International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 07 – Issue 10 || October 2022 || PP. 09-25

Modelling Prices of Petroleum Products and Their Volatilities Using Artificial Neural Networks

Faith Chebet[1*], Pius Kihara[2], Anthony Karanjah[1]
[1]*Department of Mathematics, Multimedia University of Kenya*
[2]*Department of Financial and Actuarial Mathematics, Technical University of Kenya*
[*]*chebetfk@gmailcom*

Abstract: Petroleum market is an uncertain market with regard to the price volatilities of petroleum products. This therefore brings out the need to develop a models that would help to forecast the prices and that which will forecast price volatilities in order to improve on the certainty of making future decisions relating to sales in the Kenya's petroleum market. Petroleum products, mainly petrol, kerosene and diesel, are important in driving the economies of all countries in the world but despite this, petroleum products prices have been going through fluctuations and instability, often affecting the efficiency of the same in propelling growth. The study used secondary data from Kenya National Bureau of Statistics library for the period 2011 to 2021 which was segmented into training data in the model and test data. This study showed that ANN outperformed the Auto Regressive Integrated Moving Average (ARIMA) model for predicting Kenyan petrol, diesel, and kerosene prices. This was based on the test data's Mean Squared Error (MSE) performance measures, which comprised 20% of the data. ANN also showed acceptable level of accuracy in the prediction of volatilities of prices of petrol, diesel and kerosene.

Keywords: Artificial Neural Networks(ANN), Kenya National Bureau of Statistics(KNBS), Auto Regressive Intergrated Moving Average(ARIMA), Mean Square Error(MSE) and Root Mean Square Error(RMSE).

1 Introduction

Kenya's economy largely depends on the energy sector. When vision 2030 was launched, the two main types of energy in Kenya, petroleum and electricity, were anticipated to be the prime movers of the modern sector of the Kenyan economy. As at 2020, the yearly demand of petroleum fuels stood at 4.7 million tones, all imported either as crude oil for processing at the Kenya Petroleum Refineries Limited or as refined petroleum products (EPRA, 2022)(4). Petroleum market in Kenya is regulated by Energy and Petroleum Regulatory Commission (EPRA) whose among its mandate is to fix fuel prices at the 15th of every month.

Because pricing of fuels is highly volatile, a fuel user who expects to use large amounts of any fuel and expects the prices to rise in the future will enter into a forward contract with a fuel distributor to buy the fuel at a pre-agreed future price which is lower than what they anticipate the future price to be. However, the timing on when to enter into forward contracts depends on the users' ability to accurately forecast. As a result, a model that can be turned into a credible instrument for predicting fuel prices is necessary.

Forecasting of fuel prices and their volatilities will require use of non linear models to achieve precise predictions. There are two categories of such models. The time series models used for non linear predictions such as the Auto Regressive Integrated Moving Average (ARIMA) and the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models. These models assume linearity in order to predict non linear time series. On the other hand are the non parametric models such as the Kernel models, Local polynomial models, the splines and the Artificial Neural Networks (ANN). These models do not conform to any distribution and do not assume specific data points to fit a model and hence are the greatest at capturing non linearity. The model's suitability for making a good prediction model will be determined by its timeliness to forecast, reliability, acceptable level of accuracy, useful output, and ease of use (Arienda et al, 2015)(1).

2 Related Works

Price prediction and price volatility prediction is a vast field of research. There has been research in this field including forecast models. These predictions have been modelled either using time series models, the non parametric techniques or both. In other countries both the time series models and the neural networks have been applied. Most of which conclude that ANN models are better predictive models as compared to traditional time series models. In Kenya's oil industry, however, most of the models applied in price prediction and price volatility prediction are the times series models. The following are some conclusions by researchers as far as prediction in the petroleum industry is concerned. Farjamnia et al. (2007)(5) found that the ANN model produced more accurate daily oil price predictions in Iran than the auto regressive integrated moving average model. Likewise, Kee Wei Yee and Humaida Banu Samsudin (2021)(7) compared the forecasting performance of ARIMA and ANN models in forecasting palm oil prices in Malaysia. The mean absolute percentage error (MAPE) for both models were below 10% but ANN model gave more accurate predictions than ARIMA model. Bildirici and Ersin, (2015) (3) also concluded that as compared to the findings of the baseline GARCH family model, the logistic smooth transition autoregressive (LSTAR) based and neural network augmented models demonstrated considerable gains in terms of estimating the daily returns of oil prices in Brent. Uniformly, Kristjanpoller and Minutolo (2016)(8) concluded that an hybrid of ANN and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) increased price volatility prediction precision by 30% as compared to only using GARCH model as in their previous research. With these comparisons, it means that ANN models have better predictive accuracy as compared to traditional time series models. Some researchers in Kenya have also ventured into price prediction in the petroleum industry. Ndei S.(2006)(11) made short term prediction of crude oil prices in Kenya using Univariate Box-Jenkins Auto-Regressive Integrated Moving Average (UBJ-ARIMA). He concluded that the accuracy of this model decreased rapidly with change of information. Nyongesa and Wangala(2016)(10) also made a prediction of diesel prices and their volatility using ARIMA model for 5 months. Based on these researches, the predictions were made using time series models which are limited to short term predictions. Short-term forecasts are insufficient for buyers, sellers, and policymakers as a tool. Therefore, emphasis on a more reliable long term prediction model is needed. Furthermore, Mwikamba and Aiyenga (2019) (9) determined that inflating prices, such as prices of petroleum products, are best predicted using non-linear models. Comparisons between the traditionally used time series models and the Artificial Neural Networks in other countries have validated the advantage of ANN predictive models. The emerging artificial intelligence algorithms such as Artificial neural networks have been applied in prediction of prices in the energy sector. These algorithms are mainly an improvement of the traditional time series models. Azadeh et al. (2012)(2) proposed a flexible algorithm for optimum long-term oil price forecasting in noisy, uncertain, and complex situations, based on artificial neural networks (ANN) and fuzzy regression (FR). The flexibility of these algorithms allowed long term prediction rather than short term prediction of oil price. Gupta and Nigam (2020)(6) identified that the use of ANNs continuously capture the volatile pattern of crude oil prices if the optimal lag and number of the price controlling delay effect is used. ANN models therefore have the advantage of better prediction accuracy for non-linear data and long term prediction because of continuous flexibility of capturing volatile data.

3 Methodology

The research methodology used in the study is summarized below. The data used was Kenyan petroleum products prices in Kenyan shillings per litre. The petroleum products considered are petrol, diesel and kerosene. Each of these three products have a different model in each case. R packages are used for development of the models.

3.1 Research Design

Secondary data was obtained from the KNBS library and the EPRA for the period January 2011 to December 2021. The era that was considered is after gazetting the energy act(2006) on petroleum pricing. The data that was acquired was monthly data since EPRA adjusts prices of all petroleum products monthly. Using the Pareto principle, data was split into two; 80% of the data was used for training and the remaining 20% was used for testing.

Monthly prices of petroleum products and their volatilities from the data are non-linear and hence prediction of their future values required the use on non-linear models. With the assumption of linearity, future values are cramped to be linear functions of past data by using the ARIMA model. On the other hand, non linear function can be approximated using ANNs with two layers of trainable weights. This is without the assumption of non-linearity. ANNs are function generators that generate a data series as an output based on a learnt function or data model. Since an ANN network can efficiently approximate a continuous function to the necessary level of precision, ANNs were used in both price prediction and price volatility prediction.

3.2 Proposed Auto Regressive Integrated Moving Average (ARIMA) model

The proposed ARIMA predictive model was given by:

$$\hat{Y}_t = c+\alpha_1(Y_{t-1})+...+\alpha_p(Y_{t-p})+\beta_1(\epsilon_{t-1})+...+\beta_q(\epsilon_{t-q})+\epsilon_t \tag{1}$$

Where $\hat{Y}_t$ is the predicted price at time t, Y_{t-1}, ..., Y_{t-p} are the lagged prices, ϵ_{t-1}, ..., ϵ_{t-p} are the lagged white noises, ϵ_t is the white noise at time t and α_1, ..., α_p, β_1,..., β_q are constants.

Developing the ARIMA model involves three steps. The first step is model identification which involves making the data stationary by identifying the degree, d, and identifying the order p and order q. We then use the

Akaike's Information Criterion (AIC) to find the values of p and q. For that reason, differencing is not used to find the value of d because the data on which the likelihood is computed changes as a result of the differencing, the AIC values of models with different orders of differencing are not comparable. We therefore choose a value of d with the help of the Kwatkowski-Phillips-SchmidtShin (KPSS) test. AIC will then be used to determine which statistical model is the best. The AIC formula is as follows:

$$AIC = -2log(L) + 2(p + q + k + 1) \quad (2)$$

Where L denotes the data's likelihood, $k = 1$ if $c \neq 0$, and $k = 0$ if $c = 0$. The value of p and q will be obtained by minimizing AIC formula.

After tentative model identification, the next step will be to estimate the parameters in the model. For Auto Regressive (AR) process, Moving Average(MA) process and Auto Regressive Moving Average (ARMA) process, we estimate the values of c, $\alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q$. The method that was used to estimate the parameters is maximum likelihood estimation. This was done with the help of R statistical package.

Once the parameters have been estimated, the adequacy of the model is assessed by using residual analysis, we calculate the residuals from the fitted model and plot the auto correlation function (ACF) to see if the tentative model was correct by checking for stationarity characteristics. We also use the the Ljung-Box test to test for correlations on the residuals. If the test determines that there are no correlations, then the tentative model is considered adequate.

3.3 Proposed Artificial Neural Network (ANN) Model

The models that were used here are the feed forward neural networks. These networks are made up of layers, each of which has a number of nodes connected to nodes in the next layer. The models used have one hidden layer considering that the prediction of price and volatility in price of petroleum products is a simple linear relationship between the lagged values and the predicted value. Thus, the ANN predictive model for prices was given by:

$$\hat{Y}_t = \omega_0 + \sum_{j=1}^{q} \omega_j . g(\omega_{0j} + \sum_{i=1}^{p} \omega_{ij} Y_{t-i}) + \epsilon_t \quad (3)$$

Where $\hat{Y}_t$ is the predicted price at time t, Y_{t-i}; $i = 1,2,...,p$, are the lagged prices, ω_0 and ω_j are constant terms of the output and the hidden layers respectively , ω_{0j} and ω_{ij} are the connection weights between the inputs and the hidden neurons, p is the number of input nodes, q is the number of hidden nodes and $g(.)$ is the sigmoid activation function which transforms a real valued input into a range of between 0 and 1. Similarly, the price volatilities prediction model was given:

$$\hat{\sigma}_t^2 = \phi_0 + \sum_{h=1}^{q} \phi_h . g(\phi_{0h} + \sum_{i=1}^{p} \phi_{ih} \sigma_{t-i}^2 + \phi_{rh} \epsilon_{t-1}) \quad (4)$$

Where $\hat{\sigma}_t^2$ is the output vector, σ_{t-i}^2 is the input matrix with time lags of $t-i$; $i = 1,2,...,p$, $g(.)$ is the sigmoid function, ϕ_0 and ϕ_h are constant terms in the output and hidden layers respectively, ϕ_{0h} and ϕ_{ih} are connection weights between the inputs and the hidden neurons, p is the number of input nodes and q is the number of hidden nodes.

Developing an ANN predictive model involves creating a network topology and training the network. Creating a network topology involves choosing the number of input neurons, the number of hidden layers, the number of hidden neurons in the hidden layer, and the number of output neurons. The input neurons are the lagged variables $Y_{t-1},..., Y_{t-p}$ for price prediction and $\sigma_{t-1}^2,...,\sigma_{t-p}^2$ for price volatility prediction. The output layer has one neuron which is the value of $\hat{Y}_t$ for price prediction and $\hat{\sigma}_t^2$ for price volatility prediction. The hidden layer as depicted by the model is one with hidden neurons between the value of input neurons and the sum of the input and output neurons. Using the rule of thumb, the number of hidden neurons is the sum of the output neurons and two thirds the number of input neurons. The hidden neurons contain a combination of inputs and respective weights and a bias value.

Training of the network involves inputting the training and target data, the training function, type of activation function, and transfer function. The training is repeated with the weights adjusted in order to reduce error. Epoch is the name given to each run through a whole data collection. A continual interaction(k) with the environment adjusts the weights. The next weight to be used is an adjustment of the previous weight given by the equation:

$$\omega_{ij}(k + 1) = \omega_{ij}(k) + \triangle\omega_{ij}(k) \quad (5)$$

Where $\omega(k)$ is the previous value of the weight vector, $\omega(k + 1)$ is the adjusted result from the previous weight vector, $\triangle$ is the adjustment function in solving a learning problem and $\triangle\omega(k)$ is the environment stimuli vector.

After weight adjustment, the resulting error signal will be given by:

$$e_i(k) = d_i(k) - y_i(k) \quad (6)$$

Where y_i is the neuron response in the i^{th} iteration and d_i is the resulting trained value from the adjusted weight. The error is measured using the performance function which is analogous to Mean Square error(MSE). The best model is obtained by training data with different network structure each with different epoch size. The one with the least performance function is then considered the best model. These weight adjustments and error measure was done with the help of R statistical package.

The adequacy of the model is assessed by using residual analysis, we calculate the residuals from the fitted model and plot the auto correlation function (ACF) to see if the tentative model was correct by checking for stationarity characteristics. We also use the the Ljung-Box test to test for correlations on the residuals. If the test determines that there are no correlations, then the tentative model is considered adequate.

3.4 Model Testing and Validation

To test the reliability of the ANN in price prediction, we compare the performance of the best ANN model on the test data with the performance of ARIMA model on the test data. To compare the two models, we use the mean square error (MSE) and the root mean square error (RMSE) of the results. The MSE is given by:

$$(1/n)\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 \quad (7)$$

Where n is the number of data points, Y_t is the actual value and $\hat{Y}_t$ is the predicted value. RMSE is given by the square root of MSE. The one with the lowest MSE and RMSE values is considered the more reliable model for prediction.

4 Empirical Results and Discussions

This section gives the resulting ARIMA and ANN models for predicting the prices of the petroleum products. Since each product is independent of each other, each product therefore had its own model in each case. The comparison of the performance between the ARIMA model and the ANN model for price prediction of each petroleum products is also described. In addition, the ANN prediction models for monthly volatility of prices of each of the three petroleum products are also discussed.

4.1 Modelling Prices of Petroleum Products Using ARIMA Model

This section shows how the ARIMA models were obtained. For all the tests in this section, we used the significance level of 0.05.

4.1.1 ARIMA Predictive Model for Petrol Price

With the help of the KPSS test for stationarity, we determined the value of the differencing order, d. Table 1 below shows the results of the KPSS test:

Data	P value	Conclusion
Original	0.01	Reject the null hypothesis (not stationary)
First differencing	0.1	Fail to reject the null hypothesis (stationary)

Table 1: Kwiatkowski–Phillips–Schmidt–Shin(KPSS) Test Results for Petrol Prices Series

As shown in table 1 above, the first differencing made the series stationary and hence the value of $d = 1$. After selecting the value of d, we therefore choose the value of orders p and q with the help of AIC value criterion. the AIC values are shown in table 2 below:

Since ARIMA(0,1,1) has the lowest AIC value, it was then considered the best model for forecasting petrol Prices. The these ARIMA (0,1,1) was selected to be the model for forecasting. data was then fitted in the model ARIMA (0,1,1) and the Ljung-Box test on the residuals resulted to a p-value of 0.9971 Which meant the residuals had no correlation. From these ARIMA (0,1,1) was selected to be the model for forecasting.

Model	AIC Value
ARIMA(0,1,0)	566.9987
ARIMA(0,1,1)	553.3565
ARIMA(0,1,1)(1,0,0)[12]	554.3016
ARIMA(0,1,1)(0,0,1)[12]	554.2288
ARIMA(0,1,1)(1,0,1)[12]	556.3881
ARIMA(1,1,1)	555.4777
ARIMA(0,1,2)	555.4777
ARIMA(1,1,0)	556.0273

Table 2: Akaike's Information Criterion (AIC) Values for a Set of Models- Petrol Prices

4.1.2 ARIMA Predictive Model for Diesel the KPSS test:

We calculated the value of the differencing order, *d*, using the KPSS test for stationarity. Table 3 below shows the results of the KPSS test:

Data	P value	Conclusion
Original	0.01	Reject the null hypothesis (not stationary)
First differencing	0.1	Fail to reject the null hypothesis (stationary)

Table 3: KPSS Test Results for Diesel Prices Time Series

As shown in table 3 above, the first differencing made the se- *q* with the help of AIC value criterion. the AIC values are ries stationary and hence the value of $d = 1$. After selecting shown in table 4 below: the value of *d*, we therefore choose the value of orders *p* and q with the help of AIC value criterion. the AIC values are shown in table 4 below:

Model	AIC Value
ARIMA(0,1,0)	558.8908
ARIMA(1,1,0)	555.4241
ARIMA(1,1,0)(1,0,0)[12]	557.5387
ARIMA(1,1,0)(0,0,1)[12]	557.5369
ARIMA(1,1,0)(1,0,1)[12]	Inf
ARIMA(2,1,0)	557.456
ARIMA(1,1,1)	557.4824
ARIMA(0,1,1)	555.4693
ARIMA(2,1,1)	558.8162

Table 4: Akaike's Information Criterion (AIC) for a set of models - Diesel Prices

Since ARIMA (1,1,0) has the lowest AIC value, it was then considered the best model for forecasting diesel Prices. The data was then fitted in the model ARIMA (1,1,0) and the Ljung-Box test on the residuals resulted to a p-value of 0.9451, indicating that the test was insignificant and that the residuals lacked correlation. As a result, ARIMA (1,1,0) was chosen as the forecasting model.

4.1.3 ARIMA Predictive Model for Kerosene

The value of the differencing order, d, was established using the KPSS test for stationarity. Table5 shows the results of the KPSS test:

As shown in table 5 above, the first differencing made the series stationary and hence the value of d = 1. After selecting the value of d, we therefore choose the value of orders p and q with the help of AIC value criterion. the AIC values are shown in table 6 below:

Data	P value	Conclusion
Original	0.01	Reject the null hypothesis (not stationary)
First differencing	0.1	Fail to reject the null hypothesis (stationary)

Table 5: KPSS Test Results for Kerosene Prices Time Series

Model	AIC Valuel
ARIMA(0,1,1)	601.66
ARIMA(0,1,0)	602.9056
ARIMA(0,1,1)(0,0,2)[12]	599.99
ARIMA(0,1,1)(0,0,1)[12]	600.67
ARIMA(0,1,1)(1,0,2)[12]	601.57
ARIMA(0,1,1)(1,0,1)[12]	601.43
ARIMA(0,1,0)(0,0,2)[12]	600.8
ARIMA(1,1,1)(0,0,2)[12]	600.62
ARIMA(0,1,2)(0,0,2)[12]	601.98
ARIMA(1,1,0)(0,0,2)[12]	600.06
ARIMA(1,1,2)(0,0,2)[12]	602.97

Table 6: Akaike's Information Criterion (AIC) for a Set of Models- Kerosene Prices

It was decided that ARIMA (0,1,1) (0,0,2) [12] was the best model for predicting kerosene prices because it has the lowest AIC value. The data was then fitted in the model ARIMA (0,1,1) (0,0,2) [12] and the Ljung-Box test on the residuals resulted to a p-value of 0.9846, demonstrating the insignificance of the test and the absence of correlation in the residuals. The forecasting model chosen was ARIMA (0,1,1) (0,0,2)[12] forecast accuracy. This measure is obtained by applying the trained model on the test data. It is given by the formula:

$$Accuracy = 1 - |(1/n)\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)}{\sum_{t=1}^{n} Y_t}|$$

Where n is the number of data points, Yt is the actual value and ˆ Y_t is the predicted value.

4.2 Modelling Prices of Petroleum Products Using ANN models

The ANN prediction models varies depending on the number of input lags, p, from which we also calculate the number of hidden neurons. The best model is the one with the highest

4.2.1 ANN Predictive Model for Petrol Prices

The first step is to choose the best ANN model. Table 7 below shows some of the model combinations used for predicting petrol prices and their accuracy:

Number of Lags	Number of hidden Layers	Accuracy
2	2	0.9822105
3	3	0.9810008
4	4	0.9820479
5	4	0.9822269
6	5	0.9857805
12	9	0.9783426

Table 7: ANN Model Identification- Petrol Prices

From table 7, the best model for predicting petrol prices is the one with 6 input lags and 5 hidden layer nodes since it has the highest accuracy as compared to the other combinations.

Its structure is shown in figure 1 below.

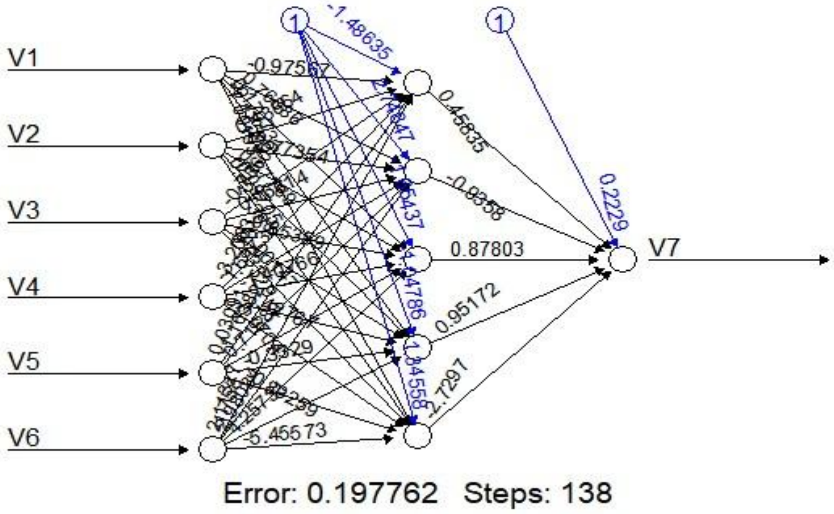

Figure 1: Structure of the Predictive ANN Model for Petrol Prices

As shown in figure 1, 6 price lags are used to predict the 7th price. The 7th price is indicated by the value V 7 and the input lags are indicated by the values V 1, V 2, ..., V 6. The network topology was achieved after 138 iteration steps and has an error of 0.199762.

4.2.2 ANN Predictive Model for Diesel Prices

The accuracy of some of the model combinations used to forecast diesel prices is shown in the table 8 below.

Number of Lags	Number of hidden Layers	Accuracy
2	2	0.9990676
3	3	0.99982
4	4	0.9966798
5	4	0.9994499
6	5	0.9991558
12	9	0.9913483

Table 8: ANN Model Identification- Diesel Prices

From table 8, the best model for predicting diesel prices is the one with 3 input lags and 3 hidden layer nodes since it has the highest accuracy as compared to the other combinations.

Its structure is shown in figure 2 below.

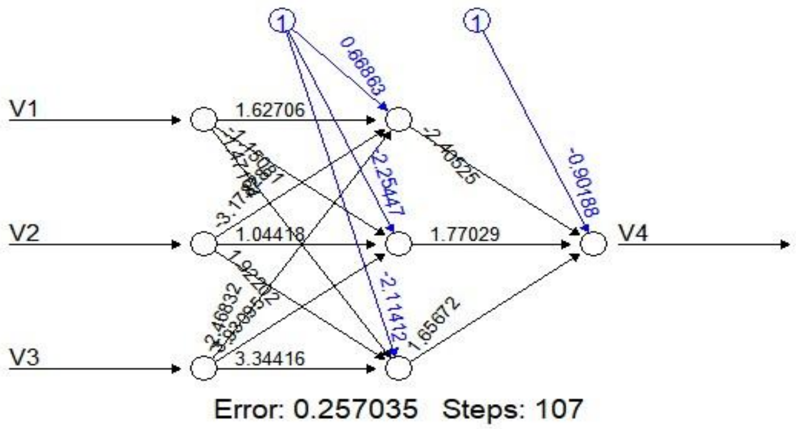

Figure 2: Structure of the Predictive ANN Model for Diesel Prices

As seen in figure 2, the fourth price is predicted using three price lags. The fourth price is denoted by V 4 and the values V 1, V 2 and V 3 denote the input lags. After 107 iterations, the network topology was obtained whose error term is 0.257035.

4.2.3 ANN Predictive Model for Kerosene Price

The accuracy of some of the model combinations used to forecast kerosene prices is shown in the table 9 below.

Number of Lags	Number of hidden Layers	Accuracy
2	2	0.9911573
3	3	0.9832892
4	4	0.9835756
5	4	0.9836891
6	5	0.981707
12	9	0.9732276

Table 9: ANN Model Identification- Kerosene Prices

From table 9, the best model for predicting diesel prices is the one with 2 input lags and 2 hidden layer nodes since it has the highest accuracy as compared to the other combinations.

Its structure is shown in figure 3 below.

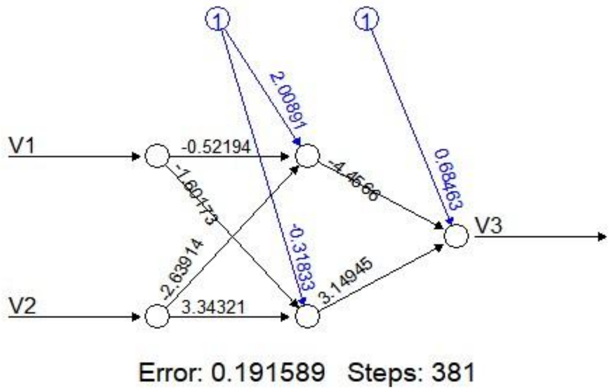

Figure 3: Structure of the Predictive ANN Model for Kerosene Prices

As seen in figure 3, the third price is predicted using two price lags. The third price is denoted by V 3 and the values V 1 and V 2 denotes the input lags. The network topology was obtained after 381 iterations, with an error term of 0.191589.

4.3 Performance of the Predictive Models

The second specific objective was to compare the performance of the ANN models with the performance of ARIMA models in predicting the pricing of petroleum products. The test data was used to compare the prediction performance of the ARIMA model and ANN model in each case of the petroleum products.

4.3.1 Petrol Models

Table 10 below shows the actual values and the predicted values produced by the ARIMA model, ARIMA(0,1,1), and the ANN model with 6 input lags and 5 hidden layer nodes. The data in table 10 can be represented in a comparative plot shown in figure 4

Sample Period	Actual Value	Predicted Value(ARIMA)	Predicted Value(ANN)
Oct 2019	108.83	115.0367	112.31412
Nov 2019	110.99	115.0367	109.44405
Dec 2019	109.91	115.0367	111.80031
Jan 2020	110.61	115.0367	110.38491
Feb 2020	112.58	115.0367	112.02050
Mar 2020	112.07	115.0367	112.53318
Apr 2020	94.09	115.0367	112.85602
May 2020	84.58	115.0367	94.57256
Jun 2020	90.34	115.0367	86.57255
Jul 2020	101.37	115.0367	89.02690
Aug 2020	104.83	115.0367	106.09272
Sep 2020	106.30	115.0367	111.07427
Oct 2020	108.13	115.0367	108.34592
Nov 2020	106.72	115.0367	106.98970
Dec 2020	107.69	115.0367	106.44984
Jan 2021	107.86	115.0367	107.93431
Feb 2021	116.03	115.0367	108.29723
Mar 2021	123.66	115.0367	115.70646
Apr 2021	123.66	115.0367	120.23666
May 2021	127.21	115.0367	120.12106
Jun 2021	127.98	115.0367	119.63718
Jul 2021	127.98	115.0367	120.73430
Aug 2021	127.98	115.0367	121.52502
Sep 2021	135.54	115.0367	120.76629
Oct 2021	130.54	115.0367	123.76341
Nov 2021	130.54	115.0367	122.78922
Dec 2021	130.54	115.0367	-

Table 10: Sample Results of ANN and ARIMA Predictive Models for Petrol Prices

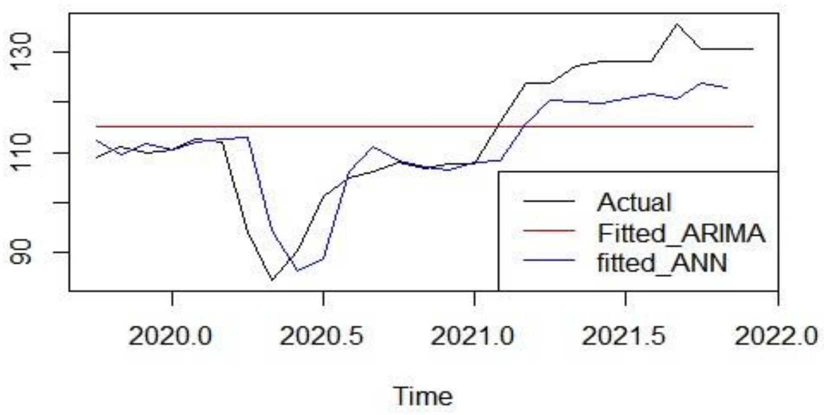

Figure 4: Graph of Predicted Values of ARIMA and ANN Models Against Actual Petrol Prices.

From figure 4, we see that the ARIMA model produced constant forecasts whose accuracy decreases with time lapse. On the other hand, ANN model forecasts shows adjustments to fit the actual data. This results to performance statistics given by mean square error (MSE) and root mean square error (RMSE) shown in table 11.

Model	MSE	RMSE
ARIMA	171.24	13.08587
ANN	51.39798	7.169239

Table 11: Performance Statistics for Petrol Prices Prediction Models

The MSE and RMSE values of the ANN model in table 11 are smaller than those of the ARIMA model. It is therefore concluded that the ANN model is better at forecasting petrol prices than the ARIMA model.

4.3.2 Diesel Models

Table 12 below shows the actual values and the predicted values produced by the ARIMA model, ARIMA(1,1,0), and the ANN model with 3 input lags and 3 hidden layer nodes. The data in table 12 can be represented in a comparative plot shown in figure 4

Sample Period	Actual Value	Predicted Value(ARIMA)	Predicted Value(ANN)
Oct 2019	102.82	103.6678	104.59047
Nov 2019	105.10	103.6149	103.53141
Dec 2019	102.28	103.6029	104.84094
Jan 2020	102.81	103.6002	104.12903
Feb 2020	105.37	103.5995	102.88593
Mar 2020	102.93	103.5994	105.35750
Apr 2020	98.84	103.5994	104.61106
May 2020	79.67	103.5994	100.05256
Jun 2020	75.88	103.5994	81.58607
Jul 2020	92.81	103.5994	72.92252
Aug 2020	95.57	103.5994	88.72709
Sep 2020	95.45	103.5994	102.50998
Oct 2020	93.85	103.5994	98.02015
Nov 2020	91.64	103.5994	95.37034
Dec 2020	92.75	103.5994	92.28320
Jan 2021	97.33	103.5994	92.72166
Feb 2021	102.84	103.5994	98.68565
Mar 2021	108.58	103.5994	104.41981
Apr 2021	108.58	103.5994	107.75910
May 2021	108.58	103.5994	108.03639
Jun 2021	108.58	103.5994	107.15259
Jul 2021	108.58	103.5994	107.15259
Aug 2021	108.58	103.5994	107.15259
Sep 2021	116.51	103.5994	107.15259
Oct 2021	111.51	103.5994	109.70960
Nov 2021	111.51	103.5994	109.43783
Dec 2021	111.51	103.5994	-

Table 12: Sample Results of ANN and ARIMA Predictive Models for Diesel Prices

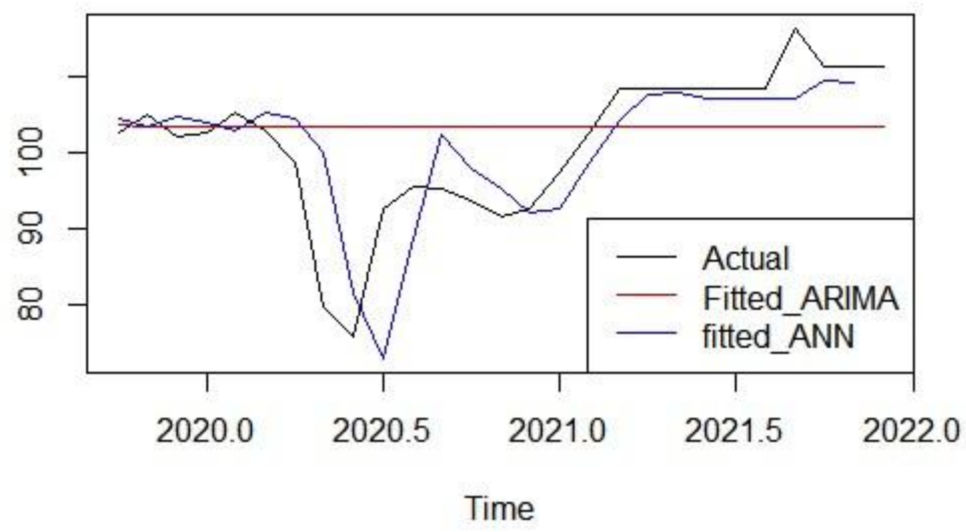

Figure 5: Graph of Predicted Values of ARIMA and ANN Models Against Actual Diesel Prices.

From figure 5, we see that the almost constant forecasts generated by the ARIMA model have decreasing accuracy over time. ANN model forecasts, on the other hand, display changes to fit the actual data. This results to performance statistics given by MSE and (RMSE) shown in table 13.

Model	MSE	RMSE
ARIMA	93.28624	9.65848
ANN	45.72903	6.762324

Table 13: Performance Statistics for Diesel Prices Prediction Models

The ANN model in table 13 has lower MSE and RMSE values than the ARIMA model. As a result, it can be said that the ANN model outperforms the ARIMA model at predicting diesel prices.

4.3.3 Kerosene Models

Table 14 below shows the actual values and the predicted values produced by the ARIMA model, ARIMA(0,1,1)(0,0,2), and the ANN model with two input lags and 2 hidden layer nodes. The data in table 12 can be represented in a comparative plot shown in figure 6.

Sample Period	Actual Value	Predicted Value(ARIMA)	Predicted Value(ANN)
Oct 2019	101.94	100.43266	99.88007
Nov 2019	104.53	99.34225	100.14442
Dec 2019	102.81	98.86193	101.58559
Jan 2020	104.46	97.63857	100.50182
Feb 2020	103.65	97.59334	101.51065
Mar 2020	96.72	96.55321	100.99038
Apr 2020	78.59	96.82008	96.23874
May 2020	81.08	96.72665	75.85094
Jun 2020	63.79	95.13057	81.93865
Jul 2020	66.41	94.26974	60.09416
Aug 2020	84.60	93.59490	65.77699
Sep 2020	84.09	88.13713	87.35358

Oct 2020	84.67	88.10749	84.94789
Nov 2020	82.58	88.25526	85.71713
Dec 2020	84.50	88.88228	83.04703
Jan 2021	88.07	89.76765	85.68842
Feb 2021	93.37	89.69086	89.58731
Mar 2021	98.78	90.72274	94.54704
Apr 2021	98.78	90.32605	98.50667
May 2021	98.78	89.91073	98.19707
Jun 2021	98.78	89.65788	98.19707
Jul 2021	98.78	89.97419	98.19707
Aug 2021	98.78	90.92620	98.19707
Sep 2021	111.74	91.60599	98.19707
Oct 2021	104.46	91.69470	104.67258
Nov 2021	105.46	91.69470	101.04234
Dec 2021	106.46	91.69470	101.95279

Table 14: Sample Results of ANN and ARIMA Predictive Models for Kerosene Prices

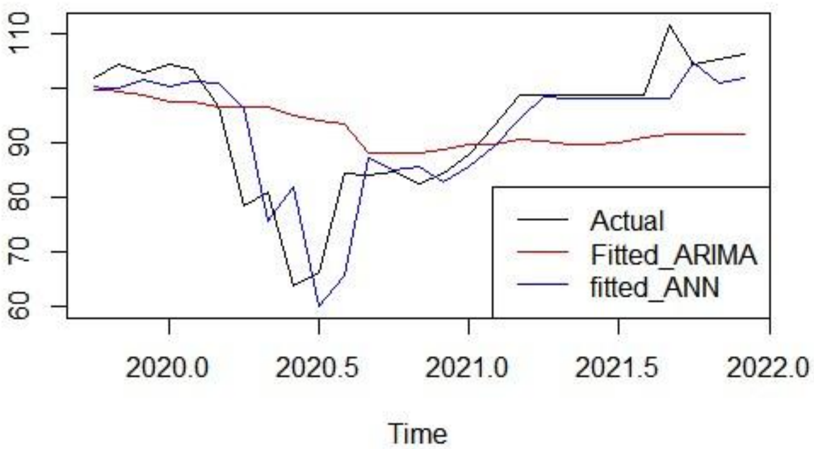

Figure 6: Graph of Predicted Values of ARIMA and ANN Models Against Actual Kerosene Prices.

From figure 6, we see that the linear forecasts produced by the ARIMA model tries to adjust with time but do not follow the pattern of the actual data. On the other hand, ANN model forecasts show adjustments to fit the actual data. This results to performance statistics given by mean square error(MSE) and root mean square error(RMSE) shown in table 15.

Model	MSE	RMSE
ARIMA	150.1414	12.25322
ANN	52.26468	7.229432

Table 15: Performance Statistics for Kerosene Prices Prediction Models

The ANN model's MSE and RMSE values in table 15 are lower than those of the ARIMA model. It follows that the ANN model performs better at predicting kerosene prices than the ARIMA model.

4.4 Modelling Monthly Volatilities of Prices of Petroleum Products Using ANN models

Before Modelling the volatilities were first calculated from the prices of petrol, diesel and kerosene. This was done in excel using the formula given in equation (9) below:

$$\sigma_t^2 = sd(Y_t, Y_{t-1}) * \sqrt{2} \quad (9)$$

Where σ_t^2 is the monthly volatility at time t, $sd(Y_t, Yt_{-1})$ is the standard deviation of the prices Y_t and Y_{t-1}. After obtaining the volatilities, the data undergoes preparation including normalization from which the models are developed. The ANN prediction models varies depending on the number of input lags, p, from which we also calculate the number of hidden neurons. The best model is the one with the highest forecast accuracy. This measure is obtained by applying the trained model on the test data. It is given by the formula given in equation (8)

4.4.1 ANN Predictive Model for Monthly Volatilities of Petrol Prices

Table 16 below shows some of the model combinations used for predicting monthly volatilities of petrol prices and their accuracy:

Number of Lags	Number of hidden Layers	Accuracy
2	2	0.7399163
3	3	0.7363572
4	4	0.7340871
5	4	0.7465204
6	5	0.7833321
12	9	0.67907

Table 16: ANN Model Identification- Petrol Prices' Volatilities

From table 16, the best model for predicting monthly volatilities of petrol prices is the one with 6 input lags and 5 hidden layer nodes since it has the highest accuracy as compared to the other combinations. Its structure is shown in figure 7 below.

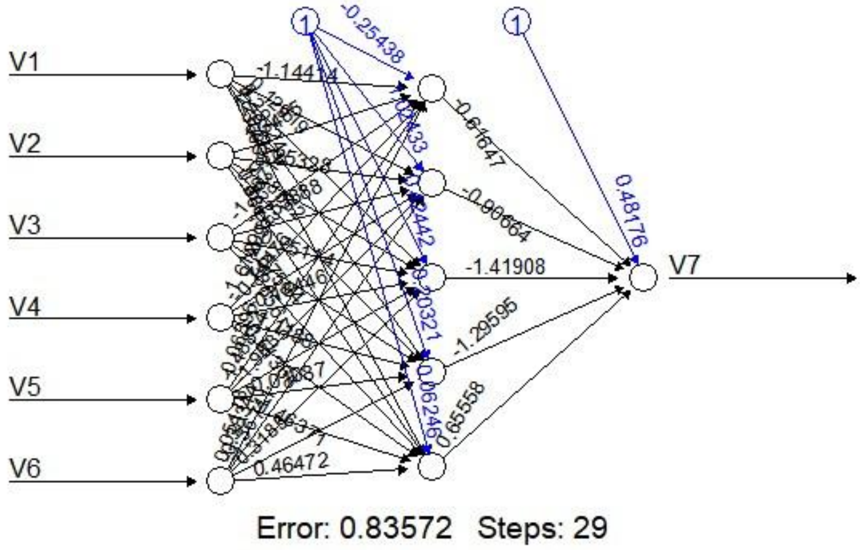

Figure 7: Structure of the Predictive ANN Model for Volatilities of Petrol Prices

As shown in figure 1, 6 price volatility lags are used to predict the 7th price volatility. The 7th volatility is indicated by the value V 7 and the input lags are indicated by the values V 1, V 2, ..., V 6. The network topology was achieved after 29 iteration steps and has an error of 0.83572. The model also resulted to MSE value of 18.51108 and the RMSE value of 4.302451 which are less than the MSE and RMSE values of the ANN predictive model for petrol prices. Figure 8 below shows how the values fitted by the network topology in figure 7 compare to the actual data.

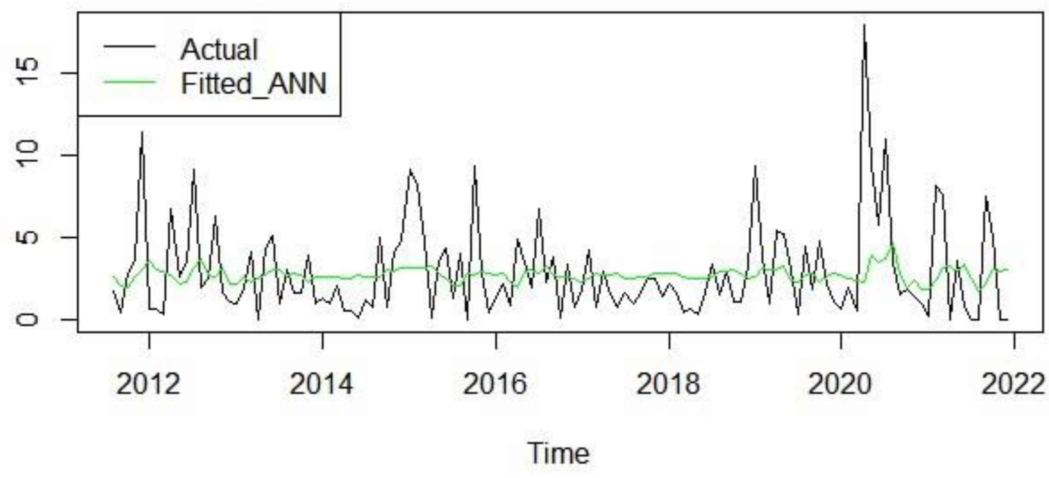

Figure 8: ANN Model Fitting for Volatilities of Petrol Prices

4.4.2 ANN Predictive Model for Volatilities of Diesel Prices

The accuracy of some of the model combinations used to forecast diesel prices is shown in the table 17 below.

Number of Lags	Number of hidden Layers	Accuracy
2	2	0.6920981
3	3	0.7021073
4	4	0.6799982
5	4	0.6854076
6	5	0.6694158
12	9	0.6725295

Table 17: ANN Model Identification- Diesel Prices' Volatilities

From table 17, the best model for predicting volatilities of diesel prices is the one with 3 input lags and 3 hidden layer nodes since it has the highest accuracy as compared to the other combinations. Its structure is shown in figure 9 below.

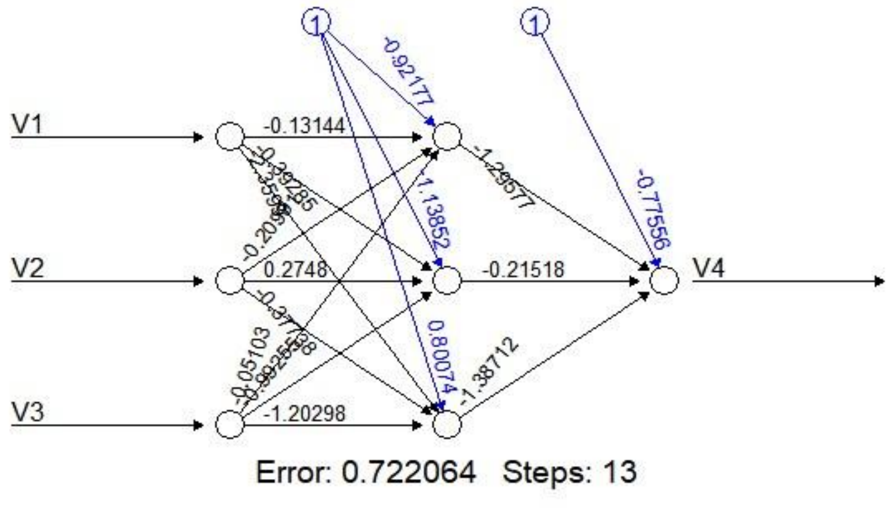

Figure 9: Structure of the Predictive ANN Model for Volatilities of Diesel Prices

As seen in figure 9, the fourth price is predicted using three price volatility lags. The fourth price volatility is denoted by $V\ 4$ and the values $V\ 1$, $V\ 2$ and $V\ 3$ denotes the input lags. After 13 iterations, the network topology was obtained whose error term is 0.722064. The model also resulted to MSE value of 23.02922 and the RMSE value of 4.798877 which are less than the MSE and RMSE values of the ANN predictive model for diesel prices. Figure 10 shows how the values fitted by the network topology in figure 9 are compared to the actual data.

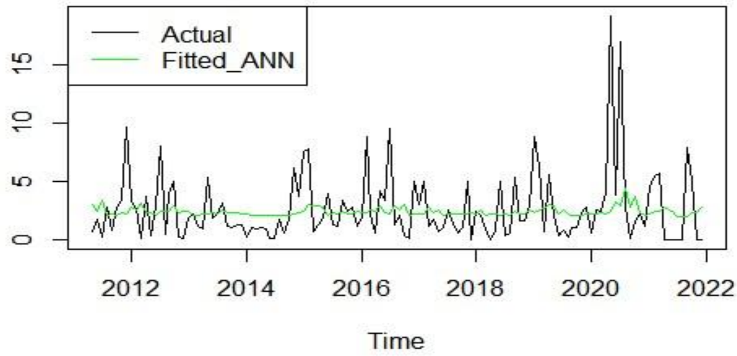

Figure 10: ANN Model Fitting for Volatilities of Diesel Prices

4.4.3 ANN Prediction Model for Volatilities of Kerosene Prices

The accuracy of some of the model combinations used to forecast kerosene prices is shown in the table 18 below.

Number of Lags	Number of hidden Layers	Accuracy
2	2	0.6425442
3	3	0.6419467
4	4	0.6512593
5	4	0.6168637
6	5	0.6370818
12	9	0.6068634

Table 18: ANN Model Identification- Kerosene Prices' Volatilities

From table 18, the best model for predicting volatilities of kerosene is the one with 4 input lags and 4 hidden layer nodes since it has the highest accuracy as compared to the other combinations. Its structure is shown in figure 11 below.

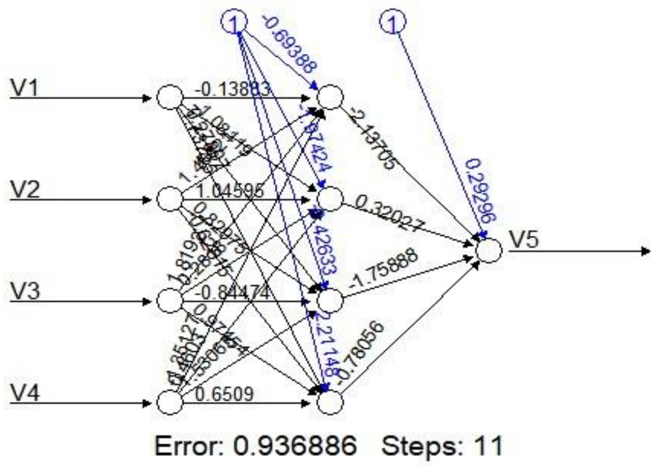

Figure 11: Structure of the Predictive ANN Model for Volatilities of Kerosene Prices

As seen in figure 11, the fifth price volatility is predicted using four price volatility lags. The fifth price is denoted by V 5 and the values V 1,..., V 4 denote the input lags. The network topology was obtained after 11 iterations, with an error term of 0.936886. The model also resulted to MSE value of 37.6775 and the RMSE value of 6.1382 which are less than the MSE and RMSE values of the ANN predictive model for kerosene prices. The comparison between the values predicted by the network topology in figure 11 and the actual volatilities of kerosene prices is shown in figure 12.

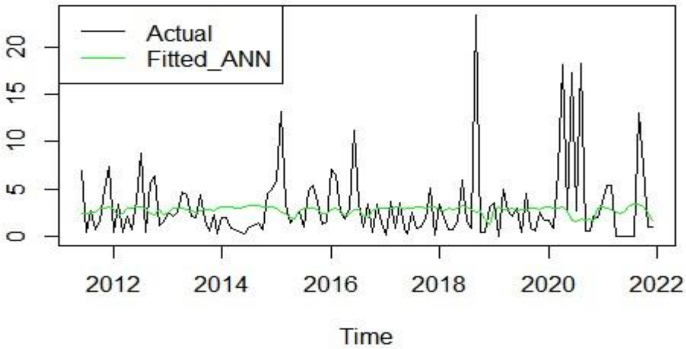

Figure 12: ANN Model Fitting for Volatilities of Kerosene Prices

5 Summary and Recommendations

The best ARIMA predictive model for petrol prices was ARIMA(0,1,1) while the best ANN predictive model had a configuration 6:5; 6 input price lags and 5 hidden layer neurons. The best ARIMA model for predicting diesel prices was ARIMA(1,1,0) and the best predictive ANN model had a configuration 3:3; 3 input price lags and 3 hidden layer neurons. Kerosene prices' best predictive models were ARIMA(0,1,1)(0,0,2)[12] which had seasonal MA(1) term and ANN model with configuration 2:2; 2 input price lags and 2 hidden layer neurons. In every instance, the ANN models outperformed the ARIMA models in terms of MSE and RSME.

ARIMA predictive models for petrol, diesel and kerosene prices had MSE values of 171.24, 93.2862 and 150.1414 respectively. On the other hand, ANN predictive models for petrol, diesel and kerosene prices had MSE values of 51.3980, 45.7290 and 52.2647 respectively. Because of this, ANN models are more accurate predictors of prices of petrol, diesel and kerosene than ARIMA models. The ANN predictive model for volatilities of petrol prices had the configuration 6:5; 6 input price volatilities and 5 hidden layer neurons. The ANN predictive model for volatilities of diesel prices had the configuration 3:3; 3 input price volatilities and 3 hidden neurons. Kerosene price volatilities resulted to an ANN predictive model with configuration 4:4; 4 input layers and 4 hidden layer neurons. These models had MSE value of 18.51108, 23.02922 and 37.6775. These indicate an out-performance of the ANN predictive models for price volatilities of petrol, diesel and kerosene as compared to the ANN predictive models for prices of petrol, diesel and kerosene.

This study used purely lagged values to develop models. Instead of using the lagged values as the input values, this study thus suggests a similar study but use of factors that affect the price of the products such as price of crude oil, supply rate, demand rate and exchange rates and inflation as the input variables. Further, this study considered ANN models for prediction. However, there are other non-parametric predictive models and even hybrid models. Such models should be considered in future research. Moreover, the literature gave some ANN-hybrid models on predicting price volatility. Therefore, another topic of research may be a comparison of ANN predictive models for price volatility and ANN-hybrid predictive models for price volatility.

References

[1] Arienda, Asana and Constantino, *Operations Management Forecasting*, Share and Discover Knowledge on SlideShare, https://www.slideshare.net/Euwinx/operationsmanagement-forecasting, (2015, February 17).

[2] Azadeh, A., Moghaddam, M., Khakzad, M., & Ebrahimipour, V., *A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting*, Computers Industrial Engineering, 62(2), (2012), 421-430.

[3] Bildirici, M., & Ersin, O.,¨ *Forecasting volatility in oil prices with a class of nonlinear volatility models: Smooth transition RBF and MLP neural networks augmented GARCH approach.*, Petroleum Science, 12(3), (2015), 534-552.

[4] Energy and Petroleum Regulatory Commission, *Energy and Petroleum Statistics report*, Complex Analysis and Operator Theory, (2022).

[5] Farjamnia, Naseri and M. Ahmadi., *Oil Price Forecasting; A Comparison between ARIMA and ANN Models*, *https://ijer.atu.ac.ir/article*$_3$629.*html*? *lang*= *en*,(2007).

[6] Gupta, N., & Nigam, S., *crude oil price prediction using artificial neural network*, Procedia Computer Science, 170, (2020).

[7] Kee Wei Yee, S., & Humaida Banu Samsudin, *Comparison between Artificial Neural Network and ARIMA Model in Forecasting Palm Oil Price in Malaysia*, Asian Journal of Management, 5, (2021).

[8] Kristjanpoller, W., Minutolo, M. C., *Forecasting volatility of oil price using an artificial neural networkGARCH model*, expert Systems with Applications, 65, (2016), 233-241.

[9] Mwikamba and Aiyenga., *An Artificial Neural Network Model for Forecasting Inflation in Kenya*, Applied Mathematical Sciences, 2019.

[10] Nyongesa, D. N., & Wagala, A., *Non linear time series modelling of the diesel prices in Kenya*, International Journal of Academic Research in Economics and Management Sciences, 5(4) (2016).

[11] Ndei S., *Short-term Forecasting of Crude Oil Prices in Kenya*, (2006).