that adjusts for the number of predictors in the regression model. The Hausman test with the Chi-square value of 24.45 and the p-value of 4.4310667577614615e-17, which is less than the 5% chosen significant level (p<.05), shows that the null hypothesis is therefore rejected. Accept the alternate hypothesis and conclude that the Fixed Effects model is suitable for the panel data analysis with exogenous and endogenous variables affecting bank ROA among deposit money banks in Nigeria.

References

[1] Guohua Feng, Jiti Gao, and Bin Peng, "An Integrated Panel Data Approach to Modelling Economic Growth," *Journal of Econometrics*, vol. 228, no. 2, pp. 379-397, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Andreas G. Koutoupis, and Theodore Malisiovas, "The Effects of the Internal Control System on the Risk, Profitability, and Compliance of the US Banking Sector: A Quantitative Approach," *International Journal of Finance & Economics*, vol. 28, no. 2, pp. 1638-1652, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Chris Humphrey, *Financing the Future: Multilateral Development Banks in the Changing World Order of the 21st Century*, Oxford University Press, 2022. [Google Scholar] [Publisher Link]

[4] Larissa Batrancea, Malar Mozhi Rathnaswamy, and Ioan Batrancea, "A Panel Data Analysis of Economic Growth Determinants in 34 African Countries," *Journal of Risk and Financial Management*, vol. 14, no. 6, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Christopher J. Hopwood, Wiebke Bleidorn, and Aidan G.C. Wright, "Connecting Theory to Methods in Longitudinal Research," *Perspectives on Psychological Science*, vol. 17, no. 3, pp. 884-894, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Udochukwu Chikaodili Nkemdilim et al., "Financial Statement Analysis and Shareholders' Investment Decisions of Deposit Money Banks in Nigeria," *SADI International Journal of Management and Accounting*, vol. 11, no. 1, pp. 11-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Leo Vashkor Dewri, "A Critical Assessment of Interrelationship Among Corporate Governance, Financial Performance, Refined Economic Value Added to Measure Firm Value and Return on Stock," *Journal of the Knowledge Economy*, vol. 13, no. 4, pp. 2718-2759, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Francesco Giordano, Marcella Niglio, and Maria Lucia Parrella, "Testing Spatial Dynamic Panel Data Models with Heterogeneous Spatial and Regression Coefficients," *Journal of Time Series Analysis*, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jonathan Kropko, and Robert Kubinec, "Interpretation and Identification of Within-Unit and Cross-Sectional Variation in Panel Data Models," *PloS One*, vol. 15, no. 4, pp. 1-22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Tobias Rüttenauer, and Volker Ludwig, "Fixed Effects Individual Slopes: Accounting and Testing for Heterogeneous Effects in Panel Data or Other Multilevel Models," *Sociological Methods & Research*, vol. 52, no. 1, pp. 43-84, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Badi H. Baltagi, "Hausman's Specification Test for Panel Data: Practical Tips," *Essays in Honor of Subal Kumbhakar*, vol. 46, pp. 13-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Deli Yuan et al., "Profitability Determining Factors of Banking Sector: Panel Data Analysis of Commercial Banks in South Asian Countries," *Frontiers in Psychology*, vol. 13, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Laura Liu, Hyungsik Roger Moon, and Frank Schorfheide, "Forecasting with Dynamic Panel Data Models," *Econometrica*, vol. 88, no. 1, pp. 171-201, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Aman Ullah, Tao Wang, and Weixin Yao, "Modal Regression for Fixed Effects Panel Data," *Empirical Economics*, vol. 60, pp. 261-308, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Md. Nezum Uddin, "Study on Bank Efficiency in Bangladesh: A Panel Data Analysis," *International Journal of Research and Innovation in Social Science*, vol. 3, no. 12, pp. 2454–6186, 2019. [Google Scholar] [Publisher Link]

[16] Elisabeta Jaba, Ioan-Bogdan Robu, and Christiana Brigitte Balan, "Panel Data Analysis Applied in Financial Performance Assessment," *Romanian Statistical Review*, vol. 65, no. 2, pp. 3–20, 2017. [Google Scholar] [Publisher Link]