Wavelet Transform as an Alternative to Power Transformation in Time Series Analysis

C.J. Ogbonna^{1,a*}, C.J. Nweke^{2,b}, E.C. Nwogu^{1,c}, I.S. Iwueze^{1,d}

¹Department of Statistics, Federal University of Technology, Owerri, Nigeria

²Department of Mathematics/Statistics/Computer Science & Informatics, Federal University, Ndufu-Alike, Ikwo, Nigeria

^ad_ogbonna59@yahoo.com, ^bcj_nweke@yahoo.com, ^cecnwogu@yahoo.com,
^disiwueze@yahoo.com

Keywords: Transformation, Wavelet, Filter and Forecast

Abstract. This study examines the discrete wavelet transform as a transformation technique in the analysis of non-stationary time series while comparing it with power transformation. A test for constant variance and choice of appropriate transformation is made using Bartlett's test for constant variance while the Daubechies 4 (D4) Maximal Overlap Discrete Wavelet Transform (DWT) is used for wavelet transform. The stationarity of the transformed (power and wavelet) series is examined with Augmented Dickey-Fuller Unit Root Test (ADF). The stationary series is modeled with Autoregressive Moving Average (ARMA) Model technique. The model precision in terms of goodness of fit is ascertained using information criteria (AIC, BIC and SBC) while the forecast performance is evaluated with RMSE, MAD, and MAPE. The study data are the Nigeria Exchange Rate (2004-2014) and the Nigeria External Reserve (1995-2010). The results of the analysis show that the power transformed series of the exchange rate data admits a random walk (ARIMA (0, 1, 0)) model while its wavelet equivalent is adequately fitted to ARIMA (1,1,0). Similarly, the power transformed version of the External Reserve is adequately fitted to ARIMA (3, 1, 0) while its wavelet transform equivalent is adequately fitted to ARIMA (0, 1, 3). In terms of model precision (goodness - of - fit), the model for the power transformed series is found to have better fit for exchange rate data while model for wavelet transformed series is found to have better fit for external reserve data. In forecast performance, the model for wavelet transformed series outperformed the model for power transformed series. Therefore, we recommend that wavelet transform be used when time series data is non-stationary in variance and our interest is majorly on forecast.

1.0 Introduction

In several organizations, managerial decisions are largely based on the available information of the past and present observations and possibly on the process that generate such observations. A time series data provides such information. Time series is used to represent the characterized time course of behavior of wide range of several systems which could be biological, physical or economical. The utility of the time series data lies in the result of the time series analysis. Such analysis will be helpful in achieving the aim for collection of such data which could be for description (exposing the main properties of a series), explanation (revealing the relationship between variables of a series especially when observations are taken on two or more variables), forecasting (prediction of the future values of a series) and control (taking appropriate corrective actions) [1].

To analyze any time series data, time series analysis techniques are adopted. The commonly used techniques are: descriptive technique, probability models technique and spectral density analysis technique. The inference based on the descriptive method and probability models is often referred to as analysis in time domain while inference based on spectral density function is referred to analysis in frequency domain [2,3,4,5].

All these models assume that the error component of the series (e_t) is normally distributed with zero mean and constant variance (σ^2) or that the series is normally distributed with constant mean (μ) and constant variance (σ^2). When any study data violates any or all of these assumptions, the series is subjected to transformation. Transformation helps to (i) stabilize the variance of a series, (ii) make the seasonal effect when present additive and (iii) make the data normally distributed [1]. One of the transformations commonly used is the power transformation developed by [6]. [7] noted that the power transformation (i) changes the scale of the original series, (ii) may introduce bias in the forecast especially when data have to be transformed back to its original scale and (iii) often the transformed series have no physical interpretation. [1] argued that transformation alone may not be helpful when variance changes through time in the absence of trend. In such case (i.e., when variance changes through time in the absence of trend), he recommended that a model that allows for changes in variance should be considered.

Wavelet method is one such method that allows for changes in variance which has been found to be useful in time series analysis. It involves decomposition, de-noising and reconstruction of series. Decomposition involves breaking down time series into two main components namely the detail and the smooth components, de-noising deals with the removal of the non significant components of the series, while reconstruction involves recovering of the original series devoid of noise. Unlike the power transformation, wavelet method does not change the scale of the series, poses no problem in its interpretation and does not rely on the assumption of any underlying distribution of the study data. Additionally, wavelet transformation method allow for decomposition of a series without knowing the underlying functional form of the series [8]. Could this lead to an improved model and forecast performance from those based on power transformation? This and other related questions are what this study intends to address. Therefore, the objective of this study is to examine the precision (in terms of goodness-of-fit) and forecast performances of the models for wavelet transformed series while comparing it with models for power transformed time series.

2.0 Literature Review

Various transformations exist, but the power transformation developed by [6] is often used. This transformation requires a correct choice of the transformation parameter often denoted as (λ). [4] suggested using a maximum likelihood value for the choice of the value of λ that results in the smallest residual sum of squares. [9] proposed a Bayesian method to choose the value of λ for a given model structure. The correct choice of the transformation parameter (λ), the simultaneous transformation and fitting of the model of a given series are the noticed limitations in the use of [6] power transformation. To remedy these limitations, [10] have shown how to apply Bartlett's transformation to time series data. Accordingly, they regress the natural logarithms of the group standard deviation ($\hat{\sigma}_i, i = 1, 2, \dots, m$) against the natural logarithms of the group means ($\bar{X}_i, i = 1, 2, \dots$) of time series data arranged chronologically in equal groups and determine the slope (β) of the relationship. [11] derived a confidence interval for the index of a power transformation that stabilizes the variance of a time series. They claimed that the confidence interval for the minimum coefficient of variation can also be used to construct confidence interval for any coefficient of variation. [12] used Box-Cox transformation approach to transform a streamflow time series data to turn the non-Gaussian heavy tailed distribution to a nearly Gaussian series. [13] applied log-transformation in time series modeling of US macroeconomic data. He demonstrated that the claim previously made concerning improvement in forecast accuracy following bias correction for the transformed data were not generally well founded. [14] in his work applied differencing and Box-Cox transformation on a non-stationary series prior to application of ARMA model. He discovered that the technique was not optimal for forecasting of the study data. Therefore, he recommended a method for modeling time series data with unstable roots and changing parameters. The appropriateness in context of out-of-sample forecast of vector autoregressive models (VAR) when the series is log-transformed was examined by [15]. The

examination was on out-of-sample forecast of GDP of multi-country set-up. They discovered that the forecast performed better when the series is transformed. [16] in their work highlighted that before application of ARMA model, achieving variance stationarity via power transformations contributes to improvement in forecasts especially for long forecasting horizons. [17] applied power transformation to achieve variance stationarity in modeling of Nigeria external reserve. Using Bartlett's version of power transformation, they discovered that the appropriate transformation for the study data was logarithmic transformation. They further fitted ARIMA (2, 1, 0) to the transformed series.

[18] investigated the potential of wavelet methods in the analysis of biological sequences as a complement method to those currently in use. He used the maximal overlap discrete wavelet (MODWT) to extract the relevant structural features from the data. [19] compared the discrete wavelet transform and maximal overlap discrete wavelet transform of Dow Jones index of US stock market. They used a unit root test to establish the non-stationarity of the study series and different wavelet transform families were examined. They discovered that the discrete wavelet transform outperformed the maximal overlap discrete wavelet transform. [20] examined the modeling and forecasting using wavelet while comparing it with ARIMA and X-II modeling approach. The study data were US dollar against DM exchange rate data and ten steps ahead forecast was made. The outcome of the forecast using average percentage forecast error (APE) criterion show that wavelet approach is the best. [21] studied the performances of wavelet-ARIMA and wavelet-ANN models for temperature data in the Northeastern Bangladesh. Their findings indicate that the performance based on the predictive capability was more effective in wavelet-ARIMA model than in the wavelet-ANN model. [22] compared the wavelet method of time series and other existing method like ARIMA and Census X-12 methods. They came to a conclusion that when a series is characterized with medium and long term structure, wavelet reduces the error forecast substantially, but increases the error forecast when such data have only short term structure.

3.0 Methodology

This Section discusses the methodology employed in this study. Sub-section 3.1 presents the methods of evaluation for possible transformation and choice of appropriate transformation (using Bartlett test for constant variance and Discrete Wavelet Transform (DWT)). Sub-section 3.2 discusses the Augmented Dickey-Fuller (ADF) Unit Root Test, Sub -sections 3.3 and 3.4 present the ARMA model and model selection criteria, respectively, while the forecast accuracy measures are discussed in Sub-section 3.5.

3.1 Methods of Evaluation for Possible Transformation and Choice of Appropriate Transformation

3.1.1 Bartlett's Transformation

[10] have shown how to apply [23] transformation techniques to time series data without considering the time series model structure. According to [10], the Bartlett transformation for time series data involves regressing the natural logarithm of the group standard deviations ($\hat{\sigma}_i$, $i=1,2,\dots,m$) against the natural logarithm of group means ($\bar{X}_i$, $i=1,2,\dots,m$) of series that is arranged chronologically in equal groups and determine the slope β , of the relationship:

$$\text{Log}_e \hat{\sigma}_i = \alpha + \beta \text{Log}_e \bar{X}_i + \varepsilon_i \quad (1)$$

Application of Eq. 1 will lead to a power transformation which is given by:

$$Y_t = \begin{cases} \log_e X_t, & \beta = 1 \\ X_t^{1-\beta}, & \beta \neq 1 \end{cases} \quad (2)$$

where Y_t is the transformed series at time t , X_t is the original value of the series at time t , α , β are the regression coefficients given in Eq. 1 and ε_t is zero mean white noise with constant variance (σ^2) .

To test for the significance of the slope β , the following hypothesis is formulated. That is, $H_0: \beta=0$ against the alternative: $H_1: \beta \neq 0$. The test statistic follows t - distribution with $n-1$ degrees of freedom and is given by:

$$t = \frac{\hat{\beta} - 0}{\sqrt{\text{Var}(\hat{\beta})}} \quad (3)$$

When β is significantly different from zero, then the series is non-stationary in variance and appropriate transformation using Eq. 2 is applied. [10] further suggested that the value of $\hat{\beta}$ should be substituted directly for the transformation instead of possibly using an approximation.

3.1.2 Discrete Wavelet Transform (DWT)

This Section considers wavelet transformation method for transforming time series data (in time domain) to wavelet coefficients (data in the wavelet domain). It involves decomposition, denoising and reconstruction. These steps are briefly discussed in what follows.

(a) Discrete Wavelet Decomposition

Given a real valued time series $\{x_t, t \in T\}$ of a dyadic length, $n=2^J$ (a length that is of power of 2), where J is a positive integer, and choosing an appropriate transformation filters, we compute the wavelet coefficients. For a sequence x_t , the J -step (level) wavelet transformation is given by:

$$w = W_J x_t \quad (4)$$

where W is column vector (length, $n=2^J$) comprising of a set of wavelet coefficients (detail coefficients ($d_{k,j}$)) and a set of averages (scaling or smooth coefficients ($v_{k,j}$)), W_J is the J -step $n \times n$ real valued orthogonal matrix often called “filter bank” defining the discrete wavelet transform (DWT) and satisfying $W_J^{-1} = W_J^T$ and $W_J^T W_J = I_n$, x_t is column vector with value of the original series at time t .

The matrix W_J is built using a finite list of numbers called filters [wavelet filters (h_l) and scaling filters (g_l)] and is partitioned into two equal part such that the first $\frac{n}{2}$ rows is a matrix of wavelet filters which used to produce the detail coefficients ($d_{k,j}$) and the last $\frac{n}{2}$ rows is a matrix of scaling filters is used to produce the smooth coefficients ($v_{k,j}$) which serve as an input data in the next stage of convolution. The first smooth coefficients (v_1) is convolved with the wavelet filter (h_l) and scaling filter (g_l) to produce the next detail coefficient (d_2) and smooth coefficient (v_2) and so on. This process is repeated up to:

$$J = \text{Log}_2(n) \quad (5)$$

Therefore, for J steps, we let $r=1,2,\dots,J$ the r^{th} step wavelet transformation becomes:

$$W_r = \begin{bmatrix} \begin{bmatrix} H_r \\ G_r \end{bmatrix} \cdot H_{r-1} \dots H_1 \\ G_{r-1} \\ \vdots \\ G_1 \end{bmatrix} \quad (6)$$

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

$$W_1 = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix} \quad (6a)$$

$$W_2 = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} * H_1 \\ G_1 \end{bmatrix} \quad (6b)$$

$$W_3 = \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} * H_2 * H_1 \\ G_2 \\ G_1 \end{bmatrix} \quad (6c)$$

The partition matrix H_k and G_k is a circulant matrix of the order $(2^{J-k} \times 2^{J-k+1})$, $k = 1, 2, 3, \dots$ with entries given by:

$$h_l, l = (M-1) - 2(i-1) + (j-1) \bmod 2^{J-k+1} \quad (7)$$

at the position (i, j) . Alternatively, after the determination of first row, other i^{th} row can as well 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). G_k has the same dimension with entries in Eq. 7 replaced with the smooth filters (g_l). The relationship between the wavelet details filters (h_l) and the scaling or smooth filters (g_l) can be represented thus:

$$g_l = (-1)^{l+1} h_{L-1-l}, \quad l = 0, 1, \dots, L-1 \quad (8)$$

where L is the wavelet filter width and:

$$h_l = \begin{cases} h_l, & 0 \leq l \leq L-1 \\ 0, & L \leq l \leq n-1 \end{cases} \quad (9)$$

Using Eq. 7 and Eq. 9, H_k can be populated as represented in Eq. 10:

$$H_k = \begin{bmatrix} h_{M-1} & h_M & h_{M+1} & h_{M+2} & \cdot & \cdot & \cdot & h_{M+n-4} & h_{M+n-3} & h_{M+n-2} \\ h_{M+n-3} & h_{M+n-2} & h_{M+n-1} & h_M & \cdot & \cdot & \cdot & h_{M+n-6} & h_{M+n-5} & h_{M+n-4} \\ h_{M+n-5} & h_{M+n-4} & h_{M+n-3} & h_{M+n-2} & \cdot & \cdot & \cdot & h_{M+n-8} & h_{M+n-7} & h_{M+n-6} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ h_M & h_{M+1} & h_{M+2} & h_{M+3} & \cdot & \cdot & \cdot & h_{M+n-3} & h_{M+n-2} & h_{M-1} \end{bmatrix} \quad (10)$$

Similarly, G_k matrix is expressed as Eq. 10 using g_l of Eq. 8.

The detail filters and scaling (smooth) filters in Eq. 7 and Eq. 8, respectively, satisfy the following conditions as given in [24]:

(a) It must have a unit norm or energy:

$$\sum_{l=0}^{L-1} h_l^2 = 1 \quad (11)$$

(b) It must sum up to square root of two:

$$\sum_{l=0}^{L-1} h_l = \sqrt{2} \quad (12)$$

(c) It must be orthogonal to its even shifts:

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \begin{cases} 1, & n=0 \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

However, when n is not dyadic in length, i.e., $n \neq 2^j$ a non-orthogonal discrete wavelet transform known as Maximal Overlap Discrete Wavelet Transform (MODWT) is adopted. It involves replacing w , W_J and J in Eq. 4 with $\tilde{w}$, $\tilde{W}_{J_0}$ and J_0 , respectively, to obtain:

$$\tilde{w} = \tilde{W}_{J_0} x_t \quad (14)$$

where $\tilde{w}$, $\tilde{W}_{J_0}$ and J_0 are the MODWT version of w , W_J and J given in Eq. 4. The entries of the partition matrix (H_k and G_k) of the matrix $\tilde{W}_{J_0}$ have detail filters values:

$$\tilde{h}_l = \frac{h_l}{\sqrt{2}} \quad (15)$$

and smooth filters values given by:

$$\tilde{g}_l = \frac{g_l}{\sqrt{2}} \quad (16)$$

where h_l are the detail filters given in Eq. 7 and g_l are the smooth filters given in Eq. 8.

The value of $J_0 < J$ and is given by:

$$J_0 = \text{Log}_2 \left[\frac{n}{L-1} + 1 \right] \quad (17)$$

where J_0 is the wavelet decomposition level for MODWT, n is the number of observations, and L is the wavelet filter width.

The detail and smooth filters of MODWT have the following properties as given in [25]:

$$\tilde{h}_l = \begin{cases} \tilde{h}_l, & 0 \leq l \leq L-1 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

$$\sum_{l=0}^{L-1} \tilde{h}_l^2 = \frac{1}{2} \quad (19)$$

$$\sum_{l=0}^{L-1} \tilde{h}_l = 1 \quad (20)$$

$$\sum_{l=0}^{L-1} \tilde{h}_l \tilde{h}_{l+2n} = 0 \quad (21)$$

(b) Data De-noising via wavelet

The detail coefficients of the wavelet decomposition are de-noised using a threshold proposed by [26] known as universal threshold given by:

$$\lambda_u = \sigma \sqrt{(2 \log_e n)} \quad (22)$$

where λ_u is the Universal threshold, n is the number of observations and $\hat{\sigma}$ is the estimate of the noise level at the finest scale.

The estimate of the noise level given in Eq. 22 is given by:

$$\hat{\sigma} = \frac{\text{median}(|d_{J-1} - \text{median}(d_{J-1})|)}{0.6745} \quad (23)$$

where $\hat{\sigma}$ is given in Eq. 22, d_{J-1} is the finest scale detail wavelet coefficient and 0.6745 is a scale factor. The threshold wavelet coefficients are obtained using either the hard thresholding rule given by:

$$\delta_{\lambda}^H = \begin{cases} 0, & \text{if } |d_{j,k}| \leq \lambda \\ d_{j,k}, & \text{if } |d_{j,k}| > \lambda \end{cases} \quad (24)$$

where $d_{j,k}$ is the detail wavelet coefficient at j -th level and k th scale, λ is the universal threshold given in Eq. 22. The wavelet coefficient could as well be obtained using soft thresholding rule known as wavelet shrinkage which is given by:

$$\delta_{\lambda}^S = \begin{cases} 0, & \text{if } |d_{j,k}| \leq \lambda \\ d_{j,k} - \lambda, & \text{if } d_{j,k} > \lambda \\ d_{j,k} + \lambda, & \text{if } d_{j,k} < -\lambda \end{cases} \quad (25)$$

where $d_{j,k}$, λ are given in Eq. 22 and Eq. 23.

(c) Wavelet Reconstruction using Multiresolution Analysis (MRA)

Given the discrete wavelet transform of the series x_t as $w = Wx_t$, the transformed version of x_t is given by:

$$y_t = W_J^{-1} w^* \quad (26)$$

where y_t is the wavelet transformed series at time t , W_J^{-1} is the matrix inverse of the matrix W_J given in Eq. 4, w^* is the de-noised version of the wavelet (detail and smooth) coefficients in Eq. 4. Eq. 26 can further be expanded to give:

$$y_t = \sum_{q=1}^J D_q + S_J \quad (27)$$

where D_j is the q th level wavelet details for $q = 1, 2, \dots, J$ and S_J is the J^{th} smooth coefficient.

The detail coefficient in Eq. 27 is obtained using:

$$D_q = W_q^{-1} w_q^* \quad (28)$$

while the J^{th} smooth is given by:

$$S_J = V_J^{-1} v_J^* \quad (29)$$

where W_q , is given in Eq. 28 and w_q^*, v_J^* are de-noised version of wavelet detail and smooth coefficients, respectively.

3.2 Augmented Dickey-Fuller Unit Root Test.

Having arrived at a transformed series (Y_t) which is stable in variance either through Bartlett's transformation or DWT, another challenge is to ascertain the stationarity in mean of the transformed series (Y_t). One of the ways to test for this is using Augmented Dickey-Fuller Unit root test. The test was developed by [27] and can be illustrated thus; given a model:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + e_t \quad (30)$$

where $\Delta Y_t = Y_t - Y_{t-1}$, Y_t is the transformed series at time t , $\beta_1, \beta_2, \delta, \alpha_i$ ($i = 1, 2, \dots, m$) are constants and e_t is a white noise with zero mean and constant variance (σ_e^2). The null hypothesis is: $H_0: \delta = 0$ against the alternative: $H_1: \delta < 0$.

The estimated t-value follows the τ (tau) statistic. If the null hypothesis is rejected, the implication is that the series is stationary in mean.

If we do not reject the null hypothesis, and assume that the trend is stochastic and the transformed series Y_t is not seasonal, then the series is made stationary by differencing. This will result to a stationary series given by:

$$Z_t = \nabla^d Y_t, \quad d \geq 1, \quad (31)$$

where Z_t is the differenced series at time t , Y_t is the transformed series at time t given in Eq. 2 or Eq. 29, ∇ is the differencing operator such that $\nabla Y_t = (1 - B)Y_t = Y_t - Y_{t-1}$ and d is the number of differencing.

3.3 The Probability (ARMA) Model

Given the stationary series $\{Z_t, t \in T\}$ obtained in Equation (31), Autoregressive Moving Average (ARMA) Process involves the combination of an autoregressive (AR) and moving average (MA) process. It is of the order (p, q) and is given by:

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (32)$$

or

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (33)$$

That is,

$$\phi(B)\tilde{Z}_t = \theta(B)e_t \quad (34)$$

where Z_t is the stationary series at time t , $\tilde{Z}_t = Z_t - \mu$ is deviation of the process from some origin, $\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ are constants, e_t is a white noise process, $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ and $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ are polynomial of degree p and q , respectively, in B .

We require the root of characteristic equation $\phi(B) = 0$ and $\theta(B) = 0$ to lie outside unit circle for the process to be stationary and invertible, respectively.

3.4 Model Precision (goodness-of-fit) Measures

Assuming that an ARMA model of k parameters is fitted to a stationary time series (Z_t) given in Eq. 31. To assess the precision of the model in terms goodness-of-fit, the following information criteria will be used:

(a) Akaike Information Criterion (AIC)

[28] introduced a criterion for measurement of goodness-of-fit of a model. The information is given by:

$$AIC = -2(\text{Maximum Likelihood}) + 2k \approx n \text{Log}_e(\hat{\sigma}_e^2) + 2k \quad (35)$$

where n is the number of observations, $\hat{\sigma}_e^2$ is the estimate of the error variance and k is the number of independent parameters estimated in the model ($k = p + q + 1$).

(b) Bayesian Information Criterion (BIC).

[29] developed an extension of Akaike Information criterion using Bayesian approach. It is given by:

$$BIC = n \text{Log}_e(\hat{\sigma}_e^2) + k \text{Log}_e(n) \quad (36)$$

where $n, \hat{\sigma}_e^2, k$ have the same meaning as given in Eq. 35.

(c) Schwartz's Bayesian Criterion (SBC)

[30] suggested using Bayesian criterion for model selection. It is given by:

$$SBC = n \text{Log}_e(\hat{\sigma}_e^2) + k \text{Log}_e(n) + k \quad (37)$$

where $n, \hat{\sigma}_e^2, K$ have the same meaning as given in Eq. 35.

The model with a minimum value of information in Eq. 35, Eq. 36 and Eq. 37 are adopted as a model with the best fit.

3.5 Forecast Accuracy Measures

Let the l -step-ahead forecast error be:

$$e_t(l) = X_{t+l} - \hat{X}_t(l) \quad (38)$$

The two commonly used statistics are:

a. Measures based on Errors

(i) Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{m} \sum_{l=1}^m e_l^2} \quad (39)$$

(ii) Mean Absolute Deviation

$$MAD = \frac{1}{m} \sum_{l=1}^m |e_l| \quad (40)$$

b. Measures Based on Percentage Errors

Mean Absolute Percentage Error given by:

$$MAPE = \left(\frac{1}{m} \sum_{l=1}^m \left| \frac{e_l}{X_{n+l}} \right| \right) \times 100\% \quad (41)$$

4.0 Results and Discussion

The time series plot of the monthly exchange rate and the monthly external reserve are as displayed in Fig. 1 and Fig. 2, respectively.

4.1 Tests for Constant Variance and Choice of Appropriate Transformation

The results of the application of the Bartlett's transformation techniques in Eq. 1 to the study data (exchange rate and the external reserve) are as shown in Table 1.

Table 1: Results of the application of Bartlett's transformation

Series	Slope(β_1)	P-value
Exchange rate (Original data)	-9.134	0.003
Exchange rate (Power transformed) series)	0.114	0.661
Exchange rate (Wavelet transformed) series)	-1.589	0.125
External reserve (Original data)	0.5285	0.012
External reserve (Power transformed) series)	0.0529	0.881
External reserve (Wavelet transformed) series)	0.3741	0.441

Level of significance (α)=0.05

From Table 1, it could be seen that the original series of the exchange rate and external reserve admits transformation which suggest variance heterogeneity of the series. While the direct substitution of the value of slope ($\beta = -9.124$) in Eq. 2 was used for the transformation of the exchange rate data, the square root transformation which is appropriate when the slope (β) equals 0.5 was used for the external reserve data. To further confirm if stationarity in variance had been achieved, the same test (Bartlett's test for constant variance) was carried out on the transformed (power and wavelet) series. The results show that no further transformation is required in both study data.

4.2 Wavelet Decomposition, De-noising and Reconstruction of Time Series Data

The results of application of wavelet based transformation (decomposition, de-noising and reconstruction) are as displayed in Table 2 while the plot of the wavelet coefficients are shown in Fig. 3 and Fig. 4 for the exchange rate and external reserve respectively. The result is the decomposition based on MODWT of the Duabachies 4 (D4) with its filter matrix represented in Eq. 10. The threshold used was the universal threshold given in Eq. 24 while the de-nosing was done using the soft threshold given in Eq. 27. The wavelet reconstruction was done using the multiresolution analysis given in Eq. 29. From the result in Table 2, the wavelet filter width used for both exchange rate and external reserve is 4 which is Daubechies wavelet (D4).

Fig 3: Wavelet coefficient plot for exchange rate

Fig 4: Wavelet coefficient plot for external reserve

Table 2: The results of the wavelet transformation analysis

Series	N	Wavelet Width (L)	Decomposition Level (J_0)	Noise Level ($\hat{\sigma}$)	Threshold value (λ)
Exchange rate	132	4	4	1.8398	5.7494
External Reserve	192	4	5	188.3039	610.6098

From Table 2, it could be seen that noise (signal-to-noise) level for exchange rate is 1.8398 with the threshold value of 5.7494 while the noise level for external reserve is 188.3039 with the threshold value of 610.6098. The noise level was estimated using Eq. 25 while the decomposition level was estimated using Eq. 19.

4.3 Augmented Dickey-Fuller Unit Root Test for Stationarity

The result of the application of Augmented Dickey-Fuller (ADF) unit root test for mean stationarity for the (power and wavelet) transformed series under study (exchange rate and external reserve) are as shown in Table 3. From Table 3, the ADF statistic value for the (power and wavelet) transformed series for exchange rate data is (-2.8680) and (-2.4204) with p-value of 0.1764 and 0.3674 respectively, suggesting non-stationarity in mean. The corresponding ADF values for first differenced series are -11.4714 and -6.4848 with p-value of 0.000 for the power and wavelet transformed series respectively.

Table 3: Results for Augmented Dickey-Fuller Unit Root Test for Stationarity

Trans-formation	Series	ADF	P-value	Max Lag	AIC	D.W	Remark
Power Transformed series.	Exchange rate (level)	-2.8680	0.1764	4	0.7598	1.8859	Non-stationary
	Exchange rate (First Differenced)	-11.4714*	0.0000	4	0.8145	1.8700	Stationary
Wavelet Transformed Series.	Exchange rate (level)	-2.4204	0.3474	12	3.6609	1.9899	Non-Stationary
	Exchange rate (First Differenced)	-6.4848*	0.0000	12	3.6978	2.1704	Stationary
Power Transformed series.	External reserve (level)	-1.3508	0.8719	3	5.9272	1.9818	Non-Stationary
	External reserve (First differenced)	-9.1424*	0.0000	3	5.9266	1.9945	Stationary
Wavelet Transformed Series.	External reserve (level)	-1.3524	0.8714	3	4.8069	1.9858	Non-stationary
	External reserve (First differenced)	-6.8824*	0.0000	3	4.8062	2.006	Stationary

* - significant at $\alpha = 0.05$.

Similarly, for the external reserve data, the ADF value for the (power and wavelet) transformed series are -1.3508 and -1.3524 with p-value of 0.8719 and 0.8714 respectively. This again suggest non-stationarity in mean while the corresponding ADF value for the first differenced series are -9.1424 and -6.8824 with p-value of 0.0000 for the (power and wavelet) transformed series, respectively.

4.4 Model Identification and Parameter Estimation

The results for the estimates of parameters for the models of exchange rate data and external reserve data are as shown in Table 4. The results in Table 4 suggest that the power transformed series of the exchange rate data admits a random walk. This is evident based on the first difference of the power transformed series is found to be a white noise. On the wavelet transformed series, ARIMA (1, 1, 0) is adequately fitted to the data with $\phi_1 = 0.5241$ with a p-value of 0.0000.

For the external reserve data, the result shows that the power transformed series admitted ARIMA (3, 1, 0) with $\phi_1 = 0.0411$, $\phi_2 = 0.4174$ and $\phi_3 = -0.4369$ whereas the wavelet transformed series has ARIMA(0,1,3) with $\theta_1 = -0.1762$, $\theta_2 = -0.5175$ and $\theta_3 = 0.1814$ that fits the data adequately. The values of the AIC, BIC and SBC were estimated using Eq. 35, Eq. 36 and Eq. 37, respectively.

Based on the value of the information criteria in Table 4, it could be observed that the model for power transformed series has a minimum value than the model for the wavelet transformed series of the exchange rate data. This suggests that the exchange rate data in the period under review has a better fit than the model for the wavelet transformed. Contrary to the result in exchange rate data, the information value for the models of wavelet transformed series was smaller than the model for the power transformed series, thus suggesting a better fit for wavelet transformed series.

Table 4: Parameter estimate of the study data for the power and wavelet transformed Series

Model	Exchange rate		External Reserve	
	Power Transformation	Wavelet Transform	Power Transformation	Wavelet Transformation
Parameter	ARIMA(0,1,0)	ARIMA(1,1,0)	ARIMA(3,1,0)	ARIMA(0,1,3)
ϕ_1	-	0.5241 (0.0000)	0.0411 (0.5321)	-
ϕ_2	-	-	0.4174 (0.0000)	-
ϕ_3	-	-	-0.4369 (0.000)	-
θ_1	-	-	-	-0.1762 (0.013)
θ_2	-	-	-	-0.5175 (0.000)
θ_3	-	-	-	0.1814 (0.0110)
σ_e^2	0.1219	4.1546	27.6408	11.1710
AIC	-275.8012	191.000	645.2987	471.3577
BIC	-271.9184	197.7621	650.5759	484.3876
SBC	-272.9184	199.7621	654.5759	488.3876

() – p-value at $\alpha=0.05$ level of significance

4.5 Model Diagnostic Check

The results of the descriptive statistics for the residuals of the models of the study data for power and wavelet transform series are as shown in Table 5. The results in Table 5 show that the error means are not significantly different from zero and the distribution of each of the residual is normal or approximately normal. This is an indication of the adequacy of the model that is fitted to the series.

Table 5: Descriptive statistics of the residuals for the fitted models

	Exchange Rate		External Reserve	
	Power transformed	Wavelet transformed	Power transformed	Wavelet transformed
Statistics	ARIMA(0,1,0)	ARIMA(1,1,0)	ARIMA(3,1,0)	ARIMA(0,1,3)
Mean	0.01134	-0.0075	-0.5206	-0.0218
Median	0.0000	0.0645	-0.6195	-0.0375
Skewness	0.0454	-0.1483	0.1637	-0.3411
Kurtosis	0.5190	0.1175	0.2187	0.4278
Variance	0.1219	4.1546	27.6408	11.1710
Minimum	-1.3273	-5.5453	-14.5606	-9.9889
Maximum	1.1239	5.7251	16.2719	8.0023

4.6 Forecast and Forecast Accuracy Measures

The results of the forecast and their corresponding accuracy measures made on the study data (exchange rate and external reserve) are as shown in Table 6 and Table 7. The results show that the root mean square error (RMSE) of 54.4304 for the forecast of the model for power transformed series is greater than the RMSE of 42.8487 for the forecast of the model for the wavelet transformed series for exchange rate. Also it could be observed that mean absolute deviation (MAD) and mean absolute percentage error (MAPE) followed the same trend with root mean square error on the same series. This suggests that the model for the wavelet transform series outperformed the model for the power transformed series in terms of forecast.

Table 6: Forecast and Forecast accuracy measures for exchange rate

		Power Transformed	Wavelet Transformed
Month	Actual	ARIMA (0,1,0)	ARIMA (1,1,0)
January	254.39	242.9206	254.6295
February	274.79	242.9206	254.8396
March	295.6	242.9206	254.9567
April	294.24	242.9206	255.0249
May	304.79	242.9206	255.0675
June	306.06	242.9206	255.0967
July	306.41	242.9206	255.1189
August	307.21	242.9206	255.1374
September	302.55	242.9206	255.1540
October	302.26	242.9206	255.1696
November	299.38	242.9206	255.1846
December	295.39	242.9206	255.1994
RMSE		54.4304	42.8487
MAD		52.3352	40.2475
MAPE		0.1749	0.1339

Table 7: Forecast and Forecast accuracy measures for external reserve

		Power Transform	Wavelet Transform
Month	Actual	ARIMA(3,1,0)	ARIMA (0,1,3)
January	33131.83	32791.5937	31172.7260
February	33246.07	33023.8641	30567.9085
March	33221.80	33826.6969	31038.8945
April	32835.33	34042.7094	31251.0026
May	32100.81	34575.1521	31463.8330
June	31890.91	34620.1442	31677.3857
July	32521.71	35040.1037	31891.6606
August	32914.97	35131.9519	32106.6578
September	31740.23	35584.4603	32322.3773
October	32594.69	35749.3762	32538.8190
November	32125.22	36200.5816	32755.9830
December	32639.78	36384.1658	32973.8692
RMSE		2616.874	1311.455
MAD		2261.028	1024.686
MAPE		0.0699	0.0312

Equivalently, for the external reserve, the root mean square error of the forecast for the model of the power transformed series is found to be 2261.02 and 1311.455 for the model of the wavelet transformed series. Again, the root mean square error is smaller in model for the wavelet transform than the power transformation model. The same trend is seen in the other forecast accuracy measures. This again suggests that the model for wavelet transformed series outperformed the model for the power transformed series in terms of forecast. The plot of the actual and forecasts for power transformed series and wavelet transformed series are as displayed in Fig. 5 and Fig. 6 for exchange rate and external reserve, respectively.

Fig 5: Plot of the actual and forecast from power and wavelet transformed series - Exchange rate

Fig 6: Actual and forecast for power and wavelet transformed series - External reserve

5.0 Conclusion

Based on the results of this study, it has been shown that discrete wavelet transform (DWT) can serve as a transformation tool for a time series data and can serve as a better alternative to power transformation especially when a series is characterized with discontinuities and heterogeneous variance. Comparatively, on the study data, the models for the wavelet transformed series outperformed the models for the power transformation especially in forecast.

We therefore recommend that discrete wavelet transform should be used as an alternative to power transformation when a time series is characterized with discontinuities and heterogeneous variance.

References

- [1] C. Chatfield, The Analysis of Time Series: An introduction. Sixth Edition, Chapman & Hall/CRC Press Company Boca Raton London New York Washington, D.C., 2004.
- [2] F. E. Croxton, D.J. Cowden, S. Klein, Applied General Statistics. Third Edition, Prentice-Hall of India Private limited, New Delhi, 1967.
- [3] W.S. Wei, Time Series Analysis: Univariate and Multivariate Methods, Second Edition, Pearson Addison-Wesley Publishing Company, Inc. USA, 1990.
- [4] G.E.P. Box, G. M. Jenkins, G.C. Reinsel, Time series analysis: Forecasting and Control, 3rd edition. Prentice-Hall, Inc, Englewood Cliffs, 1994.
- [5] J. D. Hamilton, Time Series Analysis. Princeton University Press Princeton, New Jersey, 1994.
- [6] G.E.P. Box, D. R. Cox, An Analysis of Transformations. J. Royal Statistical Society. 26 (1964) 211-252.
- [7] J.W. Osborne, Notes on the use of data transformations. Journal of Practical Assessment, Research and Evaluation. 8(6) (2002).

-
- [8] J.B. Ramsey, The Contributions of Wavelets to the Analysis of Economic and Financial data, *Phil. Trans. R. Soc. Lond. A*.357 (1999) 2593-2606.
 - [9] C.W.S. Chen, J.C. Lee, On selecting a power transformation in time series analysis, *Journal of Forecasting*. 16 (1997) 343-354.
 - [10] A. C. Akpanta, I. S. Iwueze, On the applying the Bartlett transformation method to time Series Data, *Journal of Mathematical Sciences*. 20(3) (2009) 227-243.
 - [11] V.M. Guerrero, R. Perera, Variance stabilizing power transformation for time series, *Journal of Modern Applied Statistical Methods*. 3(2) (2004) 356-369.
 - [12] A.T. Cahill, Determination of changes in stream flow variance by means of wavelet-based test, *Water Resource Research*. 38(60) (2002) 1065-1068.
 - [13] John Haywood, On Log-Transformations, Vector Autoregressions and Empirical Evidence, The GSBGM working paper series, School of Econometrics and Finance, Victoria University of Wellington, New Zealand, 2000.
 - [14] C. Grillenzoni, Forecasting unstable and non stationary time series, *International Journal of Forecasting*. 14 (1998) 469-482.
 - [15] J. Mayr, D. Ulbricht, Log versus level in VAR forecasting. Discussion papers, Deutsches Institut for wirtschaftsforschung, Berlin, 2014.
 - [16] S. Makridakis, M. Hibon, ARMA models and the Box-Jenkins methodology, Working paper series. INSEAD, Funtainebleau, France, 1995.
 - [17] I.S. Iwueze, E.C. Nwogu, V.U. Nlebedim, Time series modeling of Nigeria external reserve. *CBN Journal of Applied Statistics*. 4(2) (2013) 111-128.
 - [18] M. Vannucci, Non-decimated wavelet analysis of biological sequences: Application to protein structure and genomics, *The Indian Journal of Statistics, Series B*. 63B(2) (2001) 218-233.
 - [19] A.A.A. Dghais, M.T. Ismail, A comparative study between discrete wavelet transform and maximal overlap discrete wavelet transform for testing stationarity, *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*. 7(12) (2013) 1159-1163.
 - [20] H. Wong et al., Modelling and forecasting by wavelet and the application to exchange rate, *Journal of Applied Statistics*. 30(5) (2003) 537-553.
 - [21] A.H. Nury, K. Hasan, J.B.Md. Alam, Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh, *Journal of King Saud University – Science*. (2015) 1-15.
 - [22] S. Schluter, C. Deuschle, Wavelet-based forecasting of ARIMA time series – An empirical comparison of different methods, *Managerial Economics*. 15(1) (2014) 107-131.
 - [23] M.S. Bartlett, The use of transformation, *Biometrika*. 3 (1947) 39-52.
 - [24] B. Vidakovic, Statistical Modeling by Wavelets, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1999.
 - [25] D.B. Percival, A.T. Walden, Wavelet Methods for Time Series Analysis, Cambridge UK: Cambridge University Press, 2000.
 - [26] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika*. 81(3) (1994) 425-455.
 - [27] S.E. Said, D.A. Dickey, Testing for unit roots in ARMA models of unknown order, *Biometrika*. 71(3) (1984) 599-607.

- [28] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, Proc. 2nd International Symposium on Information Theory, Eds. B. N. Petrov and F. Csaki, Akademiai Kiado, Budapest, 1973, pp. 267-281.
- [29] H. Akaike, A Bayesian Analysis of Minimum AIC Procedure, Ann. of Inst. of Stat. Math. 30A (1978) 9-14.
- [30] G. Schwartz, Estimating the Dimension of a Model, Ann. Statist. 6 (1978) 461-464.