

Modeling Sugarcane Yields in the Kenya Sugar Industry: A SARIMA Model Forecasting Approach

Mwanga D.^{1,*}, Ong'ala J.², Orwa G.¹

¹Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

²Department of Mathematics, Masinde Muliro University of Science and Technology, Kenya

Abstract The purpose of this study was to fit a model that forecasts quarterly sugarcane yields in Kenya. Seasonal ARIMA models are explored and tested. Seasonal ARIMA(2,1,2)(2,0,3)₄ is found to be the best model that fits quarterly sugarcane yields from 1973-2015. Sugarcane yields data collected quarterly from 1973-2014 is used for modeling and SARIMA (2,1,2)(2,0,3)₄ model is fit and 2015 quarterly forecasts are compared against the actual quarterly yields in 2015. If all factors are held constant, the model predicted a drop in sugarcane yields in 2016 to 60 (95% CI: 34.58, 84.69) tonnes of cane per hectare (tch) in 2016, 54 (95% CI: 26.24, 82.43) tch in 2017 and 51.48 (95% CI: 21.51, 81.45) tch in 2018. A steady increase would be observed again from 2020-2024.

Keywords Seasonal ARIMA, Exponential smoothing, Forecast, Sugarcane yields

1. Introduction

Historical Background

Sugarcane is an important commercial crop grown in four major sugarcane growing belts in the Kenya such as; Central Nyanza, Western Kenya, South Nyanza and the Coast. Before independence, the sugar industry in Kenya was dominated by the private sector. Historically, the growing of sugarcane in Kenya started with the involvement of the Kenya Government at the turn of the century, with the establishment of experimental farms at Mazeras and Kibos, whose sole activity was to evaluate sugarcane and other introduced crops. Subsequently, large production of sugarcane started in 1923 when a sugar factory was built at Miwani in Nyanza Province, Kisumu District and at Ramisi in Coast Province, Kwale District. Today sugarcane is grown for white sugar production in Nyando, South Nyanza, Mumias, Nzoia and Busia by small, large scale farmers and sugar factories.

Sugarcane yields have declined in the past decade with average tones cane per hectare (tch) dropping from 74 tch in 2004 to 61 tch in 2014 [1]. The total area under cane was approximately 211,342 hectares in 2014 compared to 213,920 hectares in 2013 [2]. This decline has had adverse effects directly on the livelihood of sugarcane farmers who directly depend on sugarcane farming. Sugarcane farmers have experienced many challenges over time and many have

since threatened to pull out of sugarcane farming and explore other profitable farming practices [3]. There are many studies that have been done in order to help secure the industry from falling which includes studies that have been conducted by Kenya Agricultural and Livestock Research Organization-Sugar Research Institute (KALRO-SRI), the institution mandated to carry out research in sugar and sugarcane. They range from studies on best management practices, drainage and irrigation, farmers' trainings, trainings of trainers (ToTs), coping strategies to the challenges facing the industry, studying and releasing new and improved varieties, development of a synchrony model to help the millers bridge the gap between factory crushing capacity and harvesting time and many others. However, the adoption of these technologies by the farmers and the sugarcane millers has however not been very encouraging [4] [5]. This is generally contributed by the low motivation farmers have developed towards sugarcane farming more often associating it with poverty and feeling that sugarcane farming is not productive.

Forecasting Methods in the Kenya Sugar Industry

There are limited studies in Kenya that have explored the forecasting methods on sugarcane production in the Kenya with the view to improve or develop new methods. This necessitates the need to develop models and methodologies that would be useful to predict sugarcane yields and its components.

Since sugarcane was first grown in Kenya, sugarcane yield is estimated using conventional approaches through biennial field surveys by the millers and the sugar directorate. Their methodology is based on visual physical assessment [2, 6]. In this method, a monthly productivity index ranging

* Corresponding author:

mtaimwanga@gmail.com (Mwanga D.)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

between 0 to 5 is applied to sample cane crop from the age of one month while considering the parameters; crop vigour, crop colour, crop density, weed status, pests and diseases at the time of the assessment. The estimated yield is then used to project sugarcane production for the current and the subsequent year. Mulianga *et al.* (2013) explored the suitability of the Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectrometer (MODIS) obtained for six sugar management zones, over nine years (2002–2010), to forecast sugarcane yield on an annual and zonal base. They took into account the characteristics of the sugarcane crop management (15-month cycle for a ratoon, accompanied with continuous harvest in Western Kenya), the temporal series of NDVI was normalized through an original weighting method that considered the growth period of the sugarcane crop (wNDVI), and correlated it with historical yield datasets. They found out that results when using wNDVI were consistent with historical yield and significant, while results when using traditional annual NDVI integrated over the calendar year were not significant [7].

Time Series modeling

A time series is a sequence of data points, typically consisting of successive measurements made over a time interval. Time series analysis accounts for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted for. Sugarcane yields when taken at equal time intervals over a period of time constitute a time series data which could then be analyzed using time series techniques. Presently, statistical techniques of time series analysis have been widely discussed in the literature and there is a great variety of circumstances of research in which they can be used, especially in studies involving time dependent data.

Time series models were first introduced by Box and Jenkins in 1960 hence the name Box-Jenkins Model [8]. Originally, Box and Jenkins methodology involved three iterative steps viz; model selection, parameter estimation, and model checking [8-10]. Recent development to this was to add a preliminary stage of data preparation and a final stage of model application which is forecasting [11]. Data preparation in this sense involves transformations and differencing if the data under study require that it be done to satisfy the Box and Jenkins assumptions before modeling. Model selection involves using graphs and model selection tools such as Akaike Information Criteria and Bayesian Information Criteria to identify the model that best fits the data. Parameter estimation means finding the values of the model coefficients which provides the best fit to the data. Model checking means testing for assumptions to identify areas of model inadequacy which in this case involves testing whether residuals are white noise. Once the model has been selected, estimated and checked, it is then used for forecasting [12].

A time series Y_t is a function of any or all of the four components; trend T_t , seasonal S_t , cyclic C_t and a random

term ε_t [13]. This is represented in equation 1.

$$Y_t = f(T_t, S_t, C_t, \varepsilon_t) \quad (1)$$

Y_t can be multiplicative (equation 2) or additive (equation 3) depending on the series under study.

Multiplicative

$$Y_t = T_t * S_t * C_t * \varepsilon_t \quad (2)$$

Additive

$$Y_t = T_t + S_t + C_t + \varepsilon_t \quad (3)$$

Cyclic components can only be noted when the period is long. More often in a shorter period of time, the cyclic component is hidden in the trend and therefore modeled together with the trend.

Seasonal ARIMA Models (SARIMA)

Often time series possess a seasonal component that repeats itself after every S observations ($S=12$ for monthly and $S=4$ for quarterly). In order to deal with seasonality, ARIMA processes have been generalized: SARIMA models then have been formulated [14].

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA model. It is denoted as [15]

$$\text{ARIMA}(p, d, q)(P, D, Q)_S \quad (4)$$

Where p =non-seasonal AR order, d =non-seasonal differencing order, q =non-seasonal MA order, P =seasonal AR order, D =seasonal differencing order, Q =Seasonal differencing order, and S is the number of periods in a season.

The first part of the brackets in equation 4 is the non-seasonal part of the model while the second part is the seasonal part.

This is represented more formally by equation 5;

$$\Phi(B^S)\varphi(B)(y_t - \mu) = \Theta(B^S)\theta(B)\alpha_t \quad (5)$$

Where S is the number of periods per season, y_t is the time series observation at time t , α_t is white noise, μ is the mean of the series, Φ is seasonal AR parameters, φ is non-seasonal AR parameters, Θ is the seasonal MA parameters, θ is non-seasonal MA parameters, and B is the back shift operator.

The back shift operator B can be simplified in AR and MA terms as;

$$\varphi(B)y_t = 1 - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p} \quad (6)$$

Or

$$\theta(B)\alpha_t = 1 + \theta_1 \alpha_{t-1} + \dots + \theta_q \alpha_{t-q} \quad (7)$$

The non-seasonal components are;

$$\text{AR: } \varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p \quad (8)$$

$$\text{MA: } \theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (9)$$

The Seasonal components are;

$$\text{Seasonal AR (SAR): } \Phi(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS} \quad (10)$$

$$\text{Seasonal MA (SMA): } \Theta(B^S) = 1 + \Theta_1 B^S + \dots + \Theta_Q B^{QS} \quad (11)$$

In other words, the non-seasonal part is multiplied by the

seasonal part of the ARIMA model. On the left hand side of equation 5, the seasonal and non-seasonal AR components multiply each other, and on the right hand side, the seasonal and non-seasonal MA components multiply each other.

SARIMA Models are an extension of the common ARIMA models to account for the presence of seasonal terms in the time series [14] and the Box Jenkins methodology is followed when applying it. In Box Jenkins methodology, before fitting a time series model the time series object must be stationary. A time series is stationary if it has constant mean and constant variance, otherwise if not stationary, then it is transformed to stationary by differencing which is simply computing the differences between consecutive observations [12].

SARIMA models have gained popularity in most in modeling data collected either monthly or quarterly. Kibunja *et al.* (2014) applied SARIMA model to forecast precipitation in the Mount Kenya Region collected monthly recognizing the fact that it is predicting rainfall amounts in Kenya which fluctuates from year to year through empirical observation of the atmosphere alone is difficult [16]. They found SARIMA (1,0,1)(1,0,0)₁₂ as the best model to forecast rainfall in the Mt Kenya region. SARIMA have also been applied in tourism sector to model tourist accommodation demands in Kenya [17]. They used quarterly data and found SARIMA (1,1,2)(1,1,1)₄ as the most suitable. Gikungu *et al.* (2015) applied SARIMA to forecast inflation rates in Kenya collected quarterly. They found that SARIMA (0,1,0)(0,0,1)₄ was the most suitable in forecasting quarterly inflation rates [18]. Another study that has applied SARIMA in forecasting was done by Fannoh *et al.* (2012) in Liberia. They found ARIMA(0,1,0)(2,0,0)₁₂ as the best model that fits monthly inflation rates in Liberia [19]. Many other studies in other sectors not captured here have also used SARIMA modeling approach with success.

In the sugar and sugarcane sector however, the use of SARIMA models is not evident. Nevertheless, ARIMA models have been applied both in Kenya and other countries. Kumar *et al.* (2014) fit an ARIMA (2, 1, 0) and forecast annual sugarcane production data in India from 1950 to 2012. This model was able to forecast sugarcane production in India and was able to predict an increase in sugarcane production in 2013 and a sharp decrease in 2014 [20]. The study used data in annual scale which may not have captured the variations that exist in the dataset taken at quarterly time intervals. This is so because sugarcane harvested during rainy season may not be exposed to the same conditions as sugarcane harvested during dry season. This needs to be taken into account. Therefore, taking the data quarterly will put all those concerns into consideration.

In previous work [21] it is indicated that ARIMA (4,1,1) is the best model to predict future adoption of KEN 83-737 cane variety. This method succeeded to predict a drop in

adoption of KEN 83-737 in 2012 and 2013. This study however did not consider the production of sugarcane as a whole which is necessary for the industry to have the general view of the direction the sugarcane yields is taking or will take for necessary decision on policy recommendations.

2. Materials and Methods

The data used in this study are secondary data on sugarcane yields collected from Agriculture, Food and Fisheries Authority - Sugar Directorate (AFFA-SD) Year books of Statistics. Sugar Directorate collects and analyzes data on sugar and sugarcane from all the sugarcane growing zones and mills in the Kenya sugar Industry annually. The data is then published in the yearbooks of statistics and shared with agriculture research institutions and other stakeholders for consumption. This study focuses on quarterly sugarcane yield in terms of tones cane per hectare (tch) collected from the Sugar Directorate yearbooks of statistics from 1973-2015.

The data were available annually from 1973-2015 and quarterly from 1999-2015. Quarterly data from 1973-1998 were obtained through interpolation using the “zoo” package [22] available in the R software using the available annual data. The “zoo” package has the ability to interpolate missing data without interfering with the underlying trends in the time series. The time series is explored to identify any underlying patterns or behaviors. This is done by decomposing the time series to extract the trend, seasonality and the cyclic components using classical and Seasonal and Trend decomposition using Loess (STL) [23] approaches.

Box-Jenkins SARIMA models are explored following all the stages required in Box and Jenkins modeling technique and the best predictive model chosen that will forecast the quarterly sugarcane yields in the Kenya Sugar Industry. Model selection is based on the bias corrected Akaike Information Criteria (AICc) [24] where the model with minimum AICc is selected. Data from 1973 – 2014 is used to fit the SARIMA model and data from the four quarters of 2015 are used to check the adequacy of the forecast. Model diagnostic checking is done by analyzing the residuals. Ljung-Box test for independence of the residuals is done where residuals should be uncorrelated and look like white noise [25].

3. Results and Discussion

Quarterly sugarcane yields data from 1973-2014 had an average of 76.68 tch (SD=16.51). There was larger variation in the earlier years (1973-2000) with an average of 82.19 tch (SD=17.02) compared to 65.64 tch (SD=7.52) from 2001-2014. Figure 1 shows the plot of the time series.

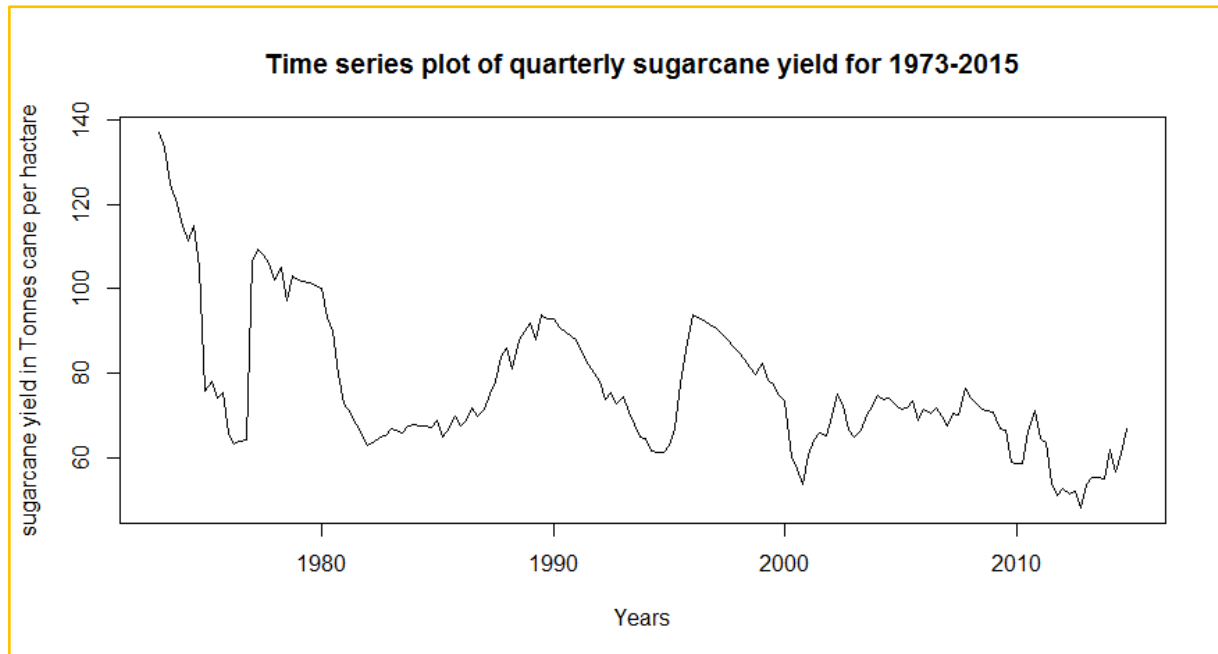


Figure 1. Plot of quarterly sugarcane yields

The time series was decomposed using Seasonal – Trend Decomposition based on Loess (STL) [23] approach to extract its components viz. trend, seasonal and cyclic and plot in figure 2 was obtained. In the quarterly scale, the decomposition shows that the seasonal component changes slowly as shown in the second panel such that similar pattern is seen for consecutive years but years far apart may exhibit different patterns. The remainder shown in the fourth panel is the random effect showing an indication of higher variation in the earlier years (1973-1978) which reduces as year progresses. This also indicates presence of seasonality.

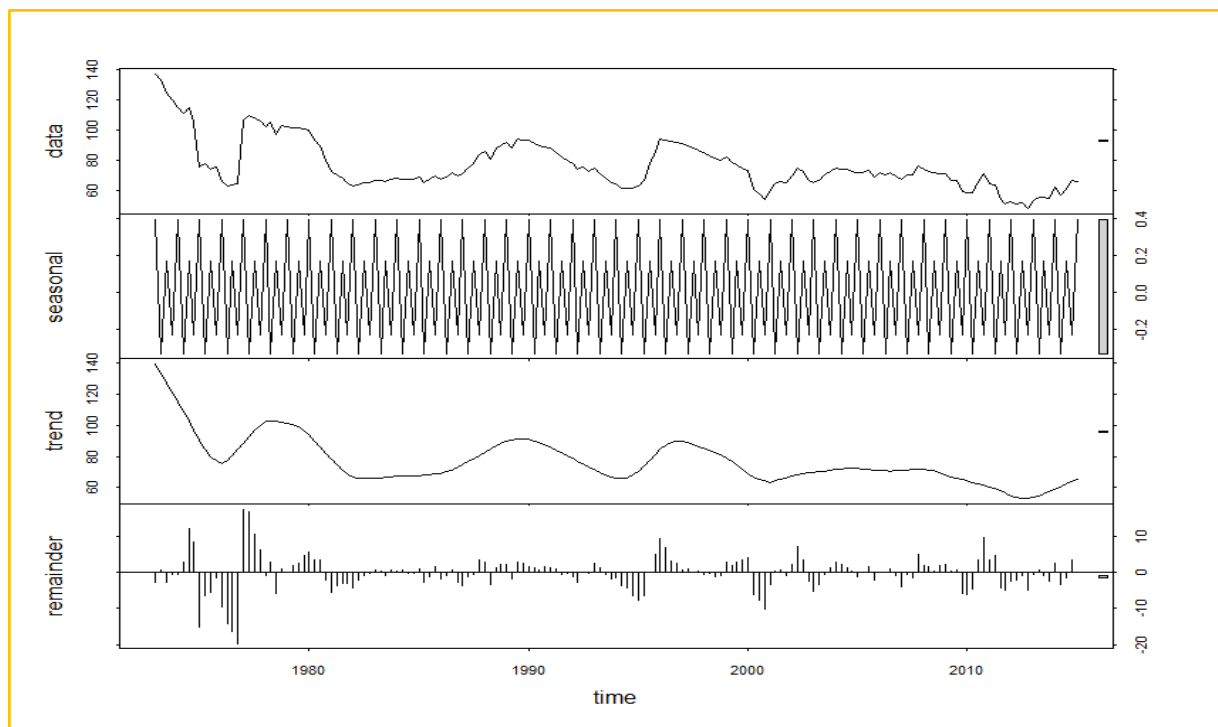


Figure 2. Decomposition of time series using STL

The relative size of the grey bars on the right hand side of each plot shows that trend is the dominating component. The large grey bar in the second panel shows that the variation in the seasonal component is small compared to the variation in the data and trend.

A time series linear regression line model was fit in the time series with trend as the explanatory variable. The results indicate the trend is significant and indicates a drop in sugarcane yield (figure 3) by 0.20 for every succeeding quarter ($p < 0.001$, 95 CI: -0.24, -0.16).

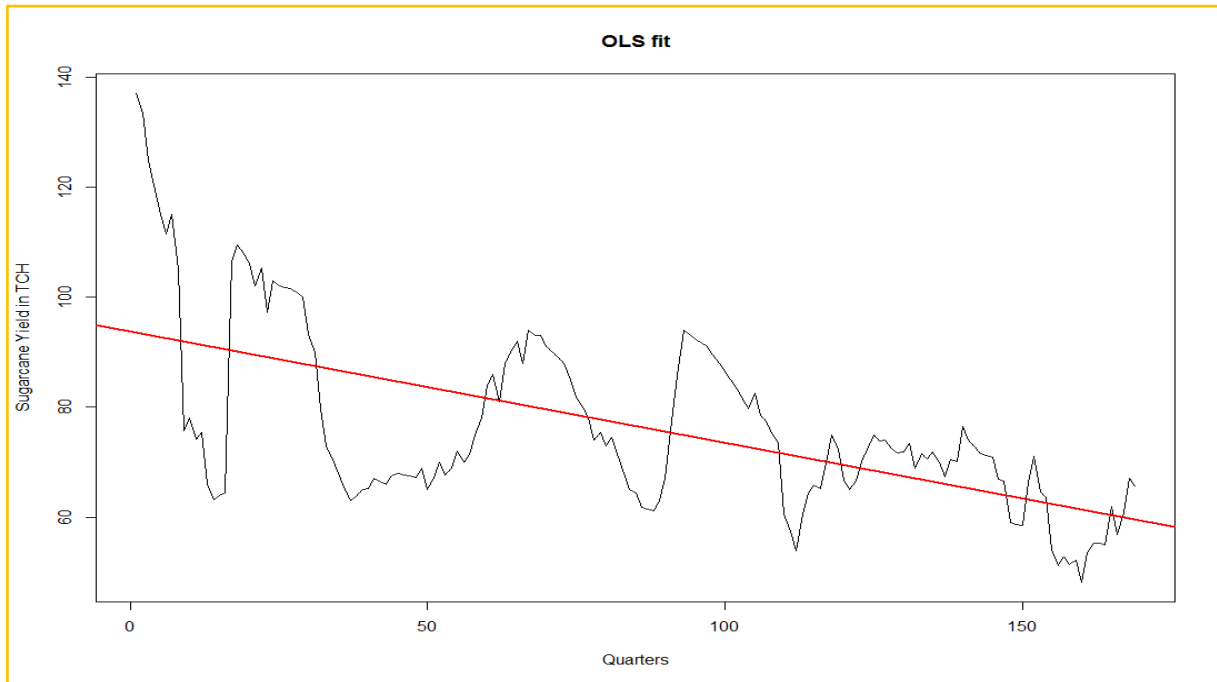


Figure 3. Trend plot of sugarcane yields from 1973-2015

Model identification

Box and Jenkins (1976) described that a time series model requires that the time series object be stationary. Augmented Dickey Fuller Test indicated the time series is not stationary (Dickey-Fuller = -3.2917, p -value = 0.0753). Stationarity of the time series was achieved through differencing once. This indicates that the non-seasonal order of differencing in the SARIMA model is 1.

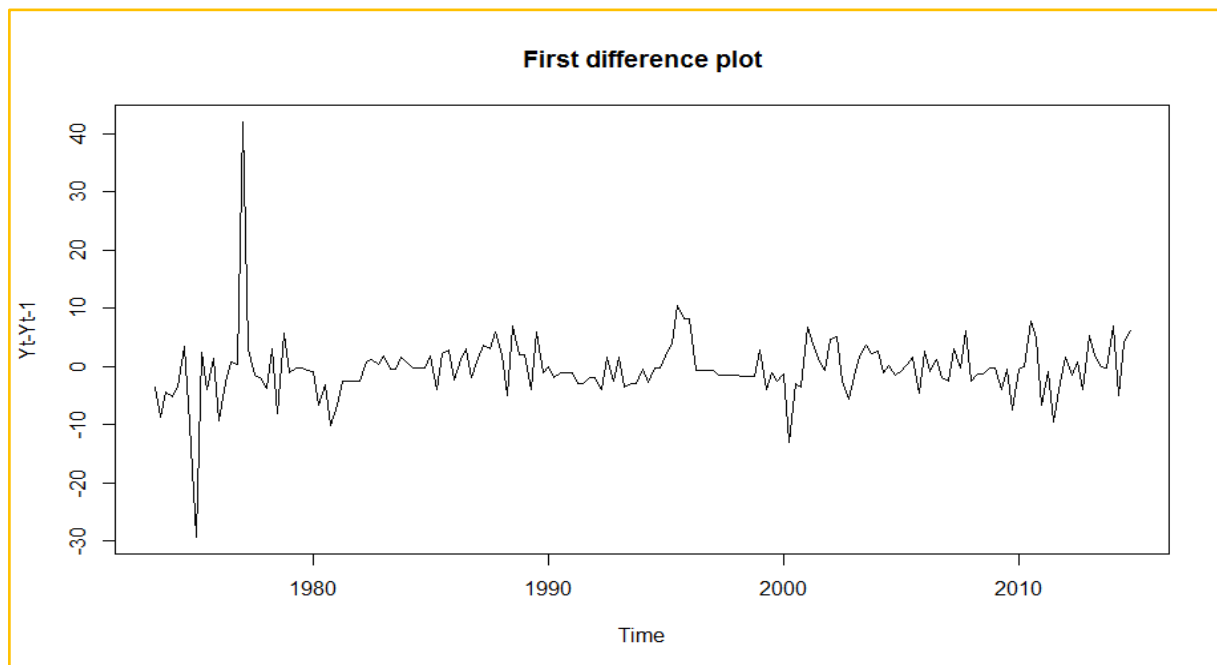


Figure 4. Plot of the first difference for the time series

In figure 4 it is evident that the time series has become stationary after first differencing despite high variation in the earlier years.

The autocorrelation function and the partial autocorrelation function plots (figure 5) suggest anAR2, AR4, or MA2.

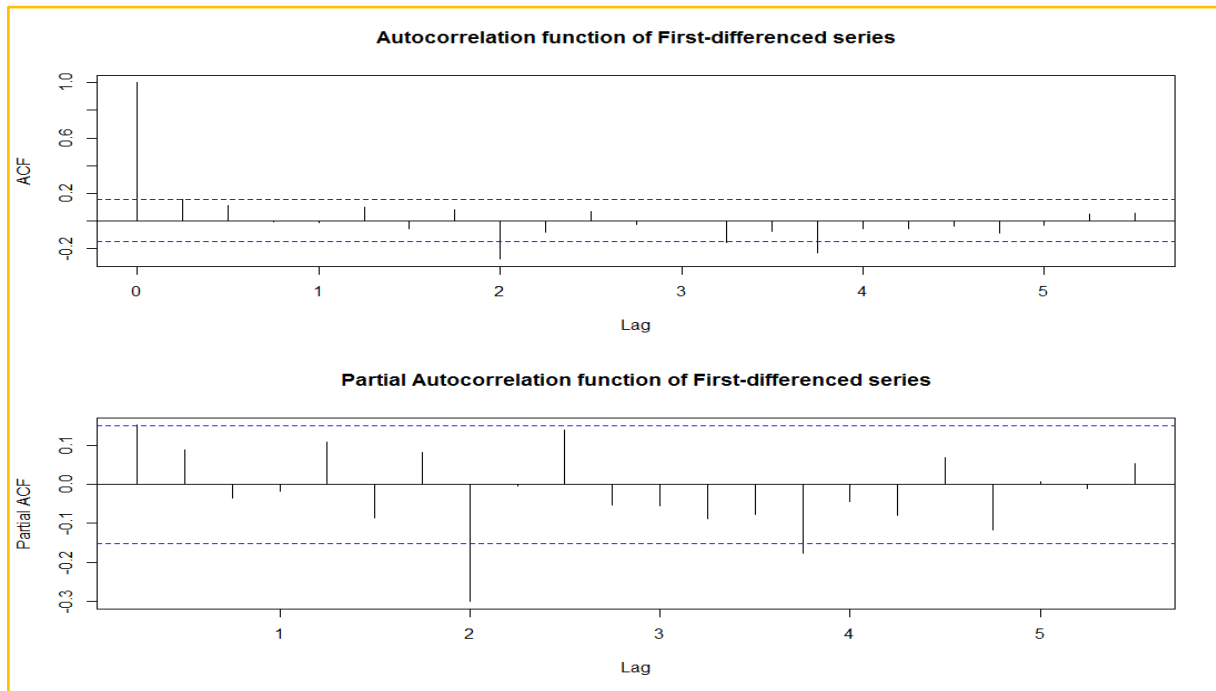


Figure 5. Estimated ACF and PACF for the differenced time series

Due to the complexity of identifying the combination of the order of the seasonal and non-seasonal AR and MA components that has minimum AICc, model parsimony was used in choosing the order of these components. The basic principle of parsimony means that the simplest possible model is chosen subject to the model adequacy constraints [26]. Table 1 shows some of the models that were considered together with their AICc and forecast accuracy measures.

Table 1. AICc and forecast accuracy measures of the considered models

(p,d,q)(P,D,Q)[S]	AICc	RMSE	MAPE	(p,d,q)(P,D,Q)[S]	AICc	RMSE	MAPE
(0,1,0)(2,0,3)[4]	1035.42	4.96	4.31	(0,1,2)(3,0,0)[4]	1035.38	5.05	4.35
(0,1,0)(2,0,4)[4]	1035.63	4.85	4.21	(1,1,0)(0,0,2)[4]	1034.22	5.10	4.38
(0,1,0)(0,0,0)[4]	1049.41	5.45	4.25	(1,1,0)(0,0,3)[4]	1036.27	5.10	4.38
(0,1,0)(0,0,3)[4]	1038.08	5.16	4.49	(1,1,0)(1,0,2)[4]	1036.30	5.10	4.38
(0,1,0)(1,0,2)[4]	1038.09	5.16	4.50	(1,1,0)(1,0,3)[4]	1035.02	5.04	4.35
(0,1,0)(2,0,0)[4]	1036.70	5.17	4.54	(1,1,0)(2,0,0)[4]	1032.97	5.08	4.39
(0,1,0)(2,0,1)[4]	1037.83	5.15	4.56	(1,1,0)(2,0,1)[4]	1034.99	5.07	4.40
(0,1,0)(2,0,3)[4]	1035.42	4.96	4.31	(1,1,0)(2,0,3)[4]	1034.05	4.95	4.32
(0,1,1)(2,0,0)[4]	1034.04	5.09	4.42	(1,1,1)(0,0,2)[4]	1035.20	5.08	4.34
(0,1,1)(2,0,3)[4]	1034.54	4.95	4.32	(1,1,1)(2,0,0)[4]	1033.85	5.06	4.35
(0,1,1)(0,0,3)[4]	1037.06	5.11	4.40	(2,1,0)(0,0,2)[4]	1034.45	5.07	4.33
(0,1,1)(1,0,2)[4]	1037.07	5.11	4.40	(2,1,0)(2,0,0)[4]	1033.49	5.05	4.34
(0,1,1)(2,0,1)[4]	1036.01	5.09	4.43	(2,1,0)(2,0,1)[4]	1035.48	5.05	4.35
(0,1,2)(0,0,2)[4]	1034.08	5.06	4.33	(2,1,0)(2,0,3)[4]	1035.74	4.94	4.30
(0,1,2)(0,0,4)[4]	1035.33	5.00	4.40	(2,1,1)(0,0,2)[4]	1036.37	4.94	4.30
(0,1,2)(2,0,0)[4]	1033.41	5.05	4.34	(2,1,2)(2,0,3)[4]	1024.88	4.59	3.98
(0,1,2)(2,0,2)[4]	1037.17	5.04	4.35	(2,1,2)(2,0,4)[4]	1036.12	4.84	4.33
(0,1,2)(2,0,3)[4]	1035.42	4.93	4.27	(2,1,2)(3,0,0)[4]	1035.65	4.97	4.32

The model with minimum AICc (1024.88) was identified to be SARIMA(2,1,2)(2,0,3)₄

Model Estimation

The second stage in fitting Box and Jenkins models is model parameters estimation. The table of SARIMA(2,1,2)(2,0,3)₄ model parameter estimates and their standard errors is given in the table 2.

Table 2. SARIMA(2,1,2)(2,0,3)₄ model parameter estimates

Model components	Estimate	Std.Error	Pr(> z)
ar1	-1.1275	0.0720	< 0.0001***
ar2	-0.7693	0.0666	< 0.0001***
ma1	1.2518	0.0256	< 0.0001***
ma2	1.0000	0.0327	< 0.0001***
sar1	1.3640	0.1081	< 0.0001***
sar2	-0.6489	0.1168	< 0.0001***
sma1	-1.5769	0.3606	< 0.0001***
sma2	0.6272	0.4164	0.1320
sma3	0.2088	0.1779	0.2406

Notes: *** significant at 0.001

When these coefficients are substituted back in equation 5, the model is then represented as in an equation 12;

$$\begin{aligned} \hat{Y}_t = & \mu + 1.13y_{t-1} + 0.77y_{t-2} - 1.36y_{t-4} - 1.54y_{t-1}y_{t-4} - 1.03y_{t-2}y_{t-4} + 0.65y_{t-8} + 0.73y_{t-1}y_{t-8} \\ & + 0.5y_{t-2}y_{t-8} - 1.25\alpha_{t-1} - \alpha_{t-2} + 1.58\alpha_{t-4} + 1.98\alpha_{t-1}\alpha_{t-4} + 1.58\alpha_{t-2}\alpha_{t-4} - 0.63\alpha_{t-8} \\ & - 0.79\alpha_{t-1}\alpha_{t-8} - 0.63\alpha_{t-2}\alpha_{t-8} - 0.21\alpha_{t-12} - 0.26\alpha_{t-1}\alpha_{t-12} - 0.21\alpha_{t-2}\alpha_{t-12} \end{aligned} \quad (12)$$

Where \hat{Y}_t the estimate of sugarcane yield at time t , μ is the mean time series process, y_{t-i} is the past observations and α_i are the random shocks.

Model diagnostic checking

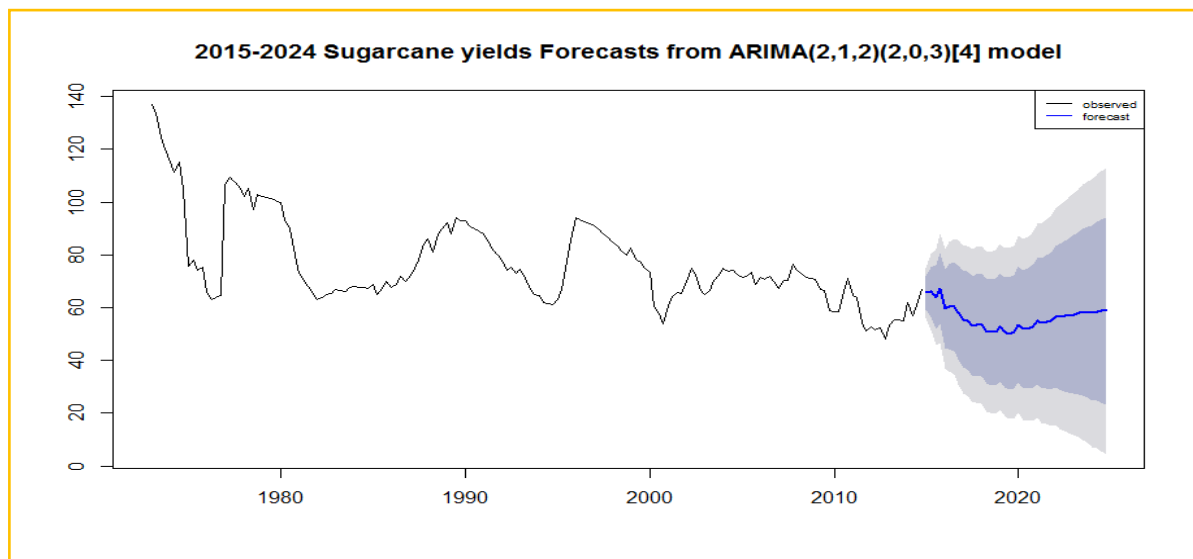
The third stage in Box and Jenkins modeling is model diagnostic checking that requires that the residuals look like white noise and uncorrelated. The Ljung Box test for autocorrelation indicate independence of the residuals ($\chi^2 = 0.5430$, $df = 1$, $p\text{-value} = 0.4612$). Shapiro Wilk's test for normality also showed that the residuals were approximately normal ($W = 0.9891$, $p\text{-value} = 0.2234$).

Forecasting

In this study SARIMA(2,1,2)(2,0,3)₄ was identified as the best fit and thus used for forecasting. Forecasts and actual values for the all the quarters of the year 2015 together with their 95% confidence intervals are given in the in table 3.

Table 3. Forecasts and observed from SARIMA(2,1,2)(2,0,3)₄ for the year 2015

Year Qtr	Actual	Forecast	Low 95%	High 95%	Error
2015 Q1	65.73	65.60	56.08	75.13	0.13
2015 Q2	67.35	65.96	51.68	80.23	1.39
2015 Q3	68.64	63.95	45.64	82.26	4.69
2015 Q4	63.90	67.25	46.54	87.96	-3.35

**Figure 6.** A 10- year forecast from SARIMA(2,1,2)(2,0,3)₄

In a 10-year quarterly forecasting if all sugarcane production factors remained constant, this model predicts a drop in quarterly yields from the year 2016 – 2019 then followed by a steady rise from 2020-2024 (Figure 6). The actual annual yield for 2016 is 62.9 tch which is a drop from 66.4 tch in 2015. This is consistent to what the model predicted.

The table 4 below shows the 10-year quarterly sugarcane yields forecasts

Table 4. A 10-year sugarcane yields forecasts by SARIMA(2,1,2)(2,0,3)₄

Year Qtr.	Point Forecast	Low 95%	High 95%	Year Qtr.	Point Forecast	Low 95%	High 95%
2015 Q1	65.60	56.08	75.13	2020 Q1	53.67	20.21	87.13
2015 Q2	65.96	51.68	80.23	2020 Q2	51.83	17.61	86.04
2015 Q3	63.95	45.64	82.26	2020 Q3	52.05	17.10	87.00
2015 Q4	67.25	46.54	87.96	2020 Q4	52.65	17.02	88.28
2016 Q1	59.56	36.83	82.28	2021 Q1	54.94	18.25	91.63
2016 Q2	60.32	35.89	84.75	2021 Q2	54.24	16.51	91.98
2016 Q3	60.42	34.73	86.12	2021 Q3	54.53	15.76	93.31
2016 Q4	58.24	30.88	85.60	2021 Q4	54.89	15.14	94.64
2017 Q1	55.49	27.55	83.44	2022 Q1	56.52	15.52	97.51
2017 Q2	55.09	26.66	83.51	2022 Q2	56.41	14.16	98.65
2017 Q3	53.05	24.04	82.06	2022 Q3	56.78	13.32	100.24
2017 Q4	53.40	23.87	82.93	2022 Q4	57.02	12.40	101.63
2018 Q1	53.41	23.53	83.29	2023 Q1	57.49	11.59	103.40
2018 Q2	50.88	20.62	81.14	2023 Q2	57.97	10.81	105.13
2018 Q3	50.66	20.08	81.24	2023 Q3	58.28	9.89	106.67
2018 Q4	50.74	19.80	81.67	2023 Q4	58.25	8.66	107.83
2019 Q1	52.56	21.17	83.96	2024 Q1	57.99	7.23	108.75
2019 Q2	50.74	18.90	82.57	2024 Q2	58.64	6.73	110.56
2019 Q3	50.20	17.88	82.51	2024 Q3	58.79	5.76	111.83
2019 Q4	50.79	18.04	83.54	2024 Q4	58.69	4.55	112.83

4. Conclusions

The main objective of this study was to identify a model that fits quarterly sugarcane yields data and forecast future yields based on the past values. This study found SARIMA(2,1,2)(2,0,3)₄ as the best model which had the lowest AICc and therefore fit to the quarterly sugarcane yields data from 1973-2014. The four quarters of the year 2015 were used to check the adequacy of the model. If all factors remain constant the model predicted a fall in yields until the year 2020 before starting to steadily rising again. Seasonal ARIMA models are proving to be good candidates of modeling time series with seasonal patterns and can be applied in any sector.

ACKNOWLEDGEMENTS

I acknowledge the Sugar Directorate and for making available the sugarcane yields data available in their annual yearbooks of Statistics and the Sugar Research Institute where from the individual Yearbooks of Statistics were accessed.

REFERENCES

- [1] AFFA - Sugar Directorate. *Yearbook of Statistics*. Nairobi: Government Printers, 2014.
- [2] AFFA-Sugar Directorate. *Yearbook of Statistics*. Nairobi: Government Printers, 2013.
- [3] Kweyu, W.W. *Factors influencing withdrawal of Farmers from Sugarcane farming: A Case of Mumias District, Kakamega County, Kenya*. 2012.
- [4] *Adoption of improved sugarcane varieties in Nyando Sugarcane zone*. Odenya, J. O., et al. Nairobi: s.n., 2010. 12th KARI Conference.
- [5] *A Baseline Survey on the status of Sugarcane Production Technologies in Western Kenya*. Jamoza, J. E., Amolo, R. A. and Muturi, S. M. 2013, International Journal of Sugarcane Technologists.
- [6] Kenya Sugar Board. Kenya Sugar Board Strategic Plan 2009. *Kenya Sugar Board*. [Online] 2009. [Cited: 07 15, 2017.] <http://www.kenyasugar.co.ke>.

- [7] *Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI*. Mulianga, B, et al., et al. 2013, Remote Sensing, pp. 2184-2199.
- [8] Box, G. E.P. and Jenkins, G. M. *Time Series Analysis: Forecasting and Control*, 2nd ed. San Francisco: Holden-Day, 1976.
- [9] *An Application of ARIMA model to real estate prices in Hong Kong*. Tse, R.Y. 1997, Journal of Property Finance, pp. 152-163.
- [10] *Modeling Multiple Time Series with Applications*. Tiao, G. and Box, G. E. P. 1981, Journal of the American Statistical Applications, pp. 802-816.
- [11] *Forecasting: Methods and applications*. Madrikadis, S., Wheelright, S. C. and Hyndman, R. J. 1998, New York: Wiley & Sons.
- [12] *Box-Jenkins modeling*. Hyndman, R. J. 2001.
- [13] Hyndman, J.R. Time Series Components. *Otexts*. [Online] 2017. [Cited: July 15, 2017.] <https://www.otexts.org/fpp/6/1>.
- [14] *ARIMA and SARIMA models*. Gerolimetto, M. 2010.
- [15] Pennsylvania State University. Stat510. *Applied Time Series Analysis*. [Online] 2017. [Cited: 07 15, 2017.] <https://onlinecourses.science.psu.edu/stat510/node/67>.
- [16] *Forecasting Precipitation Using SARIMA model: A case study of Mt. Kenya Region*. Kibunja, H. W., et al., et al. 2014, Mathematical Theory and Modeling.
- [17] *Time Series Modeling of tourist accommodation demands in Kenya*. Otieno, G., Mung'atu, J. and Orwa, G. 2014, Mathematics Theory and Modeling.
- [18] *Forecasting inflation rates in Kenya using SARIMA model*. Gikungu, S. W., Waititu, A. G. and Kihoro, J. M. 2015, American Journal of Theoretical and Applied Statistics, pp. 15-18.
- [19] *Modeling Inflation Rates in Libaria: SARIMA approach*. Fannoh, R., Orwa, G. O. and Mung'atu, G. J. 2012, International Journal of Science and Research, pp. 1360-1367.
- [20] *An Application of Time Series ARIMA forecasting model for predicting sugarcane production in India*. Kumar, M. and Anand, M. 2014, Studies in Business Research, pp. 81-94.
- [21] *Application of Time Series Model for Predicting Future Adoption of Sugarcane Variety: KEN 83-737*. Ong'ala, J. O. and Mwanga, D. M. 2015, Scholars Journal of Physics, Mathematics and Statistics, pp. 196-204.
- [22] *zoo: S3 infrastructure for regular and irregular time series*. Zeileis, A. and Grothendieck, G. arXiv preprint math/0505527, 2005.
- [23] *STL: A Seasonal-Trend Decomposition Procedure based on Loess*. Cleveland, R. B., et al., et al. 1990, Journal of Official Statistics, pp. 3-73.
- [24] *A New Look at the Statistical Model Identification*. Akaike, H. 1974, IEEE Transactions on Automatic Control, pp. 716-723.
- [25] Box, G. E.P. and Jenkins, G. M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.
- [26] *Parsimony, Model Adequacy and Periodic Correlation in Time Series Forecasting*. McLeod, A. I. s.l.: The International Statistical Review, 1993.
- [27] *Time Series Analysis: Forecasting and Control*. Box, G. E., Jenkins, G. M. and Reinsel, G. 1994.
- [28] Coghlan, A. *A Little Book of R for Time Series: Release 0.2*. 2017.
- [29] Hyndman, Rob J. ARIMA Modeling in R; How does auto.arima () work? *Otexts*. [Online] 2017. [Cited: April 28, 2017.] <https://www.otexts.org/fpp/8/7>.
- [30] Wawire, N. W., et al., et al. *Technology Adoption Study in the Kenya Sugar Industry*. Kisumu: Kenya Sugar Research Foundation, 2006.
- [31] AFFA - Sugar Directorate. *Yearbook of Statistics*. Nairobi: Government printers, 2015.
- [32] *Automatic time series forecasting: The forecast package for R*. Hyndman, R. J. and Khandakar, Y. 3, s.l.: Journal of Statistical Software, 2008.
- [33] Kenya Agricultural and Livestock Research Organization [KALRO]. Historical Background. *Sugar Research Institute*. [Online] 2016. [Cited: Nov 15, 2016.] www.kalro.org/Sugar_Research_Institute.