

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

Maths 2020; 5(6): 01-06

© 2020 Stats & Maths

www.mathsjournal.com

Received: 12-08-2020

Accepted: 07-10-2020

Cornelius Kipkeny Koech

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

Joel Cheruiyot Chelule

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

Ayubu Anapapa

University of Eldoret,
Department of Mathematics and
Computer Science, Eldoret,
Kenya

Herbert Imboga

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

Corresponding Author:

Cornelius Kipkeny Koech

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

Time series analysis of Nandi county government revenue using seasonal autoregressive integrated moving average (SARIMA) model

**Cornelius Kipkeny Koech, Joel Cheruiyot Chelule, Ayubu Anapapa and
Herbert Imboga**

DOI: <https://doi.org/10.22271/math.2020.v5.i6a.599>

Abstract

In Kenya, there are two levels of government namely the National Government (NG) and County Governments (CGs). NG's revenue is mainly from taxation, borrowing, and grants among other sources while CGs rely on allocation from NG as well as Own Source Revenue (OSR). CG of Nandi has been using basic descriptive statistics to analyze OSR data such as line and bar graphs with minimal forecasting capabilities. This study focuses on the Nandi county revenue analysis from 2013/14 financial year (FY) to the 2018/19 FY. The analysis showed that the Nandi county revenue department collected an average monthly revenue amounting to Ksh. 19.29 million (Ksh. 231.5 million annually). This study performed a time series analysis on the revenues using Seasonal Auto-Regressive Integrated Moving Average (SARIMA). The best SARIMA model SARIMA(0,0,0)(1,0,0)[12] chosen proved to fit well in the data. The study also projected the revenue of the CG of Nandi going into the future. The average amount of monthly revenues forecasted is Ksh. 34.97 million. This means that the county government of Nandi has a potential of raising an average of Ksh. 34.97 million monthly, Ksh. 104.91 million quarterly and Ksh. 419.64 million annually.

Keywords: Time series analysis, nandi county, county governments, SARIMA, revenue

1. Introduction

The constitution of Kenya requires that the NG allocates the CGs a minimum of fifteen percent (15%) of the income from the last audited books of accounts. The amount allocated alone is not enough for CGs to finance their budgets and thus the need for OSR. The CG of Nandi has been raising its OSR through a number of sources referred to as Revenue Streams. They include business permits, liquor licensing, plot rents, land rates, house and stall rents, parking fees, market fees, cess, slaughter fees. The revenues are collected daily, monthly and annually. This study analyzed the CG of Nandi OSR from 2014/15 to 2019/2019 FY and forecasted the amounts to be collected in four years to come.

1.2 Statement of the Problem

County Government of Nandi has been using simple descriptive statistical techniques such as charts, bar, and line graphs to represent data on revenue collections. The techniques highlighted above have not been able to inform the CG of the seasonal trends associated with the collection of revenues over the years, neither does it inform on the amount of revenues likely to be collected in the future. This study, therefore, used a time series model known as Seasonal Autoregressive Integrated Moving Average (SARIMA) model to analyze the County Government of Nandi revenue, came up with a model that projected the amount of revenues to be collected in the future. This eliminates the problem of planning for the unknown and an unpredictable future.

1.3 Objectives of the Study

1.3.1 General Objective

To Perform Time Series Analysis of Nandi County Government Revenue Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Model.

1.3.2 Specific Objectives

1. To analyze patterns of Nandi County revenue for the period 2014/15 to 2018/19 FY.
2. To fit a SARIMA model for the Nandi county revenue data for the period 2014/15 to 2018/19 FY
3. To forecast, using the fitted model, Nandi county revenue to be collected in the next four years.

1.4 Limitations of the study

1. COVID-19 pandemic to some extent interfered with the study forecast timetable more significantly the 2nd and 3rd quarters of 2020.
2. Time factor in putting multiple pieces of information per revenue stream together with the aim of forecasting revenues per stream.

2. Literature Review

According to a research conducted by the National Treasury in 2018, OSR collections by counties are up to four times below the minimum potential, shining the spotlight on inefficiencies by the devolved units. In the financial year 2017/18, CGs targeted to raise Sh49.2 billion in OSR but only collected Sh32.5 billion, similar to collections in 2016/17. This was a massive Sh92.2 billion below the potential. Nevertheless, OSR realized in the financial year 2017/18 was better (66 percent) than in 2016/17 (56.4 percent), which had a higher target (Sh. 57.7 billion). Some counties are notably doing well in matters of OSR collections. A good example is Nakuru County that in 2017/18 FY collected Ksh. 2.28 Billion up from Ksh. 1.96 Billion in 2016/2017 FY. The CG of Nandi OSR target for 2019/2020 FY stands at KShs 376.8 million. If achieved, this will be a significant percentage of the NG allocation.

Makananisa (2015) [6] modeled and forecasted Annual Tax Revenue of the South African Taxes Using Time Series Holt-Winters and ARIMA/SARIMA Models. Makananisa noted that the models performed well in analyzing and forecasting South African taxes which were in the categories of Personal income tax, Value Added Tax, Corporate Income Tax, and Total Income Revenues. Both the Holt winters model and SARIMA Models performed well against personal income tax and value-added tax data. However, he noted that the Holt-Winters model outperformed the SARIMA model for the volatile Corporate Income Tax data, and for Total Tax Revenue, the SARIMA model out-performed the Holt-Winters model. This clearly shows that the type of model to be used greatly relies on the type and nature of data in consideration.

Braimoh *et al.* (2018) used seasonal and Bilinear Autoregressive Integrated Moving Average (BARIMA) models in the modeling Nigeria's Gross Domestic Product (GDP). They used the two models which proved positive after analyzing the results. They were able to forecast the GDP of Nigeria. However, they noted that the BARIMA model lacked the seasonal part of the SARIMA model because the bilinear time series model is the aggregation of the linear and non-linear part of the autoregressive and moving average process and thus outperformed by the SARIMA model. They also noted that the use of Akaike Information Criterion to compare the performances of the two models revealed better performance of BARIMA model than SARIMA model. This shows that the BARIMA model fits well with financial and economic data only that it lacks the seasonal component.

Adanacioglu *et al.* (2012) [1] used a SARIMA model to analyze tomato prices at wholesale level in Turkey. Tomato

prices for the period from 2000 to 2010 were analyzed. The results obtained from this study showed that the prices of tomatoes in Turkey did not show any trend towards an increase or a decrease, in other words, the prices exhibited a stationary structure. The forecasts predicted from the SARIMA (1,0,0) (1,1,1)12 model which was chosen to determine the course of the prices of the next 3 years showed that any significant changes would not occur in Real Tomato Prices by the end of 2014.

3. Research Methodology

3.1 Data Source

This study used the secondary data obtained from the CG of Nandi, Revenue department. The data majorly comprised the amounts of revenue collected monthly from 2014/15 to 2018/19 FY.

3.2 Method of Analysis

Time series analysis was used to model the CG of Nandi revenue collected for a period of four years and forecasted the amount projected to be collected in four years to come. The general form of SARIMA as proposed by Box and Jenkins (1976) [3] and Kendall and Ord(1990) was used. Data were analyzed using R-Software.

The ARIMA model is for non-seasonal non-stationary data. Box and Jenkins have generalized this model to deal with seasonality. Their proposed model is known as the Seasonal ARIMA (SARIMA) model. In this model seasonal differencing of appropriate order is used to remove non-stationarity from the series. Each of the autoregressive and moving average processes has both the seasonal and non-seasonal parts. However, the distribution of the correlogram is the determining factor for the choice of the class of ARIMA or SARIMA model. The reason for the use of SARIMA (p, d, q)(P, D, Q)12 in this research work is that the revenue data were recorded on monthly basis. The seasonality period is 12. The general form of seasonal model SARIMA(p, d, q) (P, D, Q)s is given by:

$$\Phi_p(B^S)\varphi(B)\nabla_s^D\nabla^dX_t = \Theta_q(B^S)\theta(B)W_t \tag{3.1}$$

Where $\{W_t\}$ is the nonstationary time series, $\{W_t\}$ is the usual Gaussian white noise process. s is the period of the time series. The polynomials $\varphi(B)$ and $\theta(B)$ of orders p and q represent the ordinary autoregressive and moving average components. The seasonal autoregressive and moving average components are $\Phi_p(B^S)$ and $\Theta_q(B^S)$, where P and Q are their orders. ∇^d and ∇_s^D are ordinary and seasonal difference components. B is the backshift operator. The expressions can be shown as follows;

$$\varphi(B) = 1 - \varphi_1B - \varphi_2B^2 - \dots - \varphi_pB^p \tag{3.2}$$

$$\Phi_p(B^S) = 1 - \Phi_1B^S - \Phi_2B^{2S} - \dots - \Phi_pB^{pS} \tag{3.3}$$

$$\theta(B) = 1 + \theta_1B + \theta_2B^2 \dots + \theta_qB^q \tag{3.4}$$

$$\Theta_q(B^S) = 1 + \Theta_1B^S + \Theta_2B^{2S} + \dots + \Theta_qB^{qS} \tag{3.5}$$

$$\nabla^d = (1 - B)^d \tag{3.6}$$

$$\nabla_s^D = (1 - B^S)^D \tag{3.7}$$

$$B^kX_t = X_{t-k} \tag{3.8}$$

This study focused on the amount of revenue collected monthly. The seasonal period of the series $s = 12$. Equation (3.1) may therefore be written as:

$$\Phi_P(B^{12})\varphi(B)\nabla_S^D\nabla^dX_t = \Theta_Q(B^{12})\theta(B)W_t \quad (3.9)$$

The general forecasting equation is given by:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q} \quad (3.10)$$

The mathematical formulation of a SARIMA $(p,d,q) \times (P,D,Q)_s$ model in terms of lag polynomials is given below;

$$\Phi_P(Ls)\phi_P(L)(1-L)^d(1-L^s)^dY_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t \quad (3.11)$$

$$\Phi_P(Ls)\phi_P(L)Z_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t \quad (3.12)$$

Here Z_t is the seasonally differenced series.

Forecasting comprises four steps i.e Model Identification, Parameter estimation, diagnostic checking, and forecasting (Gujarati, 2004). This process is iterative in that it will have to be repeated until a desirable model of the data is identified.

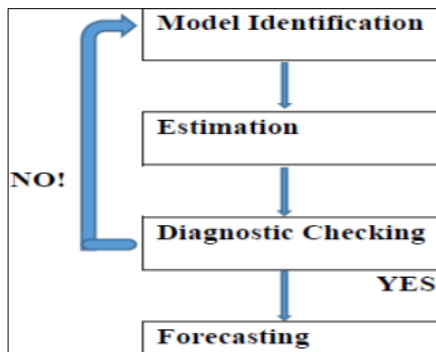


Fig 1: The Box-Jenkins Procedure (Box and Jenkins, 1976, p. 19)

4. Data Analysis, Presentation and Interpretation

4.1 Descriptive Data Analysis

The monthly revenue data in millions were presented as follows;

Table 1: Nandi County Monthly Revenues

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2014							19.9	10.1	13.5	14.0	17.8	9.0
2015	17.8	21.1	27.8	24.6	11.6	21.5	17.9	12.1	13.2	10.2	29.0	17.4
2016	11.3	11.8	26.5	20.8	32.3	17.0	20.4	69.9	11.3	10.7	13.5	8.0
2017	10.1	12.6	20.1	7.9	11.3	49.0	6.5	5.1	8.5	11.3	14.9	11.3
2018	19.3	17.5	41.0	16.8	15.1	30.7	13.9	16.1	11.4	23.8	10.6	7.7
2019	16.6	19.9	25.9	73.8	32.0	34.7						

The summary of the data is given as follows;

Table 2: Descriptive Statistics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Monthly	5.10	11.30	16.35	19.29	21.20	73.80
Quarterly	20.10	41.92	53.10	57.87	67.08	140.50
Annual	197.9	208.7	219.5	231.4	244.7	286.3

The summary shows that the average amount of revenue that CG of Nandi collected per month is Ksh. 19.29 million. The least and the maximum amounts ever collected is Ksh. 5.10 million and Ksh. 73.80 million. The average amount collected quarterly was Ksh. 57.87 million. The least and the highest

amount collected quarterly was Ksh. 20.10 million and 140.50 million respectively. Yearly, the average amount collected was Ksh. 231.4 million. The least and the maximum amount ever collected was Ksh. 197.9 and 286.3 million respectively.

4.2 Time Plot for Nandi County Revenues

Time plot for monthly, quarterly, and yearly amounts of revenues collected is given as follows;

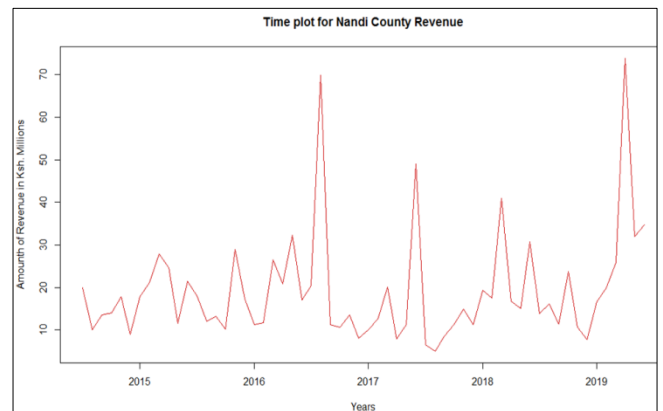


Fig 2: Time Plot for Nandi County Monthly Revenues

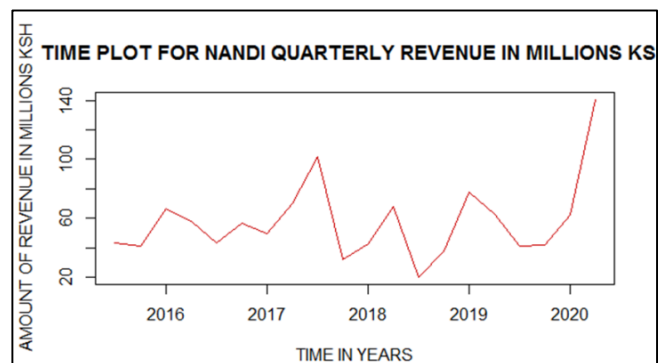


Fig 3: Time Plot for Nandi County Quarterly Revenues

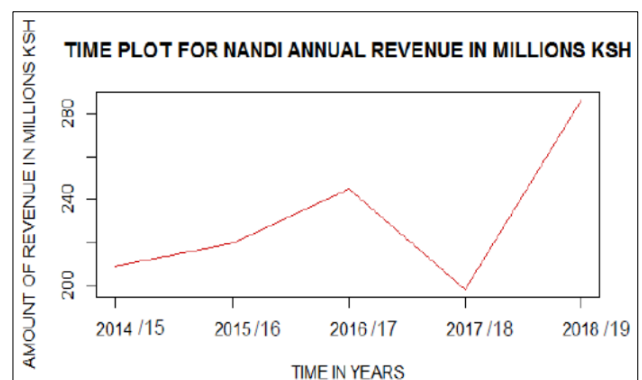


Fig 4: Time Plot for Nandi County Annual Revenues

The time plot for monthly revenues indicates that the revenues collected increased and decreased through different months of the year. This shows that revenue collection follows a particular trend. The trend is not different from the quarterly revenues trend. The amounts of revenues collected increased and declined at different quarters of the respective years. The yearly revenues are seen to rise steadily from 2014/15 FY to 2016/17 FY. The revenues declined sharply in the 2017/18 FY before rising again steadily in the 2018/19 FY.

4.3 Decomposition of the Time Series

The monthly revenue time series were decomposed in an attempt to check the existence of a seasonal trend.

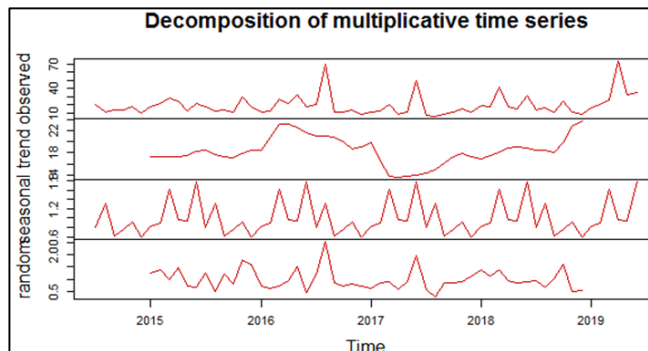


Fig 5: Decomposed Plot for Monthly Time Series

The figure above shows how the monthly time series were decomposed into random, seasonal, and trend components. The trend component shows the behavior of the revenues as they were collected, whether they were rising or declining. In this particular case, the revenues rose from 2014/15 FY to 2016/17 FY, declined in 2017/18 FY, and rose steadily thereafter. The random components represent all the components that did not fit in the trend and the seasonal part. The seasonal components show the periodic behavior of the revenues within a yearly cycle. This confirms the significance of choosing the SARIMA model.

4.4 Box Plot

The Box-Plot was plotted to study the seasonal pattern for revenue collection in Nandi County.

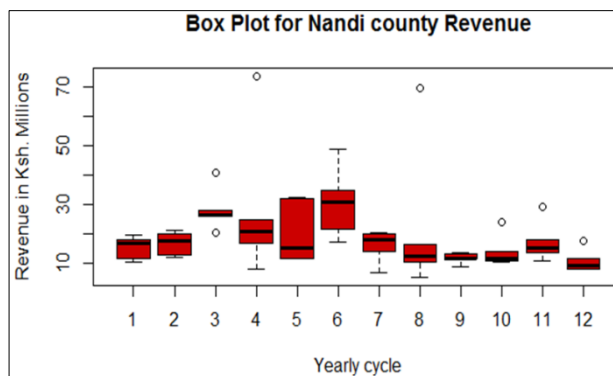


Fig 6: Box Plot for Nandi County Monthly Revenue

The Box plot clearly shows the yearly seasonal behavior of revenue collection in the respective months. In general, the months of May and June recorded the highest level of revenues with June recording the best. The months of September and December recorded the least amounts of revenue in general. The seasonal trend shows that the amounts of revenue collection increased steadily from January to March. It decreases slightly in April and rises sharply in May and June. From July, the revenues decrease steadily to its worst in September, rises slightly in October and November, and drops again in December. This cycle repeats annually and may be used to predict the expected future seasonal pattern.

4.5 Test of Hypothesis

It was established whether the mean amounts of revenues collected were the same for all the months across the five years.

H_0 : No difference in the mean amount of monthly revenue collected

H_1 : There is a significant difference in the mean amount of monthly revenue collected.

Table 3: ANOVA Table for Nandi County Monthly revenues

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
data\$Revenue_Monthly	6	372.1041	22326.246	1.984	0.0292
Residuals	189	1457927	804		

Since p-value <0.05, we, therefore, reject the null hypothesis. There is a statistically significant difference in the mean amount of revenue collected per month.

4.6 Trend Differencing for the Amount of Monthly Revenue Collected

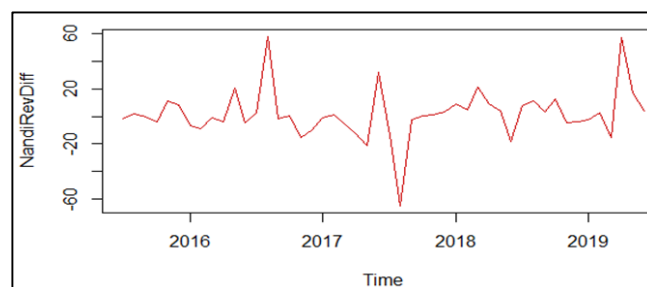


Fig 7: Time Plot for First Differenced Nandi County Monthly Revenue Data

Trend differencing was performed using the first differencing method to remove the trend component in the original data. The observations of the mean move about the mean and its variability is almost constant. The amounts of revenue collected monthly now look stable and the data can be said to be stationary.

4.7 Choosing the Best Model

4.7.1 Model Identification

The first figure shows the autocorrelation plot the Nandi county monthly revenue of the first differencing at various lags. The second figure shows the Partial autocorrelation plot of the monthly revenue also at different lags.

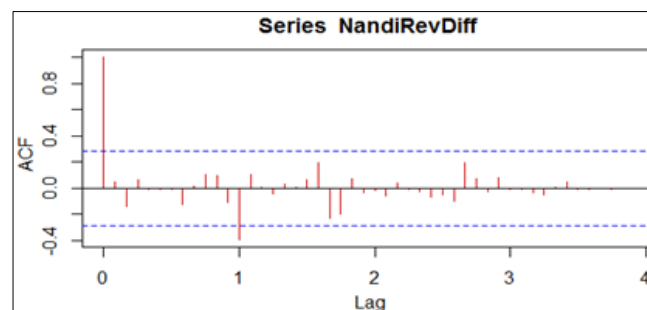


Fig 8: ACF for Nandi County Monthly Revenues Data

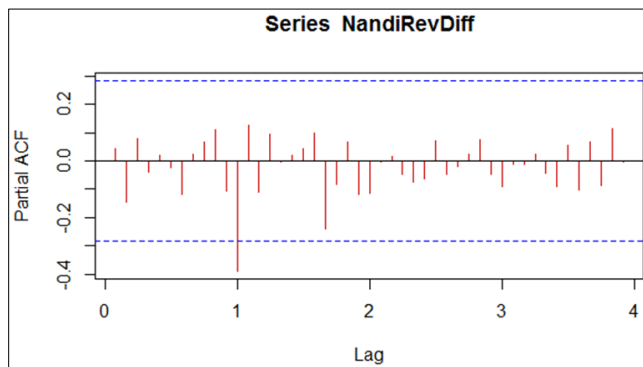


Fig 9: PACF for Nandi County Monthly Revenue Data

From the plots, the only significant autocorrelation is at lag 1, showing an MA (1) pattern. Also, partial autocorrelations are significant, showing an AR (1). Several models were suggested but auto.arima function was used to choose the best SARIMA model with respect to the least AIC value. The table below shows the various models which were identified before choosing the most suitable model.

Table 4: AIC Table for Various SARIMA/ARIMA Models

MODEL	AIC Value
ARIMA(2,0,2)(1,0,1) [12]	Inf
ARIMA(0,0,0)	417.8147
ARIMA(1,0,0)(1,0,0) [12]	412.8947
ARIMA(0,0,1)(0,0,1) [12]	Inf
ARIMA(0,0,0)	416.2003
ARIMA(1,0,0)	419.7187
ARIMA(1,0,0)(1,0,1)[12]	Inf
ARIMA(1,0,0)(0,0,1)[12]	Inf
ARIMA(0,0,0)(1,0,0)[12]	411.1182
ARIMA(0,0,0)(1,0,1)[12]	Inf
ARIMA(0,0,0)(0,0,1)[12]	Inf
ARIMA(0,0,1)(1,0,0)[12]	412.7845
ARIMA(1,0,1)(1,0,0)[12]	413.371
ARIMA(0,0,0)(1,0,0)[12]	409.5628
ARIMA(0,0,0)(1,0,1)[12]	Inf
ARIMA(0,0,0)(0,0,1)[12]	Inf
ARIMA(1,0,0)(1,0,0)[12]	411.2737
ARIMA(0,0,1)(1,0,0)[12]	411.1368
ARIMA(1,0,1)(1,0,0)[12]	411.7661

The best model was chosen as SARIMA (0,0,0)(1,0,0)[12] with zero mean. The model had the least AIC value of 409.56.

4.7.2 Parameter Estimation

After identification of the best model, it was evidenced that SARIMA (0,0,0)(1,0,0) [12] was the best model. The estimated parameters and model fits are shown below.

Table 5: Parameter Estimates

Coefficients	Estimates	Standard Errors
Sar1	-0.4182	0.1288

Table 6: Model Fit

	Values
R ²	266.2
Log likelihood	-202.78
AIC	409.56
AICc	409.83
BIC	413.31

4.7.3 Model Diagnostic Stage

4.5.3.1 Ljung-Box Test

Ljung-Box Test was run with the following hypothesis.

H₀: The SARIMA model is suitable for the data
 H₁: The SARIMA model is unsuitable for the data.
 The results were as follows;

Table 7: Ljung-Box Test

	Chis-value	Df	p-value
Monthly Revenue	2.4151	5	0.7892

We fail to reject H₀. We conclude that the model was suitable for the data.

4.5.3.2 Train and Test Method

The Train and Test method was also used to test the suitability of the model. A new model was fit while omitting the data for the last six months of the last financial year.

Table 8: Forecast Values for Model Diagnostic

	Point Forecast	Lo 95	Hi 95
Jan 2019	0.2241967	-30.80892	31.25732
Feb 2019	-0.3169471	-31.35007	30.71617
Mar 2019	1.0800874	-29.95303	32.11321
Apr 2019	-1.5538472	-32.58697	29.47927
May 2019	-0.3114252	31.34454	30.72169
Jun 2019	0.6438592	-30.38926	31.67698

The same model was used to forecast the amounts of revenues for the final. These projected amounts were then compared with the original data. The original data for the last six months is as follows;

Table 1: Original Data for Model Diagnostic

	Jan	Feb	Mar	Apr	May	Jun
2019	19.9	16.6	25.9	73.8	32.0	34.7

When the original data is compared with the projected data, there is a slight variance. The projected values move around the mean. The mean of the original and the forecasted data are 33.8 and 31.0 respectively. The variance of the projected data from the original data is -2.8 which is a small difference. This leads to the conclusion that the SARIMA model was suitable for the data.

4.8 Forecasting the Amounts of Revenues

The monthly amounts of Nandi County Government revenues were projected using the fitted model SARIMA(0,0,0)(1,0,0)[12].

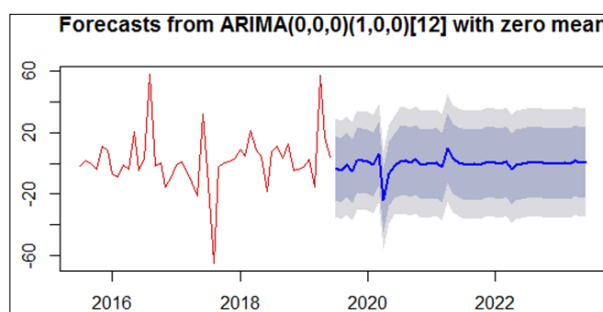


Fig 10: Forecast Plot for Nandi County Monthly Revenues

The forecasts show the amounts of revenues collected by the Nandi County government will continue displaying the seasonality which was witnessed in the data. An overview of the forecasted amounts in millions are given in the table below;

Table 10: Forecast Values for Nandi County Revenues

Year	Forecasted revenues in ksh. Millions			
	April	June	September	December
2020	8.1425984953	30.3068260088	35.170534538	34.0337796868
2021	44.631832611	35.3629084364	34.899556799	35.3749396888
2022	30.942908439	34.8191073869	35.278158453	35.0793566701
2023	36.932800990	35.3118018324	-	-

The average amounts of monthly forecasted revenues for the projected four years is Ksh. 34.97 million per month. The least forecasted amount is Ksh. 8.14 million and the highest possible amount is Ksh. 44.63 million.

5. Summary, Conclusions and Recommendations

The analysis showed that the Nandi county revenue department collects an average monthly revenue amounting to Ksh. 19.29 million. The revenues collected followed a yearly seasonal trend. The revenues increases from January to its peak in June. Then it declines steadily to its worst in September, rises slightly, and falls again in December. This is a cycle that revenues collected follow annually.

The SARIMA model SARIMA (0,0,0)(1,0,0)[12] used to forecast the Nandi County OSR proved to fit well in the data. This is attributed to the fact that the data (amounts of revenues collected monthly) follow a seasonal trend. The amounts forecasted also followed the same seasonal trend as the original data. The average amount of monthly revenues forecasted is Ksh. 34.97 million. This means that the county government of Nandi has a potential of raising an average of Ksh. 34.97 million monthly, Ksh. 104.91 million quarterly and Ksh. 419.64 million annually.

The SARIMA model is recommended in forecasting any kind of income for government or private institutions with seasonal trends. The study highly recommends to all the counties in Kenya to consider applying the time series SARIMA model in forecasting their OSR. Researchers are also invited to extend their studies to the factors causing seasonal trends in revenue collections. Further research should also consider using other models such as VAR and comparing its efficiency with SARIMA.

6. References

1. Adanacioglu H, Yercan M. An analysis of tomato prices at wholesale level in Turkey: an application of SARIMA model. *Custos e@ gronegocio on line* 2012;8(4):52-75.
2. Botric V, Vizek M. The Israeli Tax System: VAT Revenue Forecasting in Israel Israel. Ministry of Finance, State Revenue Administration, 2008.
3. Box GE, Jenkins GM, Reinsel GC, Ljung GM. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
4. Chang X, Gao M, Wang Y, Hou X. Seasonal autoregressive integrated moving average model for precipitation time series. *Journal of Mathematics & Statistics* 2012;8(4).
5. Koirala TP. Government revenue forecasting in Nepal. *NBR Economic Review* 2011;4(2).
6. Makananisa MP. Forecasting annual tax revenue of the South African taxes using time series Holt-Winters and ARIMA/SARIMA Models (Doctoral dissertation), 2015.
7. Nyoni T. Modeling and Forecasting Naira/USD Exchange Rate in Nigeria: a Box-Jenkins ARIMA approach, 2018.
8. Tebbs M. STAT 520 Forecasting and Time Series (University of South Carolina Department of Statistics. Spring, 2010).