Wavelet-based hybrid models to improve commodities price forecasting : an evidence for the monthly cocoa' price

Romaric KPANOU *

Abstract

This paper propose a new hybrid wavelet-based method to forecast integrated time series. That one consist, at first by using unit-root test, to select a suitable level for the Daubechies' MODWT-based time series decomposition into only stationary detail coefficients and a non-stationary scaling coefficient. After that, dealing with detail coefficients, time series is organized to yield the both irregular and regular components. So, the regular component is predicted by the regression on wavelet coefficients method, while a suitable non-linear model is fitting for the most irregular component. In sequel, a suitable model is identify and fitting for the scaling coefficient that represent the non-stationary information from the time series. Finally, predicted time series is getting by summing the three fitted components. Using the monthly international cocoa price, we show that the proposed method is more accurate than several others proposed in literature and based on wavelets. However, more investigations shall be done on the wavelet filter to use and the choice of suitable decomposition level.

Keywords: Forecasting, non-stationary, Wavelets, cocoa price

1 Introduction

Commodity prices fluctuations are crucial for the both importer and exporter countries. In fact, the knowledge of these prices is using to make decision in international trade. Some African countries derive their main resources from the foreign commodities exchange. So, commodities price fluctuations management is a veritable issue for these countries. So, each moment, it is important for those one to have an idea about the future prices level before buying or selling. In order to address this concern, several methodologies are designed to examine price fluctuations and guide as result the decision [1]. Based-on the historical data,

*Institut de Mathématiques et de Sciences Physiques de l'Université d'Abomey-Calavi, BP 613, Porto-Novo, Benin. Email: romaric.kpanou@imsp-uac.org

there are time series analysis among these methodologies. Recently, more and more hybrid time series analysis models are born with the aim of improving forecasts. This survey focuses on approaches combining wavelets.
These approaches are based on the main idea to get a new spectral representation of a stationary process by replacing the Fourier basis by wavelets one. Because of the introduction of both localisation and scaling factors, the wavelet transform allows to have the time series variability in time by change of scale this will have an immediate effect on the size of the frequency analysis. Following that idea, Guy P. Nason et al [2] defined the *Locally Stationary Wavelet (LSW)* processes. Their definition allows to get the first wavelet-based approach for time series forecasting which was proposed by Fryzlewicz et al [3].

Using the Haar "à trous" wavelet transform, Olivier Renaud et al [4] designed Multi-resolution Auto-regressive Model (MAR) to predict by direct regression on non-decimated wavelet components according to its past values. We can qualify this approach as the second approach combined wavelet for forecasting. For the MAR, they used only wavelet component lagged in dyadic way. They showed that, using the Haar wavelet to transform an autoregressive process $\{X_t\}$, the statistic $N^{1/2}(\hat{\alpha}_N - \alpha)$ converges to centered Gaussian distribution. Let us note that α is least square estimator of the MAR model. These results were extended in various directions by Aminghafari and Poggi [5].
There are a third approach that combine Wavelets and the Artificial Neural Networks (ANN) model for forecasting. That third approach is most use to handle non-linearity problem and wavelets are currently used as preprocessing ([6], [7]).
These three approaches are extended to various ways and the main results are provided, using simulated or real data. Otherwise, when authors deal with real-life data the model or procedure that they use to get accurate forecasts depend on field that is thanking account. M. Rhif, Manel [8] offer a good literature review about that. Furthermore, combining common time series models with above approaches and others statistics tools like Principal component analysis (PCA), several hybrid-models are developed by researchers to make forecasts more and more accurate. We can refer to Nitin Singh, S. R. Mohanty [9], Gabralla, Lubna A Abraham, Ajith [10] who propose a literature review about that.
For this study, we consider some results from the literature about commodities prices forecasting during the ten latter years. The appended Table 8 summarizes the various publications on this subject. From these studies, using wavelets, it is possible to decompose series of commodities price, whether agricultural commodities price (such as rice, corn, cocoa, coffee) or energy commodities price (crude oil, electricity , gas) and apply time series models on resulting components to improve the forecasting of future values.
The main purpose of this paper is to propose an forecasting approach, using wavelets and that is similar to which of Saâdaoui F. et al. [11] but based on forecasting by applying ordinary least square on wavelets named "prediction by regression on wavelet coefficients". The proposed method will be compared to various approaches proposed in the literature namely ARIMA/GARCH, ANN, Wavelet-ARIMA, Wavelet-GARCH and Wavelet-ANN.

According to empirical ways, we remark that authors have widely focused on energy commodities prices (for example, electricity price is the most used) more than agricultural (for example cocoa price isn't almost used) ones. So, using the international monthly cocoa price data, we have shown that the proposed method predict more suitable data than those benchmarks. However, it (the proposed method) returns most accurate forecasts only for four first ahead periods. In the following we present successively: methodology and describe the (in the second Section), empirical results (in the third section) and conclusion (in the fourth section).

2 Methodology

Here we present the basic tools of the proposed approach, the approach itself, the benchmark approaches and data.

2.1 On wavelet transform

When we consider the time-discrete time series, two wavelet transform are mostly use namely the Discrete Wavelet Transform (DWT) that is decimate and the Maximum Overlap Discrete Wavelet Transform (MODWT) that isn't. Let's note that the decimate discrete wavelet transform (DWT) is not translation invariant and this is why the non-decimated DWT named MODWT or Stationary Wavelet Transform (SWT) was proposed by Coifman and Donoho (1995) among others. This last wavelet transform is redundant and it can be considered as an extension of the DWT based on the 'à trous algorithm' [12].

Whether using DWT or MODWT, it is possible to recover the initial time series from wavelets components or coefficients of the decomposition scales, by an additive way named "the multi-resolution analysis".

Let's mention that at opposite to the MODWT that is defined for any sample size N, the DWT of a given level J_0 needs that the sample be an integer multiple of 2^{J_0}. However, there a link between the DWT filters and those of the MODWT. Indeed, if $\{h_{j,l}\}$ and $\{g_{j,l}\}$ are the DWT wavelet and scaling filters respectively, then those of the MODWT are defined by:

$$\tilde{h}_{j,l} = h_{j,l}/2^{j/2} \qquad \text{and} \qquad \tilde{g}_{j,l} = h_{j,l}/2^{j/2}.$$

Where $j = 1, \cdots, J$ indicates the decomposition levels and $l = 1, \cdots, L$ the filter length. So, these last filters allows to non-decimate the resulting component and conserve by the same way the signal energy.

Hence, let $\{X_t\}_{t\in}$ be a time series of any given sample size N. At the decomposition level j, the MODWT wavelet and scaling components are defined respectively as follows:

$$\widetilde{W}_{j,t} = \sum_{l=0}^{N-1} \tilde{h}^{\circ}_{j,t} X_{t-l mod N} \qquad (1)$$

and

$$\widetilde{V}_{j,t} = \sum_{l=0}^{N-1} \tilde{g}^{\circ}_{j,t} X_{t-l mod N}, \qquad (2)$$

where $\{\tilde{h}^{\circ}_{j,l}\}$ (respectively $\{\tilde{g}^{\circ}_{j,l}\}$) is the periodized version[1] of $\{h_{j,l}\}$ (respectively $\{g_{j,l}\}$) to length N; whe have: $\tilde{h}^{\circ}_{j,l} = \sum_{n=0}^{\infty} \tilde{h}_{l+nN}$ (respectively $\tilde{g}^{\circ}_{j,l} = \sum_{n=0}^{\infty} \tilde{g}_{l+nN}$), for all integer n.

So the components are obtained by convolution of the MODWT filters and the time series.

Note that from Equations 1 and 2, any MODWT component at each scale will have the same length as the original signal $\{X_t\}$.

Now, for any vector $\mathbf{X}$ of length N ($\mathbf{X} = [X_0, X_1, \cdots, X_{N-1}]^T$) , we can express the above Equations 1 and 2 in matrix notations as

$$\widetilde{\mathbf{W}}_j = \widetilde{\mathcal{W}}_j \mathbf{X} \qquad \text{and} \qquad \widetilde{\mathbf{V}}_j = \widetilde{\mathcal{V}}_j \mathbf{X},$$

where j is a given level of decomposition, $\widetilde{\mathcal{W}}_j$ and $\widetilde{\mathcal{V}}_j$ are each $N \times N$ orthogonal transform matrix.

In virtue of the MODWT-based multi-resolution analysis, Percival [12] recover the initial time series $\{X_t\}$ as follows:

$$\mathbf{X} = \sum_{j=0}^{J} \widetilde{\mathcal{W}}_j^T \widetilde{\mathbf{W}}_j + \widetilde{\mathcal{V}}_J^T \widetilde{\mathbf{V}}_J. \qquad (3)$$

Where $\widetilde{\mathcal{D}}_j = \widetilde{\mathcal{W}}_j^T \widetilde{\mathcal{W}}_j$ is the jth level MODWT detail coefficient and $\widetilde{\mathcal{S}}_J = \widetilde{\mathcal{V}}_J^T \widetilde{\mathbf{V}}_J$ is the MODWT scaling or smoothed coefficient of the last level J. For more details, see Walden and Percival [12], p.96 – 97, 169, 171, 203.

After choosing an appropriate wavelet transform, it is importance to choose the kind or family wavelet functions in order to get a suitable coefficients from time series data decomposition. When one discuss about of the wavelet function selection, there are the both qualitative (see Philippe Masset [13] p21,22.) and quantitative approaches (refer to W.K. Ngui et al [14]). In this study we deal with the commodities price data that are often non-stationary. Therefore, it is better to use a wavelet mother that can handle easily the non-stationarity.

Moreover, using wavelets coefficients, the forecasting task involves the reconstruction of the data after working on the decomposition scales separately. Adding the criteria defined

[1]The periodized version is necessary to handle the boundary problem.

by Philippe Masset [13], we will use Daubechies wavelets. Indeed, it has good abilities in processing the non-stationary time series when one use the MODWT. These abilities are given in Walden and Percival [12] (p.304, 305) and repeated by Li Zhu et al [15]. Li Zhu et al [15] illustrate this abilities or property with non-stationary and long-range dependence (LRD) time series whose $d-$order backward differentiation is stationary. However, one can take advantage from Daubechies wavelets to handle the non stionarity whatever short or long-range time series. So, we'll use an 'extremal phase' Daubechies filter of width $L > 2d$, where "d" is the time series integration order (see Walden and Percival [12], p.286 − 288).

After data decomposition with a suitable wavelet filter, one will apply the appropriate method on the signal components to predict it. So in the sequel we'll present models and the proposed approach.

2.2 Prediction through ordinary least square on wavelet coefficients

From the above matrix Equation 3, we have $\mathbf{X} = \sum_{j=0}^{J} \widetilde{\mathcal{D}}_j + \widetilde{\mathcal{S}}_J$.

Letting X_t the t^{th} element of time series; $d_{j,t}$ (respectively $s_{J,t}$) the t^{th} element of the wavelet component (respectively scaling component) obtained at the j^{th} (respectively J^{th}) level of decomposition. We will have :

$$X_t = \sum_{j=1}^{J} d_{j,t} + s_{J,t}. \tag{4}$$

Starting from that Equation 4 and Recalling that the one-step forward prediction, of a time series follow $\{X_t\}$ that follows an AR(p) process, can be written as:

$$\widehat{X}_{N+1} = \sum_{k=1}^{p} \widehat{\phi}_k X_{N-(k-1)},$$

Renaud et al [4] assume that for $t = N + 1$ the reconstruction Equation 4 of the time series could be rewritten as :

$$\widehat{X}_{N+1} = \widehat{s}_{J,N+1} + \sum_{j=1}^{J} \widehat{d}_{j,N+1}. \tag{5}$$

So to predict X_{N+1}, it suffice to predict detail components $d_{j,N+1}$ and scaling $s_{J,N+1}$ component at each level $j = 1, \cdots, J$. Otherwise, due to redundancy it is past values, delayed in a dyadic way starting from N, that will be used. So :

$$d_{j,N-2^j(k-1)} \text{ and } s_{J,N-2^J(k-1)}$$

for $k = 1, \cdots, r_J, r_{J+1}$, where r_j i selected as the order of an AR process fitted on the coefficients of level j [5]. So we'll have :

$$\widehat{d}_{j,N+1} = \sum_{k=1}^{r_j} \hat{a}_{j,k} d_{j,N-2^j(k-1)}. \quad \text{and} \quad \widehat{s}_{J,N+1} = \sum_{k=1}^{r_J+1} \hat{a}_{J+1,k} s_{J,N-2^J(k-1)}.$$

Let us remark that the values on the right hand sides of the two previous equations are provided by wavelet transform of time series observed values, so they depend only on the past values of time series itself.

Then, given the N observations $X_1, \cdots, X_N$, the prediction equation of X_{N+1}, is given as follows 6:

$$\widehat{X}_{N+1} = \sum_{j=1}^{J} \sum_{k=1}^{r_j} \hat{a}_{j,k} d_{j,N-2^j(k-1)} + \sum_{k=1}^{r_{J+1}} \hat{a}_{J+1,k} s_{J,N-2^J(k-1)}. \tag{6}$$

To deal with a non-stationary time series, authors propose to separate non-stationary component from original time series before applying the above described method on only detail components. Hence, the scaling component will be added to the non-stationary part to yield a signal which can be predicted using a deterministic or stochastic way; see Aminghafari et al [5] and Renaud et al [4].

Starting from this point, we will propose an approach to deal with stochastic non-stationary time series, using the method of the regression on wavelet component, on some detail components provided by wavelet-based time series transform.

2.3 Proposed method

As mentioned above, when one use Daubechies wavelet filters to transform a stochastic non-stationary time series of order d, the MODWT gives only stationary details components or coefficients, if the length L of Daubechies filters is such that $L \geq 2d$. However, in computational way, this property is not always true. In fact, one can choose a decomposition level $J_0 \leq [log_2(N)]$ which yield some non-stationary details coefficients. For that, we propose to use a unit-root test to select a suitable level J_0 that allows to get only stationary detail or wavelet coefficients. Thereby one will predict some details coefficients using the regression on wavelet coefficients method assuming that $r_{J_0+1} = 0$[2] in the above equation 6. When we decompose a time series using wavelet, the scaling coefficient can be seen as the original series main trend, while detail components might represent the noise of the original series. So, fluctuations are not concentrate in all detail coefficients as same manner. In fact, the very first coefficients are rougher than those which follow. Then, it isn't suitable to apply an autoregressive model on all these coefficients as mentioned by the method of the "prediction by regression on wavelet coefficients". As proposed by Saâdaoui, Foued and Rabbouch, Hana [11] the idea is to group all irregular information from the original signal

[2]Replacing J by J_0.

in one and same component and regular one.
Depends on the kind and length of data, After MODWT-based decomposition, one can have several detail coefficients which are rougher and can't be suitably fitted by a AR model. All those will be grouped or summed to yield one signal. Suppose that the number of these coefficients is n. Starting from the equation 4 and by virtue of multiresolution analysis, we can define the original time series $\{X_t\}$ as follows :

$$\begin{aligned} X_t &= \sum_{j=1}^{n} d_{j,t} + \sum_{j=n+1}^{J_0} d_{j,t} + s_{J_0,t} \\ &= IR_t + R_t + s_{J_0,t}. \end{aligned}$$

The most regular detail coefficients defined by $R_t = \sum_{j=n+1}^{J_0} d_{j,t}$ will be fitted by using the regression on wavelet coefficients above described. In this procedure, let us denote by $\hat{e}_j$ residuals after fitting AR(r_j) models on the detail component d_j of level j. These residuals will be added to the more rougher components ($IR_t = \sum_{j=1}^{n} d_{j,t}$) to yield the new signal $\{F_t\}$ which represent the most irregular part of original one. We'll have a signal defined by

$$F_t = \sum_{j=1}^{n} d_{j,t} + \sum_{j=n+1}^{J_0} \hat{e}_{j,t}.$$

This previous time series will be fitted by a nonlinear method.
Now, it remains to fit the smoothest coefficient $s_{J_0,t}$ using an appropriate model. As mentioned, the present approach is suitable to predict stochastic non-stationary time series. So, let us note that apart from the presence of a unit-root into original series, since the decomposition level J_0 ($J_0 < log_2(N)$) is selected to get only stationary detail coefficients, the scaling coefficient which is less smoother than usual ($J_0 = [log_2(N)]$[3] for the usual case) could not be suitably fitted by polynomial.
Finally the fitted values (respectively forecasts) are computed by summing the fitted values from each of three components (respectively forecasts) of time series. It means (for fitted values) $\hat{X}_t = \widehat{IR}_t + \hat{R}_t + \hat{s}_{J_0,t}$. The Figure 4 summarize the above described approach. This new approach will be compared to some models widely used by authors and quoted in the literature review, which are presented in the following subsection.

2.4 Benchmark models and performance measures

Here, we present some models which will use as benchmark for the proposed one. We will use Autoregressive Integrated Moving Average/General Autoregressive Conditionally heteroscedastic (ARIMA/GARCH), Artificial Neural Networks (ANN), Wavelet-ARIMA, Wavelt-GARCH and Wavelet-ANN.

[3] $[log_2(N)]$ denotes the integer part of $log_2(N)$.

2.4.1 ARIMA and GARCH models

ARIMA

The underlying process of a given time series $\{X_t\}$ is an Autoregressive Integrated Moving Average denoted by ARIMA($p,\ d,\ q$) if $Y_t = (1-B)^d X_t$ is a causal ARMA($p,\ q$) process which is defined as follows:

$$\phi(B)Y_t = \phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t, \quad \varepsilon_t \sim WN(0,\sigma^2),$$

where d is a nonnegative integer, B is backward shift operator, $\phi(z)$ and $\theta(z)$ are polynomials of degrees p and q respectively and $\phi(z) = 0$ for $|z| > 1$;

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \text{ and } \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

have no common factors.

After estimating the ARIMA model parameters, residuals are used for testing for Autoregressive Conditional Heteroscedasticity effects using the ARCH-Lagrange Multiplier (ARCH-LM) (see Engle [16]). If residuals are conditionally heteroscedastic then we'll fit a GARCH model for residuals to estimate the conditional variance.

GARCH model

The process $\{\epsilon_t\}$ is said to be ARCH(q) if

$$\epsilon_t|\psi_{t-1} \sim N(0,h_t) \tag{7}$$

$$h_t = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 \tag{8}$$

where ψ_{t-1} is a given information available up to time $t-1$, $a_0 > 0$, $a_i \geq 0$ for all i and $\sum_{i=1}^{q} a_i < 1$ are required to be satisfied to ensure non negativity and finite unconditional variance of stationary $\{\epsilon_t\}$ series.

The generalized ARCH (GARCH) model 8 is designed by adding to the conditional variance equation 8, the linear combination of its own values lagged in time. Hence, we have:

$$\epsilon_t = \varepsilon_t h_t^{1/2}$$

$$h_t = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} b_j h_{t-j}, \tag{9}$$

where $\varepsilon_t \sim IID(0,1)$.

If $a_0 > 0,\ a_i \geq 0, \quad i = 1,2,\cdots,q. \quad b_j \geq 0 \quad b_j \geq 0, \quad j = 1,2,\cdots,p$, then the conditional variance is positive. The stationary condition of the GARCH(p , q) process is

$$\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1.$$

2.4.2 Artificial Neural Networks (ANN)

Following [17] we can say that ANN propose a complex nonlinear relationships between the response variable and its predictors. Both responses (or outputs) and their predictors (or inputs) are named "neurons" and they are organized in layers. So the bottom layers contain inputs and the top one outputs. Nevertheless there may also have intermediate layers denoted as "hidden layers". The Figure 1 shows the artificial neural networks with one hidden layer and five nodes on that layer.

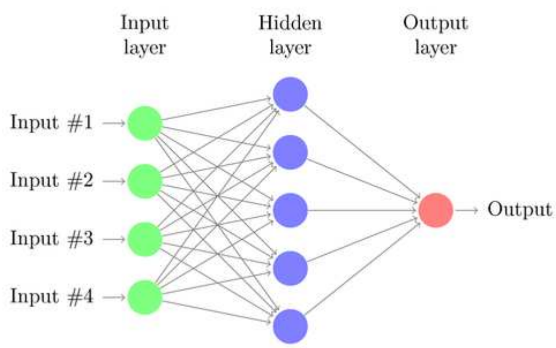


FIGURE 1: A simple neural network whose input layer containing four neurons and hidden layer five neurons [17].

In this kind of neurons network, each node of actual layer receives the information from nodes on previous layer. This ANN architecture is the so-called a "multiplayer feed-forward network".
About the mathematical formulation of ANN model, note that one assign a "weight" to each predictor. So one use a weighted linear combination to compute the information receive by each node from its previous. Finally, one use the nonlinear function to modify result before being output. To illustrate, starting from the above Figure 1 the inputs information received by the hidden neuron k can be expressed as

$$y_k = a_k + \sum_{i=1}^{4} w_{i,k} x_i.$$

After that, this result computed in each hidden neuron will be modified with a nonlinear function; for example:

$$s(y) = \frac{1}{1 + e^{-y}}$$

before giving the input for the next layer. Note that the coefficients $a_1, \cdots, a_5$ and $w_{1,1}, \cdots, w_{4,5}$ are "learned" from the data. For more details we can refer to [17].

2.4.3 Wavelet combining with the above presented models

A part of models presented above, we will use also as benchmarks, their combinations with wavelets that are developed by authors and presented in the literature review. We will use three packages developed in "R Software" by Kumar Paul, Sandipan Samanta, Ankit Tanwar and Anjoy Priyanka, basing on these approaches. The first one, denoted "WaveletARIMA", is packed for "prediction by direct regression on wavelet coefficients" approach proposed by Aminghafari and Poggi [5]. The second is denoted "WaveletGARCH" and it is based on the approach proposed by Paul Ranjit in [18]. The last one is "WaveletANN" based on approach proposed by Anjoy and Paul [19].

2.4.4 Assessing forecast performance

We can note that in the the literature the Mean Squared Error or Deviation (MSE or MSD) and its root square (RMSE or RMSD) are widely used to evaluate effectiveness of forecasts. It is given by square root of the mean square error between observed values and their forecasts. It seems that RMSE is more suitable for assessing forecast of commodities price [20]. So we will use the RMSE in this sturdy. Other measure that is more simple to understand and compute is the Mean Absolute Error (MAE). It is accurate to assess forecast performance on a single time series. Since, that measure depend on scale, it is not suitable to compare forecasting performance across series [21].

Since we use only one time series to compare different forecasting models or approaches, we will use also the MAE. For the same reasons we add the Mean Absolute Scaled Error or Deviation (MASE) (for more detail on this measure, see Rob J. Hyndman [21]).

Denoting by N the length of the analyzed time series, x_t the actual value of the time series at date t, and $\hat{x}_t$ its fitted value at the same date, the formulas to calculate these three measures are defined as follows.

$$MAE = \frac{1}{N}\sum_{t=1}^{N} |x_t - \hat{x}_t| \,, \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2}, \tag{11}$$

$$MASE = \text{Mean}(|q_t|), \tag{12}$$

where the scaled error (q_t) is defined as:

$$q_t = \frac{x_t - \hat{x}_t}{\frac{1}{N-1}\sum_{i=2}^{N} |x_i - x_{i-1}|}.$$

Since the all framework is define it is important to test it empirically. So, the following section is devoted to what data will be used, where they come from and how they will be organized to test the proposed accuracy.

2.5 Data and descriptive statistics

In this section we present successively the data provenance, how they will be organized and descriptive statistics.

2.5.1 Data and their provenance

In this study we are interesting to the price of Cocoa grain. The Data come from the Word Bank dataset and are accessible via the following web link [22]. Those are the monthly price of Cocoa grain. They are average of daily price that is provided by the International Cocoa Organization (ICCO). Data are expressed in nominal US dollars per kilogram ($/kg) and cover the period from 1960/01 to 2020/12[4]; so the length of data is $N = 732$. In add, there are no missing values. For more details, we can refer to World Bank web site [22].

2.5.2 Data: Training and Test

In this paper, we separate the series into two parts. The first one data is used for the model building and the second is used as test. More precisely data training portion covers the period from 1960/01 to 2020/03 (723 observations). The remaining portion, of data, covering the period from 2020/04 to 2020/12 (10 observations) is used for the evaluation of forecasting and we will interest to ten (10) months ahead forecasts horizon. Let now, describe data after applying time series models.

2.5.3 Descriptive statistics

The Table 1 presents some data descriptive statistics.

TABLE 1: Descriptive statistics

Min	Max	Q1	Q2	Q3	Mean	Std.dev
0,259	4,3629	1,011	1,565	2,297	1,707	0,885

We see that, from 1960 to 2020, the monthly price of one kilogram of the cocoa grain has fluctuated between \$ US 0, 26 and \$ US 4, 36 around an average value of 1, 71. So the range of data is 4, 10. It is noted that 25% of cocoa prices observed are less than \$ US 1, 01% and 75% are less than \$ US 2, 30. So, 50% of prices are located between \$ Us 1, 01% and \$ US 2, 30. Otherwise the fact that the average is slightly higher than the median (Q_2) show that the series is slightly spread to the right and is therefore not symmetrical. Throughout the time, if all monthly price are supposed to be equal to the mean, the monthly actual price would have been missed at \$ US 0, 88 more or less. So the error would be just under 52% of the average value.

[4]We have chosen this period because starting 1960 the both terminal markets of New York and London have already opened. In add, extensive production techniques already existed and we can say that rises of African production have less impacted production.

3 Results

We present here results from benchmarks approaches, those of the proposed approach and compare their accuracy.

3.1 From benchmarks models and approaches

When we observe the data representation (Figure 5 in appendix), the histogram confirm that the series is asymmetric. In add, the graph of the evolution let see that the series is likely to be non-stationary. This fact is well confirmed by the autocorrelation function of the series which is positive and weakly decreasing does not converge. The stepwise Augmented Dickey and Fuller (ADF) test, allows us to conclude that the underlying process is integrated with drift. In add, there are no seasonality component. Hence, for data fitting, the first candidate model is non-seasonal ARIMA.

3.1.1 ARIMA fitted for cocoa price data

After the first difference of the data they are stationary. Hence, the model ARIMA(2, 1, 2) is identified after carrying out a grouping task. Results are presented in the Table 2 .

TABLE 2: ARIMA model fitting for data

Parameters	**Estimator**	**Stand. error**	t**−Stat**
ar_1	−0.8522	0.1220	−7.036
ar_2	−0.4339	0.1171	−3.810
ma_1	1.1122	0.1122	9.978
ma_2	0.6134	0.0958	6.501
N	723		
$\hat{\sigma}^2$	0.01257	L-B=3.5161	p−val=0.97
LL	557.48	J-Ber(2)=375.18	p−val$< 2.2e-16$

By observing the Table 2, the t−Statistic' absolute value attached to each estimated parameter is greater than the critical value of 1.644854. Then, parameters are significant. Otherwise, we note that the Ljung-Box and Jacque-Berra test show that errors are correlated and non-normal. So we check residuals heteroscedasticity for trying to apply GARCH model.

3.1.2 Hybrid ARIMA-GARCH fitted for cocoa price data

Checking for the presence of conditional heteroscedasticity into residuals series, the Lagrange-Multiplier test allows to reject the null hypothesis of absence of heteroscedasticity (See Table

9 in appendix, for more details.). After several tasks for selecting the suitable ARIMA-GARCH model for data, we have finally find the best in the class of ARIMA(2, 1, 0)-GARCH(3, 0) models. The Table 3 present estimated parameters.

TABLE 3: ARIMA(2, 1, 0)-GARCH(3, 0) model fitted for data

	Normal distribution			**Student distribution**		
	Estimate	Std. Error	t value	Estimate	Std. Error	t value
ar1	1.182312	0.013320	88.7595	1.188820	0.006611	179.8177
ar2	−0.183838	0.013234	−13.8910	−0.190994	0.006553	−29.1458
omega	0.004797	0.000450	10.6651	0.001957	0.000510	3.8373
alpha1	0.148506	0.055210	2.6898	0.313508	0.082939	3.7800
alpha2	0.444878	0.076013	5.8526	0.484448	0.090176	5.3723
alpha3	0.117188	0.046229	2.5349	0.201044	0.061874	3.2493
Shape				3.877842	0.389057	9.9673
N	723			723		
LL	609.2015			700.4328		
AIC	−1.6686			−1.9182		
Ljung-Box(1)	4.778	p-value	$2.883e-02$	1.199	p-value	0.2735
ARCH LM (8)	0.52636	p-value	0.9797	1.0921	p-value	0.9103
Jera Berra (2)	759.18	p-value	$2.2e-16$	763.46	p-value	$2.2e-16$

Whether the distribution, all model parameters are significant. Otherwise, the ARCH-LM test show absence of heteroscedasticity in residuals. However, when parameters are estimated using the student distribution, it can seen that residuals are autocorrelated (see the $p-$value of the Ljung-Box test). As results, we will use parameters estimated using normal distribution.

3.1.3 ANN fitted for cocoa price data

Fanally for ANN model fitting for data, the "nnetar" function in R-software is used to fit neural networks. We use $p = 2$ lagged values on inputs layer as suggested by the Autocorrelation function which cut off at lag 2. The estimated noise variance is $\sigma^2 = 0.004103$. Before compare the above model, let's apply the proposed method for data. It is that concern the next subsection.

3.1.4 Others models combining wavelet coefficients fitted for data

Concerned models (Wavelet combined with ARIMA, GARCH and ANN) are hybrid. So, we won't have to present parameters for these models. However, we will check their accuracy using the prediction performance measures. That is the purpose of the next section. Let's mention that we have used the same wavelet filter (d4: 4 width Daubechies filter) and five (5) decomposition level, for all models or approaches.

3.2 Proposed approach applying

We start with data decomposition, before fitting a suitable model for yielded components.

3.2.1 Wavelets-based data decomposition

Since we use 723 observation as training, the full data decomposition based on MODWT must need $[log_2(723)] = 9$. Since data are integrated of order $d = 1$, as mentioned above (in the methodology) we use Daubechies filters of width $L = 4$ ($L \geq 2d$). For suitable level selecting for the proposed approach, we separate stationary information from non-stationary one. The Table 4 show wavelets coefficients and its stationary testing results using the Phillips-Perron test for the null hypothesis of unit root.

TABLE 4: Wavelet coefficients Stationary text

Coef.	$DF-$**Stat**	p**-value**
d_1	−41.437	0.01
d_2	−12.630	0.01
d_3	−8.6887	0.01
d_4	−7.2748	0.01
d_5	−4.1633	0.01
d_6	−2.3593	0.4262
d_7	−1.1773	0.9101
d_8	−0.18746	0.99
d_9	1.45070	0.99
s_9	1.18740	0.99

We note that coefficients are stationary until the 5^{th} decomposition level and non-stationary after that. So we use five (5) decomposition levels for our data. The Figure 6 (in appendix) show these coefficients. As expected, we note that the first one coefficient is the most roughest. In add, the scaling coefficient is the smoothest, but less coarse than that of decomposition using all 9 levels (see figure 7 in appendix).
Since data have already suitable decomposed, successively, we fit an AR(r_j) for each detail coefficient of level $j = 2, \cdots, 5$ and compute its residuals; add these residuals to the first

detail component and find suitable model for fitting resulting time series; at the end we fit a suitable model for the scaling coefficient.

3.2.2 Regression on regular wavelet coefficients

Starting by AR models selection, the Figure 8, in the appendix, show the partial autocorrelation functions for detail coefficients. Thus we select $r_2 = 8$, $r_3 = 15$, $r_4 = 2$, $r_5 = 5$. All parameters were significant and could be used to forecast the future detail coefficients values. However, the information that is computing here, is essentially linear. It remain some linear and mainly non-linear features in residuals from these regressions. Now, we will add these residuals to the first detail coefficients to yield the most irregular information from the time series.

3.2.3 Suitable model for modified version of first detail coefficient

When we add residuals from these autoregressive models to the first wavelet coefficient, the resulting series stay stationary and thus eligible for ARMA models. After selection of the best ARMA which is ARMA(6, 6), we note that the residuals, from that model , were autocorrelated and non-normal. Moreover, the square of residuals presents some volatility and ARCH-LM test conduct us to fit a GARCH model for residuals. Finally, investigating the best model to fit for this time series, fitting ARMA-GARCH and ANN models for it, allows us to choose ANN model as the best. The below Table 5 present the comparison criteria [5] for models.

TABLE 5: Comparison of models fitting for modified first wavelet coefficient

Model	AIC	MAE	MASE	RMSE
ARMA(6, 6)-GARCH (1, 1)	−3.4126	0.03166834	0.6598684	0.04490782
ARMA(6, 6)-GARCH (2, 1)	−3.4099	0.03166834	0.6598683	0.04490781
ARMA(6, 6)-GARCH (2, 2)	−3.4083	0.03160391	0.6585259	0.04479535
ARMA(6, 6)-GARCH (2, 3)	−3.4123	0.03168773	0.6602724	0.04491529
ANN		0.02112703	0.4402207	0.02880275

By observing the Table 5 we remark that, according AIC criteria, ARMA(6, 6)-GARCH(1, 1) is the best among GARCH models, but according to performance measures we will select GARCH(2, 2). However, ANN model fit better the series, according to all measures.

[5] We use forecasting performance measures to choose the most suitable model.

3.2.4 Suitable model fitting for the scaling coefficient

Let's now, continue with the fitting suitable model for the scaling coefficient. As expected, the scaling coefficient is non-stationary. Investigations reveal that the ANN model is suitable for this coefficient in our case. See the Figure 9in Appendix.

Finally, the last step of the main approach is to compute the final fitted values. As mentioned above, by virtue of multiresolution analysis, this task will do by summing fitted values from three above defined components. The below Figure 2 show the both actual and fitted values of data, from 2010 to 2020.

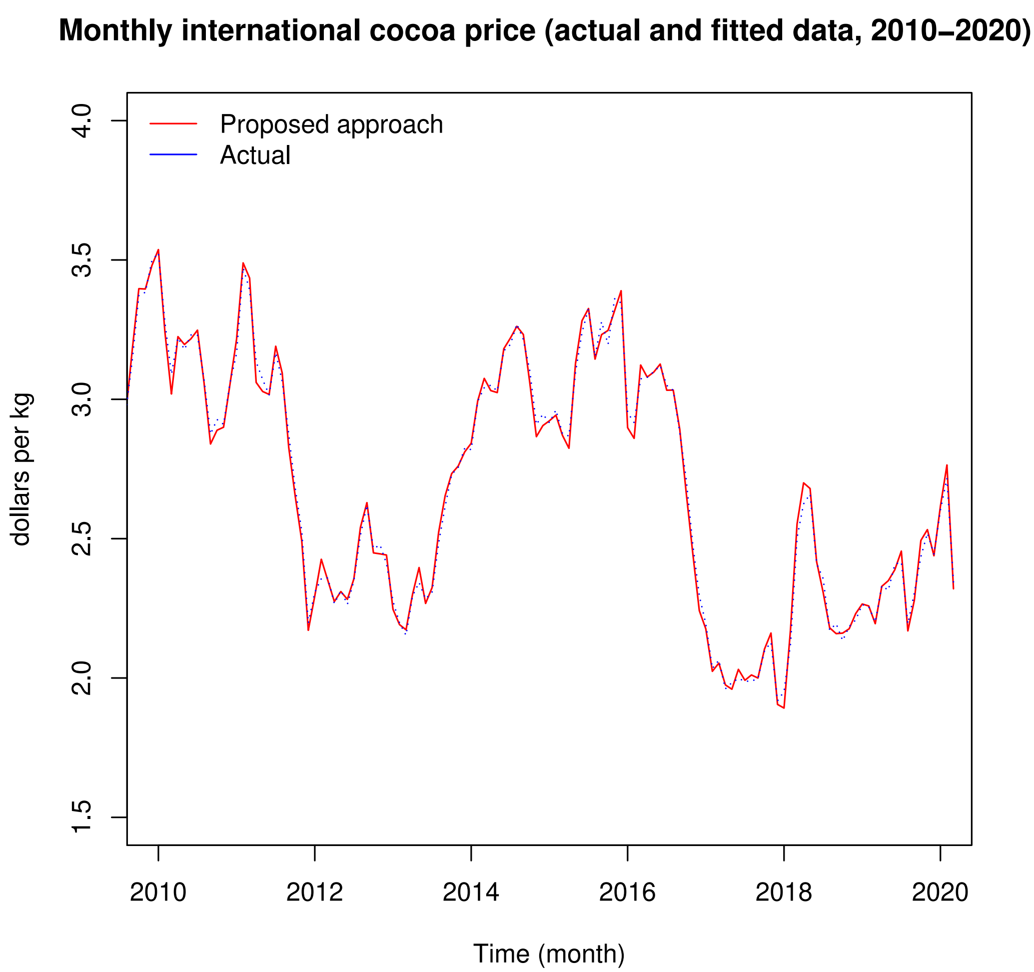


FIGURE 2: Actual and proposed model fitting values for monthly international cocoa price data

We can said that the model builds here is hybrid. Indeed, it combine parametric (AR model) and non parametric approach (Wavelets), but also semi-parametric (ANN) model.

Now, let us compare it to benchmarks.

3.3 Forecasting and models performance

In this section we compare different model to show which of them is better to predict the monthly cocoa price and which can yield the most accurate forecasts.

3.3.1 Accuracy' measurement : fitted and training data

After fitting models for data, the three different accuracy measures are reported in the Table 6. Note that these values are calculated using the training part of data.

TABLE 6: Accuracy measures : fitted and training data

Model	**MAE**	**RMSE**	**MASE**
ARIMA	0.07687242	0.11269280	0.9794505
ARIMA-GARCH	0.07652182	0.11390230	0.9749833
ANN	0.07523247	0.11225230	0.9585554
WAVELET-ARIMA	0.07856195	0.11409480	1.0009770
WAVELET-GARCH	2.50501700	0.26784400	2.5050170
WAVELET-ANN	0.10027960	0.21293350	1.2947780
Proposed	0.02042713	0.02807942	0.2602671

Overall, by observing these values, it is noted that the proposed model is the one that best fits the data, and Wavelet-GARCH less fit the data, regardless the accuracy measure. Except the proposed approach, when we compare the model two by two, it means the initial model and its association with the wavelets, it seems that the current model is more suitable than its association with wavelets. This result seems to be contrary to what is noticed in the literature.

When we consider non-wavelet-associating models, we find that neural networks are better suited to data, regardless measures. After ANN we can find ARIMA-GARCH in second position. However, RMSE show that ARIMA can be more preferred.

Now, when we consider only wavelet-associating models except one proposed, we note that Wavelet-ARIMA is the best, regardless the measure. So, it seems that, if one doesn't use the proposed approach to wavelet-based forecast the monthly cocoa price, then it is better to use Wavelet-ARIMA.

Finally, since ANN are mainly include in the proposed method in combination of wavelets

and autoregressive model, we can say that they are more suitable for fitting "monthly cocoa'bean price". However, the use of wavelets improve their performance as show by the proposed approach.
Let's note that results find here are likely same as those of Kamu Assis et al [23] whom find ARIMA-GARCH model as the best model. However, these one didn't use wavelets nor neural networks. Indeed, even if we don't use the same data sample as these authors, we note that, despite of all considerations of wavelets and neural networks, we will select also hybrid ARIMA-GARCH.
The performance of a model depends on both its ability to fit the pass data values and to forecast the future ones. The accuracy of a model is more appropriate by its ability to forecast well future values of the studied phenomenon. When we deal mainly with uni-variate time series, it comes often that the model fit best the data without to yield the best forecast . The next subsection deals with forecasts accuracy.

3.3.2 Accuracy' measurement - forecasts and test data

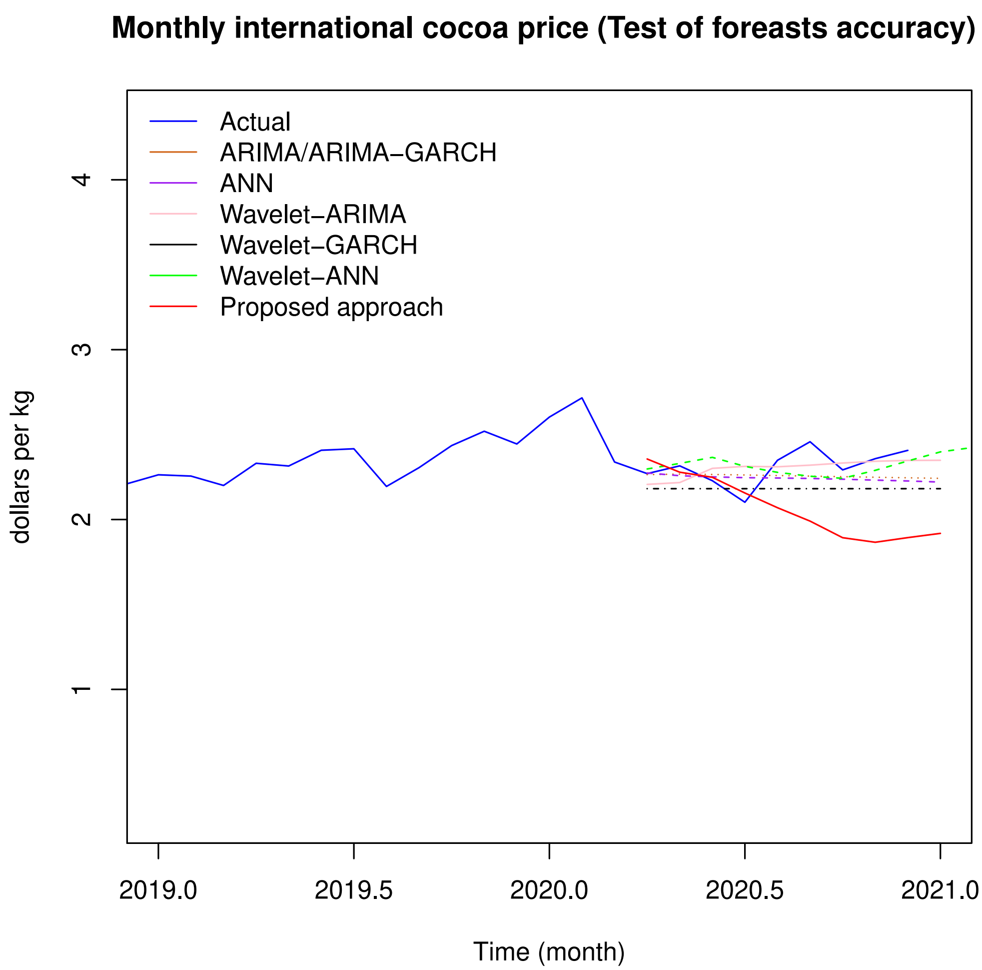


FIGURE 3: Monthly cocoa price: Forecasts and test data

Figure 3 shows ten ahead (from 2020/04 to 2021/01) forecasts values calculated using models. The actual values for the period January 2018 to December 2020 are also shown. It seems that, forecasts values from proposed approach are the most suitable only for some months until July 2020, but after this month, others approaches are more accurate. We compute the forecast accuracy measures for two periods. The first take account the first four ahead forecasts, it means from 2020/04 to 2020/07 and the second from 2020/04 to 2020/08. Table 7 present results.

TABLE 7: Forecasting approaches/model Comparison

	1-4 Month ahead			1-5 Month ahead		
Model	**MAE**	**RMSE**	**MASE**	**MAE**	**RMSE**	**MASE**
ARIMA	0.07599441	0.08624271	0.8768586	0.07778558	0.08598577	0.6137656
ARIMA-GARCH	0.06290106	0.08584970	0.7257814	0.06827449	0.08664759	0.5387185
ANN	0.05730325	0.07880024	0.6611913	0.06669400	0.08450732	0.5262477
WAVELET-ARIMA	0.11133170	0.12601070	1.28459700	0.09668896	0.11398930	0.7629223
WAVELET-GARCH	0.08752250	0.09292477	1.00987500	0.10344600	0.11178180	0.8162386
WAVELET-ANN	0.09761483	0.12700060	0.09761483	0.09229149	0.11794690	0.7282242
Proposed	0.04944269	0.05513630	0.57049250	0.09539560	0.13425110	0.7527171

Overall, regardless the accuracy measure, it seems that models combining-wavelets give forecasts less suitable than those non-combining-wavelet.
Regard MAE and RMSE, it is obvious that the proposed approach can be the best to forecast the "monthly Cocoa'price" only for four months following the end on training data. According to the MASE, it is the Wavelet-ANN model that can be the best for the same period. Even if the proposed models is not suitable according to the MASE measure, the figure show that its forecasts values are the most neighbor of actual values in this period.
Now, when we take account one month ahead in add the previous period (four months during which, proposed model forests are suitable), it seems that the ANN model can be recommended to analysis data and forecasts the future values of "cocoa prices".

4 Conclusion

Based-on literature results, we proposed in this paper a hybrid model to forecast stochastic non-stationary time series. This approach combine wavelet, autoregressive and ANN model. Using the international monthly cocoa price we showed that the proposed approach is more accurate than those combine wavelets and current models namely ARIMA, GARCH and ANN. However, the proposed approach can be best to forecast up to fourth period after those of given data.
The main lack of the proposed approach is that it is uni-variate approach. Otherwise its robustness is not tested using others similar data like those obtained by Monte-Carlo simulations.

Acknowledges: We would like to thank the German Academic Exchange Service (DAAD) scholarship grants implemented by Institut de Mathématiques et de Sciences Physiques

(IMSP) de l'Université d'Abomey-Calavi, Benin for financial support.

References

[1] G. A K and U. A, "Beyond linear methods of data analysis: Time series analysis and its applications in renal research," *Nephron Physiol*, vol. 124, p. 14, 2013.

[2] G. P. Nason, R. von Sachs, and G. Kroisandt, "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 62, no. 2, pp. 271–292, 2000.

[3] P. F. I, S. I. B. V. A. N. Bellegem, and R. V. O. N. Sachs, "Forecasting non-stationary time series by wavelet process modelling," *Annals of the Institute of Statistical Mathematics*, vol. 55, no. 4, pp. 737–764, 2003.

[4] O. Renaud, J.-L. Starck, and F. Murtagh, "Prediction based on a multiscale decomposition," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 01, no. 02, pp. 217–232, 2003.

[5] M. Aminghafari and J.-M. Poggi, "Forecasting time series using wavelets," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 05, no. 05, pp. 709–724, 2007.

[6] A. Aussem and F. Murtagh, "Combining neural network forecasts on wavelet-transformed time series," *Connection Science*, vol. 9, no. 1, pp. 113–122, 1997.

[7] Z. Qinghua and B. Albert, "Wavelet networks," *IEEE Transactions on Neural*, vol. 3, no. 6, 1992.

[8] M. Rhif, A. Ben Abbes, I. R. Farah, B. Martínez, and Y. Sang, "Wavelet transform application for/in non-stationary time-series analysis: A review," *Applied Sciences*, vol. 9, no. 7, 2019.

[9] N. Singh and S. R. Mohanty, "A review of price forecasting problem and techniques in deregulated electricity markets," *Journal of Power and Energy Engineering*, vol. 3, pp. 1–19, 2015.

[10] L. Gabralla and A. Abraham, "Computational modeling of crude oil price forecasting: A review of two decades of research," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 5, pp. 729–740, 2013.

[11] F. Saâdaoui and H. Rabbouch, "A wavelet-based hybrid neural network for short-term electricity prices forecasting," *Artificial Intelligence Review*, 2019.

[12] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2000.

[13] P. Masset, “Analysis of financial time-series using fourier and wavelet methods,” *SSRN Electronic Journal*, pp. 1–36, 2011.

[14] W. Ngui, M. Salman Leong, L. M. Hee, and A. M. Abdelrhman, “Wavelet analysis: Mother wavelet selection methods,” *Applied Mechanics and Materials*, vol. 393, pp. 953–958, 2013.

[15] Z. Li, W. Yanxin, and F. Qibin, “Modwt-arma model for time series prediction,” *Applied Mathématical Modelling*, vol. 38, pp. 1859–1865, 2013.

[16] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation,” *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[17] J. H. Rob and A. George, “Forecasting: Principles and practice,” 2018. OTexts.com/fpp3.

[18] R. K. Paul, “Arimax-garch-wavelet model for forecasting volatile data,” *Model Assisted Statistics and Applications*, vol. 10, no. 3, pp. 243–252, 2015.

[19] P. Anjoy and R. K. Paul, “Wavelet based hybrid approach for forecasting volatile potato price,” *Journal of the IndIan SocIety of agrIcultural StatIStIcS*, vol. 71, no. 1, pp. 7–14, 2017.

[20] R. Füss, C. Hoppe, and D. G. Kaiser, *Review of Commodity Futures Performance Benchmarks*, ch. 7, pp. 167–202. John Wiley & Sons, Ltd, 2008.

[21] R. Hyndman, “Another look at forecast accuracy metrics for intermittent demand,” *Foresight: The International Journal of Applied Forecasting*, vol. 4, pp. 43–46, 01 2006.

[22] W. Bank, “The word bank,” Jannuary 2021. https://www.worldbank.org/en/research/commodity-markets.

[23] A. Kamu, A. Ahmed, and R. Yusoff, “Forecasting cocoa bean prices using univariate time series mode,” *International Refereed Research Journal*, vol. 1, no. 1, pp. 71–80, 210.

[24] M. Shafie-Khah, M. P. Moghaddam, and M. K. Sheikh-El-Eslami, “Price forecasting of day-ahead electricity markets using a hybrid forecast method,” *Energy Conversion and Management*, vol. 52, no. 5, pp. 2165–2169, 2011.

[25] K. Sukiyono and Nabiu, “Selecting an accurate cacao price forecasting model,” *Journal of Physics: Conference Series*, vol. 1114, p. 012116, 11 2018.

[26] E. Haven, X. Liu, and L. Shen, "De-noising option prices with the wavelet method," *European Journal of Operational Research*, vol. 222, no. 1, pp. 104–112, 2012.

[27] S. Annamareddi and B. Gopinathan, Sudheer Dora, "A simple hybrid model for short-term load forecasting," *Journal of Engineering (United Kingdom)*, vol. 2013, 203.

[28] M. Khalid, M. Sultana, and F. Zaidi, "Prediction of agriculture commodities price returns using arma and wavelet," *Journal of Natural Sciences Research*, vol. 4, pp. 2224–3186, 01 2014.

[29] J. Jin and J. Kim, "Forecasting natural gas prices using wavelets, time series, and artificial neural networks," *PLOS ONE*, vol. 10, pp. 1–23, 11 2015.

[30] S. Rathod, K. Singh, R. Paul, S. Meher, G. Mishra, B. Gurung, M. Ray, and K. Sinha, "An improved arfima model using maximum overlap discrete wavelet transform (modwt) and ann for forecasting agricultural commodity price," *Jonal of the Indian Society of Agricultural statistics*, vol. 71, no. 2, pp. 103–11, 2017.

[31] S. Bunrit, N. Kerdprasop, and K. Kerdprasop, "Multiresolution analysis based on wavelet transform for commodity prices time series forecasting," *International Journal of Machine Learning and Computing*, vol. 8, pp. 175–180, 2018.

[32] R. E. A. B. M. Antonio J. Conejo; Miguel A. Plazas, "Day-ahead electricity price forecasting using the wavelet transform and arima models," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035–1042, 2004.

[33] K. Thomas, A. Andrew, P. David, and C. Monica Rivas, "An improved wavelet–arima approach for forecasting metal prices," *Resources Policy*, vol. 39, pp. 32–41, 2014.

[34] Z. Tan, J. Zhang, J. Wang, and J. Xu, "Day-ahead electricity price forecasting using wavelet transform combined with arima and garch models," *Applied Energy*, vol. 87, no. 11, pp. 3606–3610, 2010.

[35] K. He, L. Yu, and K. Keung, "Crude oil price analysis and forecasting using wavelet decomposed ensemble model," *Energy*, vol. 46, no. 1, pp. 564–574, 2012.

[36] S. Anbazhagan and N. Kumarappan, "Day-ahead deregulated electricity market price forecasting using recurrent neural network," *IEEE Systems Journal*, vol. 7, no. 4, pp. 866–872, 2013.

[37] A. Shabri and R. Samsudin, "Daily crude oil price forecasting using hybridizing wavelet and artificial neural network model," *Hindawi Publishing Corporation*, 2014.

[38] R. Jammazi and C. Aloui, "Crude oil price forecasting : Experimental evidence from wavelet decomposition and neural network modeling," *Energy Economics*, vol. 34, no. 3, pp. 828–841, 2012.

[39] H. T. Nguyen and I. T. Nabney, "Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models," *Energy*, vol. 35, no. 9, pp. 3674–3685, 2010.

[40] W. L and S. M, "A hybrid model for day-ahead price forecasting," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1519–1530, 2010.

[41] Y. Pang, W. Xu, L. Yu, J. Ma, K. K. Lai, S. Wang, and S. Xu, "Forecasting the crude oil spot price by wavelet neural networks unsing oecd petroleum iventory levels," *New Mathematics and Natural Computation*, vol. 07, no. 02, pp. 281–297, 2011.

[42] A. Shayeghi, H Ghasemi, "Day-ahead electricity prices forecasting by a modified cgsa technique and hybrid wt in lssvm based scheme," *Energy Conversion and Management*, vol. 74, pp. 482–491, 2013.

[43] R. Paul, "Arimax-garch-wavelet model for forecasting volatile data," *Model Assisted Statistics and Applications*, vol. 10, pp. 243–252, 07 2015.

[44] T. Sujjaviriyasup, "Predicting prices of agricultural commodities in thailand using combined approach emphasizing on data pre-processing technique," *Songklanakarin Journal of Science and Technology*, vol. 40, no. 1, pp. 75–78, 208.

[45] A. Shabri, M. Fahmi, and A. Hamid, "Wavelet-support vector machine for forecasting palm oil price," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 15, no. 3, pp. 398–406, 2019.

5 Appendix

Related to Literature review

TABLE 8: Literature review

Year	Publication	Forecasting Model	Explanation	Application	Performance
2003	Renaud et al. [4]	Multiresolution Autoregressive (MAR)	Forecast future value as linear combinaison of MODWT coefficients whose components are dydically lagged.	Similated data, Yearly minimal water levels of the Nile river level, hourly rates of data access on a web site.	
2010	Assis, K. et al [23]	ARIMA/GARCH	ARIMA in first time and GARCH on error from this.	Monthly Cocoa price	RMSE, MAPE, U-STATISTICS
2011	Shafie-Khah, M. et al. [24]	Wavelet-ARIMA-RBFN	After DWT, apply ARIMA on each component and apply ANN-RBFN on error from ARIMA model estimation.	Electricity price of market of mainland Spain	RMSE, MAPE
2018	Sukiyono et al. [25]	ARIMA	Common ARIMA procedure	Monthly Cocoa price	MAD, MAPE
2012	Emmanuel Haven et al [26]		De-nosing data using wavelets	Simulated data using Black-Scholes model (1973) and Monte Carlo Procedure	
2013	Annamareddi S. et al. [27]	Double exponential smoothing	Application of the Double exponential smoothing on all the both approximation and detail coefficients, after wavelets decomposition and de-nosing	Electrical load data from California energy market	RMSE, MAPE

Table 8: Literature review (continued from previous page)

Year	Publication	Forecasting Model	Explanation	Application	Performance
2014	Khalid M. et al [28]	Wavelet-ARMA	Apply ARMA process after wavelet decomposition	Monthly wheat, rice, Barely and Maize price.	RARE, MSE, MAD
2015	Jin, Junghwan and Kim, Jinsoo [29]	ARIMA, GARCH and ANN	ARIMA or ANN for approximation component, and GARCH or ANN for detail components	Henry Hub weekly natural gas spot price	RMSE, MAE, MAPE
2017	Santosha et al. [30]	ARFIMA-MODWT	Coefficients of the ARFIMA model was estimated using MODWT and residuals from that was modeled by ANN.	Daily wholesale price of mustard in Mumbai market	MAPE
2018	Bunrit S. et al.[31]	Wavelet-ARIMA	Applications of ARIMA model on each wavelet coefficients, after decomposition	Gold and rubber price data	MAPE
2004	Antonio et al.[32]	Wavelet-ARIMA	Same manner as above	Hourly electricity price from market of mainland Spain	MAPE
2014	Thomas K. et al[33]	Wavelet-ARIMA	Same manner as above but comparing results with several type of wavelets	monthly nominal cash prices of aluminium, copper, lead and zinc traded on the London Metal Exchange (LME)	MAE, RMSE

TABLE 8: Literature review (continued from previous page)

Year	Publication	Forecasting Model	Explanation	Application	Performance
2014	Khalid M. et al [28]	Wavelet-ARIMA	Same manner but using directly the ARMA model	Monthly international prices of wheat, rice, Barely and Maize	MARE, MSE, MAD
2010	Tan Z. et al. [34]	ARIMA-GARCH and GARCH	Application of ARIMA-GARCH on approximation coefficient, and GARCH on detail coefficients, after wavelet decomposition.	Daily-electricity prices : Market-clearing price in Spanish and PJM market	MAPE
2012	Kin Keung Lai [35]	Wavelet Decomposed Ensemble Model (WDE)	Similar to Wavelet-ARIMA but the conditional mean forecast at each scale was assumed follow an ARMA processes	USWest Texas Intermediate crude oil (WTI) and European Brent crude oil (Brent)	MSE
2017	Anjoy P. and Ranjit Kumar Paul [19]	Wavelet-GARCH	GARCH or ARIMA model applied on wavelet coefficient	Monthly potato price data of three markets, namely Haldwani, Agra and Lucknow of Uttar Pradesh	RMAPE, RMSPE
2013	Anbazhagan S. and Kumarappan N. [36]	ANN using Elman network		Hourly he electricity market of mainland Spain	MAPE

TABLE 8: Literature review (continued from previous page)

Year	Publication	Forecasting Model	Explanation	Application	Performance
2014	Shabri Ani Samsudin, Ruhaidah [37]	Wavelet Artificial Neural Network (WANN)	Application of ANN models on series obtained by additing suitable wavelet detail components with one approximation series.	West Texas Intermediate (WTI) and Brent crude oil spot daily prices	RMSE, MAPE
2015	Jammazi and Aloui [38]	Harr a Trous wavelet multilayer back propagation neural network (HTW-MBPNN)		Monthly Crude oil price	MAE, MSE
2013	Anbazhagan S. and Kumarappan N. [36]	Elman network model		electricity price forecasting	MAPE
2014	Shabri Ani Samsudin, Ruhaidah [37]	Wavelet Artificial Neural Network (WANN)	After wavelet decomposition, select detail coefficients that are more related to original time series. A new series obtained by adding approximation series and DS component is used as input into ANN model.	WTI and Brent crude oil.	MAPE, RMSE

TABLE 8: Literature review (continued from previous page)

Year	Publication	Forecasting Model	Explanation	Application	Performance
2012	Jammazi and Aloui [38]	HTW-MBPNN	By applying the Harr to Trous wavelet decomposition to the time series, the MBPNN is then implemented for both the original and the smoothed version of the time series.	Monthly WTI crude oil price	MAE, MSE
2019	Saadaoui F. et al. [11]	Wavelet-based nonlinear-fitting ARFIMA-ANN (W-NF-ARFIMA-ANN)	Use wavelet coefficient to decompose time series into nonlinear, trend and multiple seasonal effects, and to exactly extrapolate them over the time scale.	hourly power prices from the Nord Pool Exchange	
2010	Nguyen Hang T. and Nabney Ian T. [39]	Multi-variate Linear regression, ANN, GARCH	After wavelet transform, use linear regression, ANN or GARCH depends on component features, for forecasting. Use Kalman filter to update the parameters continuously on the test set.	Electricity demand and gas price forecasts	MAPE, NMSE, MAE

TABLE 8: Literature review (continued from previous page)

Year	Publication	Forecasting Model	Explanation	Application	Performance
2010	Wu et al. [40]	ARMAX, GARCH and AWNN	Use ARMAX to forecast linear relationship between the time series and an explanotory variable followed by GARCH to simulate nonconstant variances of residuals, and applies AWNN to forecast nonlinear and non-stationary impacts.	Electricity price and power system loads as an explanatory	AMAPE
2011	Ye et al. [41]	Wavelet Neural Network (WNN)	WNN was used to model the nonlinear relationship between explanatory and independent variable.	WTI monthly crude oil spot price	RMSE, MAPE and MAE
2013	Shayeghi et al. [42]	Chaotic Gravitational Search Algorithm Least Square Support Vector Machine (CGSA-LSSVM)	DWT is used to remove high-order components of input data, use the novel CGA method to remove the Gaussian noise from ouput of LSSVM.	Electricity prices	RMSE, MAE and MAPE

TABLE 8: Literature review (continued from previous page)

Year	Publication	Forecasting Model	Explanation	Application	Performance
2015	Ranjit Kumar Paul [43]	ARIMAX-GARCH-WAVELET	Forecaste exogenous variable using wavelets. So residuals offitted ARIMAX model are used for testing the presence of ARCH effect.	Annual wheat yield data of Kanpur district ofUttar Pradesh, India. Exogenous variable is the maximum temperature at the Critical Root Initiation (CRI).	MAPE, RMSE, RMAPE
2018	Thoranin Sujjaviriyasup [44]	Support Vector Regression (SVR)	Using SVR to forecast details and approximation coefficient after wavelet decomposition	Agricultural commodity price: rice, cassava, sugarcane, and coffee prices	MAE, MAPE
2019	Shabri A. et al. [45]	Wavelet-SVM like above	Use DWT to decompose time-series and apply SVM for forecasting	palm oil price	RMSE, MAE

Diagram of the proposed approach

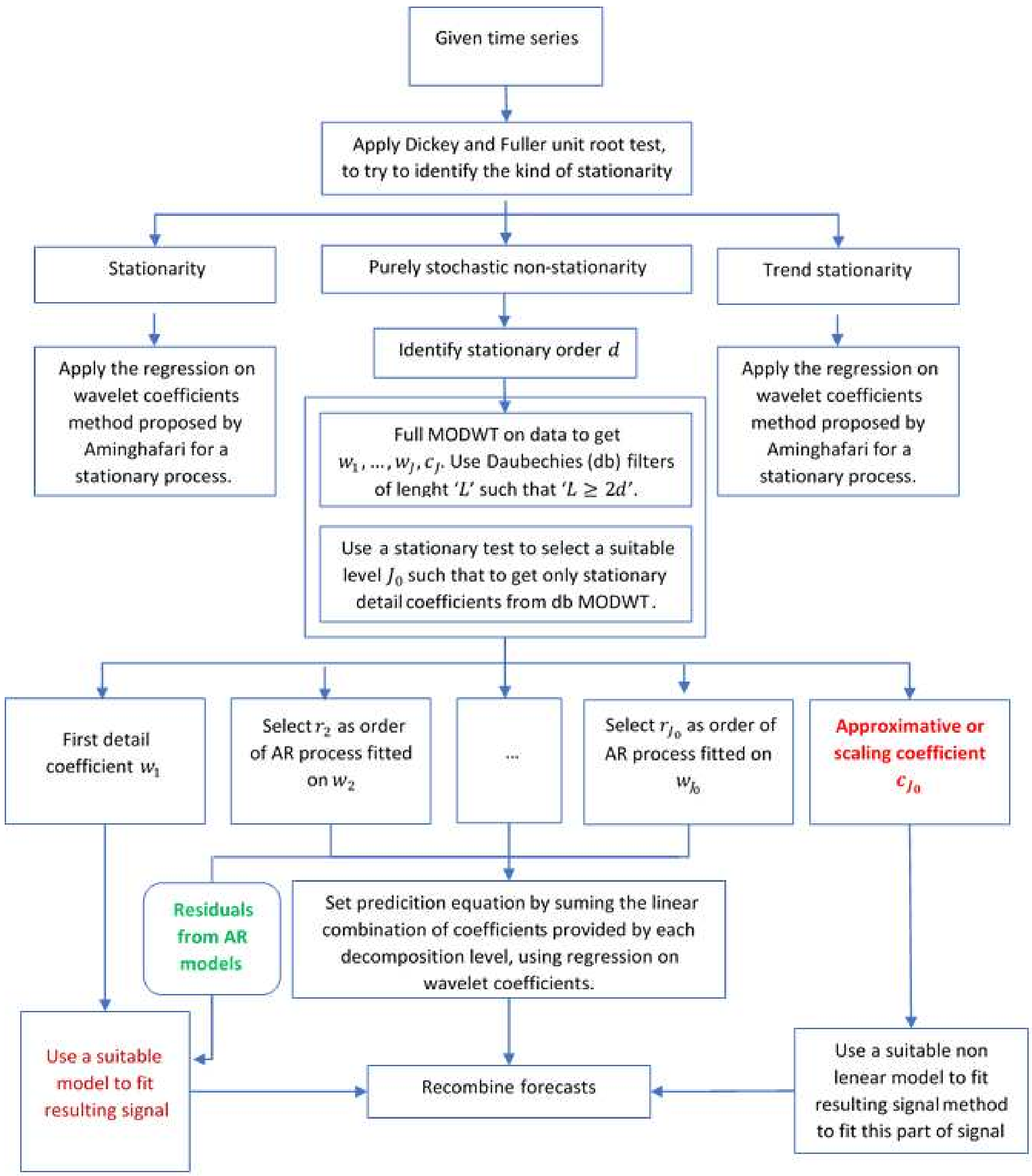


FIGURE 4: Proposed approach

Data visualization

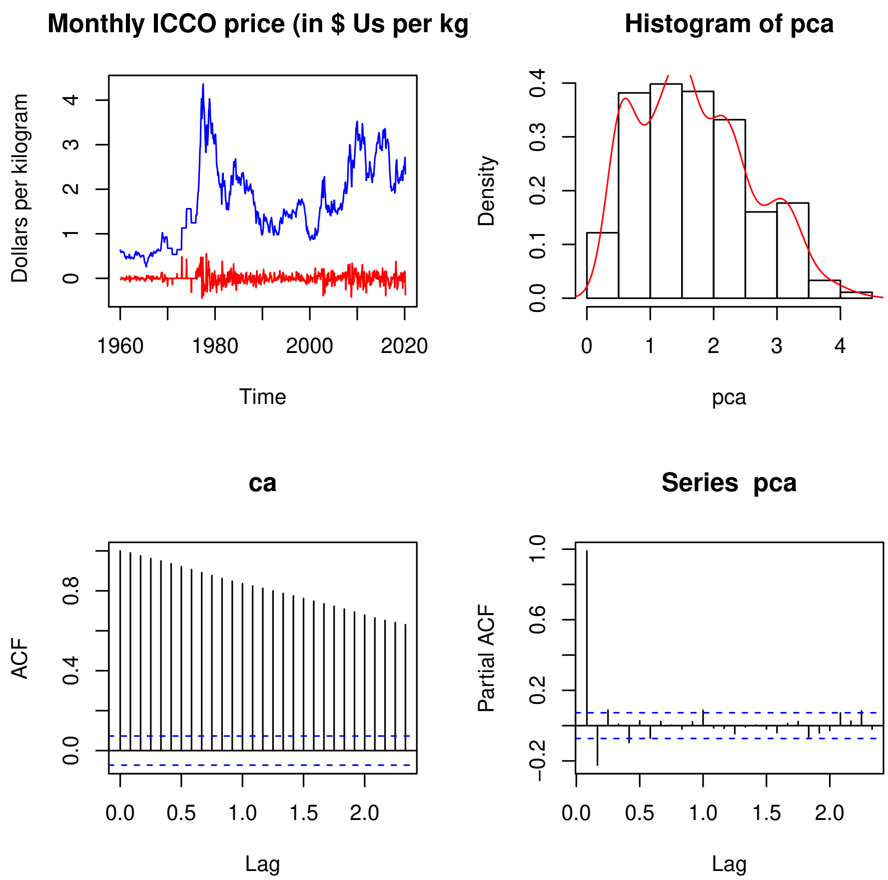


FIGURE 5: Cocoa price data visualization

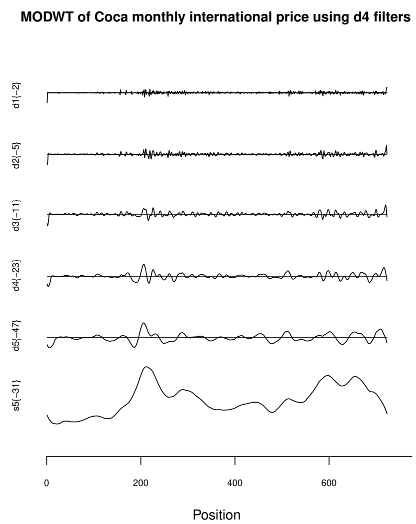
ARCH test on residuals from ARIMA model

TABLE 9: ARCH-LM test

Order	$F-$Stat.	$p-$value
4	493.2	0.00
8	214.4	0.00
12	133.5	0.00
16	93.2	$2.64e-13$
20	71.1	$6.01e-08$
24	58.1	$7.12e-05$

Whatever the regression order, the $p-$value of the statistic is likely equal to zero. So, there are enough evidence to reject the null hypothesis of absence of heteroscedasticity.

Wavelets-based decomposition



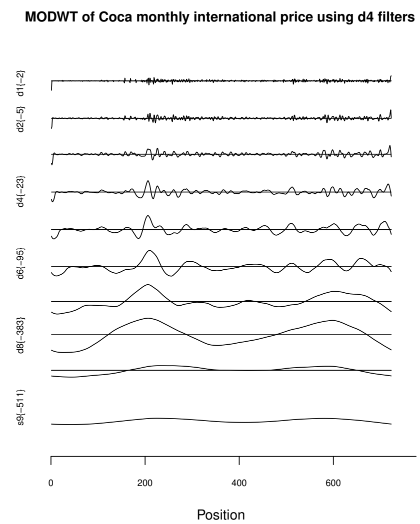


FIGURE 6: Coef. MODWT - D4 - $J = 5$ levels FIGURE 7: Coef. MODWT - D4 - $J = 9$ levels

.

Wavelet coefficients ACF and PACF

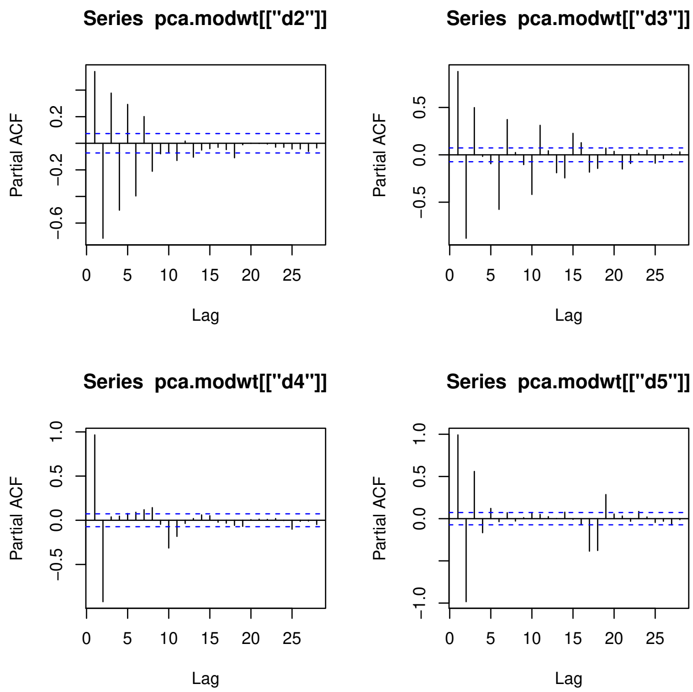


FIGURE 8: PACF of detail coefficients : d_2 - d_5

Fitting models for scaling coefficient

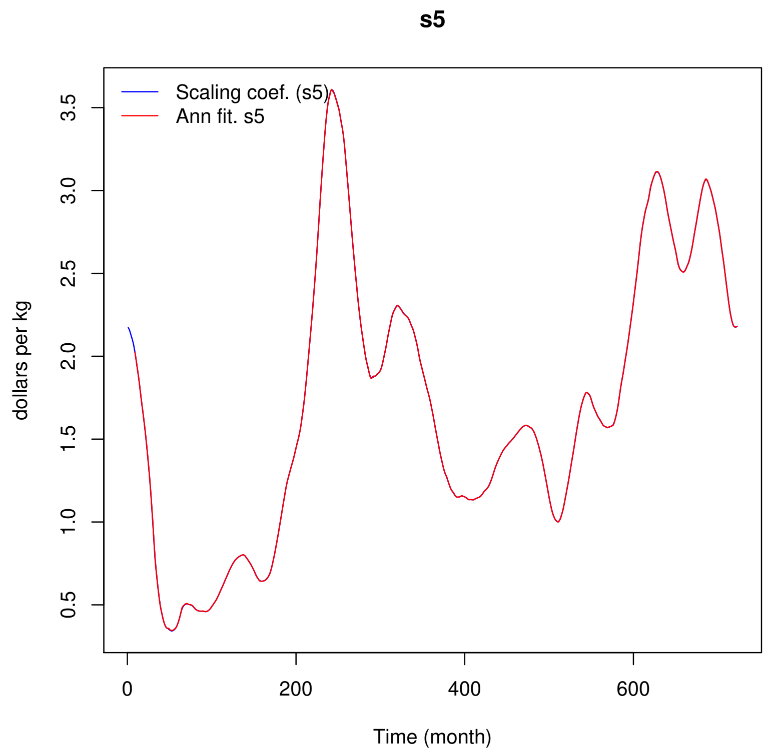


FIGURE 9: Actual and predicted values from ANN fitting for the scaling coefficient