Vol. 6, Issue: 2022

Article

ISSN: 2312-2668

International Journal of Information Sciences Management (IJIMS)

Forecasting Domestic Tax Revenues in Kenya Using Sarima & Holt-Winters Methods

Nelson Kirimi Micheni

Edwin Benson Atitwa

Patrick Kimani

Forecasting Domestic Tax Revenues in Kenya Using Sarima & Holt-Winters Methods

Nelson Kirimi Micheni
Postgraduate Student, Department of Mathematics and Statistics, University of Embu, Embu, Kenya.
kirimi.nelson1@gmail.com

Edwin Benson Atitwa
Senior Lecturer, Department of Mathematics and Statistics, University of Embu, Embu, Kenya.
atitwa.benson@embuni.ac.ke

Patrick Kimani
Senior Lecturer, Department of Mathematics and Statistics, University of Embu, Embu, Kenya.
patkim32@gmail.com

Abstract

Forecasting of tax revenues is an important factor in fiscal planning. Underestimation and overestimation of tax revenues lead to unstable economies. The study sought to find suitable Holt-Winters and SARIMA models that could be used to forecast Domestic tax revenues in Kenya. The study utilized the Domestic tax revenues collected in Kenya between Jan 2015 to December 2020. Analysis of data was done using R-software (version 4.1.0) where SARIMA and Holt-Winters time series forecasting methods were applied to the revenue data. SARIMA(0,1,1)(0,1,1)[12] model was found to be the best model since it had the least Bayesian Information Criterion (BIC=1236.49) and least forecasting errors (MAPE=6.9, MASE=0.37).The multiplicative Holt-Winters method was slightly superior to the additive method due to its lower error (MAPE=7.43). The study recommends the use of the two methods to forecast Domestic taxes in Kenya be used to capture the Domestic taxes revenues with high precision.

Keywords: Kenya Revenue Authority(KRA); Domestic Taxes Revenues; Holt-Winters (HW); Seasonal Autoregressive Integrated Moving Average (SARIMA).

Introduction

Taxation is an instrument used by institutions and governments to collect revenues to finance government expenditures. This is vital in the provision of funds for development projects and infrastructure such as education and healthcare to citizens of the country (Harelimana, 2018). This has therefore necessitated the need for governments to collect revenue through taxation as well as sealing loopholes to ensure each individual or organization gives their share to the economy. Domestic taxes are direct and indirect taxes that are levied on gains accrued in a country. They are enforced under the laws of the country where one is a resident. They include Pay As You Earn, Value Added taxes, Income taxes, rental Income, and Capital gains taxes among others.

The Kenyan budget has recorded huge growth in recent years. For instance, it grew from Kshs 1.45 trillion in FY 2012/2013 to Kshs. 2.3 trillion in FY 2016/2017 and Ksh. 3.02 trillion in FY 2019/2020. According to ICPAK (2016), this growth of budget estimates has occasioned discussions around the overall capability of budget absorption capacity at the county and national government together with its ability to raise the revenue to fund the expenditure plans considering the increasing national debt levels. To aid this, tax analysis and revenue forecasting are important tools that can be utilized to provide better budget planning and monitoring processes hence leading to better decision-making by the government. It also contributes to the creation of measures to enhance tax revenues, promote government investments, improve tax efficiency and equity which leads to increased economic growth.

In an article by Amadala (2021), the Kenya national treasury was briefed on the need to have a stronger revenue forecasting capacity. According to the Institute of Economic Affairs (IEA Kenya), this would prevent over-ambitious budget plans which would drive the country away from constant borrowing. The revenue forecasts rose from 5.5% to above 19% between 2014 to 2020. This increase in government revenue forecast was criticized by IEA and hence the need to improve the overall predictability of public funds for better budgeting.

The revenue forecasting task lies with the ministry of finance in Kenya. Numerous tax reforms have been carried out by the Kenyan government, aimed at increasing tax revenues to fund the government budget, without excessive borrowing. This is fueled by the urge for internal revenue mobilization efforts by countries to meet their public expenditures (Signé, 2016). Despite these efforts, corruption, fraud, and tax evasion schemes have been noted to be the major setbacks in revenue collection. Lilian (2015), established that Ksh.639 billion were lost annually due to tax evasion schemes by multinational companies.

Background Studies

The revenue forecasting techniques can be classified into qualitative and quantitative methods (Jenkins *et al.,*2000). The qualitative methods are also known as judgmental forecasts such as consensus forecasting. These methods are prone to bias due to human judgment. This method is widely used where data does not exist or is not representative. Studies on forecasting accuracy reveal a reduced forecast bias arising from consensus forecasts. However, this method is time-consuming hence leading to lag that occurs during forecast preparation and forecast use.

Quantitative methods are commonly used currently due to the availability of data from improved data capturing and advancements in technology. According to Makridakis *et al*.(2008), historical occurrences are assumed to continue into the future. However, there could be unexpected changes in the future since they are not constant. This, therefore, explains the importance of performing short-term forecasts compared to long-term forecasts.

Time-series methods assume that the predictor variables are incorporated in the historical occurrences that one wishes to investigate. As the actual realization of the study variable is made the model is thus revised to ensure a better future outcome. Consequently, if incorrect historical data was used, the future occurrence is expected to follow the wrong path.

According to Pinyck and Rubinfeld (1998), the ARIMA, SARIMA, and the Holt-Winters methods provide better short-term forecasts. Unfortunately, they do not explain the components that drive the variables of interest.

Ergüven *et al.* (2015) used the SARIMA and the structural time series model and concluded that the SARIMA model produces more accurate short-term forecasts than the structural time series model. Similarly, Chang and Liao (2010) forecasted the monthly outbound tourism departures of three significant destinations from Taiwan using SARIMA models.

Saayman and Saayman (2008) study used the standard ARIMA, Holt-Winters exponential smoothing, and SARIMA models to predict the monthly tourist arrivals in South Africa. The study revealed that SARIMA resulted in the best model compared to the rest.

Otu *et al.*(2014) predicted Nigeria's monthly inflation using the SARIMA model. The study involved 120 observations taken between November 2003 to October 2013. The study found SARIMA(1,1,1)(0,0,1)12 to be the most appropriate model to predict the inflation rate in the following first quarter of 2014.

Susan *et al.* (2015) used the SARIMA model to predict the inflation rate in Kenya. The SARIMA (0,1,0)(0,0,1)4 model was found to be the most appropriate model to predict the inflation rate in Kenya using the sample data. The predictability ability was checked using RMSE and MAE and found the model appropriate. This study also suggested policies that could be applied in Kenya to ensure a single-digit inflation rate was achieved.

The local authority had difficulties predicting their future revenues to assist in the budget-making process. According to Pelinescu *et al.*(2010), the Holt-Winters multiplicative and additive models were used to predict the total revenue of the local authority. This improved management and decision-making for the local income and expenditures. The Holt-Winters model was recommended by the study to forecast the multi-annual budgets.

Rahman *et al.* (2016), used monthly revenue data from the Bangladesh Bridge Authority between July 1998 to July 2016 to forecast monthly revenue using the best additive and multiplicative seasonal model of the Holt-winters method. The additive Holt-Winters method was found accurate and used to predict monthly revenue to January 2021.

The study seeks to establish suitable SARIMA and Holt-Winters time series models that can be used to forecast Domestic taxes revenues in Kenya. The models can describe time series data with non-stationary traits in different seasons hence ideal for tax revenue studies.

Research Design

Data source

The study used secondary data from the Kenya Revenue Authority on the monthly revenues collected from various tax heads and the total Domestic taxes. The data is based on facts and figures of monthly revenue data covering 72 months from January 2015 to December 2020.

SARIMA Model

Seasonal Method Autoregressive Integrated Moving Average (SARIMA) method is a time series forecasting method for stochastic data that has seasonal occurrences. It is denoted as ARIMA $(p, d, q)\ (P, D, Q)\ s$

where:

(p, d, q)-represents the non-seasonal part of the model

(P, D, Q)s-represents the seasonal part of the model

The SARIMA model will be expressed as follows;

$$\Phi_p(B^s)\varphi(B)\ \nabla_s^D \nabla^d Z_t = \Theta_Q\ (B^s)\theta(B)a_t \qquad (1)$$

where;

a_t is the nonstationary time series (Gaussian white noise)

s is the number of periods per season
$\varphi(B)$ and $\theta(B)$ autoregressive and moving average polynomials of orders p and q.
$\Phi_p(B^s)$ and $\Theta_Q(B^s)$ seasonal autoregressive and moving average components with orders P and Q
∇^d and ∇_s^D represents the ordinary and seasonal difference components and B is the backshift operator.
$Z_t = (1-B^d)(1-B^s)^D m_t$ represents the product of seasonal differencing D
Furthermore, Chang, *et.al.*(2012) breaks down the SARIMA model as follows:

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3 - \cdots - \varphi_p B^p$ -Non-seasonal AR of order p (2)

$\Phi_p(B^s) = 1 - \Phi_1(B^s) - \Phi_2(B^{2s}) - \Phi_3(B^{3s}) - \cdots - \Phi_p(B^{ps})$-Seasonal AR of order P (3)

$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \cdots + \theta_p B^p$ - Non-seasonal MA of order q (4)

$\Theta_Q(B^s) = 1 + \Theta_1(B^s) + \Theta_2(B^{2s}) + \Theta_3(B^{3s}) + \cdots + \Theta_Q(B^{Qs})$–Seasonal MA order Q (5)

$\nabla^d = (1-B)^d$
$\nabla_s^D = (1-B^s)^D$

The study will focus on a 12-monthly revenue collection time series. The seasonal period is therefore 12 ($s = 12$). The SARIMA model will therefore be;

$\Phi_p(B^{12})\, \varphi(B) \nabla_{12}^D \nabla^d Z_t = \Theta_Q\,(B^{12})\theta(B) a_t$

(6)

Holt-Winters (HW) Model

Exponential smoothing is a procedure that revises a given prediction in light of more current experience realized. The Holt-Winters method is derived from exponential smoothing and comprises the trend-smoothing, seasonal smoothing, and overall smoothing.

The coefficients α, β, γ are considered the three smoothing parameters, and p is the number of observations per seasonal cycle. Holt-Winters is classified into additive and multiplicative methods (Hyndman *et al.,*2008).

In instances when the time series is believed to have a linear trend that has an additive seasonal pattern, the additive Holt-Winters method is ideal. It could be suitable for modeling some tax heads. However, due to changes in economic performance, tax trend usually changes in a multiplicative manner.

For the additive HW method, the estimate L_t will represent the series level, b_t the trend, S_t the seasonal component while, F_{t+m} will be the forecast for the m periods ahead and t is the index denoting period. The multiplicative model is represented in the following equations:

$F_{t+m} = (L_t + b_t m) S_{t-s+m}$ for $m = 1, \ldots, M$ (7)

where;

$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1})$
$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$
$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma) S_{t-s}$

And the additive Holt-Winters Model is represented as:

$F_{t+m} = L_t + b_t m + S_{t-s+m}$ for $m = 1, \ldots, M$ (8)

where;

$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$
$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$
$S_t = \gamma(Y_t - L_t) + (1-\gamma) S_{t-s}$

The appropriate Holt-Winters smoothing factors (alpha=α, beta=β, and gamma= γ) for the best fit model will be chosen on the least MSE, MAE, and MAPE values.

Accuracy Measurement

Since the main goal of time series modeling is forecasting, the predictive accuracy of forecasts is evaluated to identify models with the least errors (Makridakis *et al*. 1998).

Forecast error $e_{t+h} = Y_{t+h} - \hat{Y}_{t+h}\,; h - i \leq 0 \text{ and } h - i > 0 \text{ then } \varepsilon_{t+h-i} = 0$ (9)

The following models are commonly used in evaluating model accuracy.

$$\text{Mean absolute deviation (MAD)} = \frac{\sum_{i=1}^{n} |\,\varepsilon_i\,|}{n}$$

$$\text{Mean squared error (MSE)} = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n}$$

$$\text{Root mean squared error (RMSE)} = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n}}$$

$$\text{Percentage error (PE)} = \frac{Y_t - F_i}{Y_t} * 100$$

$$\text{Mean percentage error (MPE)} = \frac{\sum_{i=1}^{n} PE_i}{n}$$

$$\text{Mean absolute percentage error(MAPE)} = \frac{\sum_{i=1}^{n} |PE_i|}{n}$$

The Bayesian Information Criterion (BIC) and Akaike's Information Criterion (AIC) are the most common criterion used in choosing the best model.

They are expressed as;

AIC=-2ln(Likelihood)+2r and BIC=-2ln(Likelihood) +r ln(T).

The AIC value increases with increase in the model parameters (r) number and smallest for the best model.

Forecasting

In time series, forecasting involves the estimation of future events using current and past data. Forecasting was performed after the models passed the outlined diagnostic tests.

Results and Discussion

The trend of the total DTD tax revenue collected between January 2015 to December 2020 displays an increasing pattern that is non-stationary. The natural logarithm and differencing were used for data smoothening.

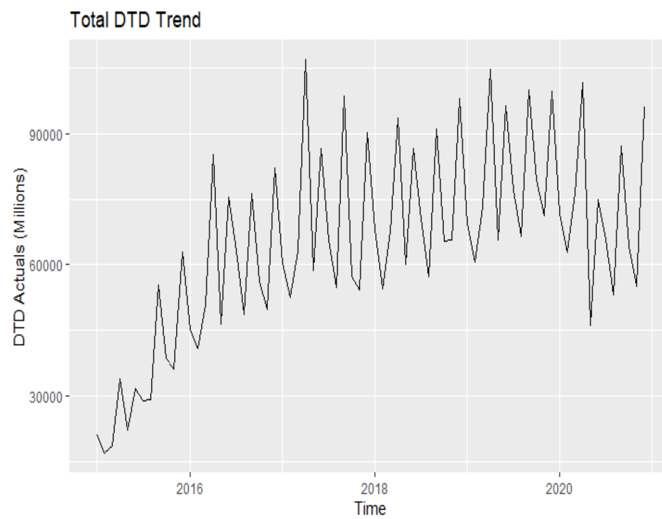

Figure 1*: Total DTD taxes from 2015 to 2021*

The revenue data was transformed at first natural logarithm to reduce the seasonality and variability. However, the revenue data was still not stationary hence the need to carry out a first differencing $(y' = y_t - y_{t-1})$. Through this, stationarity was achieved as seen in figure 2.

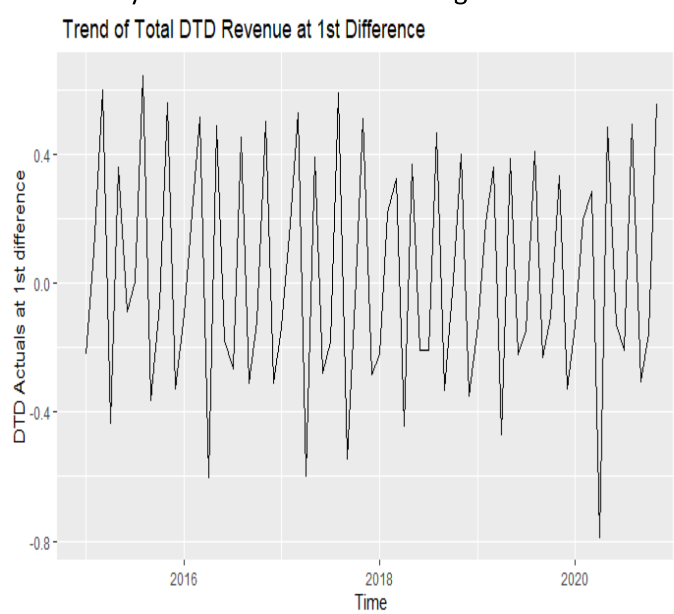

Figure 2: Natural Logarithm of DTD revenue data at First difference

The assessment of stationarity was also done through the Augmented Dickey-Fuller (ADF) and Phillip Peron (PP) tests where the null hypothesis is non-stationarity.

Vol. 6, Issue: 2022

Table 1

Stationarity tests p-values

Level	Augmented Dicky-Fuller (ADF)	Phillips-Peron (PP)	Rema
Time series at level	0.9027	0.003031 **	Both r station
Natural logarithm	0.0497	0.0000000632***	Both station
First Difference of natural logarithm	0.01	0.00000054 ***	Both station

From Table 1, the revenue data is stationary when transformed using a natural logarithm or differenced once by the significant results from the ADF and PP tests.

Diagnostic Checking

The ACF from the SARIMA and Holt-Winters residuals were found with the 95% confidence interval, to be close to the zero line with almost all spikes within the significant zone. This confirms that the residuals are independent. The histograms are fairly bell-shaped signifying normality.

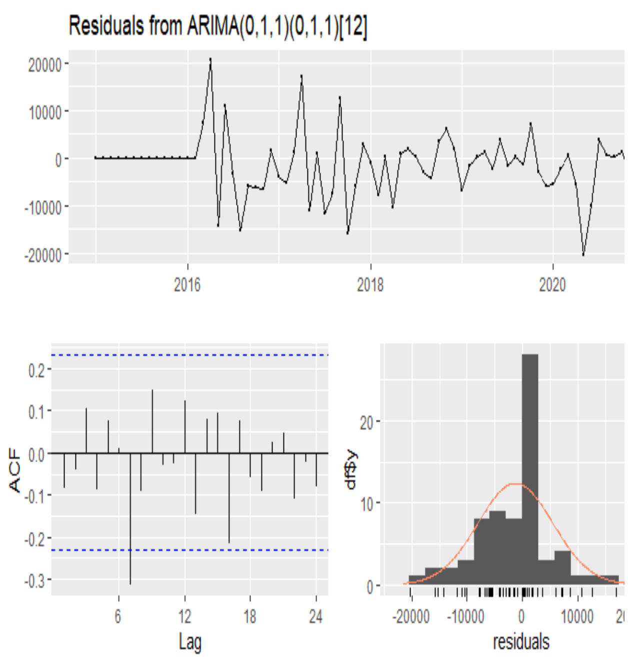

Figure 3: *SARIMA model residuals*

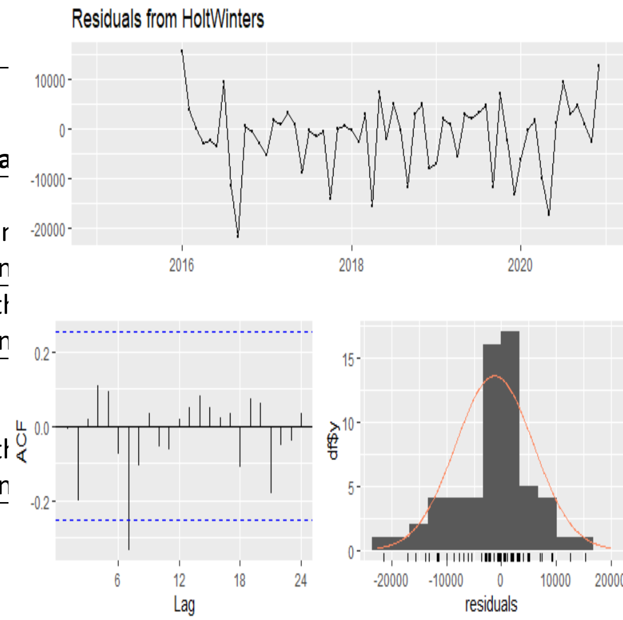

Figure 4: *Holt-Winters model residuals*

SARIMA Model

The DTD revenue collection was predicted for 12 months in 2021 using the best SARIMA model (Seasonal ARIMA (0,1,1) (0,1,1)[12]. It was selected since it produced the least AIC (1230.26), AICc (1230.70), and BIC (1236.49).

SARIMA Forecast

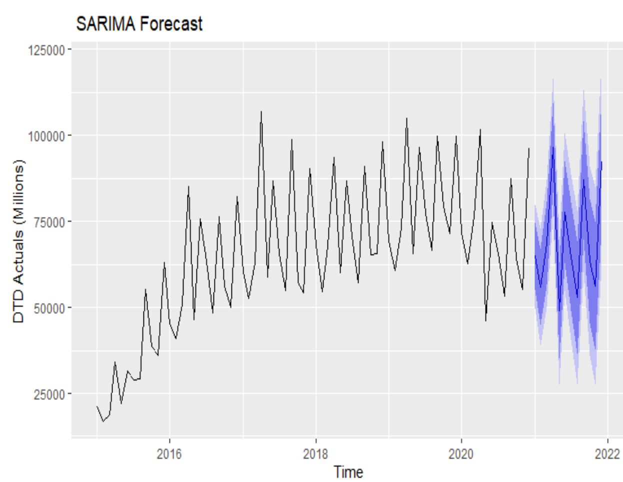

Figure 5: *SARIMA model prediction*

Holt-Winters (HW) Model

The model separates time series data into 3 components; trend, level value, and seasonal. The components are allocated weights of 0 to 1 to enable model fitting and prediction. The model is mainly used when data has a linear trend and seasonality.

Table 2

HW Constants and coefficients

		Multiplicative-Holt-Winters	Additiv Holt-Win
Smoothing constants	alpha(α)	0.6512	0.640
	beta(β)	0.0776	0.068
	gamma(γ)	1.0000	0.694
	Level Mean(a)	77578.0625	75985.0
	Growth/Trend(b)	203.1988	190.53
Seasonal smooth coefficients	s1	0.8906	-8260.2
	s2	0.7803	-16059.9
	s3	0.9332	-4205.6
	s4	1.3144	25222.9
	s5	0.7271	-17842.5
	s6	1.1872	12783.0
	s7	0.9581	-3494.3
	s8	0.7546	-15619.9
	s9	1.1983	17616.2
	s10	0.8761	-7686.8
	s11	0.7766	-14126.6
	s12	1.2376	18877.7

The fitted additive Holt-Winters models for total DTD revenues (Y_t), level mean estimate (L_t),trend estimate (b_t) and seasonal factor (S_t) are given as:

$$ln(Y_t) =(75985.04 - 190.54t) + s_t + e_t \quad (10)$$

$$b_t =0.0685(L_t - L_{t-1}) + (1 - 0.0685)b_{t-1}$$

$$S_t =0.6943+(Y_t - L_t) + (1 - 0.6943)\ S_{t-1}$$

Conversely, the resultant multiplicative Holt-Winters models can be fitted as shown below;

$$Y_t =(77578.06 - 203.20t) + sn_t + e_t \quad (11)$$

$$L_t = 0.6512(\frac{Y_t}{S_{t-s}}) +(1- 0.6512)(L_{t-1} + b_{t-1})$$

$$b_t =0.07756\ (L_t - L_{t-1}) + (1 - 0.07756)b_{t-1}$$

$$S_t =1\left(\frac{Y_t}{L_t}\right) +(1- 1)\ S_{t-s}$$

The additive and multiplicative Holt-Winters model forecasts can be represented as shown in Figures 6 and 7 below.

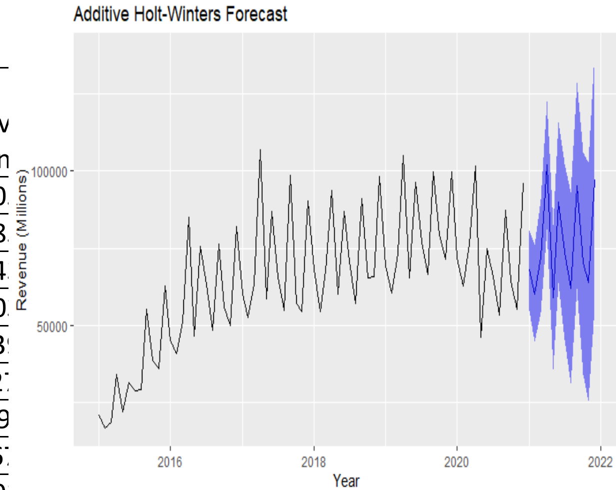

Figure 6: *Additive HW model prediction*

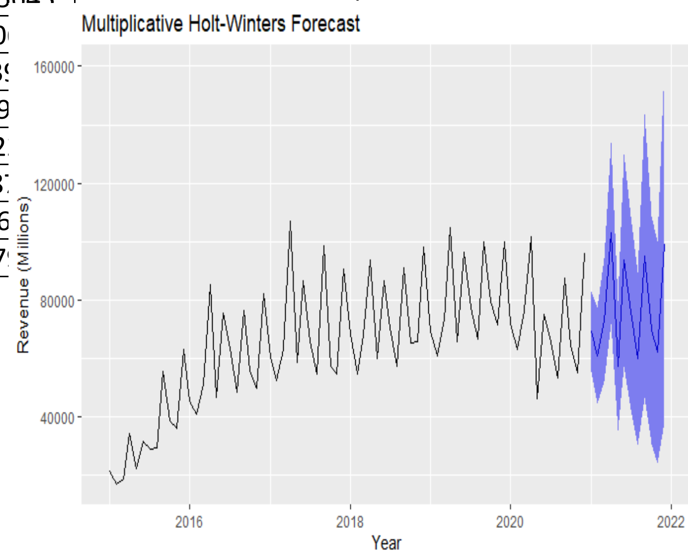

Figure 7: *Multiplicative HW model prediction*

Model Accuracy

Both SARIMA(0,1,1)(0,1,1)[12] and multiplicative Holt-Winters method produced estimates close to the original data. However, SARIMA(0,1,1)(0,1,1)[12] had the lowest error values and hence had better predictive accuracy compared to multiplicative and additive Holt-Winters models as indicated in Table 3.

Table 3

Holt-Winters Constants and coefficients

Models	MAE	MAPE	MASE
SARIMA(0,1,1)(0,1,1)[12]	4541.90	6.90	0.37
Additive Holt-Winters Method	4982.70	7.83	0.41
Multiplicative Holt-Winters Method	5126.49	7.43	0.41

The study used 6 years of domestic taxes collections from 2015 to 2020 to compare the performances of the Seasonal Autoregressive Integrated Moving Average (SARIMA) and the Holt-Winters (HW) forecasting methods. An increasing trend from 2015 to 2020 is realized from plot in figure 1. The data also revealed a seasonal pattern. The data was log-transformed and differenced to make it non-stationary (Figure 2).

Augmented Dickey-Fuller (ADF) and Phillip Peron (PP) tests established the data stationarity when transformed and differenced once (*ADF=-5.784, p=.01*). The diagnostic check for SARIMA and Holt-Winters revealed that the residuals were uncorrelated and originated from a well-specified model and can be used to forecast. The Ljung-Box test also confirms this ($\chi^2_{12} = 16.996, p = .1497$).The SARIMA(0,1,1)x(0,1,1)[12] model was found to be the best model since it had the least Bayesian Information Criterion (BIC=1236.49),Akaike's Information Criterion (AIC=1230.26) and low forecasting errors (MAPE=6.9, MASE=0.37). Regarding the Holt-Winters method, the multiplicative method was slightly superior to the additive method due to its lower error (MAPE=7.43) as shown in Table 3.

Conclusion

Revenue forecasting is an important aspect that aid in budgeting and revenue targets set by the government. It is, therefore, a key element in the government's planning process. Historical data of domestic taxes revenues collected between Jan 2015 to Dec 2020 were used in the comparison of the estimates and predictive accuracy of the SARIMA and Holt-Winters time series forecasting models. The Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Bayesian Information Criterion (BIC), Mean Absolute Squared Error (MASE), and Akaike's Information Criterion (AIC) were used to measure the model accuracy levels. Both SARIMA and Holt-Winters models were found to have good predictive ability for domestic tax revenues in Kenya. Based on the magnitude of forecasting errors, the SARIMA model was found to have a better forecasting performance compared to Additive and multiplicative Holt-Winters methods.

References:

Amadala, V. (2021). 'Treasury told to set logical budget to cut excess borrowing'. *The Star-News*

Chang, Y.W. & Liao, M.Y. (2010). A seasonal ARIMA model of tourism forecasting: The case of Taiwan.Asia Pacific Journal of Tourism Research,15(2), pp.215-221

Ergüven, M. H., Yılmaz, A., & Kutlu, D. (2015). Hybrid tourism within the context of touristic product diversification: glamping. *Journal of Academic Social Science Studies*, (41), 255-265.

Harelimana, J. B. (2018). The role of taxation on the resilient economy and development of Rwanda. *J. Financ. Mark*, *2*(01).

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of statistical software*, *27*(1), 1-22.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: the state space approach. *Springer Science & Business Media*.

ICPAK (2016). Kenya's revenue analysis 201015. A Historical Perspective to Revenue Performance in Kenya.

Jenkins, G. P., Kuo, C. Y., & Shukla, G. (2000). Tax analysis and revenue forecasting. *Cambridge, Massachusetts: Harvard Institute for International Development, Harvard* University.

Lilian, O. (2015). Kenya loses over Sh600bn every year in tax evasion.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John wiley & sons.

Otu, O. A., Osuji, G. A., Opara, J., Mbachu, H. I., & Iheagwara, A. I. (2014). Application of Sarima models in modelling and forecasting Nigeria's inflation rates. *American Journal of Applied Mathematics and Statistics*, *2*(1), 16-28.

Pelinescu, E., Anton, L. V., Ionescu, R., & Tasca, R. (2010). The analysis of local budgets and their importance in the fight against the economic crisis effects. *Romanian Journal of Economic Forecasting*, 17-32.

Pindyck, R. S., Hall, B. H., & Rubinfeld, D. L. (1998). *TSP Handbook to Accompany Econometric Models and Economic Forecasts*. McGraw-Hill/Irwin.

Rahman, M. H., Salma, U., Hossain, M. M., & Khan, M. T. F. (2016). Revenue forecasting using holt–winters exponential smoothing. *Research & Reviews: Journal of Statistics*, *5*(3), 19-25.

Saayman, A., & Saayman, M. (2008). Determinants of inbound tourism to South Africa. *Tourism economics*, *14*(1), 81-96.

Signé, L. (2016). How to Implement Domestic Resource Mobilization (DRM) Successfully for Effective Delivery of Sustainable Development Goals (SDGs) in Africa Illustrative Actionable Solutions for Policy Leaders.

Susan, W. G., Anthony, G. W., & John, M. K (2015). Forecasting Inflation Rate in Kenya Using SARIMA Model. *American Journal of Theoretical and Applied Statistics. Vol. 4, No. 1, 2015, pp.15-18. doi: 10.11648/j.ajtas.20150401.13*