Journal’s Title, vol xxx, 200xxx, no. xx, xxx-xxx

Original Research Article

Time Series Modeling of Monetary Value from Kenya’s Horticultural Export produce.

Abstract: Kenya’s horticulture sector is one of the key contributors to the country’s national income. In this paper, we apply Box-Jenkins SARIMA time series modeling approach to develop a time series model that best describes the income to Kenya’s economy from the export of horticulture produce. In the process of analysis, we considered monthly data from August 1998 to March 2023. It was found that, $SARIMA(3,1,4)(0,1,0)_{12}$ is the suitable model that describes the income from the export of Kenya’s horticulture produce.

Keywords: Horticulture, Time Series, National Income, Forecasting.

1 Introduction

Kenya’s horticulture sub-sector is among the agricultural sub-sectors that contribute significantly to the country’s national income. It is also one of the fastest growing sub-sectors in the agriculture sector despite the challenges facing it such as high cost of airfreight, reduced demand during Covid-19 pandemic due to lock downs especially in Europe, high cost of doing business due to government levies and taxes, [10]. The sub-sector employs about 350,000 directly and supports livelihoods of over six million. The income from horticultural export produce encompasses income from export of cut flowers, vegetables and fruits.

Knowledge of analysing time series data has been applied in a wide range of areas such as tourism, climate, GDP, crop yields, among others. The analysis involves the use of models such as autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), vector autoregressive (VAR) and vector autoregressive moving average(VARMA) models. Others which have been used lately include Fuzzy time series (FTS) models as seen in [7][8]. In this work, seasonal autoregressive integrated moving average(SARIMA) modeling approach which caters for the seasonality component of a time series when forecasting is adopted. This is because SARIMA models usually perform well when dealing with seasonal data. Data on monthly income to Kenya’s economy from horticultural export produce from August 1998 to March 2023 is analysed to come up with the model that can be used to forecast the income from horticulture. This is motivated by the fact that Kenya’s horticulture sector has been contributing significantly to the economy over the recent years.

The rest of the paper is organized as follows: first, literature review is discussed as given in section 2, materials and methods used in the study are given in section 3, results and discussion given in section 4 and then the conclusions are then given in section 5.

2 Literature Review

Box and Jenkins introduced the autoregressive integrated moving average (ARIMA) models in 1960, [1]. Through their innovation, time series ARIMA models have become of wide application in wide range of areas. When dealing with time series data, the series has to be stationary. If not stationary, it needs to be made stationary by the process of differencing. A stationary time series is one whose statistical properties remain unchanged over time i.e for every t and $t - s$: $E(y_t) = E(y_{t-s})$ (constant mean) and $E(y_t - \mu)^2 = E(y_{t-s} - \mu) = \sigma^2_y$ (constant variance). The operator ∇ defined by $\nabla^d = (1 - B)^d$ where B is the backward shift operator defined by $B^j y_t = y_{t-j}$ is used in the process of differencing, [3].

Further advancements brought knowledge to model time series data with seasonality which led to the multiplicative seasonal ARIMA (SARIMA) process with orders $(p, d, q)(P,D,Q)_s$ which became useful to date, [2]. In the SARIMA model, p is the order of non-seasonal AR, d is the order of non-seasonal differencing, q is the order of non-seasonal MA, P is the order of seasonal AR, D is order of seasonal differencing, Q is the order of seasonal MA and s is periods in a season.

SARIMA time series models are formed by adding the seasonal terms in the usual ARIMA model. Mathematically, a SARIMA model is given by,

$$\Phi(B_s)\varphi(B)y_t = \Theta(B_s)\theta(B)\varepsilon_t \quad (1)$$

where s is the number of periods per season, y_t is the time series observation at time t, ε_t is white noise, Φ is the seasonal AR parameters, φ is the non-seasonal AR parameters, Θ is the seasonal MA parameters, θ is the non-seasonal MA parameters and B is the back shift operator. When the regular and seasonal differencing are included, equation 1 becomes;

$$\Phi(B_s)\varphi(B)(1 - B)_d(1 - B_s)_D y_t = \Theta(B_s)\theta(B)\varepsilon_t \quad (2)$$

3

, [3] where d is the order of non-seasonal differencing and D is the order of seasonal differencing.

The non-seasonal autoregressive (AR) and non-seasonal moving average (MA) components are;

$$\varphi(B) = 1 - \varphi_1 B_1 - \cdot \cdot \cdot - \varphi_p B_p \quad (3)$$

and

$$\theta(B) = 1 - \theta_1 B_1 - \cdot \cdot \cdot - \theta_q B_q \quad (4)$$

respectively.

The seasonal autoregressive (SAR) and seasonal moving average (SMA) are;

$$\Phi(B_s) = 1 - \Phi_1 B_s - \cdot \cdot \cdot - \Phi_P B_{Ps} \quad (5)$$

and

$$\Theta(B_s) = 1 + \Theta_1 B_s + \cdot \cdot \cdot + \Theta_Q B_{Qs} \quad (6)$$

respectively.

SARIMA time series models have widely been used to analyze time series data from various sectors. For instance, Musyoki et al. [12] applied SARIMA time series model to come up with a model that describes the Kenya’s agricultural contribution to the Kenyan Gross Domestic Product (GDP). The study found that seasonal ARIMA$(1,0,0)(1,1,0)_4$ to be the best model that fits quarterly data. Kibunja et al. [9], studied on forecasting precipitation in Mt. Kenya region using SARIMA modeling. The study used monthly data and found that SARIMA$(1,0,1)(1,0,0)_{12}$ to be the best model to forecast amount of precipitation in the region. In addition, Otieno et al. [15], modeled the tourist accommodation demand in Kenya using SARIMA approach. From the quarterly data collected, SARIMA$(1,1,2)(1,1,1)_4$ model was found to be the suitable model which can then be used for forecasting. Furthermore, Choge et. al. [4] developed a SARIMA model for rainfall pattern of Uasin Gishu County in Kenya. The study came up with SARIMA$(0,0,0)(0,1,2)_{12}$ as the suitable model that fitted the data and hence the model was used for prediction.

In this paper, we develop a time series SARIMA model for the monetary value data from the export of the Kenya’s horticultural produce. The results of this study will be significant to decision makers and planners especially in the Ministry of Agriculture, specifically the horticulture sub-sector to formulate policies which can be channeled to improve the performance of the sub-sector for increased future value from the export of horticultural produce.

3 Materials and Methods

We consider monthly data from August 1998 to March 2023 for the value of export from Kenya’s horticultural produce which was obtained from www.CEICDATA.com-cental Bank of Kenya on Kenya’s Exports: domestic horticulture. The data was entered into a spreadsheet in Excel and saved as CSV format. R statistical software was then used to read the data and for further analysis. The time series was explored to identify any underlying patterns. This was achieved through decomposing the series through classical approach to extract trend, seasonality and the random components.

Box-Jenkins SARIMA modeling approach was adopted as outlined in Box and Jenkins modeling technique. The Box-Jenkins approach used involved the following stages: data preparation which involved differencing to make the data stationary, model identification, model selection and parameter estimation, diagnostic checking and finally forecasting.

Model identification was done by studying the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the stationary series. The best $SARIMA(p, d, q)(P,D,Q)_s$ model was then selected based on corrected Akaike Information Criteria(AICc) for the data up to March 2023. Maximum likelihood method was used to estimate the parameters of the model.

Model diagnostic checking was done by examining the ACF of residuals and the Ljung-Box test of residuals to check if the residuals were white noise.

4 Results and Discussion

A time series plot of the original data is as given in Figure 1,

The data was decomposed using classical approach to obtain the trend, seasonality and the irregular components as shown in Figure 2. From Figure 2, there is evidence of presence of an upward trend and a pattern repeating itself after every twelve months of a

year. The autocorrelations in the ACF plot in Figure 3 go beyond the confidence bounds
indicating that the series is non stationary.

Plot of Horticulture Export Value with Time
Year
Horticulture Export Value
2000 2005 2010 2015 2020
2000 6000 10000 14000

Figure 1: Time series plot of horticulture export value

2000 10000
observed
2000 8000
trend
−600 0 400
seasonal
−2000 1000
2000 2005 2010 2015 2020 random
Time
Decomposition of additive time series

Figure 2: Decomposed Series of AGDP

0 1 2 3 4 5
0.0 0.2 0.4 0.6 0.8 1.0
Lag
ACF
ACF of Horticultural Export Value

Figure 3: ACF Plot for Horticultural Export Value

Stationarity was achieved through the seasonal and regular differencing of the non stationary
time series whose output is as given in Figure 4.

Plot of Seasonally−trend differenced series
Year
Horticulture Export Value
2000 2005 2010 2015 2020
−6000 −2000 0 2000 6000

Figure 4: Plot of Seasonally-trend Differenced series

Identification of the Model

The autocorrelation and partial autocorrelation plots for the stationary series are given
in Figure 5 and Figure 6. The ACF plot shows that autocorrelations after lags 0,

0 1 2 3 4 5
−0.5 0.0 0.5 1.0
Lag
ACF
ACF of seasonally−trend differenced series

Figure 5: ACF of Seasonally-trend differenced Series

0 1 2 3 4 5
−0.6 −0.4 −0.2 0.0 0.2
Lag

Partial ACF

PACF of seasonally−trend differenced series

Figure 6: PACF of Seasonally-trend differenced series

1, 3 and 4 are significant while autocorrelations at other lags are within the confi8 dence bounds thus insignificant. On the other hand, the PACF plot gives significant spikes after lags 0, 1, 2, 3 and 4 while the rest are insignificant. When the ACF and PACF plots are investigated at lags multiples of the seasonality i.e $s = 12, 24, 36, \cdots$ to identify the order of the seasonal components, it is observed that the autocorrelations in both cases are insignificant. Keeping in mind that both regular and seasonal differencing were each done once, it indicates that $d = 1$ and $D = 1$. Therefore, we have $SARIMA(0,1,0)(0,1,0)_{12}$, $SARIMA(0,1,1)(0,1,0)_{12}$, $SARIMA(0,1,3)(0,1,0)_{12}$, $SARIMA(0,1,4)(0,1,0)_{12}$, $SARIMA(1,1,0)(0,1,0)_{12}$, $SARIMA(1,1,1)(0,1,0)_{12}$, $SARIMA(1,1,3)(0,1,0)_{12}$, $SARIMA(1,1,4)(0,1,0)_{12}$, $SARIMA(2,1,0)(0,1,0)_{12}$, $SARIMA(2,1,1)(0,1,0)_{12}$, $SARIMA(2,1,3)(0,1,0)_{12}$, $SARIMA(2,1,4)(0,1,0)_{12}$, $SARIMA(3,1,0)(0,1,0)_{12}$, $SARIMA(3,1,1)(0,1,0)_{12}$, $SARIMA(3,1,3)(0,1,0)_{12}$, $SARIMA(3,1,4)(0,1,0)_{12}$, $SARIMA(4,1,0)(0,1,0)_{12}$, $SARIMA(4,1,1)(0,1,0)_{12}$, $SARIMA(4,1,3)(0,1,0)_{12}$ and $SARIMA(4,1,4)(0,1,0)_{12}$ as possible models for the Kenya's monthly horticultural export produce data.

Model Selection and Estimation of Parameters

The model with the least value of AICc is considered. The values of AICc and BIC for the candidate models together with their corresponding RMSE, MAPE and MASE values are given in Table 1

Table:1 AICc, BIC, RMSE, MAPE and MASE values for the models

Model	AICc	BIC	RMSE	MAPE	MASE
$(0,1,0)(0,1,0)_{12}$	4868.3	4871.93	1282.139	14.98771	1.008753
$(0,1,1)(0,1,0)_{12}$	4732.73	4739.97	1004.228	12.97856	0.8273273
$(0,1,3)(0,1,0)_{12}$	4728.3	4742.74	989.0967	13.11745	0.8222419

$(0,1,4)(0,1,0)_{12}$	4722.01	4740.03	967.3726	12.87973	0.8062031
$(1,1,0)(0,1,0)_{12}$	4755.3	4762.54	1045.584	13.89856	0.8725567
$(1,1,1)(0,1,0)_{12}$	4727.28	4738.13	991.0418	13.08986	0.8272852
$(1,1,3)(0,1,0)_{12}$	4711.49	4729.5	950.7215	12.53814	0.7909031
$(1,1,4)(0,1,0)_{12}$	4708.69	4730.26	935.9905	12.779	0.7891897
$(2,1,0)(0,1,0)_{12}$	4733.95	4744.8	1002.967	13.29326	0.8396566
$(2,1,1)(0,1,0)_{12}$	4729.23	4743.67	990.8541	13.09985	0.8269964
$(2,1,3)(0,1,0)_{12}$	4710.17	4731.74	948.3717	12.63219	0.7893025
$(2,1,4)(0,1,0)_{12}$	4702.4	4727.52	924.064	12.17785	0.7739979
$(3,1,0)(0,1,0)_{12}$	4734.49	4748.93	1000.243	13.22092	0.8368437
$(3,1,1)(0,1,0)_{12}$	4710.65	4728.67	949.2628	12.55	0.789793
$(3,1,3)(0,1,0)_{12}$	4699.01	4724.12	918.0443	12.27734	0.7740327
$(3,1,4)(0,1,0)_{12}$	4694.36	4723	902.657	11.975	0.7532675
$(4,1,0)(0,1,0)_{12}$	4725.92	4743.93	981.3468	13.04532	0.8193576
$(4,1,1)(0,1,0)_{12}$	4727.55	4749.12	980.5648	13.02584	0.8168262
$(4,1,3)(0,1,0)_{12}$	4705.48	4734.12	933.0868	12.645	0.7879974
$(4,1,4)(0,1,0)_{12}$	4707.39	4739.54	932.7396	12.64362	0.7888484

Table 1 shows that SARIMA$(3,1,4)(0,1,0)_{12}$ with AICc value of 4694.36 is the suitable model that can be used to describe the value from the export of Kenya's horticultural produce.

Table 2 gives the parameter estimates for the selected model

Table:2 Parameter estimates for SARIMA$(3,1,4)(0,1,0)_{12}$

Model	Parameter	Parameter estimate	Std Error
$(3,1,4)(0,1,0)_{12}$	φ_1	-0.2784	0.0686
	φ_2	0.0184	0.0665
	φ_3	0.721	0.0558
	θ_1	-0.4889	0.0795
	θ_2	-0.0782	0.0213
	θ_3	-0.9726	0.0209
	θ_4	0.5397	0.0771

The model can be written in form of equation (2) as

$$(1 + 0.2784B - 0.0184B^2 - 0.721B^3)(1 - B)(1 - B^{12})y_t = (1 + 0.4889B +$$

$0.0782B^2 + 0.9726B^3$

$-0.5397B^4)\varepsilon_t$ (7)

where y_t is the value of horticultural export at time t, B is the backshift operator and ε_t is the error term.

Model Diagnostic Checking

Model checking involves checking whether the residuals are white noise, random and normally distributed. The difference sign test showed that the residuals are random ($statistic = 0.1005$, n=296, p-value=0.9199). Figure 7 gives the ACF of residuals and the p-values for Ljung-Box test statistic. The Ljung-Box test have p-values greater than 0.05 indicating that the residuals are white noise.

When 'auto.arima' function was applied, SARIMA(3,1,1)(0,0,1)$_{12}$ proved to be also a possible model to describe the data. The estimated parameters for the model are:

$\varphi_1 = 0.2671$, $\varphi_2 = 0.2141$, $\varphi_3 = 0.0433$, $\theta_1 = -0.9785$ and $\Theta_1 = 0.1672$. The SARIMA(3,1,1)(0,0,1)$_{12}$ model had AICc and BIC values of 4778.02 and 4803.83 respectively which are high than those for SARIMA(3,1,4)(0,1,0)$_{12}$. When the SARIMA (3,1,1)(0,0,1)$_{12}$ model was subjected to diagnostic checking to test if the residuals are white noise, it passed the tests i.e the residuals were white noise, random and normally distributed. The diagnostic plots for the model are given in Figure 8

Standardized Residuals
Time
2000 2005 2010 2015 2020
−4 2
0.0 0.5 1.0 1.5 2.0
−0.5 1.0
Lag
ACF
ACF of Residuals
2 4 6 8 10
0.0 0.8
p values for Ljung−Box statistic
lag
p value

Figure 7: ACF of residuals and Ljung-Box p-values

Standardized Residuals
Time
2000 2005 2010 2015 2020
−2 4

0.0 0.5 1.0 1.5 2.0
0.0 1.0
Lag
ACF
ACF of Residuals
2 4 6 8 10
0.0 0.8
p values for Ljung−Box statistic
lag
p value

Figure 8: Diagnostic plot

The two models seemed suitable and therefore to select best model, we consider the model with minimum values of AICc and BIC. The AICc and BIC values for the two models are as given in Table 3

Table: 3 AICc and BIC values for the 2 models

11

Model	AICc	BIC
SARIMA$(3,1,1)(0,0,1)_{12}$	4778.41	4803.83
SARIMA$(3,1,4)(0,1,0)_{12}$	4694.36	4723

From the AICc and BIC values, we find that SARIMA$(3,1,4)(0,1,0)_{12}$ is the best model to be used for forecasting. This is because the model has the least values of the AICc and BIC which are 4694.36 and 4723 respectively.

Forecasting

The SARIMA$(3,1,4)(0,1,0)_{12}$ model was then used to forecast up to December 2026 and the plot in Figure 9 was obtained. Figure 9 shows that the model produces forecasts with an increasing trend which agrees with the previous trend given by the data.

Forecasts from ARIMA(3,1,4)(0,1,0)[12]
2000 2005 2010 2015 2020 2025
5000 10000 15000

Figure 9: Forecasts from SARIMA$(4,1,4)(0,1,0)_{12}$

5 Conclusion

The objective of this study was to develop the best model that can be used to describe the data on value from the export of Kenya’s horticultural produce. The study obtained SARIMA$(3,1,4)(0,1,0)_{12}$ and SARIMA$(3,1,1)(0,0,1)_{12}$ as possible models for our data.

SARIMA$(3,1,4)(0,1,0)_{12}$ mmodel was identified by considering the use of the ACF and

the PACF plots while SARIMA(3,1,1)(0,0,1)$_{12}$ was identified using the ’auto.arima’ function. However, the two models were compared using AICc and BIC criteria and
12
SARIMA(3,1,4)(0,1,0)$_{12}$ was identified to be the best model since it had the minimum values of both AICc and BIC which are 4694.36 and 4723 respectively. The model was then subjected to diagnostic checks from which it was found that the model was adequate. The selected model was then used to forecast and the model predicted an increasing trend in the monthly value from export of horticultural produce of the Kenyan economy. This is important to the policy makers and planners in the horticulture sub-sector to come up with strategies on maintaining high income from the horticultural exports. As an area of study, there is need for further study be done on the individual components of the horticulture namely: value from fruits, cut flowers and vegetables using SARIMA models or any other time series models.

6 Acknowledgement

I acknowledge the Central Bank of Kenya for availing the data on www.CEICDATA.com where the data on monthly value from the export of Kenya’s horticultural produce was accessed.

References

[1] Box G. E. P. and Jenkins G. (1970). Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.

[2] Box G. E. P., Jenkins G. M. and Reinsel G. C. (1994). Time Series Analysis, Forecasting and Control, Prentice-Hall, Inc., USA

[3] Brockwell J. P. and Davis A. R. (2002). Introduction to Time Series and Forecasting, Springer-Verlag New York, 169-174

[4] Choge M., Nyongesa K., Mulati O., Makokha L. and Tireito F. (2016). Time

Series Model of Rainfall Pattern of Uasin Gishu County, IOSR Journal of Mathematics(IOSR-JM), 11(5), 77-84

[5] Government of Kenya (2009-2020). Agricultural Sector Development Strategy (ASDS)

[6] Hyndman J. R. (2017). Time Series Components, www.otexts.org/fpp/6/1

[7] Joshi, B.P., Kumar, S. (2012). A Computational Method of Forecasting Based on Intuitionistic Fuzzy Sets and Fuzzy Time Series. In: Deep, K., Nagar, A., Pant, M., Bansal, J. (eds) Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011. Advances in Intelligent and Soft Computing, vol 131. Springer, New Delhi. https://doi.org/10.1007/978-81-322-0491-6 91

[8] Joshi, B.P., Mukesh Pandey, Sanjay Kumar (2016). Use of Intuitionistic Fuzzy Time Series in Forecasting Enrollments to an Academic Institution. In: Pant, M., Deep, K., Bansal, J., Nagar, A., Das, K. (eds) Proceedings of Fifth International Conference on Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing, vol 436. Springer, Singapore. https://doi.org/10.1007/978-981-10-0448-3 70

[9] Kibunja W. E., Kihoro M. J., Orwa O. G. and Yodah O. W. (2014). Forecasting Precipitation Using SARIMA Model: A Case Study of Mt. Kenya Region, Mathematical Theory and Modeling, 4(11), 50-58

[10] Meme M. S. (2015). Export Performance of the Horticultural Sub-sector in Kenya- An Emperical Analysis; http://erepository.uonbi.ac.ke/bitstream/handle/11295/93214/.pdf

[11] Musundi S. W., M' mukiira P. M., and Mungai F. (2016). Modeling and Forecasting Kenyan GDP Using Autoregressive Integrated Moving Average (ARIMA) Models, Science Journal of Applied Mathematics and Statistics, 4(2), 64-73

[12] Musyoki M., Ong' ala J. andWawire N. (2018). Modeling Agricultural Gross Domestic Product of Kenyan Economy Using Time Series, Asian Journal of Probability and

Statistics, 2(1), 1–12
[13] Mwanga D., Ong’ala J. and Orwa G. (2017). Modeling Sugarcane Yields in the Kenya
Sugar Industry: A SARIMA Model Forecasting Approach, International Journal of Statistics and Applications, 7(6): 280-288
[14] New Partnership for Africa’s Development(NEPAD) (2003). Comprehensive Africa
Agriculture Development Programme
[15] Otieno G., Mung’atu J., and Orwa G. (2014). Time Series Modeling of Tourist Accomodation
Demands in Kenya, Mathematics Theory and Modeling, 4(10) 106-117