
MODELING AND FORECASTING OF MONTHLY GLOBAL PRICE OF BANANAS USING SEASONAL ARIMA AND MULTILAYER PERCEPTRON NEURAL NETWORK

Yeong Nain Chi

University of Maryland Eastern Shore, Princess Anne, MD, U.S.A.
Department of Agriculture, Food and Resource Sciences
e-mail: ychi@umes.edu
ORCID: 0000-0003-2263-5245

Orson Chi

University of Maryland Eastern Shore, Princess Anne, MD, U.S.A.
Cybersecurity Engineering Technology Program
e-mail: ochi@umes.edu
ORCID: 0000-0002-2801-0249

© 2021 Yeong Nain Chi, Orson Chi

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>

Quote as: Chi, Y. N, and Chi, O. (2021). Modeling and forecasting of monthly global price of bananas using seasonal ARIMA and multilayer perceptron neural network. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, 25(3).

DOI: 10.15611/eada.2021.3.02

JEL Classification: C22, C45, C53, Q11

Abstract: The primary purpose of this study was to pursue the analysis of the time series data and to demonstrate the role of time series model in the predicting process using long-term records of the monthly global price of bananas from January 1990 to November 2020. Following the Box-Jenkins methodology, ARIMA(4,1,2)(1,0,1)[12] with the drift model was selected to be the best fit model for the time series, according to the lowest AIC value in this study. Empirically, the results revealed that the MLP neural network model performed better compared to ARIMA(4,1,2)(1,0,1)[12] with the drift model at its smaller MSE value. Hence, the MLP neural network model can provide useful information important in the decision-making process related to the impact of the change of the future global price of bananas. Understanding the past global price of bananas is important for the analyses of current and future changes of global price of bananas. In order to sustain these observations, research programs utilizing the resulting data should be able to improve significantly our understanding and narrow projections of the future global price of bananas.

Keywords: bananas, global price, time series, modeling, forecasting, seasonal ARIMA, multilayer perceptron neural network.

1. Introduction

Like COVID-19, Panama Disease, also known as Tropical Race 4, has affected the global banana industry. Recently, Panama Disease has suddenly accelerated, spreading from Asia to Australia, the Middle East, Africa and Latin America, where the majority of the bananas are shipped from globally. Bananas are an economically important crop for many developing countries. Currently, the disease is present in more than 20 countries, prompting fears of a banana pandemic and shortages of the world's favorite fruit (Altendorf, 2019).

Bananas are the world's most popular fruit with different tastes, sizes and colours, and the fourth most important food crop after wheat, rice, and maize in terms of production, and is the world's favourite fruit in terms of consumption quantity (Ruiz et al., 2017). Bananas are among the most traded fruits in the world. In 2017, 22.7 million tons of bananas, excluding plantains, were traded, representing almost 20% of global production that year. The value of this trade was worth \$11 billion, which is higher than the export value of any other exported fruit (Voora, Larrea, and Bermudez, 2020).

Asia is the largest banana-producing region, while India and China were the two leading banana-producing countries. Asia-Pacific led the market with a 61% share of global consumption, while India was the world's leading producer of banana accounting for nearly 25.7% of the total output. The Philippines consolidated its position as the second largest exporter of bananas behind Ecuador (Voora et al., 2020).

Latin America and the Caribbean is the largest exporting region, responsible for approximately 80% of global exports. In 2018, global banana exports were estimated at 23.3 million tons, with Ecuador being the largest exporter of bananas accounting for 24.7% of global exports. Belgium, Costa Rica and Columbia were the other top banana exporters in the world, whereas the United States was the leading importer of bananas with an 18% share in world imports (Voora et al., 2020).

The retail value of the banana sector was estimated to be worth between \$20 billion and \$25 billion in 2016 (Voora et al., 2020). The Market Reports World (2019) forecasted that the global banana sector would experience a compound annual growth of 1.21% in consumption from 2019 to 2024, reaching a global consumption volume of 136 million tons by 2025, compared to 116.2 million tons in 2017.

According to the U.S. Bureau of Labor Statistics, the average consumer price of bananas (all fresh traditional first quality organic and non-organic yellow bananas) was \$0.53 (Per Lb. in U.S. City Average) with the standard deviation of \$0.06 (Minimum: \$0.40, Maximum: \$0.64, and Median: \$0.51). Average price is to measure the price level in a particular month, not to measure price change over time. Furthermore, the average consumer price for bananas was 39.91% higher in 2020 compared to 1990 (a \$13.29 difference in value). Between 1990 and 2020, bananas costing \$33.30 in 1990 would cost \$46.59 in 2020 for an equivalent purchase.

Compared to the overall inflation rate of 2.30% during this same period, inflation for bananas was lower.

Given this understanding of market demand and supply of bananas globally, this study would test one simple research question, namely what is the trend of the global price of bananas in the near future? Moreover, a research hypothesis was set that there is a consistent increase in the global price of bananas. Hence, the primary purpose of this study was to pursue the analysis of time series data and to demonstrate the role of the time series model in the predicting process using long-term records of the monthly global price of bananas from January 1990 to November 2020.

This paper is organized as follows. The second section briefly reviews the time series applications in the global banana market and the availability of neural networks for time series modeling and forecasting. The third section shows the data source in terms of the monthly global price of bananas. The fourth section presents the methodological approach. The fifth section demonstrates the empirical results using the seasonal ARIMA and multilayer perceptron neural network. The last section provides concluding remarks, and further discussion.

2. Literature review

Time series forecasting is the use of a model to predict future values based on previously observed values, which is one of the most applied data science techniques in business, used extensively in finance, supply chain management, production and inventory planning. It has a well-established theoretical grounding in statistics, and is important because there are so many prediction problems that involve a time component. Several articles used time series techniques to forecast banana production (Eyduran et al., 2020; Hamjah, 2014; Hossain et al., 2016), while quite a few studies focused on banana price forecasting (Fatin, Titik, and Mulyatno, 2020; Omar, Dewan, and Hoq, 2014).

Neural networks have become one of the most popular trends in machine learning for time series modeling and forecasting. Recently, there has been increasing interest in using neural networks to model and forecast banana harvest yields (Rathod, Mishra, and Singh, 2017; Rathod and Mishra, 2018; Rebortera and Fajardo, 2019). Despite the importance of banana demand and supply in the global markets, there is a lack of studies available in the technical literature on global price forecasting schemes.

3. Data

The long-term records of the monthly global price of bananas (units: U.S. dollars per metric ton, not seasonally adjusted) from January 1990 to November 2020, is available to the public from the International Monetary Fund, retrieved from FRED, the Federal Reserve Bank of St. Louis (<https://fred.stlouisfed.org/series/>

PBANSOPUSDM). The average monthly global price of bananas was \$710.59 U.S. dollars per metric ton with the standard deviation of \$272.47 (Minimum: \$250.51, Maximum: \$1,298.34, and Median: \$648.51).

4. Methodology

For time series analysis, the Box-Jenkins methodology, introduced by Box and Jenkins in 1970, refers to a systematic approach of identifying, fitting, checking, and forecasting the ARIMA model (Young, 1977), which can be applied to find the best fit to provide an adequate description to the time series. The Box-Jenkins methodology can be used as the process for estimating the seasonal ARIMA (SARIMA) model in this study based on its autocorrelation function (ACF) and partial autocorrelation function (PACF) as a means of determining the stationarity of the univariate time series and the lag lengths of the time series.

In statistics, stationarity is a property of a stochastic process. Intuitively, stationarity means that the statistical properties (i.e. mean and variance) of the time series do not change over time. Thus, the Box-Jenkins methodology starts with the assumption that the time series should be as on a stationary status. In the Box-Jenkins methodology there are four important steps, including identification, estimation, diagnostic checking and forecasting, that should be taken in consideration when applying it (Box et al., 2016). The identification step is applied to achieve stationarity and to build a suitable model using ACFs, PACFs, and to differencing. If the time series is non-stationary, it needs to be stationarized through differencing.

In the estimation step, plots and summary statistics can be used to identify trends, and autoregressive and/or moving average elements to gain an idea of the amount of differencing and the size of the lag that will be required for model identification. The next estimation step is to use a fitting procedure to find the coefficients of the model. In order to find good parameters for the model, Akaike's Information Criterion (AIC) or Bayesian Information Criterion (BIC) can be employed to determine the orders of the time series. Good models are obtained by minimizing the AIC or BIC value.

The diagnostic checking step is primarily to use the plots and statistical tests of the residual errors to determine the model fitting, and to evaluate the fit model in the context of the available data and check for areas where the model may be improved. The process is repeated until a desirable level of fit is achieved. There are many accuracy metrics applied after model identification and estimation helping in choosing the best fit model. The commonly used accuracy metrics to judge forecasts include: Mean Square Error (MSE) = $(1/n) \sum (Y_t - \hat{Y}_t)^2$, Root Mean Square Error (RMSE) = $\sqrt{\text{MSE}}$, Mean Absolute Percentage Error (MAPE) = $(1/n) \sum (|Y_t - \hat{Y}_t| / |Y_t|) \times 100$, and Mean Absolute Error (MAE) = $(1/n) \sum |Y_t - \hat{Y}_t|$.

The last step is the forecasting stage that can be applied in the forecasting process after checking the model in the previous steps. In this study, R 4.0.2 for Windows,

an open source for statistical computing and graphics supported by the R Foundation for Statistical Computing, was used as the tool to estimate the model parameters to fit the SARIMA model and the multilayer perceptron (MLP) neural network model, as well to achieve the purpose of this study.

5. Empirical results

5.1. Identification of the SARIMA model

R 4.0.2 for Windows, an open source for statistical computing and graphics supported by the R Foundation for Statistical Computing, was used as the tool to model and forecast the monthly global price of bananas from January 1990 to November 2020 for this study (Figure 1).

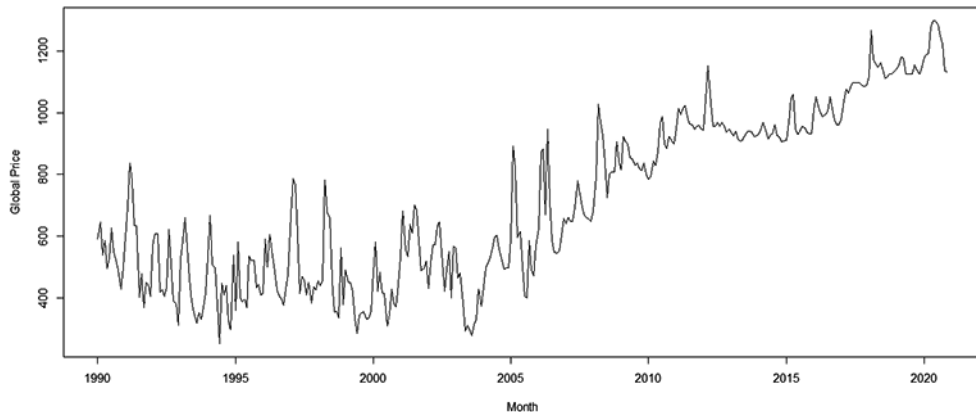


Fig. 1. Time Series Plot of Monthly Global Prices of Bananas, January 1990–November 2020

Source: own work.

The function “`decompose()`” in R can be applied to estimate the seasonal, trend and irregular components of a seasonal time series. The plots in Figure 2 showed the original time series (top), the estimated trend component (second from top), the estimated seasonal component (third from top), and the estimated irregular component (bottom). The estimated trend component showed a steady increase over time, and the estimated seasonal component definitely displayed seasonality, with a pattern recurrence occurring once every 12 months (yearly).

Decomposing the time series means separating the time series into a trend component, a seasonal component and an irregular component. Thus, one of the main objectives for the decomposition is to estimate the seasonal effects that can be used to present seasonally adjusted values. Seasonal adjustment is the estimation

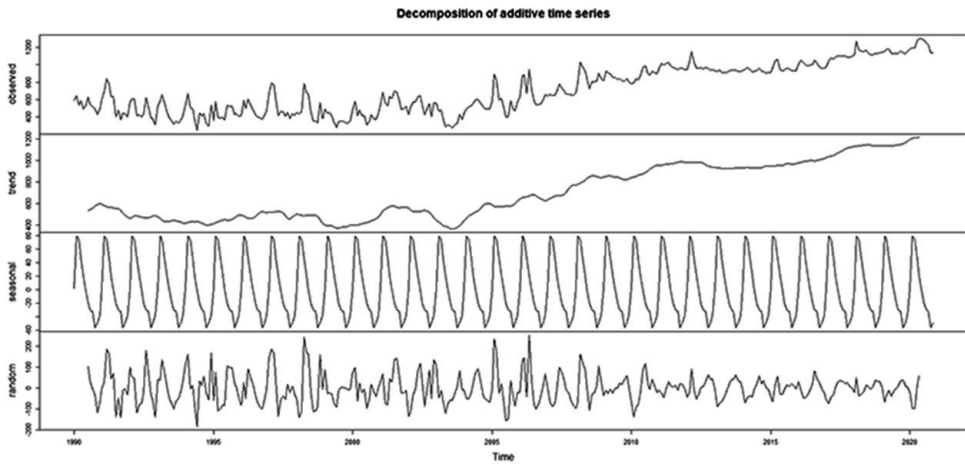


Fig. 2. Decomposition of Monthly Global Prices of Bananas, January 1990-November 2020

Source: own work.

and removal of seasonal effects that are not explainable by the dynamics of trends or cycles from a time series to reveal certain non-seasonal features. This can be done by subtracting the estimated seasonal component from the original time series. Having removed the seasonal variation, the seasonally adjusted time series only contained the trend component and an irregular component (Figure 3).

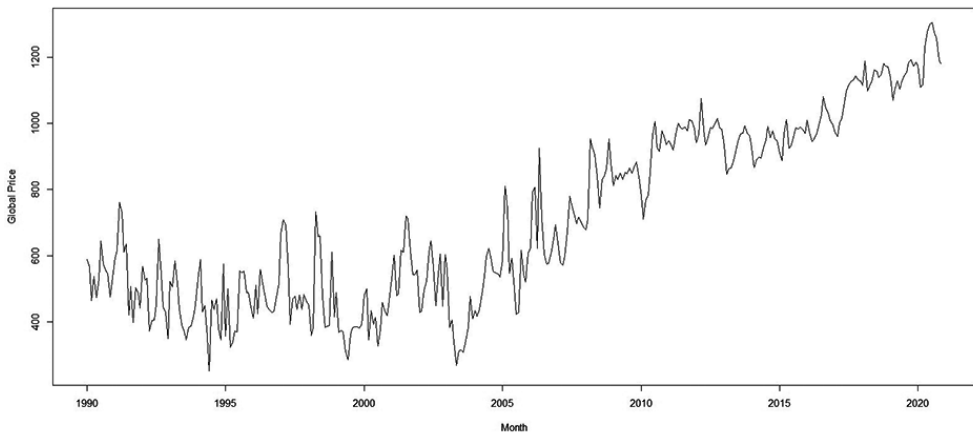


Fig. 3. Time Series Plot of Seasonal Adjusted Monthly Global Price of Bananas, January 1990-November 2020

Source: own work.

Looking at the time series plot of the data, the ACF is useful for identifying stationary or non-stationary time series. For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly. Hence, the ACF (Figure 4) of the time series, the seasonally adjusted monthly global price of bananas from January 1990 to November 2020, showed strong positive correlations at up to 26 lags that never decay to zero. Meanwhile, the test statistic of the Augmented Dickey-Fuller Test was Dickey-Fuller = -3.0039 with lag order = 7, and the p-value of the test was 0.1532. This suggested that the time series was non-stationary and need to be differenced. This was also illustrated by the single spike at the first lag, followed by small apparently random values after the first lag for the PACF (Figure 5). Typically, the ACF and PACF shown in Figures 4 and 5 were a time series that has autocorrelation at the first lag only. Since the PACF was cut off after the first lag, it seems that the time series followed the autoregressive (AR) process.

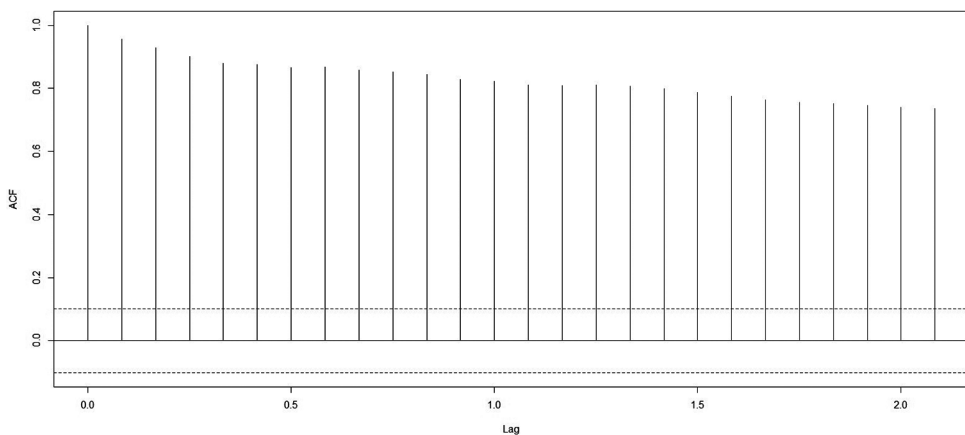


Fig. 4. ACF Plot of Seasonal Adjusted Monthly Global Price of Bananas, January 1990-November 2020

Source: own work.

In the time series analysis, differencing can be used to transform a non-stationary time series into a stationary one. When both trend and seasonality are present, one may need to apply both a non-seasonal first difference and a seasonal difference. The first difference of a time series is the time series of changes from one period to the next. Note that the graph of the first difference of the time series looked approximately stationary (Figure 6). According to the Augmented Dickey-Fuller Test, Dickey-Fuller = -10.218 with lag order = 7, and the p-value of the test was smaller than 0.01. It rejected the null hypothesis that is non-stationary, and suggested that the first difference of the time series was stationary.

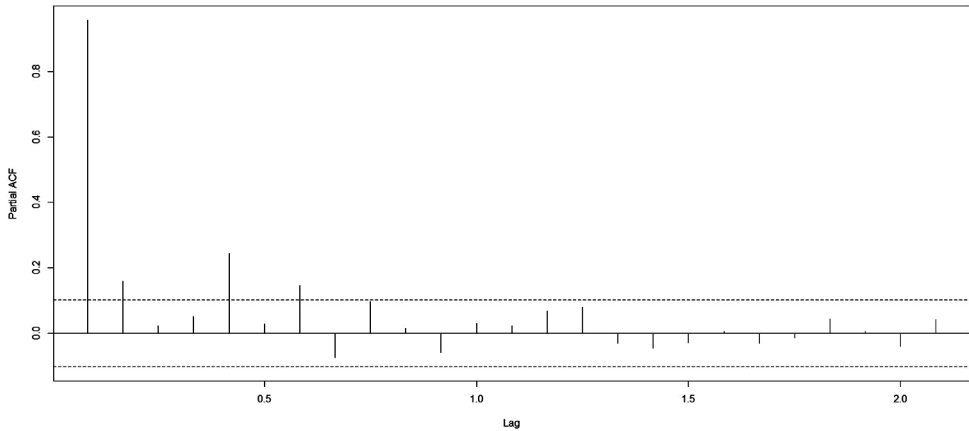


Fig. 5. PACF Plot of Seasonal Adjusted Monthly Global Price of Bananas, January 1990-November 2020
Source: own work.

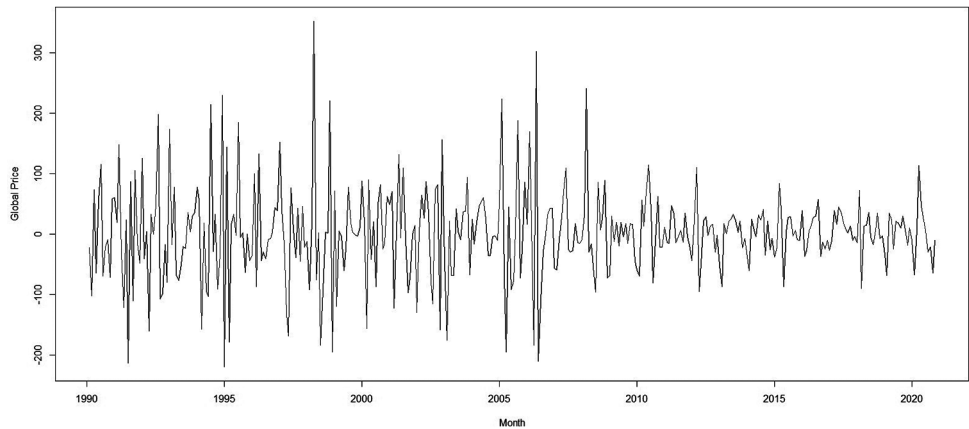


Fig. 6. Time Series Plot of First Difference of Seasonal Adjusted Monthly Global Price of Bananas, January 1990-November 2020
Source: own work.

The ACF of the first differenced time series in Figure 7 showed a significant positive spike at the first lag followed by correlations that were statistically significant. The corresponding PACF of the first differenced time series in Figure 8 presented most likely a gradual decrease, and negative after the first few lags. Since the ACF was cut off after the first lag and the PACF decreased gradually, a reasonable conclusion was that the first differenced time series followed the moving average (MA) process.

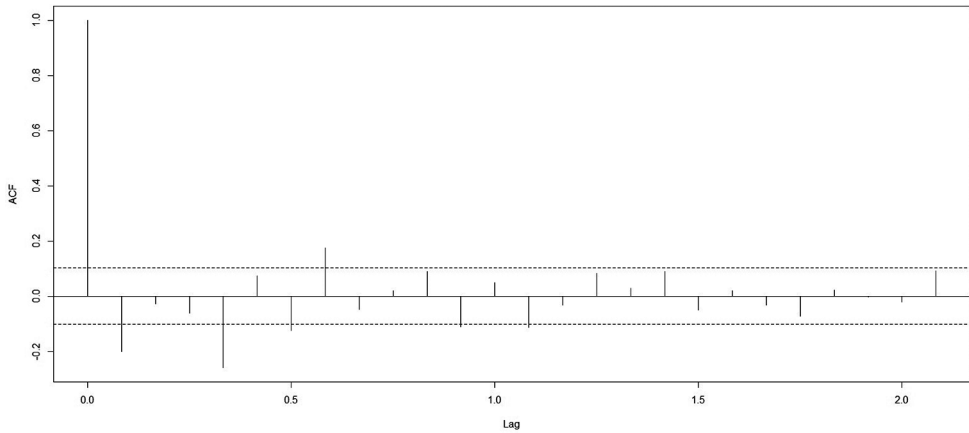


Fig. 7. ACF Plot of First Difference of Seasonal Adjusted Monthly Global Price of Bananas, January 1990-November 2020

Source: own work.

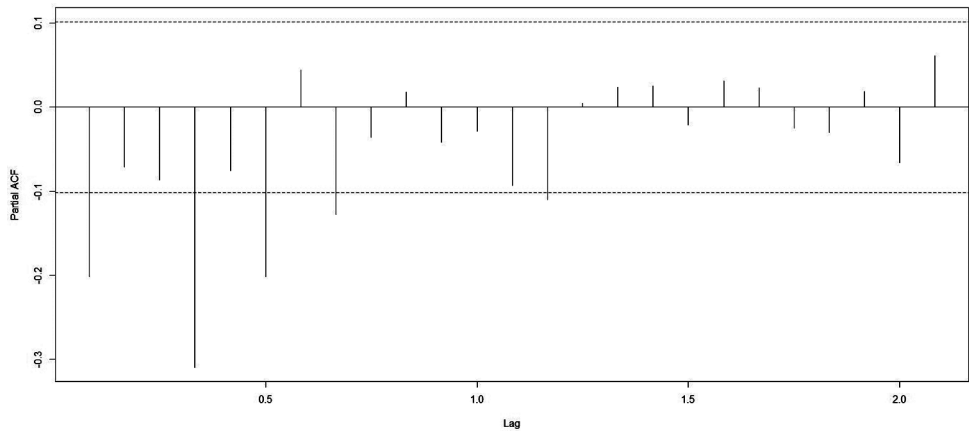


Fig. 8. PACF Plot of First Difference of Seasonal Adjusted Monthly Global Price of Bananas, January 1990-November 2020

Source: own work.

Seasonal differencing can remove seasonal trends and can also remove a seasonal random walk type of non-stationarity. Seasonal differencing is defined as the difference between a value and a value with a lag that is a multiple of seasonality. In this case, seasonality = 12 months per year is the span of the periodic seasonal behaviour. Figure 9 shows the graph of the 12th difference of the time series, which looked approximately stationary. Meanwhile, the test statistic of the Augmented Dickey-Fuller Test was Dickey-Fuller = -32.05 with lag order = 7, and the p-value

of the test was smaller than 0.01. It rejected the null hypothesis that is non-stationary, and suggested that the 12th first difference of the time series was stationary.

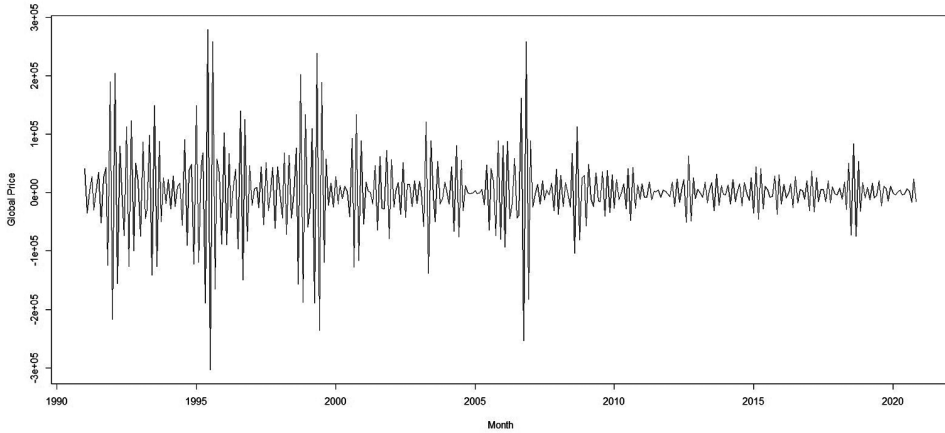


Fig. 9. Time Series Plot of 12th Difference of Seasonal Adjusted Monthly Global Price of Bananas, January 1990–November 2020

Source: own work.

Figure 10 showed that ACF demonstrated most likely a steady decay after the first few lags and bounced between being positive and negative statistically significant. Meanwhile, Figure 11 showed what PACF usually looks like a steady negative decay in the partial correlations toward zero. This is consistent with the general recommendation that if the autocorrelation at the first seasonal lag is positive, one should use an autoregressive (AR) process vs. a moving average (MA) process.

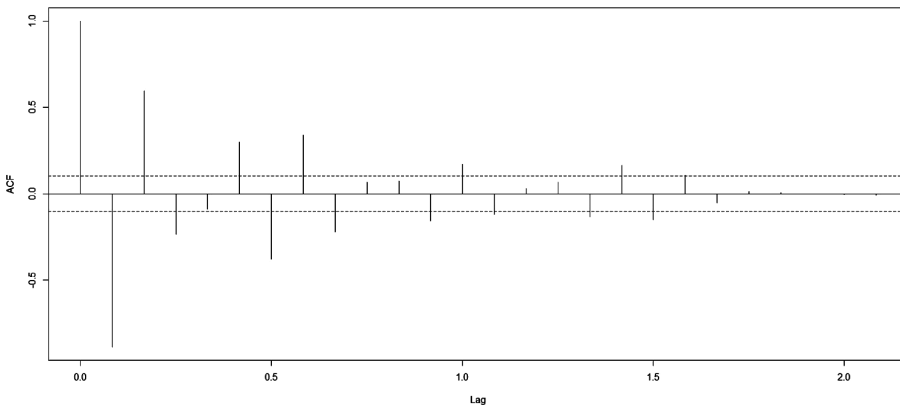


Fig. 10. ACF Plot of 12th Difference of Seasonal Adjusted Monthly Global Price of Banana, January 1990–November 2020

Source: own work.

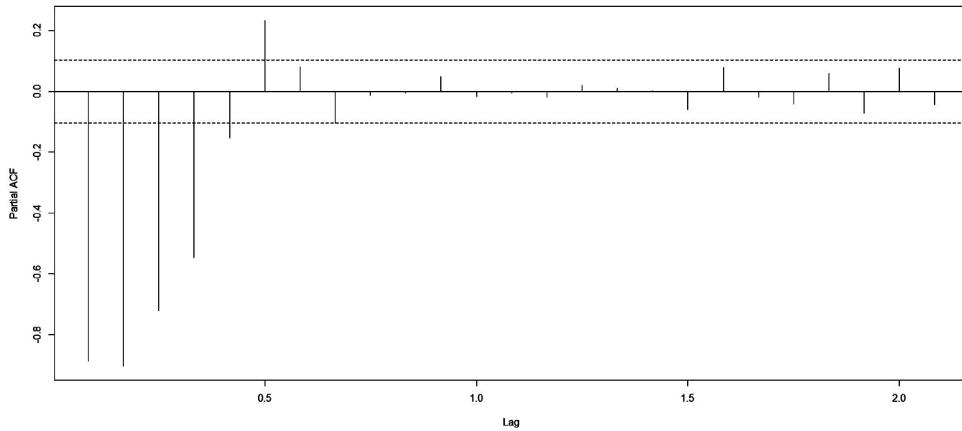


Fig. 11. PACF Plot of 12th Difference of Seasonal Adjusted Monthly Global Price of Bananas, January 1990-November 2020

Source: own work.

5.2. Estimation of the SARIMA model parameters

SARIMA is an extension of Autoregressive Integrated Moving Average (ARIMA) that explicitly supports the univariate time series with a seasonal component. One shorthand notation for the SARIMA model is ARIMA(p,d,q)(P,D,Q)[m], where p = the order of the autoregressive process (the number of lagged terms), d = the number of differences required to make the time series stationary, q = the order of the moving average process (the number of lagged terms), P = the order of the seasonal autoregressive process, D = the number of seasonal differences applied to the time series, Q = the order of the seasonal moving average process, and m = the seasonality of the model, i.e. the number of time steps for a single seasonal period.

In backshift notation B, “backshift operator” or “lag operator”, is a useful notational device when working with time series lags: $By_t = y_{t-1}$ (means “back up by one time unit”) and $B^ky_t = y_{t-k}$ (means “backshift k times”). Thus, the ARIMA(p,d,q)(P,D,Q)[m] model can be expressed in backshift notation as:

$$\phi_p(B)\Phi_p(B^m)(1-B)^d(1-B^m)^Dy_t = c + \theta_q(B)\Theta_Q(B^m)e_t, \tag{1}$$

where: $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p) = (1 - \sum_{i=1}^p \phi_i B^i)$, $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q) = (1 - \sum_{i=1}^q \theta_i B^i)$, $\Phi_p(B^m) = (1 - \Phi_1 B^m - \dots - \Phi_p B^{mp}) = (1 - \sum_{i=1}^p \Phi_i B^{mi})$, $\Theta_Q(B^m) = (1 - \Theta_1 B^m - \dots - \Theta_Q B^{mQ}) = (1 - \sum_{i=1}^Q \Theta_i B^i)$, c is a constant, and e_t is the residual error (i.e. white noise) (Montgomery, Jennings, and Kulahci, 2008).

Empirically, the choice of the model order is somewhat arbitrary. In order to find a solution, the function “auto.arima()” from the “forecast” package in R 4.0.2 for

Windows was employed to identify both the structure of the time series (stationary or not) and type (seasonal or not), and sets the model’s parameters, that takes into account the AIC, AICc or BIC values generated to determine the best fit SARIMA model.

Table 1. Comparison of selected SARIMA models

Model	AIC	BIC
ARIMA(1,1,1)(1,0,1)[12]	4207.03	4226.59
ARIMA(1,1,2)(1,0,1)[12]	4209.02	4232.50
ARIMA(2,1,1)(1,0,1)[12]	4209.02	4232.62
ARIMA(2,1,2)(1,0,1)[12]	4211.17	4238.57
ARIMA(3,1,1)(1,0,1)[12]	4205.03	4232.43
ARIMA(3,1,2)(1,0,1)[12]	4203.31	4234.62
ARIMA(4,0,1)(1,0,1)[12]	4246.97	4282.21
ARIMA(4,0,2)(1,0,1)[12]	4257.49	4296.65
ARIMA(4,1,1)(1,0,1)[12]	4195.21	4226.52
ARIMA(4,1,2)(1,0,1)[12]	4187.14	4222.36
ARIMA(4,1,2)(1,0,1)[12] with Drift	4186.52	4225.66

Source: own work.

Consequently, the ARIMA(4,1,2)(1,0,1)[12] with the drift model was selected to be the best fit model for the time series, according to the lowest AIC value (= 4186.52) in this study (Table 1). Given this option, the ARIMA(4,1,2)(1,0,1)[12] with the drift model was chosen for further forecasting process, and the parameters of the ARIMA(4,1,2)(1,0,1)[12] with the drift model were presented in Table 2. The ARIMA(4,1,2)(1,0,1)[12] with the drift model yielded the following forecasting equation:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1 - B)(1 - \Phi_1 B^{12})\hat{Y}_t = c + (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^{12}). \tag{2}$$

The ARIMA(4,1,2)(1,0,1)[12] with the drift model for the time series, seasonal adjusted monthly global price of bananas from January 1990 to November 2020, can be expressed as follows:

$$(1 + 0.2506B - 0.3871B^2 + 0.0506B^3 + 0.2801B^4)(1 - B)(1 + 0.3492B^{12})\hat{Y}_t = 1.8360 + (1 + 0.0376B + 0.6005B^2)(1 - 0.3620B^{12}). \tag{3}$$

Table 2. Parameters of the ARIMA(4,1,2)(1,0,1)[12] with the Drift Model

Parameter	Estimate	Standard Error
Constant	1.8360	1.0835
AR Lag 1	-0.2506	0.1005
AR Lag 2	0.3871	0.0964
AR Lag 3	-0.0506	0.0631
AR Lag 4	-0.2801	0.0567
Difference		
MA Lag 1	-0.0376	0.0994
MA Lag 2	-0.6005	0.1007
SAR 1	-0.3492	2.1797
SMA 1	0.3620	2.2000
Sigma ² estimated as 4653, Log Likelihood = -2083.26		
AIC = 4186.52, AICc = 4187.14, BIC = 4225.66		
Training Set Error Measure: RMSE = 67.28535, MSE = 4527.31832, MAE = 48.03735, MAPE = 8.644118		

Source: own work.

5.3. Diagnostic checking of the SARIMA model

A common task when building a forecasting model is to check that the residuals satisfy some assumptions that they are uncorrelated, normally distributed, etc. The top figure of Figure 12 showed that the residuals from the ARIMA(4,1,2)(1,0,1)[12] with the drift model did not violate the assumption of constant location and scale. The bottom right figure of Figure 12 showed that the residual histogram did not reveal a series deviation from normality in this case. The ACF plot of the residuals (the bottom left figure of Figure 12) also showed that all sample autocorrelations were within the threshold limits, indicating that the residuals appeared to be random.

The Ljung-Box Q-test (Ljung and Box, 1978) is a diagnostic tool used to test the lack of fit of a time series model. In this case, the test statistic of the Ljung-Box Q-test was $Q = 14.768$ with 15 degrees of freedom, and the p-value of the test was 0.4682 with model degrees of freedom: 9 and total lags used: 24, indicating that the residuals were random and that the model provided an adequate fit to the time series. This, combined with the Ljung-Box Q-test statistic, suggested that the ARIMA(4,1,2)(1,0,1)[12] with the drift model appropriately modelled the dynamics for this time series.

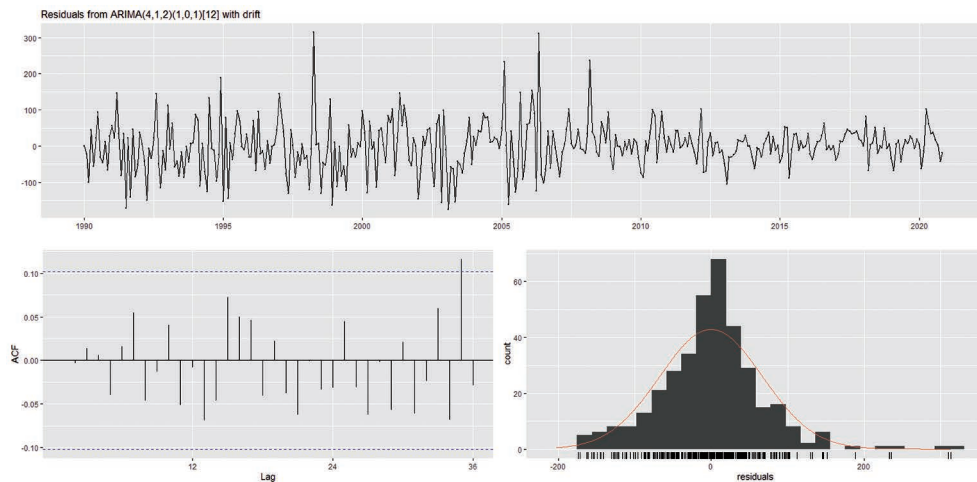


Fig. 12. Residuals of ACF and PACF

Source: own work.

5.4. Using the SARIMA model in the forecasting process

The forecasted values of monthly global price of bananas are shown in Table 3. Forecasts from ARIMA(4,1,2)(1,0,1)[12] with the drift model are shown in Figure 13, presenting a linear upward trend projected to the future. The next figure (Figure 14) illustrated that the black line represented the visuals of the monthly global price of bananas dataset without forecasting and the red line represented the visuals of the monthly global price of bananas dataset with forecasted values. The forecasting process with ARIMA(4,1,2)(1,0,1)[12] with the drift model indicated a good fit of the model for forecasting at a constant increasing rate in the short term.

Table 3. Forecasted Values of monthly global price of bananas U.S. dollars per metric ton

Date	Point forecast	95% Lower confidence limit	95% Upper confidence limit
1	2	3	4
December 2020	1197.120	1063.429	1330.811
January 2021	1212.145	1048.047	1376.244
February 2021	1234.762	1053.799	1415.726
March 2021	1238.786	1048.342	1429.231
April 2021	1244.975	1053.400	1436.551
May 2021	1242.932	1048.071	1437.794
June 2021	1241.234	1045.005	1437.464

1	2	3	4
July 2021	1241.369	1041.357	1441.381
August 2021	1241.096	1037.121	1445.072
September 2021	1243.752	1035.459	1452.045
October 2021	1244.949	1031.879	1458.019
November 2021	1247.691	1030.786	1464.596
December 2021	1249.882	1028.571	1471.194
January 2022	1251.849	1026.836	1476.862
February 2022	1254.101	1025.347	1482.856
March 2022	1255.762	1023.394	1488.129
April 2022	1257.184	1021.337	1493.030
May 2022	1258.610	1019.189	1498.032
June 2022	1260.371	1017.516	1503.226
July 2022	1262.226	1015.886	1508.565
August 2022	1264.215	1014.484	1513.946
September 2022	1266.249	1013.157	1519.340
October 2022	1268.357	1011.951	1524.762
November 2022	1270.242	1010.583	1529.901

Source: own work.

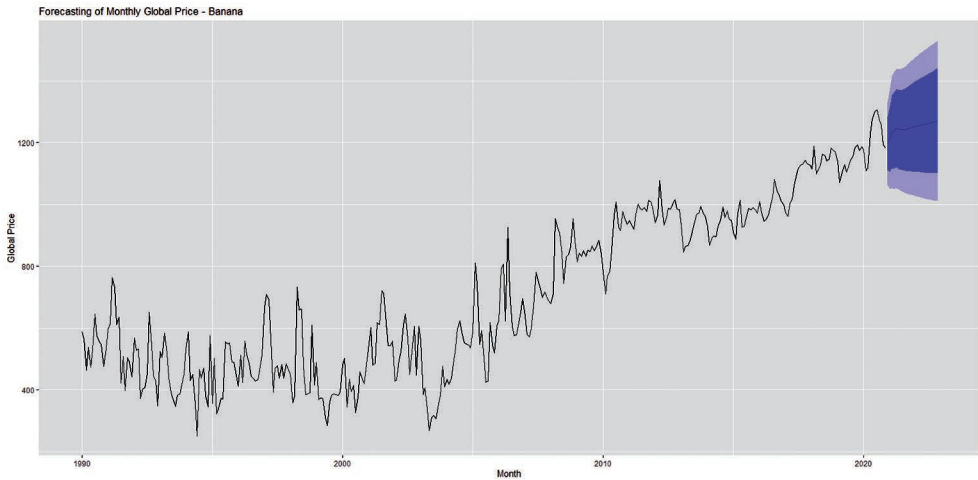


Fig. 13. Observed and forecasted monthly global price of bananas

Source: own work.

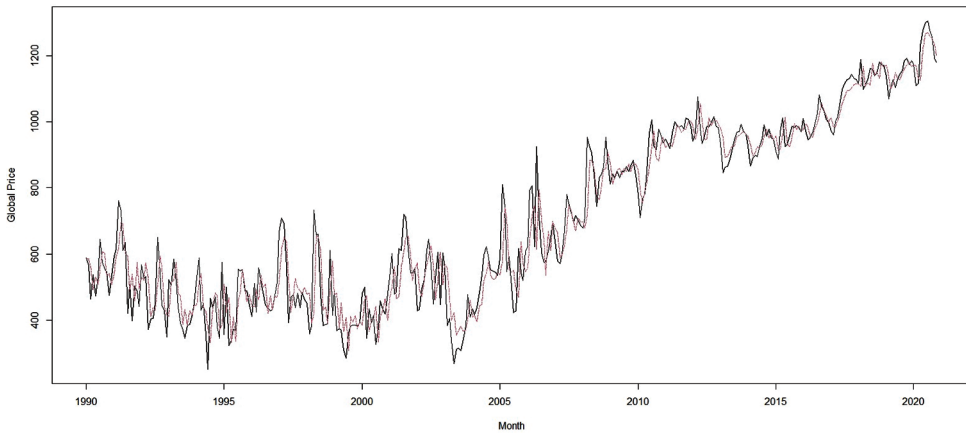


Fig. 14. Observed and forecasted monthly global price of bananas

Source: own work.

5.5. Multilayer Perceptron (MLP) neural network model

The Multilayer Perceptron (MLP) neural network is one of the most popular machine learning methods, which is able to carry out prediction tasks in a more reliable manner. MLP is a system of simple interconnected neurons, or nodes, which is a model representing a nonlinear mapping between an input vector and an output vector (Gardner and Dorling, 1998). The MLP neural network consists of three layers of nodes: an input layer, a hidden layer and an output layer (Figure 15). Except for the input nodes, each node is a neuron that uses a nonlinear activation function.

The MLP neural network is trained for multi-step-ahead prediction, on testing as well as on training data sets for short-term prediction. Thus, the optimum neural network model is proposed for the short-term prediction of the time series. The function “mlp()” from the “nnfor” package in R 4.0.2 for Windows automatically generates ensembles of networks, the training of which starts with different random initial weights.

Training process was categorized into two main steps: the first is to select the best architecture of networks, and the second is to integrate MLP neural networks with optimizers. The output indicated that the resulting MLP neural network model (Figure 15) has 5 hidden nodes, it was trained 20 times and the different forecasts were combined using the median operator. The results were compared with reference to the MSE (mean square error) = 992.9927 against ARIMA(4,1,2)(1,0,1)[12] with the drift model, MSE = 4527.31832.

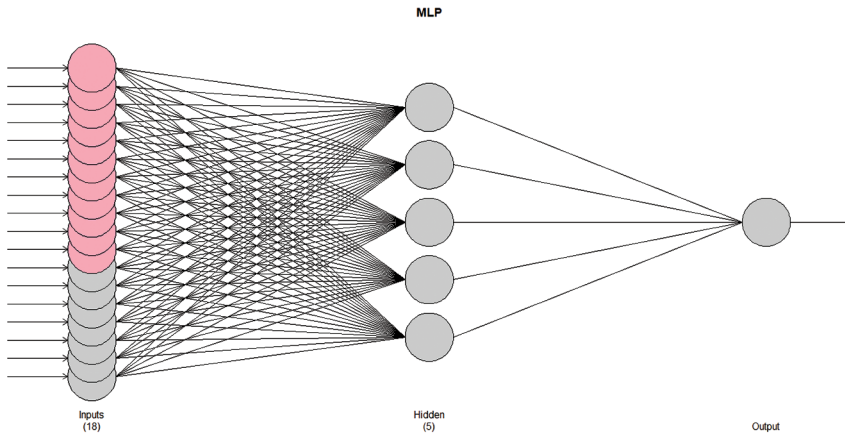


Fig. 15. Single Hidden Layer MLP Neural Network

Source: own work.

The top figure in Figure 16 showed that the residuals from the MLP neural network model did not violate the assumption of constant location and scale. The bottom right figure of Figure 16 showed that the residual histogram from the MLP neural network model did not reveal a series deviation from normality. At the same time, the ACF plot of the residuals (the bottom left figure of Figure 16) from the MLP neural network model showed that all sample autocorrelations were within the threshold limits, indicating that the residuals appeared to be random.

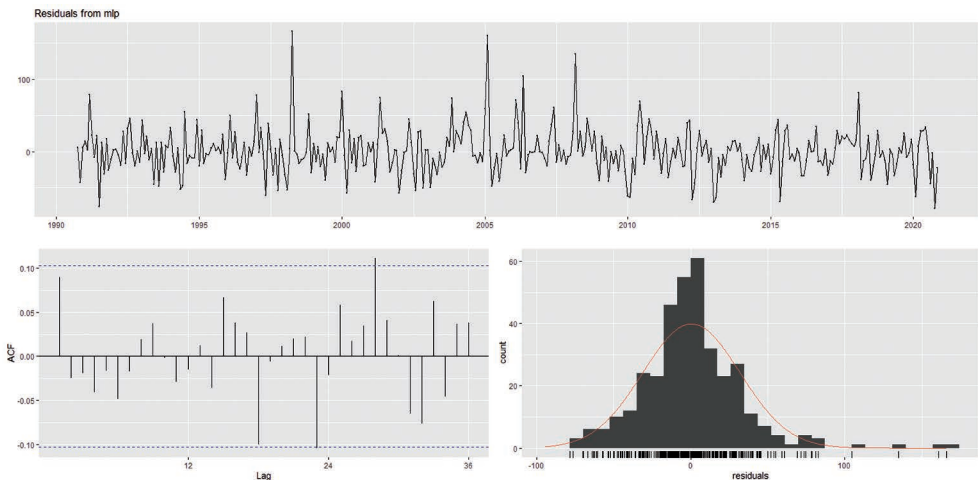


Fig. 16. Residuals of ACF and PACF

Source: own work.

Table 4. Multilayer perceptron neural network forecasted values of monthly global price of bananas U.S. dollars per metric ton

Month	2020	2021	2022	2023	2024	2025
January		1285.627	1339.981	1258.820	1278.070	1308.227
February		1451.027	1383.892	1309.732	1415.763	1410.817
March		1385.847	1367.969	1362.820	1339.679	1450.346
April		1377.963	1352.111	1265.435	1372.017	1427.567
May		1320.321	1322.902	1282.006	1341.870	1379.059
June		1281.345	1304.804	1313.639	1361.867	1372.437
July		1238.360	1312.189	1296.162	1338.052	1361.676
August		1216.296	1296.387	1250.080	1283.677	1361.079
September		1257.593	1302.307	1237.590	1263.882	1325.969
October		1235.542	1233.661	1219.605	1208.279	1254.887
November		1257.689	1302.192	1170.130	1250.265	1299.637
December	1199.071	1274.941	1300.080	1252.743	1324.093	

Source: own work.

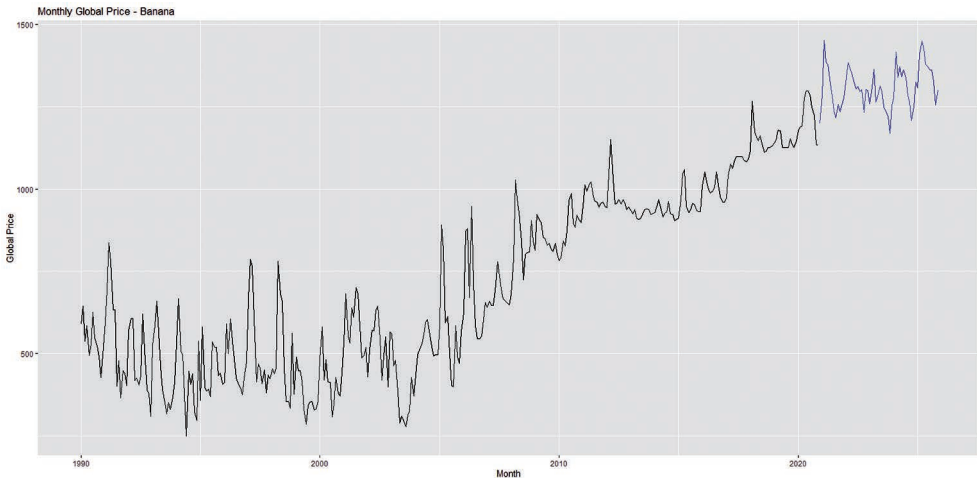


Fig. 17. Observed and forecasted monthly global price of bananas

Source: own work.

The MLP neural network forecasting the values of the monthly global price of bananas was presented in Table 4. Forecasts from the MLP neural network model are also shown in Figure 17, with a linear upward trend projected into the future. The forecasting process with the MLP neural network model indicated a good fit of the model for forecasting at a constant increasing pattern in the short term, seasonally.

6. Discussion and conclusion

Bananas rank as a leading crop in world agricultural production and trade. Assuming normal weather conditions and no further spread of banana plant diseases, bananas have seen rapidly increasing production and trade volumes in response to fast population growth in producing countries, as well as an expanding global import demand. Obviously, the future is more intensive regarding knowledge, and more understanding of the natural complexities of living systems. To bring together a wide variety of perspectives and concepts requires holistic solutions that involve working across disciplines, principles and methods to support interdisciplinarity and transdisciplinarity, to explore and formalize systems concepts, and to develop systemic methods for learning and change.

Despite the importance of banana demand and supply in the global markets, there is a lack of studies available in the technical literature on global price forecasting schemes. In this study, the best choice time series model was ARIMA(4,1,2)(1,0,1)[12] with the drift model as its lowest AIC values among other models. It was noted that this ARIMA(4,1,2)(1,0,1)[12] with the drift model provided evidence that the future monthly global price of bananas would increase over time.

Prediction is a kind of dynamic filtering, in which the past values of one or more time series are used to predict future values. The MLP neural network is one of the most popular machine learning methods, which is able to carry out prediction tasks in a more reliable manner. With time series data, lagged values of the time series can be used as inputs to a neural network, the MLP neural network was applied to time series prediction using its past values of a univariate time series in this study.

Empirically, the results revealed that the MLP neural network model performed better compared to ARIMA(4,1,2)(1,0,1)[12] with the drift model at its smaller MSE value. Hence, the MLP neural network model can provide useful information important in the decision-making process related to the impact of the changes of the future global price of bananas. Understanding the past global price of bananas is important for the analyses of the changes of the current and future global price of bananas changes.

In order to sustain these observations, research programs utilizing the resulting data should be able to improve significantly our understanding and narrow projections of the future global price of bananas. Therefore, checking the validity of the market price of bananas in the case of certain countries could be one of the main limitations of this study. Furthermore, the nonlinear autoregressive exogenous (NARX) neural network can also be used to understand not only past information of the same time series (the monthly global price of bananas), but also current and past information of the externally determined time series that influences the time series of interest (i.e. the production of bananas, climate change considerations, consumer behaviour expectation, etc.).

References

- Altendorf, S. (2019). Banana Fusarium Wilt Tropical Race 4: A mounting threat to global banana markets? In 2019 Food Outlook - Biannual Report on Global Food Markets (pp. 13-20). Rome: FAO. Retrieved from http://www.fao.org/3/ca6911en/CA6911EN_TR4EN.pdf
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). Hoboken, N.J.: John Wiley and Sons Inc.
- Eyduran, S. P., Akın, M., Eyduran, E., Çelik, S., Ertürk, Y. E., and Ercişli, S. (2020). Forecasting banana harvest area and production in Turkey using time series analysis. *Erwerbs-Obstbau*, 62, 281-291.
- Fatin, Z. N., Titik, E., and Mulyatno, S. B. (2020). The analysis of price and market integration of banana commodities in Lampung, Indonesia. *Russian Journal of Agricultural and Socio-Economic Sciences*, 3(99), 61-68.
- Gardner, M.W., and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) – A review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14), 2627-2636.
- Hamjah, M. A. (2014). Forecasting major fruit crops productions in Bangladesh using the Box-Jenkins ARIMA Model. *Journal of Economics and Sustainable Development*, 5(7), 96-107.
- Hossain, M. M., Abdulla, F., and Majumder, A. K. (2016). Forecasting of banana production in Bangladesh. *American Journal of Agricultural and Biological Sciences*, 11(2), 93-99.
- Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.
- Market Reports World. (2019). *Banana market size, share, analysis – segmented by geography – growth, trends, and forecast (2019-2024)* (Global Banana Market Research Report, Market Reports World). Retrieved from <https://www.marketreportsworld.com/banana-market-13487750>
- Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. Hoboken, N.J.: John Wiley & Sons. Inc.
- Omar, M. I., Dewan, M. F., and Hoq, M. S. (2014). Analysis of price forecasting and spatial co-integration of bananas in Bangladesh. *European Journal of Business and Management*, 6(7), 244-255.
- Rathod, S., and Mishra, G. C. (2018). Statistical models for forecasting mango and banana yield of Karnataka, India. *Journal of Agricultural Science and Technology*, 20, 803-816.
- Rathod, S., Mishra, G. C., and Singh, K. N. (2017). Hybrid time series models for forecasting banana production in Karnataka State, India. *Journal of the Indian Society of Agricultural Statistics*, 71(3), 193-200.
- Rebortera, M. A., and Fajardo, A. C. (2019). An enhanced deep learning approach in forecasting banana harvest yields. *International Journal of Advanced Computer Science and Applications*, 10(9), 275-280.
- Ruiz, A., Fobelets, V., Grosscurt, C., Galgani, P., Lord, R., Hardwicke, R., Tarin, M. P. G., McNeil, D., and Aird, S. (2017). *The external costs of banana production: A global study* (Research Report Prepared for Fairtrade International, True Price & Trucost). Retrieved from http://makefruitfair.org/wp-content/uploads/2017/07/170224_Research_Report_External_Cost_of_Bananas_-_final.pdf
- Voorla, V., Larrea, C., and Bermudez, S. (2020). Global market report: Bananas. Sustainable Commodities Marketplace Series 2019, The International Institute for Sustainable Development. Retrieved from <https://www.iisd.org/system/files/publications/ssi-global-market-report-banana.pdf>
- Young, W. L. (1977). The Box-Jenkins approach to time series analysis and forecasting: principles and applications. *RAIRO. Recherche opérationnelle*, 11(2), 129-143. Retrieved from http://www.numdam.org/article/RO_1977__11_2_129_0.pdf

MODELOWANIE I PROGNOZOWANIE MIESIĘCZNEJ GLOBALNEJ CENY BANANÓW Z WYKORZYSTANIEM SEZONOWEJ ARIMA I WIELOWARSTWOWEJ SIECI NEURONOWEJ PERCEPTRONOWEJ

Streszczenie: Podstawowym celem tego badania była analiza danych szeregów czasowych oraz wskazanie ważności modelu szeregów czasowych w procesie predykcji z wykorzystaniem długoterminowych zapisów miesięcznej ceny bananów na świecie od stycznia 1990 r. do listopada 2020 r. Zgodnie z metodologią Boxa-Jenkinsa wybrano jako najlepiej dopasowany dla szeregu czasowego model ARIMA(4,1,2)(1,0,1)[12] z dryfem, zgodnie z najniższą wartością AIC. Na podstawie wyników empirycznych stwierdzono, że model sieci neuronowej MLP działał lepiej w porównaniu z modelem ARIMA(4,1,2)(1,0,1)[12] z dryfem z mniejszą wartością MSE. Wynika z tego, że model sieci neuronowej MLP może dostarczyć użytecznych informacji, które są ważne w procesie decyzyjnym dotyczącym wpływu zmian przyszłej globalnej ceny bananów. Postrzeganie przeszłych światowych cen bananów jest ważne dla analiz zarówno bieżących, jak i przyszłych zmian światowych cen. Aby podtrzymać te obserwacje, programy badawcze wykorzystujące uzyskane dane powinny umożliwiać znaczne poprawianie wnioskowania i zawężać prognozy przyszłych światowych cen bananów.

Słowa kluczowe: banany, cena globalna, szeregi czasowe, modelowanie, prognozowanie, sezonowy model ARIMA, wielowarstwowa sieć neuronowa perceptronowa.