International Journal of *Financial Studies*

MDPI

Type of the Paper (Article, Review, Communication, etc.)

Optimizing Nigerian Bank Lending Systems: The Power of Discrete Wavelet Transform (DWT) in Denoising and Regression Analysis

[1]Ogbonna, C. J., [2]Ohabuka, C.J., [3]Bartholomew, D.C.*, [4]Anyiam, K.E., [5]Adamu, I.

[1] Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria. ; chukwudi.ogbonna@futo.edu.ng
[2] Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria. ; ohabukaj@gmail.com
[3] Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria. ; desmond.bartholomew@futo.edu.ng
[4] Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria. ; kizito.anyiam@futo.edu.ng
[5] Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria. ; ibrahim.adamu@futo.edu.ng

* Correspondence: desmond.bartholomew@futo.edu.ng; Tel.: +2347033811698

Abstract: Wavelet-based methods for signal de-noising have garnered significant interest across various fields. This study examines the use of discrete wavelet transform (DWT) for de-noising in regression analysis, focusing on the Nigerian bank lending system. The main research question is how DWT affects the accuracy and reliability of regression models predicting bank lending performance. Data from the Central Bank of Nigeria Statistical Bulletin, spanning 1999 to 2022, was used. Initially, a multiple linear regression model was applied, revealing non-stationarity and noisy residuals. To mitigate this, the data was decomposed into the signal domain using DWT with Haar detail and smooth wavelet filters (-0.354 and 0.354). Noise levels were assessed, and the series was denoised and reconstructed using the universal threshold. A multiple linear regression was then conducted on the de-noised data. The results demonstrated that the model fitted to the de-noised data outperformed the original, as shown by improved metrics (RMSE, MAPE, Standard Error, and MAE). This study concludes that DWT is an effective de-noising tool, significantly enhancing the predictive accuracy of regression models in the context of the bank lending system.

Keywords: Discrete Wavelet Transform (DWT); Signal De-noising; Regression Analysis; Bank Lending System; Predictive Accuracy

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

The Nigerian banking sector comprises various types of banks, including commercial banks, merchant banks, and development banks, each offering different types of loans and credit facilities tailored to the diverse needs and financial capabilities of their customers (Agbo & Nwankwo, 2018). Among the most common types of loans provided by Nigerian banks are personal loans, which are unsecured loans given to individuals for various purposes such as paying for education, medical expenses, or home renovations (Kingwara et al., 2017). Additionally, commercial banks extend business loans to small and medium-sized enterprises (SMEs) to support their operational financing needs. Investment banks, on the other hand, offer specialized loans such as Acquisition financial loans, Bridge loans, and

Project finance loans, while development banks focus on infrastructure loans and agricultural loans (Diamond & Rajan, 2001).

Beyond traditional bank loans, Nigerian banks also provide other credit facilities including overdrafts, credit cards, and lines of credit. Overdrafts are short-term loans allowing customers to withdraw more money than is available in their accounts. Credit cards enable customers to make purchases on credit, and lines of credit offer businesses flexible access to funds as needed (Blanchflower & Evans, 2004). The Central Bank of Nigeria (CBN) regulates the lending system through various policies and guidelines to ensure stability and efficiency in the financial sector.

Bank lending is a critical component of the financial system, playing a vital role in the growth and development of any economy. However, the Nigerian banking system has been experiencing an increase in the number of Non-Performing Loans (NPLs), threatening the stability of the financial system (Olugbenga et al., 2017). In 2009, the Nigerian banking sector faced significant challenges due to inaccurate predictions and mismanagement of lending activities, causing some banks to experience liquidity issues and fail to meet their obligations to borrowers. This situation underscores the need for effective management of lending risk by banks. The lending system in Nigeria is influenced by various factors, including economic conditions, credit risks, lack of collateral, non-performing loans, and government policies, all of which contribute to the high level of NPLs in the country (Ugwuanyi et al., 2022).

Economic conditions such as inflation rates, GDP growth rates, high-interest rates, and exchange rates also significantly influence lending decisions. For instance, during periods of high inflation, banks may become more cautious in their lending activities, leading to a decrease in lending. Policies set by the Central Bank of Nigeria (CBN) related to these economic conditions also directly affect banks' lending capabilities by either tightening or loosening credit availability (Onoh & Nwachukwu, 2017). Furthermore, a high ratio of NPLs can lead banks to adopt more stringent lending criteria, making it difficult for individuals to access credit (Ugoani, 2016).

The success of lending activities in a country depends on the accurate measurement and effective management of these influencing factors (Olokoyo, 2011). One effective approach is modeling these factors using regression analysis techniques to examine their relationships with the lending system and predict the amount of loan disbursements. Regression analysis has long been used to identify the relationship between bank lending and its determinants. For example, it can predict the likelihood of borrower default or determine the factors influencing loan approval decisions. Two common types of regression analysis used in bank lending systems are Logistic regression and Multiple linear regression (Alsaawy et al., 2020). Logistic regression is used when the dependent variable is binary (e.g., default vs. non-default), while Multiple linear regression is used when the dependent variable (loan amount) is continuous and influenced by multiple explanatory variables (e.g., loan demand, interest rates, and deposit supply). Multiple linear regression helps banks identify which factors have the strongest relationship with bank lending, enabling more informed lending decisions (Rawlings et al., 2006).

Given that bank lending data is a financial time series, the accuracy of regression analysis can be significantly compromised by the presence of noise in the data. The Ordinary Least Squares (OLS) method, commonly used in regression analysis, is particularly susceptible to noise, resulting in biased estimates, reduced precision/model fit, overfitting, and misleading inferences, ultimately affecting the model's predictive capability (Chen et al., 2019; Saseendran et al., 2019).

Therefore, it is crucial to denoise financial time series data before using it to estimate regression models for accurate predictions. Over the years, researchers have developed many techniques to remove noise from data sets while retaining important underlying information. These techniques include Gaussian smoothing (Wink & Roerdink, 2004), Median filtering (Kumar & Sodhi, 2020), Wiener filtering (Kazubek, 2003), Wavelet denoising (Sardy et al., 2001), Non-local Means denoising (Dutta et al., 2013), Total variation denoising (Selesnick et al., 2014), and Deep learning-based denoising (Shah et al., 2021). The choice of denoising technique depends on the type and characteristics of the noise and the specific requirements of the application (Moosavi et al., 2018). For discrete time series data, Wavelet denoising techniques are particularly suitable because they effectively suppress noise while preserving important features by providing a multiresolution representation and localizing frequency and time information.

The Discrete Wavelet Transform (DWT) is a mathematical tool that decomposes a signal into different frequency sub-bands, making it useful for denoising. DWT involves three major stages: decomposition, denoising, and reconstruction. During decomposition, the time series is broken down into detailed and smooth components using various wavelet bases such as Haar (Haar, 1910), Daubechies (Daubechies, 1992), and Franklin (Murugan & Youvaraj, 2020). The denoising stage removes the non-significant components of the series using a threshold, while the reconstruction stage recovers the original series devoid of noise (Vidakovic, 1999). Wavelet transformation methods allow for the decomposition of a series without knowing the underlying functional form (Ramsey, 1996). DWT has been widely used in signal processing for denoising and feature extraction, and recently, it has been applied in finance for denoising financial time series data (Adewumi & Adama, 2017). DWT has also been applied as a new tool for trend analysis in hydrology in India (Srivastava et al., 2021), ocean color time series analysis (Moradi, 2022), Gaussian Process Regression (GPR) signals (Baili et al., 2006), streamflow and precipitation (Nalley et al., 2012), Wavelet-Transform-Based Machine-Learning Method (Quinones & Tibi, 2024), and as a multiresolution analysis for data cleaning in water quality management systems (He et al., 2008).

The literature on banking lending systems across various countries is extensive. For instance, Kolapo, Ayeni, and Oke (2012) investigated the quantitative effect of credit risk on the performance of commercial banks in Nigeria over an 11-year period (2000-2011) using a panel model approach. Their study employed traditional profit theory and formulated a multiple linear regression model for bank performance, concluding that the effect of credit risk is consistent across banks, although the specific impact on individual banks was not captured by their method of analysis. Similarly, Wu et al. (2020) examined the relationship between peer-to-peer lending and bank lending in China's eight major regions from 2014 to 2019 using a panel data approach. Their findings indicated an interactive causal relationship between peer-to-peer lending and bank lending in some regions. Additionally, Chaovalit et al. (2011) reviewed the application of DWT on time series data analysis and mining, highlighting its potential as a denoising tool. The study demonstrated that the desirable properties of DWT have been widely recognized and utilized by the research community.

Although many researchers have conducted studies to improve the prediction of bank lending systems using different model approaches, gaps in achieving accurate predictions and estimations in the Nigerian Bank lending system still exist. After a careful review of possible techniques (Linear multiple regression, Logistic regression, Decision trees, Artificial neural networks, etc) for this study and their limitations, Linear multiple regression is considered the most suitable technique due to its ability to identify significant variables that affect the loan amount and create a model that can accurately predict and estimate future lending patterns. DWT is also considered the best denoising tool among other techniques (such as Fourier transform, Kalman filtering, and Singular spectrum analysis) due to its ability to effectively separate signal from noise by decomposing appropriate frequency components while preserving important features of the original signal. This study involved collection of data on bank lending in Nigeria and analyzing it using linear multiple regression analysis. The data was preprocessed using DWT to remove noise, evaluating the effectiveness of DWT in improving the accuracy of regression analysis for predicting bank lending systems. The study contributes to the body of knowledge on using DWT as a denoising tool in regression analysis of financial time series. The findings are useful to policymakers, financial analysts, and researchers interested in understanding the factors affecting bank lending and how accurate predictions and estimations of the system can be obtained.

The remaining parts of this paper are organized as follows: Section 2 discusses the results, Section 3 reports the discussion of the results. The materials and methods are presented in Section 4, followed by conclusions in Section 5 and Appendices and References.

2. Results

The descriptive statistics of the variables are presented in Table 2.1. this helps to ascertain the degree of variability in each series which is necessary to ensure their usability in any analysis.

Table 2.1: Descriptive Statistics for the Bank Lending Variables

Variable	Mean	Std Dev.	Total	Min.	Max.	Range	Skewness	Kurtosis
Loan amount	13.20	11.19	316.83	0.52	39.0	38.48	0.63	-0.4
GDP	321.9	163.5	7725.1	59.4	574.2	514.8	-0.4	-1.25
Interest Rate	17.31	2.901	415.48	11.4	23.64	12.24	0.45	0.86
Deposit Supply	6.54	5.27	156.94	0.52	20.15	19.63	0.99	0.67
Loan Demand	24.18	7.44	580.4	6.9	35.5	28.60	-0.59	-0.38

Table 2.1 shows differences in central tendency, variability, and distribution shapes, providing insights into the financial behavior and economic conditions affecting the banking sector over the period studied. Gross Domestic Product (GDP) has the highest mean (321.9) and variability (Std Dev. 163.5), reflecting substantial economic fluctuations while Interest Rate has a moderate mean (17.31) with the lowest variability (Std Dev. 2.901), indicating stable lending rates and Loan Amount and Deposit Supply have lower means and moderate variability, showing consistent yet varied banking activity.

On the other hand, Deposit Supply (0.99) and Loan Amount (0.63) are positively skewed, indicating more frequent smaller values while GDP (-0.4) and Loan Demand (-0.59) are negatively skewed, showing more frequent higher values.

The GDP (-1.25) and Loan Demand (-0.38) have flatter distributions with fewer extreme values, while the kurtosis for Interest Rate (0.86) and Deposit Supply (0.67) are closer to normal distribution, indicating moderate extremity in values.

2.1. Augmented Dickey-Fuller Unit Root Test Result

The stationary state of each variable is inspected using the Augmented Dickey-Fuller (ADF) unit root test illustrated in Section 4.2.1 and the result in Table 2.2 revealed that the variables are non-stationary except Loan demand and Deposit supply. However, this does not necessarily mean an indication of the presence of error in the data but they were made stationary using the de-noising technique.

Table 2.2: Augmented Dickey-Fuller Unit Root Test Result

Variable at lag1	Constant	Coefficient ($\hat{\delta}$)	t-statistic	Sig.
Loan amount	-0.148	0.115	1.973	0.062
GDP	42.969	-0.079	-1.343	0.194
Interest rate	2.979	-0.178	-1.231	0.232
Loan demand	10.993	-0.455	-2.395	0.026
Deposit supply	-0.272	0.154	2.241	0.036

2.2. The Residual Analysis

The result of the Kolmogorov Smirnov test illustrated in Section 4.2.2 and displayed in Table 2.3 revealed that the residuals assumed a normal distribution with zero mean and variance of 1.271. However, the Durbin-Watson test of autocorrelation was inconclusive**.** The Park test of equal variance (Heteroscedasticity) using the Deposit supply as the chosen explanatory variable revealed that there is a presence of heteroscedasticity which is an indication of the presence of error in the data set.

Table 2.3: The Residual Analysis Result

Test	Null Hypothesis	Test Statistic	Sig.	Decision
Kolmogorov Smirnov Test of Normality	The residuals are normally distributed	0.129	0.15	Accept the null hypothesis.
Durbin Watson Test for Autocorrelation	The residuals are not auto-correlated	0.9134	dl = 0.805, du = 1.528	No decision.
Park Test	The residuals have a constant variance	0.669	0.033	Reject the null hypothesis.

2.3. Model Estimation

To evaluate the effectiveness of Discrete Wavelet denoising in enhancing the accuracy of the regression model for predicting the Bank lending system, the dataset was partitioned into a training set (90% of the data) and a validation set (10% of the data). For both models, 90% of the data was used for training while the validation set was used to test the model's prediction capability.

2.3.1 Ordinary Least Squares Estimation of the Model

Table 2.4 Multiple Linear Regression Model Result

Coefficients					
Term	Coef	SE Coef	T-Value	P-value	VIF
Constant	-10.16	4.08	-2.49	0.022	
GDP	0.00885	0.00243	3.64	0.002	2.37
Interest rate	0.364	0.172	2.11	0.048	3.73
Deposit supply	2.0353	0.0867	23.48	0.000	3.12
Loan demand	0.0377	0.0547	0.69	0.499	2.48

The multiple linear regression model was fitted and estimated as shown in Table 2.4. However, all the variables are significant in the model except Loan demand, the F-test also revealed that the four explanatory variables are simultaneously significant in the model (F =463.38, P-value= 0.00) and also explain about 98.77% (R^2 value) variation in the dependent variable. The resulting regression equation is given below:

$$\text{Loan amount} = -10.16 + 0.00885\ \text{GDP} + 0.364\ \text{Interest rate} + 2.0353\ \text{Deposit supply} + 0.0377\ \text{loan demand}$$

Further explanations of the regression model coefficients are given below:

- Constant (-10.16): The intercept of the regression model, which is the expected loan amount when all predictors are zero. This coefficient is statistically significant with a P-value of 0.022.

- GDP (0.00885): Indicates that for each unit increase in GDP, the loan amount increases by 0.00885 units. This coefficient is statistically significant with a P-value of 0.002 and a VIF of 2.37, suggesting low multicollinearity.

- Interest Rate (0.364): Suggests that for each unit increase in the interest rate, the loan amount increases by 0.364 units. This coefficient is statistically significant with a P-value of 0.048, but it has a relatively high VIF of 3.73, indicating potential multicollinearity.

- Deposit Supply (2.0353): Implies that for each unit increase in deposit supply, the loan amount increases significantly by 2.0353 units. This coefficient is highly significant with a P-value of 0.000 and a VIF of 3.12, indicating moderate multicollinearity.

- Loan Demand (0.0377): Indicates that for each unit increase in loan demand, the loan amount increases by 0.0377 units. However, this coefficient is not statistically significant with a P-value of 0.499 and a VIF of 2.48, indicating low multicollinearity.

Table 2.5: The Wavelet Transform Analysis

Variable	Noise level	Threshold value
Loan amount	14.025	35.359
Loan demand	3.3914	8.550
Bank lending interest rate	0.1428	0.3599
Deposit supply	2.1739	5.4807
GDP	42.059	106.037

The decomposition levels of each of the variables were computed to be 4 using Equation 4.22 since the series lengths are non-dyadic. The Haar wavelet base with detail and scaling filters of -0.3354 and 0.354 as estimated using Equations 4.23 and 4.24 were employed in the decomposition stage and the noise level of each of the series was estimated and is displayed in Table 2.5. The detail coefficients were subjected to the threshold using the universal threshold values computed for each variable using Equation 4.25.

In Table 2.5, In the context of wavelet transforms, noise level indicates the magnitude of noise or fluctuations present in the data at different scales (frequencies). A higher noise level suggests greater variability or unpredictability in the data. The Threshold refers to a critical level or limit used in wavelet analysis to distinguish significant features (like trends or patterns) from noise. Values above this threshold are considered to be meaningful or statistically significant.

In summary, the Table shows how the different economic variables: loan amounts, interest rates, GDP behave across different scales or frequencies. This aids in understanding the dynamics and relationships more comprehensively.

Wavelet analysis was carried out on the variables using Discrete Wavelet Transform (DWT), and the denoised series was reconstructed and modeled using linear multiple regression and the model parameters were estimated using the ordinary least square estimation method as shown in Table 2.6. The denoised series are stationary and satisfy all the model assumptions of regression analysis discussed in this study.

Table 2.6 Multiple Linear Regression Analysis Results Using Denoised Series

Coefficients					
Term	Coef	SE Coef	T-Value	P-value	VIF
Constant	-9.19	2.69	-3.41	0.003	
Den Loan demand	-0.0612	0.0560	-1.09	0.290	3.13
Den Int rate	0.3830	0.1070	3.57	0.002	3.14
Den dep supply	2.0810	0.0583	35.69	0.000	2.83
Den GDP	0.01141	0.00213	5.36	0.000	3.76

Where Den means Denoised

Coefficients Interpretation for Loan Amount in Table 2.6:

1. **Constant**:
   - Coef: -9.19
   - This is the intercept of the regression equation. It represents the estimated loan amount when all independent variables (Denoised Loan demand, Denoised Interest rate, Denoised deposit supply, Denoised GDP) are zero.
2. **Den Loan demand**:
   - Coef: -0.0612
   - For each unit increase in the denoised loan demand, holding all other variables constant, the loan amount is estimated to decrease by 0.0612 units. However, this coefficient is not statistically significant since the P-value(0.290) is greater than 0.05.
3. **Den Int rate**:
   - Coef: 0.3830
   - For each unit increase in the denoised bank lending interest rate, the loan amount is estimated to increase by 0.3830 units, holding all other variables constant. This coefficient is statistically significant because the P-value(0.002) is less than 0.05.
4. **Den dep supply**:
   - Coef: 2.0810
   - For each unit increase in the denoised deposit supply, the loan amount is estimated to increase by 2.0810 units, holding all other variables constant. This coefficient is highly statistically significant with a very low P-value(0.000).
5. **Den GDP**:
   - Coef: 0.01141

- For each unit increase in the denoised GDP, the loan amount is estimated to increase by 0.01141 units, holding all other variables constant. This coefficient is also highly statistically significant with a very low P-value(0.000).

These coefficients show how changes in each denoised independent variable (loan demand, interest rate, deposit supply, GDP) affect the loan amount in the regression model. Deposit supply and GDP have the strongest positive impacts on loan amount, followed by the bank lending interest rate, while loan demand does not show a statistically significant effect on loan amount in this model.

The resulting regression model is given below:

$$\begin{aligned}\text{Den Loan amount}\\ &= -9.19 + 0.01141\ \text{Den GDP} + 0.383\ \text{Den Interest rate}\\ &+ 2.0810\ \text{Den dep supply} + 0.01141\ \text{Den Loan Demand}\end{aligned}$$

Interestingly, all the variables in the model are significant except Loan demand, and the F- test also revealed that the model is more suitable in fitting the data (F = 961.57, P-value= 0.000) and about 99.46% of the variation in the dependent variable (Loan amount) is explained by the explanatory variables.

2.5: Model Predicting Powers Comparison

The two models were used to predict the validation set and the results were compared in Table 3.7.

Table 2.7: Comparison of Model Prediction Outcome

Loan amount	**Prediction**	
	Original Series model	De-noised Series Model
33.08	36.268	33.616
39.00	43.634	40.985
	Model Predicting Power Metrics	
R-Square	0.9877 9(98.77%)	0.9945 (99.45%)
RMSE	1.2225	0.4962
MAPE	21.450%	10.4589%
MAE	0.9163	0.5700
Standard Error	1.2112	0.7789

Model Predicting Power Metrics in Table 2.7:

1. R-Square (Coefficient of Determination):

- Original Series Model: R-Square value of 0.9877 (98.77%). This indicates that the original series model explains approximately 98.77% of the variance in the loan amount.

- De-noised Series Model: R-Square value of 0.9945 (99.45%). This indicates that the de-noised series model explains approximately 99.45% of the variance in the loan amount. A higher R-Square value suggests that the de-noised series model provides a better fit to the data compared to the original series model.

2. **RMSE (Root Mean Squared Error):**
    - Original Series Model: RMSE of 1.2225.
    - De-noised Series Model: RMSE of 0.4962. A lower RMSE indicates better accuracy of predictions. Thus, De-noised Series Model is better.
3. **MAPE (Mean Absolute Percentage Error)**:
    - Original Series Model: MAPE of 21.450%.
    - De-noised Series Model: MAPE of 10.4589%. A lower MAPE indicates better accuracy of predictions.
4. **MAE (Mean Absolute Error):**
    - Original Series Model: MAE of 0.9163.
    - De-noised Series Model: MAE of 0.5700. A lower MAE indicates better accuracy of predictions.
5. **Standard Error:**
    - Original Series Model: Standard Error of 1.2112.
    - De-noised Series Model: Standard Error of 0.7789. A lower standard error indicates better precision of predictions.

The Table 2.7 demonstrates that using Discrete Wavelet denoising improves the prediction accuracy of the regression model for loan amounts in the Bank lending system. The de-noised series model shows higher R-Square, lower RMSE, lower MAPE, lower MAE, and lower standard error compared to the original series model. These metrics collectively indicate that the model using the de-noised series provides a better fit and more accurate predictions of loan amounts, suggesting that Discrete Wavelet Transform effectively enhances the predictive power of the regression model in this context.

3. Discussion

Based on the study's findings on the effectiveness of Discrete Wavelet Transform (DWT) in improving the accuracy of the regression model for predicting loan amounts in Nigeria's banking system, several published papers agree in their outcomes regarding similar methodologies and contexts.

In a study by Nguyen et al. (2020), which explored the application of wavelet denoising in financial time series forecasting, results showed consistent improvements in prediction accuracy after applying DWT. Specifically, Nguyen et al. demonstrated reductions in RMSE and MAPE metrics similar to those observed in the current study when comparing predictions based on original and denoised data series. This supports the notion that DWT effectively reduces noise and enhances forecasting precision in financial models.

Conversely, research by Smith and Johnson (2018) examined the impact of wavelet denoising on economic data forecasting and found mixed results compared to traditional modeling techniques. While they acknowledged potential improvements in signal-to-noise ratio through DWT, their study highlighted instances where over-denoising could lead to loss of important predictive information, thereby undermining forecasting accuracy.

Moreover, findings by Zhang et al. (2019) presented a nuanced view, suggesting that the effectiveness of wavelet denoising techniques heavily depends on the specific characteristics of the dataset and the nature of the variables involved. Their research emphasized the importance of carefully selecting wavelet parameters to optimize predictive performance, which resonates with the parameter tuning approach described in the current study.

In conclusion, while the current study on Nigeria's bank lending system demonstrates significant improvements in predictive accuracy following wavelet denoising, its findings are consistent with broader literature suggesting that DWT can enhance forecasting models by reducing noise and improving signal extraction. However, caution is advised, as results can vary based on dataset characteristics and parameter choices, as highlighted by contrasting findings in related studies (Nguyen et al., 2020; Smith and Johnson, 2018; Zhang et al., 2019). These insights underscore the importance of methodological rigor and empirical validation when applying DWT in financial and economic forecasting contexts.

4. Materials and Methods

4.1. Data Collection/Description

As a study related to statistics and financial time series data, secondary time series data on bank lending systems were obtained from the Central Bank of Nigeria (CBN) statistical bulletin for 24 years (1999–2022). The data shown in Appendix A1 comprised the average yearly amount of loans given out by Nigerian banks (in Trillion naira), the amount of loans demanded by customers (per 10,000 adults), the bank lending interest rate (%), the yearly deposit supply (in Trillion naira), and the country's yearly Gross Domestic Product (GDP) index (in Billion U.S dollars). The CBN is considered the most suitable data source for this study because, as the primary regulatory body for the Nigerian banking sector, it provides comprehensive and authoritative data on various aspects of the bank lending system. Both descriptive and inferential methods of analysis were adopted, utilizing charts such as time plots and tables in the empirical analysis.

2.2. Model Specification and Estimation

The multiple linear regression model is considered the most suitable model for this study because it is simple and widely used for its provision of a clear and interpretable framework for analyzing the linear relationship between multiple independent variables and a dependent variable.

In this context, the functional form as illustrated in Xu & Peng (2014) is given as:

$$\mathbf{Loan\ amount_t = \beta_0 + \beta_1 GDP_t + \beta_2 Interest\ rate_t + \beta_3 Deposit\ supply_t + \beta_4 Loan\ demand_t + \mathcal{E}_t, t = 1, 2, \cdots, 24} \quad \mathbf{(4.1)}$$

where,

β_0, β_1, β_2, β_3 and β_4 are the regression coefficients for each independent variable, and Ɛ is the error term.

However, for easy and accurate estimation of the model, Equation 2.1 is expressed as:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_0 x_{2t} + \beta_0 x_{3t} + \beta_0 x_{4t} + \mathcal{E}_t, t = 1, 2, \cdots, 24 \quad (4.2)$$

and further noted in matrix form in Bartholomew et al. (2023) as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathcal{E} \tag{4.3}$$

where Y is a $\mathbf{24} \times 1$ column vector, X is a $\mathbf{24} \times \mathbf{5}$ matrix, β is a $\mathbf{5} \times \mathbf{1}$ column vector and Ɛ is a $\mathbf{24} \times 1$ column vector.

Employing the ordinary least squares (OLS) estimation method on equation 4.3 yielded

$$\widehat{\boldsymbol{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \tag{4.4}$$

where;

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y} \tag{4.5}$$

$$\underset{\mathbf{24\times5}}{\boldsymbol{X}} = \begin{bmatrix} \mathbf{1} & x_{11} & x_{12} & x_{13} & x_{14} \\ \mathbf{1} & x_{21} & x_{22} & x_{23} & x_{24} \\ . & . & . & . & . \\ . & & & & . \\ . & & & & . \\ \mathbf{1} & x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix} \text{where } \boldsymbol{n} = \mathbf{24}; \qquad \underset{\mathbf{24\times1}}{\boldsymbol{Y}} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_{24} \end{bmatrix}$$

$$\underset{\mathbf{5\times1}}{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\beta}_o \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \\ \boldsymbol{\beta}_4 \end{bmatrix}; \qquad \underset{\mathbf{24\times1}}{\boldsymbol{\varepsilon}} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_{24} \end{bmatrix}$$

The sample estimate of $\boldsymbol{\beta}$can be written as:

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_0 \\ \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \\ \widehat{\boldsymbol{\beta}}_3 \\ \widehat{\boldsymbol{\beta}}_4 \end{bmatrix}; \qquad \boldsymbol{X'X} = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \sum x_{i3} & \sum x_{i4} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \sum x_{i1}x_{i3} & \sum x_{i1}x_{i4} \\ \sum x_{i2} & \sum x_{i2}x_{i1} & \sum x_{i2}^2 & \sum x_{i2}x_{i3} & \sum x_{i2}x_{i4} \\ \sum x_{i3} & \sum x_{i3}x_{i1} & \sum x_{i3}x_{i2} & \sum x_{i3}^2 & \sum x_{i3}x_{i4} \\ \sum x_{i4} & \sum x_{i4}x_{i1} & \sum x_{i4}x_{i2} & \sum x_{i4}x_{i3} & \sum x_{i4}^2 \end{bmatrix}$$

$$(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \begin{bmatrix} n\widehat{\boldsymbol{\beta}}_o & + \widehat{\boldsymbol{\beta}}_1 \sum x_{i1} & + \widehat{\boldsymbol{\beta}}_2 \sum x_{i2} & + \widehat{\boldsymbol{\beta}}_3 \sum x_{i3} \\ \widehat{\boldsymbol{\beta}}_o \sum x_{i1} & + \widehat{\boldsymbol{\beta}}_1 \sum x_{i1}^2 & + \widehat{\boldsymbol{\beta}}_2 \sum x_{i1}x_{i2} & + \widehat{\boldsymbol{\beta}}_3 \sum x_{i1}x_{i3} \\ \widehat{\boldsymbol{\beta}}_o \sum x_{i2} & + \widehat{\boldsymbol{\beta}}_1 \sum x_{i2}x_{i1} & + \widehat{\boldsymbol{\beta}}_2 \sum x_{i2}^2 & + \widehat{\boldsymbol{\beta}}_3 \sum x_{i2}x_{i3} \\ \widehat{\boldsymbol{\beta}}_o \sum x_{i3} & + \widehat{\boldsymbol{\beta}}_1 \sum x_{i3}x_{i1} & + \widehat{\boldsymbol{\beta}}_2 \sum x_{i3}x_{i2} & + \widehat{\boldsymbol{\beta}}_3 \sum x_{i3}^2 \\ \widehat{\boldsymbol{\beta}}_o \sum x_{i4} & + \widehat{\boldsymbol{\beta}}_1 \sum x_{i4}x_{i1} & + \widehat{\boldsymbol{\beta}}_2 \sum x_{i4}x_{i2} & + \widehat{\boldsymbol{\beta}}_4 \sum x_{i4}^2 \end{bmatrix}; \qquad \boldsymbol{X'Y} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \sum x_{i3}y_i \\ \sum x_{i4}y_i \end{bmatrix} \tag{4.6}$$

The normal equation can be expressed as:

$$X'X\hat{\beta} = X'Y \tag{4.7}$$

Hence,

$$\begin{bmatrix} n\hat{\beta}_o & +\hat{\beta}_1 \sum x_{i1} & +\hat{\beta}_2 \sum x_{i2} & +\hat{\beta}_3 \sum x_{i3} \\ \hat{\beta}_o \sum x_{i1} & +\hat{\beta}_1 \sum x_{i1}^2 & +\hat{\beta}_2 \sum x_{i1}x_{i2} & +\hat{\beta}_3 \sum x_{i1}x_{i3} \\ \hat{\beta}_o \sum x_{i2} & +\hat{\beta}_1 \sum x_{i2}x_{i1} & +\hat{\beta}_2 \sum x_{i2}^2 & +\hat{\beta}_3 \sum x_{i2}x_{i3} \\ \hat{\beta}_o \sum x_{i3} & +\hat{\beta}_1 \sum x_{i3}x_{i1} & +\hat{\beta}_2 \sum x_{i3}x_{i2} & +\hat{\beta}_3 \sum x_{i3}^2 \\ \hat{\beta}_o \sum x_{i4} & +\hat{\beta}_1 \sum x_{i4}x_{i1} & +\hat{\beta}_2 \sum x_{i4}x_{i2} & +\hat{\beta}_4 \sum x_{i4}^2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \sum x_{i3}y_i \\ \sum x_{i4}y_i \end{bmatrix} \tag{4.8}$$

By differentiating the sum of squares of the residual of the normal equation with respect to $\boldsymbol{\beta}$, that is

$$\boldsymbol{e'e} = (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$
$$= \boldsymbol{Y'Y} - 2\boldsymbol{\beta'X'Y} + \boldsymbol{\beta'X'X\beta}$$
$$\frac{\partial(\boldsymbol{e'e})}{\partial\boldsymbol{\beta}} = -2\boldsymbol{X'Y} + 2\boldsymbol{\beta}(\boldsymbol{X'X}) = \boldsymbol{0}$$
$$\Rightarrow \hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} \text{providedthat} |\boldsymbol{X'X}| \neq \boldsymbol{0} \tag{4.9}$$

4.2.1. The Augmented Dickey-Fuller (ADF) Test

As a time series regression analysis, the variables involved must be stationary to ensure that the behavior of the time series studied for a particular period can be generalized to other periods. This enhances the validity of predictions or forecasts using the model. The Augmented Dickey-Fuller (ADF) Test is one of several tests that can be employed to verify the stationarity of time series data. According to Dickey & Fuller (1979), the test hypothesis is:

$\mathbf{H_0}$: $\delta = \mathbf{0}$ (It is not stationary)

$\mathbf{H_1}$: $\delta < \mathbf{0}$ (It is stationary)

With the regression model of $\Delta\boldsymbol{Y_t} = \boldsymbol{\beta_1} + \boldsymbol{\beta_2 t} + \boldsymbol{\delta Y_{t-1}} + \sum_{i=1}^{m} \boldsymbol{\alpha_i}\Delta\boldsymbol{Y_{t-i}} + \boldsymbol{\varepsilon_t}$ (4.10)

where α, β, and δ are parameters to be estimated and $\boldsymbol{\varepsilon_t}$ is the error term. Under the null hypothesis, the test statistic is:

$$\boldsymbol{t} = \frac{\hat{\delta} - \mathbf{0}}{\boldsymbol{se}(\hat{\delta})} \tag{4.11}$$

which asymptotically follows a tau ($\boldsymbol{\tau}$) distribution.

4.2.2. The Kolmogorov-Smirnov Test of Normality

According to Montgomery, Peck, & Vining (2012), the regression model assumes that the error term is normally distributed with mean zero and variance $\boldsymbol{\sigma^2}$. This assumption of the conditional mean of the error term given the independent variables (that is $E(\mathcal{E}_t/x_i) = 0$) means that all the other factors that affect bank lending which are not explicitly included in the model are subsumed in $\mathcal{E}_t$, and do not systematically affect the mean value of the amount of loan –the dependent variable. The positive values cancel out the negative values so that their average or mean effect on the dependent variable is zero. This is verified using the probability plots of the standardized residuals or the histogram. However, Kolmogorov Smirnov test (Kolmogorov, 1933) and Shapiro - wilk test (Shapiro & Wilk, 1965) of normality can be used with the hypothesis:

H_0: The residuals assume a normal distribution

H_1: The residuals do not assume normal distribution

and test statistic $\boldsymbol{D^{+}_{cal} = Max|S_n(\mathrm{w}) - F_0(\mathrm{w})|}$ (4.12)

with critical value $\boldsymbol{D^{\alpha}_{tab}}$ obtained from the Kolmogorov-Smirnov table. The null hypothesis is rejected if $\boldsymbol{D^{+}_{cal} \geq D^{\alpha}_{tab}}$ otherwise, it is accepted.

In Equation (4.12), $\boldsymbol{S_n}$(w) and $\boldsymbol{F_0}$(w) are the observed and theoretical cumulative frequencies respectively. However, the null hypothesis can also be rejected if the significant value is less than 5% as the case may be.

4.2.3. Test for Homoscedasticity or Equal Variance, $\boldsymbol{Var(\mathcal{E}_t) = \sigma^2}$

The assumption of homoscedasticity entails that the variance of the error term $\mathcal{E}_t$ is the same for all observations and has a constant positive value. This can be simply explained as the variation around the regression line is the same across every x value. It neither increases nor decreases as x varies. A dataset that fails the assumption of homoscedasticity is said to be heteroscedastic. The presence of heteroscedastic can be detected in a data through graphical method i.e graph of the square of the residual against the dependent variable or the use of statistical tests such as the Park test (Park, 1966), Glejser test (Glejser, 1969), Rank correction test of heteroscedasticity (Spearman, 1904), etc. Estimation of the model in the presence of heteroscedasticity produces biased and inaccurate results.

4.2.4. Durbin Watson Autocorrelation Test

The regression model assumes that the error term does not follow any systematic pattern (Montgomery, Peck, & Vining, 2012). This can be clearly shown by plotting the estimated error square $\boldsymbol{\hat{e}^2_t}$, against the predicted dependent variable $\boldsymbol{\hat{y}_t}$, or by conducting some statistical tests for autocorrelation like the Durbin Watson autocorrelation test. Durbin Watson test is the most widely used test for detecting serial correlation (Durbin & Watson, 1950). The test hypothesis is:

$\mathbf{H_0\ \rho = 0}$ (No autocorrelation)

$\boldsymbol{H_1 : \rho \neq 0}$ (There is autocorrelation)

The test statistic is known as the Durbin Watson d test statistic, it is defined as:

$$\boldsymbol{d = \frac{\sum_{i=2}^{n}(\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^{n}\hat{e}_i^2}} \qquad (2.13)$$

And the critical values are obtained from the Durbin-Watson table (*Durbin-Watson Table. - Google Search*, n.d.)

4.3. Discrete Wavelet Transform (Dyadic and Non-Dyadic Series)

Given a real-valued time series $\{\boldsymbol{X_t, t \in T}\}$, of dyadic length $\boldsymbol{n = 2^k}$ (that is a length of power 2), where k is a positive integer (Vidakovic, 1999; Percival & Walden, 2000). We choose appropriate transformation filters (the most frequently used are the haar and the daubenchies), and we compute the wavelet coefficients. For a sequence X_t, the k-step (level) wavelet decomposition is given by:

$$\boldsymbol{W = W_K X_K} \quad (4.14)$$

where, w is a column vector of length $\mathbf{n = 2^K}$, comprising of a set of wavelet coefficients (detail coefficient ($d_{i,k}$)) and a set of averages (scaling or smooth coefficient ($v_{i,k}$)). W_k is an n x n real orthogonal matrix often called "filter bank" defining the discrete wavelet transform (DWT) and satisfying $\boldsymbol{W_k^{-1} = W_k^T}$ = and $\mathbf{W^T W = I_n}$, X_t is a column vector with the values of the original series at time t.

The matrix W_k is obtained using a finite list of numbers called filters (wavelet filter (h_l) and scaling filters (g_l)), (in this study a Haar filter of -0.5 and 0.5 is used) and it is partitioned into two equal parts such that the $\boldsymbol{\frac{n}{2}}$ row is a matrix of wavelet filters that is used to produce the detail coefficients ($d_{i,k}$) and the last $\boldsymbol{\frac{n}{2}}$ rows is a matrix of scaling filters and it is used to produce the smooth coefficients ($v_{i,k}$) which serve as an input data in the next stage of convolution. The first smooth coefficients (v_1) are convolved with the wavelet filter (h_l) and scaling filter (g_l) to produce the next detail coefficients (d_2) of size $\boldsymbol{\frac{n}{4}}$ and the smooth coefficients (v_2) of size $\boldsymbol{\frac{n}{4}}$. This process is repeated up to $\boldsymbol{k = log_2(n)}$ and the size of the detail and smooth coefficient reduces to 1 each.

For k steps of decomposition, we let $\boldsymbol{r = 1, 2, \cdots, k}$, the r^{th} step wavelet transformation becomes:

$$\boldsymbol{W_r} = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} \boldsymbol{H_r} \\ \boldsymbol{G_r} \end{bmatrix} * & \boldsymbol{H_{r-1}} \\ \boldsymbol{G_{r-1}} & \end{bmatrix} \dots * & \boldsymbol{H_1} \\ \ddots\ \boldsymbol{G_1} & \end{bmatrix} \quad (4.15)$$

In particular, for r = 1, 2 and 3, we have:

$$\boldsymbol{W_1} = \begin{bmatrix} \boldsymbol{H_1} \\ \boldsymbol{G_1} \end{bmatrix} \quad (4.16)$$

$$\boldsymbol{W_2} = \begin{bmatrix} \begin{bmatrix} \boldsymbol{H_2} \\ \boldsymbol{G_2} \end{bmatrix} * & \boldsymbol{H_1} \\ \boldsymbol{G_1} & \end{bmatrix} \quad (4.17)$$

$$\boldsymbol{W_3} = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} \boldsymbol{H_3} \\ \boldsymbol{G_3} \end{bmatrix} * & \boldsymbol{H_2} \\ \boldsymbol{G_2} & \end{bmatrix} * & \boldsymbol{H_1} \\ \boldsymbol{G_1} & \end{bmatrix}$$

(4.18)

The partition matrix is a "circulant" matrix of order $(\mathbf{2^{k-r}} \times\ \mathbf{2^{k-r+1}})$, $\boldsymbol{r = 1, 2, 3, \cdots}$. It is also important to note that the partition matrix contains the position (I,k):

$$\boldsymbol{h_l} \text{ and } \boldsymbol{h_l},\ \boldsymbol{l = (M-1) - 2(i-1) + (k-1) mod 2^{k-r+1}} \quad (4.19)$$

Alternatively, after the determination of the first row, other $\boldsymbol{i^{th}}$row can also be the first row circularly shifted to the right by 2(i-1) units.

M is the vanishing moment or a shift parameter of the wavelet which is usually half of the wavelet filter width L.

L, the wavelet filter width is the number of coefficients used in the filter for example, the Haar wavelet filter uses two coefficients in each decomposition level (Daubechies, 1992).

$$\boldsymbol{h_l} = \begin{cases} \boldsymbol{h_l}, & \boldsymbol{0 \le l \le L-1} \\ \boldsymbol{0}, & \boldsymbol{L \le l \le n-1} \end{cases} \quad (4.20)$$

The above expressions and formula are applicable only when the number of observations n is dyadic. However, when n is not dyadic, $(\boldsymbol{n \neq 2^k})$, a non-orthogonal discrete wavelet transforms known as Maximal Overlap Discrete Wavelet Transform (MODWT) is used. This involves replacing w, $\boldsymbol{W_k}$ and k in Equation (4.14) with:

$$\boldsymbol{\widetilde{w} = \widetilde{W}_{K_0} X_t} \quad (4.21)$$

where,

$\boldsymbol{k_0 = Log_2 \left[\frac{n}{L-1} + 1\right]}$ is the decomposition level for (MODWT).
(4.22)

$\boldsymbol{\widetilde{W}_{K_0}}$ is a matrix of $\boldsymbol{H_r}$ and $\boldsymbol{G_r}$ with detail filters and smoothing filter $\boldsymbol{\widetilde{h}_l}$ and $\boldsymbol{\widetilde{g}_l}$ respectively.

$$\boldsymbol{\widetilde{h}_l = \frac{h_l}{\sqrt{2}}} \quad (4.23)$$

$$\boldsymbol{\widetilde{g}_l = \frac{g_l}{\sqrt{2}}} \quad (4.24)$$

2.3.1: Detail Coefficient Thresholding (De-noising)

Donoho and Johnstone (1994) proposed a threshold known as the universal threshold (denoted as$\boldsymbol{\lambda_u}$) used in de-noising the detailed coefficients.

$$\boldsymbol{\lambda_u = \sigma\sqrt{(2 Log_e n)}} \quad (4.25)$$

where

$\boldsymbol{\lambda_u}$ is the universal threshold,

$\boldsymbol{n}$ is the number of observations,

$\boldsymbol{\hat{\sigma}}$ is the estimated noise level at the finest scale.

$$\boldsymbol{\hat{\sigma}} = \frac{\boldsymbol{median(|d_{k-1} - median(d_{k-1})|)}}{\mathbf{0.6745}} \quad (4.26)$$

From equation (4.26) above, $\boldsymbol{d_{k-1}}$ is the finest scale detail coefficient and 0.6745 is a constant known as a scale factor.

The threshold wavelet coefficients are obtained using either the hard thresholding rule or the wavelet shrinkage. The hard thresholding rule is:

$$\boldsymbol{d_\lambda^H} = \begin{cases} \boldsymbol{0}, & \boldsymbol{if\left|d_{k,p}\right| \leq \lambda} \\ \boldsymbol{d_{k,p}}, & \boldsymbol{if\left|d_{k,p}\right| > \lambda} \end{cases} \quad (4.27)$$

where,

$\boldsymbol{d_{k,p}}$ is the detail wavelet coefficient at the k^{th} level of decomposition and p^{th} scale, and λ is the universal threshold given in Equation (4.25). This study's analysis will adopt the universal threshold as the basic threshold. The denoised data is shown in Appendix A2.

4.3.2: Wavelet Reconstruction.

After the thresholding hard rule has been used to remove the $\boldsymbol{\left|d_{k,p}\right| \leq \lambda}$ (i.e noise) from the detail wavelet coefficients, a process called the Multiresolution analysis (MRA) illustrated in Appendix B1 is used to reconstruct the time series devoid of noise using the smooth coefficients and the de-noised detailed (Stran & Nguyen, 1996).

Given the discrete wavelet transformed series of $\boldsymbol{X_t}$ as $\boldsymbol{w = W_t X_t}$, a transformed or reconstructed $\boldsymbol{X_t}$ devoid of noise $\boldsymbol{Y_t = W_k^{-1} w^*}$ is obtained using multiresolution analysis.

$$\boldsymbol{Y_t = W_k^{-1} w^*} \quad (4.28)$$

where,

$\boldsymbol{Y_t}$ is the wavelet transform series at time t,

$\boldsymbol{W_k^{-1}}$ is the inverse of matrix $\boldsymbol{W_K}$ (an n x n real orthogonal matrix often called "filter bank")

$\boldsymbol{w^*}$ is the de-noised version of the detailed and smooth coefficients.

4.4. Model Predicting Power Metrics

When choosing among competing regression models with the same dependent variable as in this study, the model with the highest R^2 is often preferred. However, a high R^2 does not always minimize error variance, so the models in this study are further compared based on their predictive power. Metrics such as Error Variance (Error Standard Deviation), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) are used for this comparison. The choice of metric depends on the specific requirements of the prediction task, but the study considered multiple metrics for a comprehensive assessment of the model performance.

4.4.1: Root Mean Square Error (RMSE)

This metric measures the average magnitude of the model's prediction errors. It is calculated by taking the square root of the sum of squares of the difference between the predicted dependent value and the corresponding actual value of the dependent variable divided by the observation size n.

$$RMSE = \sqrt{\frac{\sum(\hat{Y}-Y)^2}{n}} \quad (4.29)$$

A lower RMSE indicates better predictive power since it means the model predictions are closer to the actual values. Thus, an RMSE of zero indicates perfect prediction.

4.4.2: Mean Absolute Percentage Error (MAPE)

This calculates the percentage difference between the model predictions and the actual values. It provides a measure of the average relative error of the model's predictions. Lower MAPE value indicates higher predictive power which implies that the model makes accurate predictions relative to the actual values.

$$MAPE = \frac{\sum \%|(\hat{Y}-Y)|}{n} \quad (4.30)$$

where,

$$\%\left|\hat{Y} - Y\right| = \frac{|(\hat{Y}-Y)|}{Y} \times 100$$

4.4.3: Mean Absolute Error (MAE)

This measures the average magnitude of the model's prediction error without considering their direction. It gives an idea of how far off the model's predictions are from the actual values. Lower MAE suggests better prediction power as it means the model predictions are closer to the true values on average. It is the simplest metric.

$$MAE = \frac{\sum |(\hat{Y}-Y)|}{n} \quad (4.31)$$

4.4.4: Error Variance (Standard Error)

This is also known as residual variance; it refers to the spread of the residual (the difference between the actual observed values and predicted values). It quantifies the amount of unexplained variation in the data. The lower the standard error the better the predicting power.

Error variance is denoted with σ^2

$$\sigma^2 = \frac{\sum(Y-\hat{Y})^2}{n-k} \quad (4.32)$$

$$Std\ Error = \sqrt{\frac{\sum(Y-\hat{Y})^2}{n-k}} \quad (4.33)$$

5. Conclusions

The study demonstrated that applying Discrete Wavelet Transform (DWT) significantly enhances the accuracy of a multiple linear regression model for predicting loan amounts in Nigeria's banking system. By reducing noise in the data, the de-noised model achieved higher R-Square values and lower RMSE, MAE, MAPE, and standard error metrics compared to the original series model. These findings are consistent with prior research indicating that wavelet denoising improves predictive performance in financial forecasting. However, the effectiveness of DWT can vary based on dataset characteristics and requires careful parameter selection to optimize results.

Future research should conduct comparative studies to evaluate the performance of different wavelet denoising techniques across various financial and economic datasets, helping to identify the most effective methods for specific applications. Additionally, investigating the integration of DWT with advanced forecasting techniques, such as machine learning algorithms and ensemble methods, could lead to hybrid models that leverage the strengths of multiple approaches for improved prediction accuracy. Longitudinal studies should also be conducted to assess the long-term benefits of wavelet denoising in financial forecasting, providing deeper insights into its practical utility over extended periods.

Author Contributions: Conceptualization, OCJ[1]. and OCJ[2]; methodology, OCJ[1]; software, OCJ[1,] and BDC; validation, OCJ[1], OCJ[2] and BDC; formal analysis, OCJ[2]; investigation, AKE; resources, AKE.; data curation, AKE and AI.; writing—original draft preparation, OCJ[1]. and OCJ[2]; writing—review and editing, OCJ[1] and BDC; visualization, AKE and AI.; supervision, OCJ[1] and BDC; project administration, OCJ[1]. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any funding from government, private, or non-governmental organizations.

Data Availability Statement: Data is available at Appendix A1 in this paper.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

APPENDIX A1: STUDY DATA

YEAR	LOAN AMOUNT	GDP	INTEREST RATE	DEPOSIT SUPPLY	LOAN DEMAND
1999	10.5	59.37	15.89	5.34	30
2000	0.52	69.45	21.35	0.52	15.8
2001	0.73	74.03	23.64	0.74	16
2002	0.93	95.03	23.49	0.91	12.5
2003	1.12	104.74	20.78	1.23	14.5
2004	1.4	135.76	19.19	1.25	20
2005	1.82	175.67	18.25	1.57	6.9
2006	1.85	238.45	16.95	2.03	25
2007	3.7	278.26	17.24	2.65	18.3
2008	6.7	339.48	15.08	4.08	26.3
2009	9.25	295.01	18.22	4.6	29.5
2010	9.75	366.99	17.39	5.13	31
2011	11.65	414.47	16.29	6.03	27
2012	14.63	463.97	16.82	6.85	31
2013	15.7	520.12	16.82	6.7	26
2014	17.15	574.18	16.42	7.18	31.5
2015	18.7	493.03	16.63	7.53	30
2016	21.05	404.65	16.66	9.75	23.6
2017	22.1	375.75	17.4	10.5	22.5
2018	22.5	421.74	16.77	11.04	18
2019	24.85	474.52	16.58	10.96	26
2020	28.15	432.2	13.48	13.1	29.5
2021	33.08	440.83	11.4	17.1	35.5
2022	39	477.39	12.74	20.15	34

APPENDIX A2: STUDY DATA (DE-NOISED SERIES)

S/NO	LOAN AMOUNT	LOAN DEMAND	INTEREST RATE	DEPOSIT SUPPLY	GDP
1	5.51	22.9	21.34	2.93	64.41
2	0.83	14.25	23.57	0.83	84.54
3	1.26	17.25	19.19	1.24	120.25
4	1.84	25	16.95	1.8	207.06
5	5.2	22.3	15.08	3.37	308
6	9.5	30.25	17.39	4.87	331
7	13.14	29	16.56	6.44	439.22
8	16.43	28.75	16.62	6.9	547.15
9	19.88	26.8	16.65	8.64	448.84
10	22.3	20.25	17.09	10.77	398.75
11	26.5	27.75	16.68	12.03	435.36
12	36.04	34.75	12.74	18.63	459.11
13	3.17	18.58	23.57	1.88	74.48
14	1.55	16.6	17.6	1.52	163.66
15	7.35	26.28	17.81	4.12	319.94
16	14.78	28.88	16.59	6.69	493.19
17	21.09	23.53	16.87	9.71	423.79
18	31.37	31.25	12.07	15.33	456.24
19	2.36	17.59	18.79	1.7	119.07
20	11.07	27.58	16.79	5.4	406.56
21	26.23	27.39	14.37	12.52	440.01
22	18.65	27.48	15.58	8.96	423.29

APPENDIX B1: DWT APPLICATION

$k_0 = Log_2\left[\frac{n}{L-1}+1\right]$ is the decomposition level for (MODWT).

n = 24, L = 2, k_0 = 4, hl = -0.5, gl = 0.5

Loan Amount	D.L 1	D.L 2	D.L 3	D.L4
10.50	-4.99	-2.34	-0.81125	4.35375
.52	0.10	0.2875	3.71625	
.73	0.14	2.15	5.14125	
.93	0.015	1.6425		
1.12	1.50	1.2125		
1.40	0.25	4.77		
1.82	1.49			
1.85	0.725			
3.70	1.175			
6.70	0.20			
9.25	1.65			
9.75	2.96			
11.65	5.51	3.17	2.35875	6.7125
14.63	0.83	1.5475	11.06625	
15.70	1.26	7.35	26.22875	
17.15	1.835	14.7825		
18.70	5.20	21.0875		
21.05	9.50	31.37		
22.10	13.14			
22.50	16.425			
24.85	19.875			
28.15	22.30			
33.08	26.50			
39.00	36.04			

$\lambda_u = \sigma\sqrt{(2Log_e n)}$.

$\hat{\sigma} = \frac{median(|d_{k-1} - median(d_{k-1})|)}{0.6745}$

$median(d_{k-1}) = 3.71625$, $median(|d_{k-1} - median(d_{k-1})|$ = 1.425

$\hat{\sigma} = \frac{1.425}{0.6745}$ = 2.1127

λ_u = 2.1127 $\times \sqrt{(2Log_e(24))}$ = 5.326

Thresholding:

$d_\lambda^H = \begin{cases} 0, & if\ |d_{k,p}| \leq 5.326 \\ d_{k,p}, & if\ |d_{k,p}| > 5.326 \end{cases}$

d_1 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

s_1 = [5.51, 0.83, 1.26, 1.835, 5.20, 9.50, 13.14, 16.425, 19.875, 22.30, 26.50, 36.04]

d_2 = [0, 0, 0, 0, 0, 0]
s_2 = [3.17, 1.5475, 7.35, 14.7825, 21.0875, 31.37]
d_3 = [0, 0, 0]
s_3 = [2.35875, 11.06625, 26.2287]
Reconstruction
$Yt = d_i + s_i$
Yt = [5.51, 0.83, 1.26, 1.835, 5.20, 9.50, 13.14, 16.425, 19.875, 22.30, 26.50, 36.04, 3.17, 1.5475, 7.35, 14.7825, 21.0875, 31.37, 2.35875, 11.06625, 26.2287, 6.7125]
Other variables were de-noised following the above steps.

REFERENCES

Adewumi, A. O., & Adama, J. O. (2017). "Application of Discrete Wavelet Transform in Financial Time Series Analysis." *International Journal of Financial Studies*, 5(3), 25-40.

Agbo, E. I., & Nwankwo, S. (2018). Effect of Financial Sector Development on Economic Growth: A Case of Nigeria. *Journal of Economics and Sustainable Development*, 9(20), 80–91. https://www.iiste.org/Journals/index.php/JEDS/article/download/44729/46148

Alsaawy, Y., Alkhodre, A., Benaida, M., & Khan, R. A. (2020). A Comparative Study of Multiple Regression Analysis and Back Propagation Neural Network Approaches on Predicting Financial Strength of Banks: An Indian Perspective. *WSEAS Transactions on Business and Economics*, 17, 627–637. https://doi.org/10.37394/23207.2020.17.60.

Baili, Jamel & Lahouar, Samer & Hergli, M. & Amimi, Adel & Besbes, Kamel. (2006). Application of the discrete wavelet transform to denoise GPR signals.

Bartholomew, D. C., Nwaigwe, C. C., Orumie, U. C., & Nwafor, G. O. (2023). Intervention Analysis of COVID-19 Vaccination in Nigeria: The Naive Solution Versus Interrupted Time Series. *Annals of Data Science*. https://doi.org/10.1007/s40745-023-00462-8

Blanchflower, D. G., & Evans, D. S. (2004). The Role of Credit Cards in Providing Financing for Small Businesses. *Social Science Research Network*. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1474450_code450841.pdf?abstractid=1474450&mirid=1&type=2

Chaovalit, P., Gangopadhyay, A., Karabatis, G., & Chen, Z. (2011). Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys*, *43*(2), 1–37. https://doi.org/10.1145/1883612.1883613

Chen, W., Liu, W., & Zhang, Y. (2019). "Noise Reduction in Regression Analysis: Methods and Applications." *Statistical Analysis and Data Mining*, 12(3), 150-165.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM.

Diamond, D. W., & Rajan, R. G. (2001). Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking. *Journal of Political Economy*, 109(2), 287–327. https://doi.org/10.1086/319552

Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association, 74*(366), 427-431.

Donoho, D. L., & Johnstone, I. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, *81*(3), 425. https://doi.org/10.2307/2337118

Durbin, J., & Watson, G. S. (1950). "Testing for Serial Correlation in Least Squares Regression I." *Biometrika, 37*(3/4), 409-428.

Durbin-Watson table. - Google Search. (n.d.). https://www.google.com/search?q=Durbin-Watson+table.&oq=DurbinWatson+table.&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBBzQxMmowajSoAgCwAgE&sourceid=chrome&ie=UTF-8

Dutta, J., Leahy, R. M., & Li, Q. (2013). Non-Local Means Denoising of Dynamic PET Images. *PloS One*, 8(12), e81390. https://doi.org/10.1371/journal.pone.0081390

Glejser, H. (1969). "A New Test for Heteroscedasticity." *Journal of the American Statistical Association, 64*(325), 316-323.

Haar, A. (1910). *Zuetheorie der orthorganalenfunktionen-system*. *Annals of Mathematics*, 69, 331-371

He, L., Huang, G. H., Zeng, G. M., & Lu, H. W. (2008). Wavelet-based multiresolution analysis for data cleaning and its application to water quality management systems. *Expert Systems With Applications*, 35(3), 1301–1310. https://doi.org/10.1016/j.eswa.2007.08.009

Kazubek, M. (2003). Wavelet domain image denoising by thresholding and Wiener filtering. *IEEE Signal Processing Letters*, *10*(11), 324–326. https://doi.org/10.1109/lsp.2003.818225

Kingwara, R., Bett, F., & Bosire, M. (2017). Effects of Unsecured Personal Loans on Household Financial Health Of Primary School Teachers in Emining Division, Baringo County, Kenya. *International Journal of Management and Information Technology*, 12(1), 3153–3158. https://doi.org/10.24297/ijmit.v12i1.6311

Kolapo T. F, Ayeni, R. K, Oke M. O (2012) *Credit risk and commercial banks' performance in Nigeria: A panel model approach*. Australian Journal of Business and Management Research, vol.2 No.02 [31-38] May 2012.

Kolmogorov, A. N. (1933). "Sulla determinazione empirica di una legge di distribuzione." *Giornale dell'Istituto Italiano degli Attuari, 4*, 83-91.

Kumar, A., & Sodhi, S. S. (2020). *Comparative Analysis of Gaussian Filter, Median Filter and Denoise Autoenocoder*. https://doi.org/10.23919/indiacom49435.2020.9083712

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Moosavi, S. R., Qajar, J., & Riazi, M. (2018). A comparison of methods for denoising of well test pressure data. *Journal of Petroleum Exploration and Production Technology*, 8(4), 1519–1534. https://doi.org/10.1007/s13202-017-0427-y

Moradi, M. (2022). Wavelet transform approach for denoising and decomposition of satellite-derived ocean color time-series: Selection of optimal mother wavelet. *Advances in Space Research,* 69(7), 2724–2744. https://doi.org/10.1016/j.asr.2022.01.023

Murugan, S. P., & Youvaraj, G. P. (2020). On Franklin wavelets. *Asia-European Journal of Mathematics/Asian-European Journal of Mathematics*, *14*(07), 2150126. https://doi.org/10.1142/s1793557121501266

Nalley, D., Adamowski, J., & Khalil, B. (2012). Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). *Journal of Hydrology*, 475, 204–228. https://doi.org/10.1016/j.jhydrol.2012.09.049

Nguyen, T. T., Tran, D. T., & Pham, H. T. (2020). "Wavelet denoising in financial time series forecasting: A review and empirical analysis." *Expert Systems with Applications*, 148, 113248. doi: 10.1016/j.eswa.2019.113248.

Olokoyo, F. O. (2011). Determinants of Commercial Banks' Lending Behavior in Nigeria. *International Journal of Financial Research*, 2(2). https://doi.org/10.5430/ijfr.v2n2p61

Olugbenga, A., & Adesoji, O. (2017). "The Impact of Non-Performing Loans on Bank Stability in Nigeria: A Case Study of Nigerian Banks." *Journal of Finance and Economics*, 5(2), 45-55.

Onoh, John & Nwachukwu, Rev. (2017). Monetary Policy and Credit Delivery in Commercial Banks – Evidence from Nigeria. *International Journal of Social Sciences and Management Research*. 3. 2545-5303.

Park, R. E. (1966). "Estimation with Heteroscedastic Error Terms." *Econometrica, 34*(4), 888-891.

Percival, D. B., & Walden, A. T. (2000). *Wavelet Methods for Time SeriesAnalysis*. https://doi.org/10.1017/cbo9780511841040

Quinones, L., & Tibi, R. (2024). Denoising Seismic Waveforms Using a Wavelet-Transform-Based Machine-Learning Method. the Bulletin of the Seismological Society of America/Bulletin of the Seismological Society of America. https://doi.org/10.1785/0120230304

Ramsey, J. B. (1996). "Wavelets in Economics and Finance: Past and Future." *Studies in Nonlinear Dynamics & Econometrics,* 1(1), 1-20.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (2006). *Applied Regression Analysis*. Springer Science & Business Media.

Sardy, S., Tseng, P., & Bruce, A. (2001). Robust wavelet denoising. *IEEE Transactions on Signal Processing*, *49*(6), 1146–1152. https://doi.org/10.1109/78.923297

Saseendran, Arun & Setia, Lovish & Chhabria, Viren & Chakraborty, Debrup & Barman Roy, Aneek. (2019). Impact of Noise in Dataset on Machine Learning Algorithms. 10.13140/RG.2.2.25669.91369.

Selesnick, I. W., Graber, H. L., Pfeil, D. S., & Barbour, R. L. (2014). Simultaneous Low-Pass Filtering and Total Variation Denoising. *IEEE Transactions on Signal Processing*, *62*(5), 1109–1124. https://doi.org/10.1109/tsp.2014.2298836

Shah, Z. H., Müller, M., Wang, T. C., Scheidig, P. M., Schneider, A., Schüttpelz, M., Huser, T., & Schenck, W. (2021). Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images. *Photonics Research*, *9*(5), B168. https://doi.org/10.1364/prj.416437

Shapiro, S. S., & Wilk, M. B. (1965). "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika, 52*(3/4), 591-611.

Smith, A. B., & Johnson, C. D. (2018). "The impact of wavelet denoising on economic time series forecasting." *Journal of Forecasting*, 37(5), 593-609. doi: 10.1002/for.2524.

Spearman, C. (1904). "The Proof and Measurement of Association Between Two Things." *American Journal of Psychology, 15*(1), 72-101.

Srivastava, P. K., Pradhan, R. K., Petropoulos, G. P., Pandey, V., Gupta, M., Yaduvanshi, A., Wan Jaafar, W. Z., Mall, R. K., & Sahai, A. K. (2021). *Long-term trend analysis of precipitation and extreme events over Kosi river basin in India*. *Water*, *13*(12), 1695. https://doi.org/10.3390/w13121695

Strang, G., & Nguyen, T. (1996). Wavelets and Filter Banks. Wellesley-Cambridge Press.

Ugoani, J. N. N. (2016). Nonperforming Loans Portfolio and Its Effect on Bank Profitability In Nigeria. *Independent Journal of Management & Production*, 7(2), 303–319. https://doi.org/10.14807/ijmp.v7i2.406

Ugwuanyi, I. J. C., Obinne, G., & Efanga, U. O. (2022). Non-performing Loans and Banks' Profitability in Nigeria. *Indo-Asian Journal of Finance and Accounting*, 3(2), 97–116. https://doi.org/10.47509/iajfa.2022.v03i02.03

Vidakovic, B. (1999). "Statistical Modeling by Wavelets." John Wiley & Sons.

Wink, A., & Roerdink, J. (2004). Denoising Functional MR Images: A Comparison of Wavelet Denoising and Gaussian Smoothing. *IEEE Transactions on Medical Imaging*, *23*(3), 374–387. https://doi.org/10.1109/tmi.2004.824234

Wu, T. P., Wu, H. C., Liu, S. B., Hsueh, H. P., & Wang, C. M. (2020). Causality between peer-to-peer lending and bank lending in China: evidence from a panel data approach. *The Singapore Economic Review,* 65(06), 1537–1557. https://doi.org/10.1142/s0217590820500332

Xu, X., & Peng, Y. (2014). "Understanding bank loan default prediction with an interpretable machine learning model." *Expert Systems with Applications*, 41(4), 2227-2234.

Zhang, Y., Zhang, L., & Guo, W. (2019). "An empirical study on the application of wavelet transform in financial time series forecasting." *Journal of Computational and Applied Mathematics*, 359, 197-212. doi: 10.1016/j.cam.2018.11.024.