SEASONAL NAÏVE MODEL INCORPORATING TREND COMPONENT FOR TAX REVENUE FORECAST IN KENYA

FREDRICK KYALO SAMUEL

BSc (Applied Statistics)

I56/CTY/PT/27569/2014

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (STATISTICS) IN THE SCHOOL OF PURE AND APPLIED SCIENCES OF KENYATTA UNIVERSITY

MARCH, 2023

DECLARATION

This project is my original research work and has not been presented for a degree award or any other award in any other University or institution of higher learning.

Signature……………………………..…… Date…………………………………

Fredrick Kyalo Samuel

Department of Mathematics and Actuarial Science

Supervisor:

I confirm that the work reported in this project was carried out by the student under my supervision.

Dr. Titus K. Kibua

Signature………………………………… Date…………………………………

Department of Mathematics and Actuarial Science,

School of Pure and Applied Sciences,

Kenyatta University

DEDICATION

I dedicate this project to my dear wife, Lilian Muthangya for her moral support during the course of undertaking the project.

ACKNOWLEDGEMENT

I thank Almighty God for good health, strength and determination while undertaking this project.

Special acknowledgement goes to my supervisor, Dr. Titus K. Kibua for his availability, professional advice and tireless support while undertaking this project.

I express my gratitude to Dr. J.K. Mung'atu and Dr. Ananda for their professional guidance in the research field which led to the successful finalization of the project.

Lastly, I acknowledge the entire academic staff of the Mathematics and Actuarial Science Department of Kenyatta University for their professional input which was critical in enriching this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

ACF	Autocorrelation Function
A-I-A	Appropriation in Aid
AIC	Akaike's Information Criteria
ANN	Artificial Neural Network
AR	Autoregression
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criteria
BRP	Budget Rationalization Programme
BVAR	Bayesian Vector Auto-regression
CIT	Corporate Income Tax
EAC	East Africa Community
GA	Genetic Algorithm
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GA-SVM	Support Vector Machine with Genetic Algorithm
GDP	Gross Domestic Product
GOK	Government of Kenya

IEA	Institute of Economic Affairs
IMF	International Monetary Fund
KIPPRA	Kenya Institute for Public Policy Research and Analysis
KNBS	Kenya National Bureau of Statistics
KRA	Kenya Revenue Authority
LSE	Least Squares Estimation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percent Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Squared Error
ME	Mean Error
MIDAS	Mixed Data Sampling
MPE	Mean Percentage Error
MSE	Mean Squared Error
NNs	Neural Networks
PAYE	Pay As You Earn
PIT	Personal Income Tax

PSO	Particle Swarm Optimization
PWC	Price Water Coopers
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SD	Standard Deviation
SVM	Support Vector Machine
TMP	Tax Modernization Programme
TTR	Total Tax Revenue
VAR	Vector autoregression
VAT	Value Added Tax

ABSTRACT

Tax revenue is largest source of government revenue in Kenya. Nevertheless, the tax revenue collection has overtime fell below the planned targets. Besides, Kenya has witnessed continuous increase in public debt since public expenditures have maintained consistent growth pattern and continually surpassed revenues. The structure of tax revenue data in Kenya exhibit seasonality fluctuations with progressive increase (trend) in monthly tax revenue collections of the year. In order to facilitate government in proper fiscal planning and long-term projections, modelling and forecasting tax revenue is desirable. The objectives of this study were to develop a seasonal naïve model incorporating the trend component for forecasting tax revenue in Kenya and use the model to forecasting tax revenue collections in Kenya for the next two years. This research used time series approach to build the model. The monthly tax revenue data comprising of 192 months spanning July 2000 to June 2016 was used in this study. The study found that seasonal naïve model with trend was appropriate model for forecasting tax revenue data since it recognized both seasonal and trend components in the data and recommended application of the developed model in forecasting tax revenue collections in Kenya. Modelling the causal relationship of tax revenue with other variables that account for seasonality such as inflation, exchange rates, public expenditure and public debt was identified for future area of study.

CHAPTER ONE

INTRODUCTION

1.1 Background

Tax is a compulsory contribution paid by individuals and corporate entities to the government which is not associated to the benefit received (Hyman, 1987). Tax revenue is a largest source of government revenue in Kenya whose key components are Income tax, VAT, Excise duty and Imports duty. During the financial year 2015/16, the tax revenue constituted 92 per cent of the total recurrent revenue which was estimated at Ksh 1,401.2 billion (KNBS, 2016).

Nevertheless, tax revenue collection has been below the planned targets. For instance, the cumulative revenue (including A-I-A) of Ksh 581.1 billion was realised during the first half of the financial year 2015/2016 against a target of Ksh 678 billion. Similarly, the ordinary revenue of Ksh 544.2 billion was collected over the same period against a target of Ksh 591.8 billion. Therefore, the total revenue was below target by 17.0 per cent while ordinary revenue fell short of target by 8.7 per cent. The performance of income tax, VAT, Excise duty and import duty were below their target by 10 per cent, 9 per cent, 6 per cent and 5 per cent, respectively (KIPPRA, 2016).

Besides, there has been continuous increase in the public debt since the public expenditures have not only continually exceeded revenues but also maintained consistent growth pattern (KNBS, 2016). The total revenue increased by Ksh 676.36 billion between the financial year 2011/12 and the financial year 2015/16 while total expenditure increased by Ksh 787.54 billion during the same period. In comparison to the GDP, the revenue averaged 19.88 per cent between year 2011 and year 2015 whilst the average expenditure was 30.12 per cent thereby contributing to a resource gap of around 10.24 per cent (KIPPRA, 2016). In addition, Kenya recorded the highest debt to GDP ratio among EAC countries which was estimated at 52.7 per cent in the year 2015 compared to Tanzania's 40.5 per cent, Burundi's 38.4 per cent, Uganda's 35.4 per cent, and Rwanda's 34.6 per cent (KIPPRA, 2016).

In order to ensure revenues are forecasted with certainty, the tax structure should be flexible and stable. The fluctuation in tax revenues may create uncertainty which negatively affect execution of the government development programmes (Todaro & Smith, 2003). Consequently, the tax system should raise enough revenue to meet public expenditure. In addition, tax system should be elastic to ensure revenues grow with GDP. The tax revenue and expenditure should grow in similar manner to guarantee macroeconomic stability (World Bank, 1990). This enables tax structure to facilitate fiscal policy through macroeconomic stabilization, distribution of the tax burden, and achievement of growth objectives (Musgrave, R. & Musgrave, P., 1989). Thus, tax buoyancy and tax elasticity become vital in determining relationship between growth in GDP and revenue to estimate country's tax productivity.

Tax buoyancy measures how change in GDP vary with tax revenue. For instance, tax buoyancy of one infers that growth of GDP by one percent would culminate in increase by one percent of the tax revenue. Similarly, tax buoyancy greater than one infers growth in GDP by one percent would increase tax revenue by more than one percent (Belinga, et al. 2014).

Tax elasticity, on the other hand, measures change in tax revenue when there is a change in GDP after controlling exogenous variables such as discretionary changes in tax policy. The tax is elastic if increase in GDP by one percent leads to increase in tax revenue by greater than one percent, while holding discretionary tax changes constant. Elasticity assumes that if tax rates or bases are not changed, the tax system remains unchanged (Osoro, 1993; Wawire, 2017).

Consequently, an elastic tax system is essential in an economy as the growth in GDP culminates in automatic increase in revenues without changing the tax rates. The tax policy makers require this information to guide the government on adequacy of future tax revenues in meeting the public expenditure without varying the existing tax rates or bases. It also acts as a guide in identifying the buoyant sectors of the economy where the government needs to concentrate when expanding the tax base and planning tax changes.

Kenya has undertaken numerous tax reforms for enhancing efficiency in tax revenue collections to fund government operations and minimize public borrowing (IEA, 2012). This is reinforced by growing need to fund public expenditure through mobilizing internal resources (IMF, 2011). The Kenya’s tax reform programme should improve productivity and efficiency in taxation, improve tax effectiveness through greater tax elasticity, and enhance tax collection and administration whilst lowering the tax rates (Wagacha, 1999).

The major tax reforms in Kenya commenced in 1986 through Sessional Paper No 1 of 1986 which sought to address the problem of destabilizing effects of the deficits which were also becoming unsustainable (GoK, 1986). The key fiscal policy proposals were Tax Modernization Programme (TMP) and Budget Rationalization Programme (BRP) which were adopted in 1986 and 1987, respectively (Moyi & Muriithi, 2003). The TMP sought to enhance elasticity of the tax system through broadening the government revenue base while the BRP sought to regulate expenditure through strict fiscal controls.

Despite continuous changes in tax system to meet objectives of the TMP, similar challenges are experienced today akin to the pre-reform challenges. Kenya remains amongst most tax unfriendly countries worldwide with low tax competitiveness, whereby Kenyan firms report approximately loss of 68.2 per cent of profit to taxes. Tax evasion is very high, whereby a tax gap of approximately 35 per cent and 33.1 per cent was reported in 2000/01 and 2001/02, respectively (KIPPRA, 2004a). Tax revenue evasion by multinational corporations results to an estimated loss of Sh639 billion annually thus hampering economic growth (Ochieng, 2015).

The tax code remains cumbersome and complex, characterized by a narrow tax base with huge dispersions, uneven and unfair taxes, and low compliance (KIPPRA, 2004b). The tax system exhibits other key challenges which include; tax structures which are not easy to administer; low productivity due unresponsiveness to growth and discretionary policy; introduce serious economic distortions despite raising little revenue; treat similar circumstances of labour and capital differently; and is skewed, selective and favours persons with the capability to evade tax enforcement and administration system (KIPPRA, 2004b).

The Kenya's tax structure is heavily skewed on VAT and income taxes which are two major sources of total tax revenue (Mutua, 2012). For instance, income tax and VAT accounted for 36.3 per cent and 25 percent of total government revenue (total taxes plus A-I-A) for the period 2005/06 to 2011/12 (Mutua, 2012). Tax revenue in Kenya is collected and reported on daily and monthly basis.

Income tax is charged for each year to all income accrued in or derived from Kenya to either resident or non-resident persons in Kenya. The income taxes in Kenya are levied on personal income, profits and capital gains and comprise of PIT, corporation tax and withholding tax (Moyi & Ronge, 2006). The PIT is charged from self-employment while PAYE is charged on employment income. The resident companies are charged corporation tax of 30 per cent on profits while non-resident companies are levied 37.5 per cent corporation tax on profits (GoK, 2009).

The PIT and PAYE are charged on a progressive scale. The lowest rate is 10 per cent while the highest rate is 30 per cent. The rate of increase in income taxes to total tax revenue has been higher in comparison with other types of taxes due to ease in administration. The law requires individuals and companies to assess themselves and remit tax to the government at regular intervals. The PAYE system makes collection easier for salaried employees, whereby the employer calculates taxes, subtracts taxes at source, and remits deducted tax to KRA, thereby enhancing tax compliance (GoK, 2009). The individual income tax returns is filled on or before 30 June of the following year. The late filling attracts a penalty of 5% of the tax due date or Ksh. 2,000 whichever is higher.

The VAT was introduced in year 1990 as a government strategy to increase tax revenue by expanding the tax base. Nevertheless, the ability of VAT to increase tax revenue is hinged on tax rate, tax base, exemptions, simplicity of the tax regime, tax administration efficiency, and tax compliance level (Glenn et al., 2000; Wawire, 2003; Moyi & Ronge, 2006). The VAT Act, 2013 Cap. 476 of the Laws of Kenya governs administration of VAT which covers the exempted goods; zero-rated goods and services including exports and essential goods; and taxable goods and services. VAT is a consumption tax charged on the use of taxable products and services supplied or imported into Kenya. It imposed at every point of the supply chain, from production to consumption, whenever there is value addition on applicable goods and services. The VAT has been stable overtime when compared with excise tax and imports duty. The VAT averaged 27.5 per cent between year 1980 and year 2018 hence constituting the largest share of taxes as a percentage of GDP (KIPPRA, 2020).

The VAT is charged to traders whose annual turnover is at least Ksh. 5 million in a 12-month period. Nevertheless, there is an option of voluntary registration of persons with less than Ksh. 5 million to be allowed to charge VAT. The normal goods in Kenya attract a standard rate of VAT of 16%. The monthly VAT returns are submitted to KRA on monthly basis, on or before the 20th day of the following month. Persons with no VAT to declare are obliged to submit a NIL return. The late filling attracts a penalty charge of Ksh. 10,000 or 5% of the tax due date. The late payment attracts a penalty charge of 5% of tax due and an interest of one per cent per month on the unpaid tax until the full tax is paid (KRA, 2021).

The excise taxes are charged with discriminatory intent (Bolnick and Haughton, 1998). The luxury goods and services often attract excise tax. In Kenya, the commodities that are levied excise duties include tobacco and cigarettes, alcoholic products, juices and soft drinks, mineral water, financial transactions, automobiles and airtime. These commodities are characterized by an income elasticity of demand of more than unit and low-price elasticity of demand which curtails consumption of the goods as price increases. The excise taxes are attractive to government as extra revenue stream to finance budget deficits due to high tax rates coupled with low administrative costs. The Excise taxes averaged 3.1 per cent as a percentage of GDP between year 1980 and year 2018 but demonstrated huge fluctuations (KIPPRA, 2020). The computation of excise tax is done annually to take into account inflation as per the Finance Act, 2020.

Import duty is governed by the provisions of the EAC Customs Management Act and is levied on imported goods at varied rates. The raw materials and capital goods attract a rate of 0% while intermediate goods and finished goods attract rate of 10% and 25% respectively. Nevertheless, Council of Ministers of the EAC partner states can prescribe a different rate of duty, depending on the agenda concerning certain industries (PWC, 2021).

Despite challenges in forecasting tax revenue exhibiting seasonality behaviour, modelling and forecasting tax revenue is generally necessary in an economy to facilitate government in proper fiscal planning and long-term projections. This enables government to manage public expenditure and public debt. Consequently, development of a suitable model with capability of producing reliable and valid tax revenue forecast estimates becomes vital.

This study aims at developing a suitable oscillatory model capable of forecasting tax revenue data exhibiting seasonality-periodic fluctuations. Time series models or stochastic models are appropriate for the studies concerning data exhibiting seasonality fluctuations. However, due to complexity and problem of optimization of numerous variables using stochastic models, time series models are preferred for developing forecasting models for data exhibiting seasonality-periodic fluctuations. The seasonality experienced in tax revenue can only be modelled using models that allow for the same structure in the data. Therefore, failure to capture these structural aspects has always been associated with suboptimal prediction of tax income, hence interfering with planning.

1.2 Problem Statement

The structure of tax revenue data in Kenya exhibit seasonality fluctuations with progressive increase (trend) in monthly revenue tax collections of the year. This implies that the data are ever increasing up to the end of the year. There is also some depicted trend within the seasonal data. For instance, cumulative collections in January 2000 is less than cumulative collections in January 2001 which is less than cumulative collections in January 2002, and so on. Consequently, the structure of the data indicates seasonal increase over time. In addition, the tax revenue data has a strong seasonality that makes its estimation using the traditional SARIMA models erroneous. The consequence is that of wrong tax revenue estimation. However, due to strong seasonality in tax revenue data, traditional SARIMA models may not work well hence the need to develop seasonal naïve models. Additionally, since the seasonal naïve models only consider the same period last year, in this study we also incorporate a trend component within the seasonal naïve models to address lack of long-term memory of seasonal naïve models since they depend on the previous realization.

1.3 Objectives

1.3.1 General Objective

To forecast tax revenue in Kenya using a seasonal naïve model incorporating a trend component.

1.3.2 Specific Objectives

(i) To develop a seasonal naïve model incorporating the trend component for tax revenue forecast in Kenya.

(ii) To determine asymptotic properties of the developed model.

(iii) To predict tax revenue for the two years in Kenya using the developed model.

1.4 Significance of the Study

The SARIMA models have been applied in numerous research studies due to their capability to forecast time series data exhibiting seasonality-periodic fluctuation behaviour. However, for the data exhibiting strong seasonality, seasonal naïve models are preferred as opposed to traditional SARIMA models. The seasonal naïve models equate each prediction to the past observed value of the same season, hence suitable for tax revenue data that exhibit similar characteristics when monthly tax revenues are considered. In addition, tax revenue exhibits some trend, whether increasing or decreasing. Therefore, a model that captures both the seasonal naïve and trend qualities will be more appropriate for modelling and forecasting tax revenues over the years.

CHAPTER TWO

LITERATURE REVIEW

2.1. Theoretical Review

The seasonal naïve model is efficient for forecasting highly seasonal data. The model assumes that the same value of period in previous year is same as the same period in current year (Fretchling, 2001). This model accounts for seasonality by setting every forecast to be same as the preceding observed value of the same period (Hyndman & Athanasopoulos, 2018).

Seasonal naïve models are better than mean and naïve models. The mean model ignores seasonality since it predicts all future values to be same as the average (or "mean") of the past data. The naïve models set the forecast to be the value of the last observed value hence ignoring seasonality as well. Consequently, the mean and naïve methods assume no changes at all between time periods (Hyndman & Athanasopoulos, 2018).

The seasonal naïve forecasting model assumes that each season's values form an independent random walk hence is also referred as *seasonal random walk* model. For instance, the model assumes that January's value this year is a random step away from January's value last year, February's value this year is a random step away from February's value last year, etc., and the mean value is the same constant at every step (Nau, 2017).

The seasonal naïve models have a constant term, seasonal differencing of one order and no other parameters hence they are special case of SARIMA model which exhibit the following important properties: average long-term trend in the past is equal to long-term trend in forecasts; the forecast seasonal pattern equals to the most recent seasonal pattern; slow response to the cyclical downturns or upturns, always looking past one year; and is inherently an additive model, which can also capture multiplicative seasonal patterns through logging and/or deflating the data (Nau, 2017).

The seasonal naïve models make the following assumptions: the average seasonal difference computed over the entire history of the time series is equal to the predicted values of all future seasonal differences; the average year-to-year trend observed in the past is equal the future trend; model is relatively stable when there are abrupt changes in the data; slow response to cyclical downturns and upturns in the data. Therefore, there are natural positive autocorrelation that depict one-step-ahead forecast errors (Nau, 2017).

2.2. Empirical Review

Botrić & Vizek (2012) compared accuracy of forecasts from time series models and official forecasts which relied on expert judgment using a case study of Croatia. Disaggregated approach was adopted to estimate separate models for each of the seven revenue sources, namely: VAT, social contributions, income tax, property tax, corporation tax, import and excises duties. The time series models – linear trend, regression, error correction, random walk and ARIMA time series models were applied to quarterly revenue data. The random walk and linear trend models were estimated for all seven aggregated revenues. The regression, error correction, and ARIMA models were not estimated for components of social contribution and excises due to unsatisfactory diagnostic tests and unreliable proxies for tax bases of the social contribution and excises taxes. The forecasts of social contribution and excises components obtained from random walk and trend models were aggregated in order to ascertain their accuracy by aggregating the forecasts. The data was transformed with view of obtaining diagnostic tests of the suggested models and best possible fit. Computation of one-year and two-year ahead forecasts was undertaken and comparison made with the official forecast values obtained from expert judgement. The study results noted that the forecasts from time series models were generally more accurate compared to official forecasts that were prepared using expert judgment. The study concluded that official forecasters would benefit from implementing forecasting framework such as adoption of ARIMA model, error correction or regression model to determine most likely future direction of the economy.

Streimikiene et al (2018) conducted study to forecast Pakistan's tax revenue for the financial year 2016/17 and analysed how working class was impacted by indirect taxes. Three different time series models were analysed for their efficiency, namely; ARIMA, AR with seasonal dummies, and VAR models using RMSE, MAE and MAPE tests. The monthly data spanning July 1985 to December 2016 was used to predict year 2017. The total tax revenue was forecasted using four components of revenues; sales tax, direct tax, federal customs duties and excise duty. The ARIMA model gave better forecast values in comparison with other models. In addition, the results indicated that indirect taxes generated huge portion of tax revenue thereby directly affecting the working class of Pakistan since it caused more inflation. Based on the study, it was concluded that there was dire need for the government to enhance system of tax collection, broaden the tax net, and come up with new tax reforms.

Chimilila (2017) employed a time series approach to model and forecast Tanzania's tax revenue and its volatility. The study applied ARMA model to forecast tax revenue and GARCH model to forecast volatility. The monthly tax revenue data from January 2000 to February 2015 comprising of 182 months was used in the study. The AIC and BIC was used to select best models while RMSE was applied to compare the forecast performance of ARMA and GARCH models. The study results indicated steady increase in tax revenue, with a persistent volatility which increased over time. The taxes from bases (income) which had high volatility had correlation with observed volatility. The linear combination of ARMA and S(12) was recommended for forecasting monthly tax revenue while GARCH(1,1) was recommended for forecasting volatility. Further, broadening consumption tax base within the existing tax portfolio to diversity taxes was recommended in order to reduce volatility and improve contribution of consumption tax to revenue collection.

Molapo et al (2019) used ARIMA, BVAR, and State Space exponential smoothing (Error, Trend, Seasonal [ETS]) models to forecast major tax revenues in South Africa, i.e. VAT, CIT, PIT and TTR using quarterly data from years 1998 to 2012. The out of sample forecasts from 2012 Quarter 2 to 2015 Quarter 1 for the three models were compared using RMSE. The study results indicated BVAR model was more accurate in forecasting major tax revenues. The ETS outperformed BVAR method for TTR forecasting. The study recommended inclusion of BVAR method to existing tax revenue forecast techniques which were being used in South Africa.

Luchko et al (2021) used ARIMA model to forecast VAT on imported goods, services and works in Ukraine. The VAT on import was noted to have significant impact on budget revenues. The study recommended ARIMA (0,0,1) (0,0,1) for forecasting VAT on imports. The study identified application of artificial intelligence and block-chain technologies in conditions of uncertain tax payment and optimization of tax administration as future areas of research studies.

Nandi et al (2014) applied univariate Time Series to forecast Bangladesh's tax revenue. The monthly tax revenue data from July 2004 to December 2012 was applied in the study. Four statistical models were applied, namely; One-Step ARIMA SARIMA method, Dynamic Forecast ARIMA SARIMA method, Holt-winter seasonal multiplicative approach, and Holt-winter seasonal additive approach. The study reviewed that Holt-Winter seasonal multiplicative approach was appropriate for revenue forecasting based on the minimum MAPE and MSE. The study compared forecasted tax revenue error projected by Ministry of Finance of Bangladesh with the four statistical techniques and established that larger errors with higher volatility in tax gaps was produced by existing judgmental tax revenue forecasting method in Bangladesh compared to the statistical methods employed in the study. The Holt-Winter multiplicative approach outperformed other models in forecasting monthly tax revenue in Bangladesh. Thus, the study recommended fiscal Authority in Bangladesh to systematically analyse several statistical models in revenue forecasting rather than relying on judgmental method.

Shahnazarian et al (2017) applied BVAR models in forecasting corporate tax in Sweden. The study results noted that external shocks, macroeconomic shocks, and financial shocks explained a significant part of the forecasting error variance of corporates' profit. In addition, the shocks in profits, tax adjustments and fiscal measures explained a major part of the forecasting error variance of the tax base for corporate tax. The combination of forecasts from BVAR, MIDAS and ARIX was noted to be a viable approach for forecasting corporate tax revenues. The study concluded that model-based input was crucial for forecasting government's corporate tax revenue and analysing the sensitivity of corporate tax revenues in alternative macro-economic environments.

Nazmi & Leuthold (1988) applied time series analysis to forecast state income tax receipts. The BIC was applied to identify appropriate ARMA model while grid search was used to identify suitable power transformation for state income tax forecasts. The RMSE, MAE and ME was used to compare performance of forecasting models. The study results indicated that ARIMA (0,1,1) $(0,1,1)_4$ with power transformation of 0.25 was superior to the other models. The time series model employed less information compared to regression models. Specifically, the regression models relied on the actual values for personal tax for the in-sample period and forecast values of personal income for out-of-sample period while time series model utilized tax receipts data only.

Lu et al (2009) applied GA-SVM time series to forecast gross tax in China from year 1990 to year 2001. The GA was used to identify the suitable training parameters of SVM, whereby the fitness was evaluated by leave-one-out method. The GA-SVM model achieved better forecasting accuracy compared to ANN and grey model in gross tax forecasting.

Ticona et al (2017) applied Hybrid Model based on GAs and NNs for a multi-step tax revenue forecast using exogenous and endogenous variables. The exogenous variables were included to enhance forecasting accuracy. The study used data from Federal Revenue of Brazil (RFB) spanning from January 2000 and December 2014. These data was divided into three data sets for different purposes of the NNs: training data (years 2000 to 2012), validation data (the year 2003) and test data (the year 2014). The study results indicated that hybrid model of GAs and NNs was efficient in forecasting Tax Revenue collections. Further, the results were noted to be more accurate compared to the indicators method applied by the RFB.

Li-Xia et al (2011) applied combination of SVM and PSO for tax forecasting. The tax data of China and Kunming (a city in China) was used for comparing forecasting performances of the proposed PSVM model and ANN model. The experimental data was divided into training and testing data whereby the tax data of China from years 1990 to 1997 was used as training data while remainder was testing data. Similarly, the training data of Kunming spanned years 1994 to 1999 while other data was used as the testing data. The MAPE was used to measure forecasting accuracy. The parameters of SVM were optimized by adoption of PSO. The training parameters of SVM were selected using PSO. The SVM represented non-linear relationship in tax forecasting. The tax forecast model was constructed using PSVM model which was noted to have good forecasting performance.

Batóg & Batóg (2021) applied time series and causal models to examine classical forecasting models for regional government revenue. The data covering year 2000 to year 2018 was applied in the study to assess the reliability of three models; time series models, dependency models and structural forecasting. The exponential smoothing with the exponential trend; and structural forecasting using the ratio of regional gross product to GDP study were most effective methods of forecasting local government revenue based on the study results.

Mamula (2015) examined the forecasting accuracy of four models in forecasting international tourism flows from German to Croatia, namely; seasonal naïve, SARIMA, Holt-Winters triple exponential smoothing, and multiple regression models. The in-sample and the out-of-sample MAPE was used to compare forecasting performances of the models. There was no significant difference upon diagnostics for the four forecasting models. The multiple regression model was noted to be more superior in forecasting of German tourists' arrivals in Croatia.

2.3. Research gap

The structure of tax revenue data in Kenya exhibit strong seasonal fluctuations with progressive increase (trend) in monthly revenue tax collections of the year. Seasonal naïve models incorporating a trend component is desirable to address the tax revenue structure in Kenya. Despite this, there are no existing research studies which have used seasonal naïve random variables incorporating trend component in tax revenue forecasting in as much as data implicates seasonal naïve models incorporating trend component. The study seeks to address this research gap.

CHAPTER THREE

METHODOLOGY

3.1. Stationarity

A time series $\{Y_t\}$ consists of series of observations Y_t ordered by time t. Stationary time series implies that change over time is constant in the time series, for instance, if it grows linearly. The joint distribution of $(Y_{t_1}, \ldots, Y_{t_k})$, does not vary over time for a strictly stationary time series $\{Y_t\}$ which is very difficult to verify empirically. A weaker stationarity is commonly used whereby both the mean of Y_t and the covariance between Y_t and Y_{t-j} don't vary with time, whereby j is an arbitrary integer. More specifically, $\{Y_t\}$ is weakly stationary if (a) $E[Y_t] = \mu$ (μ constant), and (b) $Cov\,(Y_t, Y_{t-j}) = \gamma_j$, which only depends on j. The covariance $\gamma_j = Cov\,(Y_t, Y_{t-j})$ is the lag-j autocovariance of Y_t which has two important properties: (a) $\gamma_0 =$ Var (Y_t), and (b) $\gamma_{-j} = \gamma_j$ (Tsay, 2010).

3.2. Model specification

Let's first consider the naïve model. This forecasting model is also referred to a random walk, an ARIMA (0, 1, 0) model which stipulates that the conditions today will be the same as yesterday. The random naïve model is a differencing at lag-1 and is described by:

$$Y_t = c + Y_{t-1} + a_t \tag{3.1}$$

Where Y_t is tax revenue at time t, Y_{t-1} is tax revenue at time $t-1$, $a_t \sim N(0, \sigma^2)$ is white noise series with zero mean, and the constant $c = E(Y_t - Y_{t-1})$ is the drift of the model. The residual series $a_t = Y_t - Y_{t-1}$. The naïve models do not meet the requirement of weak stationarity (Tsay, 2010).

The differencing at lag-s results to seasonal naïve models which is difference between observations y_t and the previous season's observation y_{t-s} for the period s. This is highly relevant for data with seasonality. The seasonal naïve model is described by:

$$Y_t = c + Y_{t-s} + a_t \tag{3.2}$$

Where Y_t is tax revenue at time t, Y_{t-s} is tax revenue at time $t-s$, the constant $c = E(Y_t - Y_{t-s})$, s is periods, and the white noise series $a_t \sim N(0,1)$ with zero mean. The seasonal naïve model, also called the seasonal random walk, is a special SARIMA model which is written as an ARIMA (0, 0, 0) (0, 1, 0)$_s$ model at period s (Hyndman and Athanasopoulos, 2018).

The seasonal naïve model in equation (3.2) does not take the trend component in to account. In order to incorporate the trend component, we need to develop a model for measuring trend and seasonality. The seasonal naïve model in equation (3.2) is improved by incorporating trend component as indicated in equation (3.3) below:

$$Y_t = c + Y_{t-s} + \beta t + a_t \tag{3.3}$$

Where Y_t is tax revenue at time t, Y_{t-s} is tax revenue at time $t-s$, the constant $c = E(Y_t - Y_{t-s})$ for s periods, mean βt is deterministic trend component, and white noise series $a_t \sim N(0,1)$ with zero mean. Equation (3.3) incorporates a trend component into the seasonal naïve model (Equation (3.2)) in order to deal with loss of long-term memory of Equation (3.2).

3.3. Parameter estimation

3.3.1. Least Squares Estimation

The Least Squares Estimation (LSE) method minimizes the squared differences between observed data and their expected values to estimate parameters (Brian and David, 2005). The least squares principle facilitates effective identification of coefficients by minimizing summation of the squared errors. The LSE provides the least value for sum of squared errors (Hyndman and Athanasopoulos, 2018) i.e., we choose the values of $\beta_0, \beta_1, \ldots, \beta_k$ that minimize

$$\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \ldots - \beta_k x_{k,t})^2 \tag{3.4}$$

We now define the least squares estimates of seasonal naïve model with deterministic trend in equation (3.3):

$Y_t = c + Y_{t-s} + \beta t + a_t$ where $a_t \sim i.i.d\, N(0, \sigma^2)$

Equation (3.3) can be written as

$a_t = Y_t - (c + Y_{t-s} + \beta t)$

${a_t}^2 = [Y_t - (c + Y_{t-s} + \beta t)]^2$

$\sum_{t=1}^{T} a_t^2 = \sum_{t=1}^{T} [Y_t - (c + Y_{t-s} + \beta t)]^2$

Let Q = $\sum_{t=1}^{T} a_t^2 = \sum_{t=1}^{T} [Y_t - (c + Y_{t-s} + \beta t)]^2$

$$Q = \sum_{t=1}^{T} [Y_t^2 - 2Y_t(c + Y_{t-s} + \beta t) + (c + Y_{t-s} + \beta t)^2]$$

$$Q = \sum_{t=1}^{T} [Y_t^2 - 2cY_t - 2Y_tY_{t-s} - 2Y_t\beta t) + (c + Y_{t-s} + \beta t)^2]$$

$$Q = \sum_{t=1}^{T} [(Y_t^2 - 2cY_t - 2Y_tY_{t-s} - 2Y_t\beta t) + ((c + Y_{t-s} + \beta t)(c + Y_{t-s} + \beta t))]$$

$$Q = \sum_{t=1}^{T} [Y_t^2 - 2cY_t - 2Y_tY_{t-s} - 2Y_t\beta t) + (c^2 + cY_{t-s} + c\beta t + cY_{t-s} + Y_{t-s}^2 + \beta tY_{t-s} + c\beta t + \beta tY_{t-s} + \beta^2 t^2)]$$

$$Q = \sum_{t=1}^{T} [Y_t^2 - 2cY_t - 2Y_tY_{t-s} - 2Y_t\beta t + c^2 + 2cY_{t-s} + 2c\beta t + Y_{t-s}^2 + 2\beta tY_{t-s} + (\beta t)^2)]$$

$$Q = \sum_{t=1}^{T} Y_t^2 - 2c\sum_{t=1}^{T} Y_t - 2\sum_{t=1}^{T} Y_t\, Y_{t-s} - 2\sum_{t=1}^{T} Y_t\, \beta t + Tc^2 + 2c\sum_{t=1}^{T} Y_{t-s} + 2c\sum_{t=1}^{T} \beta t + \sum_{t=1}^{T} Y_{t-s}^2 + 2\sum_{t=1}^{T} \beta t\, Y_{t-s} + \sum_{t=1}^{T} (\beta t)^2$$

$$\frac{\partial Q}{\partial \beta} = -2\sum_{t=1}^{T} tY_t + 2c\sum_{t=1}^{T} t + 2\sum_{t=1}^{T} t\,Y_{t-s} + 2\beta\sum_{t=1}^{T} t^2$$

But $\frac{\partial Q}{\partial \beta} = 0$ which implies that:

$$-2\sum_{t=1}^{T} tY_t + 2c\sum_{t=1}^{T} t + 2\sum_{t=1}^{T} t\,Y_{t-s} + 2\beta\sum_{t=1}^{T} t^2 = 0$$

$$2\beta\sum_{t=1}^{T} t^2 = 2\sum_{t=1}^{T} tY_t - 2c\sum_{t=1}^{T} t - 2\sum_{t=1}^{T} t\,Y_{t-s}$$

$$\beta\sum_{t=1}^{T} t^2 = 2[\sum_{t=1}^{T} tY_t - c\sum_{t=1}^{T} t - \sum_{t=1}^{T} t\,Y_{t-s}]/2$$

$$\beta\sum_{t=1}^{T} t^2 = \sum_{t=1}^{T} tY_t - c\sum_{t=1}^{T} t - \sum_{t=1}^{T} t\,Y_{t-s}$$

$$\hat{\beta} = \frac{\sum_{t=1}^{T} tY_t - c\sum_{t=1}^{T} t - \sum_{t=1}^{T} tY_{t-s}}{\sum_{t=1}^{T} t^2} \quad (3.5)$$

3.3.2. Asymptotic properties

We define the r^{th} moment of a discrete random variable Y by:

$$m'_r = E(Y^r) = \sum_y y^r f(y); \quad \text{for r= 1, 2,3, …} \quad (3.6)$$

The r^{th} moment of continuous random variable Y is given by:

$$m'_r = E(Y^r) = \int_{-\infty}^{\infty} y^r\ f(y)dy; \quad (3.7)$$

whereby E is the expectation, and f (y) denotes probability density function of Y.

The mean or expectation of Y is the first moment which measures the central location of the distribution.

Let the mean or expectation of Y be denoted by μ_y

We can define the r^{th} central moment of discrete random variable Y by:

$$m_r = E[(Y - \mu_y)^r] = \sum_y (y - \mu_y)^r f(y); \qquad \text{for r= 1, 2,3, …} \tag{3.8}$$

Similarly, for continuous random variable Y, the r^{th} central moment is defined by:

$$m_r = E[(Y - \mu_y)^r] = \int_{-\infty}^{\infty} (y - \mu_y)^r \ f(y)dy; \tag{3.9}$$

as long as the integral exists.

We denote the second central moment, also called the variance of Y, by σ_y^2 , which measures variability of Y. The standard deviation of Y, denoted by σ_x is the positive square root of variance. The normal distribution is uniquely determined by the mean and variance, which are the first two moments of a random variable (Tsay, 2010).

3.3.2.1. Mean of seasonal naïve model with deterministic trend

We now define the mean of seasonal naïve model with deterministic trend in equation (3.3):

$$Y_t = c + Y_{t-s} + \beta t + a_t$$

where βt is deterministic trend that does not vary with time, $t \geq s$, and $a_t \sim \ i.i.d\ N(0, \sigma^2)$

Equation (3.3) can be written as:

$$Y_t = (c + Y_{t-s} + a_t) + \beta t$$

$$Y_s = (c + Y_0 + a_s) + \beta t \qquad \text{when } t = s$$

$$Y_{2s} = (c + Y_s + a_{2s}) + \beta t \qquad \text{when } t = 2s$$

$$Y_{2s} = [c + (c + Y_0 + a_s) + a_{2s}] + \beta t$$

$$Y_{2s} = [2c + Y_0 + a_s + a_{2s}] + \beta t$$

$Y_{3s} = [3c + Y_0 + a_s + a_{2s} + a_{3s}] + \beta t$ when $t = 3s$

.
.
.

$Y_{ts} = [tc + Y_0 + a_s + a_{2s} + a_{3s} + \ldots + a_{ts}] + \beta t$

$Y_{ts} = tc + Y_0 + \sum_{i=1}^{t} a_{is} + \beta t$ when $t = ts$ (3.10)

By putting $s = 1$ in equation (3.10), we obtain:

$$Y_t = tc + Y_0 + \sum_{i=1}^{t} a_i + \beta t \tag{3.11}$$

The mean of the model is given by:

$E(Y_t) = E(tc + Y_0 + \sum_{i=1}^{t} a_i + \beta t)$

$E(Y_t) = [E(tc) + E(Y_0) + \sum_{i=1}^{t} E(a_i) + E(\beta t)]$

$$E(Y_t) = tc + \beta t \quad = f(t) \tag{3.12}$$

since $Y_0 = 0$; $\sum_{i=1}^{t} E(a_i) = 0$; tc = constant; βt = deterministic constant; and $a_t \sim$ $i.i.d\ N(0, \sigma^2)$

3.3.2.2. Variance of seasonal naïve model with deterministic trend

We now define the variance of seasonal naïve model with deterministic trend in equation (3.3):

$Y_t = c + Y_{t-s} + \beta t + a_t$

Equation (3.3) can be written as equation (3.11):

$$Y_t = tc + Y_0 + \sum_{i=1}^{t} a_i + \beta t$$

The variance of the model is given by:

$Var(Y_t) = Var(tc + Y_0 + \sum_{i=1}^{t} a_i + \beta t)$

$Var(Y_t) = Var(tc) + Var\ (Y_0) + Var\ (\sum_{i=1}^{t} a_i) + Var\ (\beta t)$

$$Var(Y_t) = \sum_{i=1}^{t} Var\ (a_i) = t\sigma^2\ = f(t) \qquad (3.13)$$

since $Y_0\ = 0$; and $\beta t\ =$ deterministic constant; and $a_t \sim\ i.i.d\ N(0, \sigma^2)$

We conclude that the seasonal naïve model with deterministic trend component is not stationary since the mean (equation 3.12) and variance (equation 3.13) are functions of time, t

3.4. Model Checking

3.4.1. Akaike and Bayesian Information Criteria

The selection of the final model is done by a penalty function statistic such as AIC or AICc or BIC (Akaike, 1974); (Schwarz, 1978); and (Sakamoto, Ishinguro, & Kitagawa, 1986). Given several competing models, the one with lowest AIC, AICc or BIC value, after ranking, is considered as the best model. The information criterion determines the closeness of the fitted values to the true values of a given model.

Generally, the AIC, AICc and BIC take the following form:

$$AIC =\ 2k - 2log(L) \quad or \quad 2k\ -\ nlog\ (\frac{RSS}{n}) \qquad (3.14)$$

$$AICc\ =\ AIC + \frac{2k(k+1)}{n-k-1}$$

(3.15)

$$BIC\ =\ -2log\ (L) + klog(n) \quad or \quad log\ (\sigma_e^2) + \frac{k}{n}\ log\ (n)$$

(3.16)

Whereby k denotes number of parameters in the model; n is the number of observations; RSS indicates the Residual Sum Squares for the estimated model; L is the maximized value of the likelihood function for the estimated model; and σ_e^2 is the error variance.

The AICc is the modification of AIC with a second order correction for small sample sizes (Hurvich &Tsai 1989). The AICc should be used irrespective of the sample size because AICc converges to AIC as n gets large (Anderson, Burnham, & White, 1998).

3.4.2. Ljung-Box Portmanteau test

The serial correlation in the observations are measured by Ljung-Box Portmanteau test (Tsay, 2010). Let ρ_j be the autocorrelation, the test can be set up with the following hypothesis:

$$H_o = \rho_1 = \ldots = \rho_m = 0$$
$$H_a = \rho_j \neq 0 \quad for\ some\ j \in (1, \mathrm{m})$$

The test statistic, Q is calculated by:

$$Q(m) = T(T+2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{T-j} \tag{3.17}$$

whereby T denotes the number of observations; m indicates the chosen lag; and $\hat{\rho}_j$ denotes the sample autocorrelation. The $Q(m)$ is asymptotically χ^2- distributed with m degrees of freedom, under the null hypothesis.

If $Q(\mathrm{m}) > \chi^2{}_m(\alpha)$ for the chosen level α, the H_o is rejected and the model is considered inadequate. The test can be applied on residuals of a fitted model whereby $Q(m)$ is instead asymptotically $\chi^2{}_{m-p}$ under the null hypothesis, where p is the corresponding number of parameters in the fitted model.

3.4.3. Model comparison criteria

The following measures of forecast accuracy are considered whereby the forecast model with the smallest error is considered among competing models.

3.4.3.1. Mean Error

The average of all the errors in a dataset is the mean error which is given by:

$$ME = \frac{1}{n}\sum_{i=1}^{t} y_i - x_i \tag{3.18}$$

Where ME is Mean Error; x_i is true value; y_i is prediction and n denotes total number of data points.

3.4.3.2. Mean Absolute Error

The mean absolute error measures the average magnitude of the errors in a set of forecasts, irrespective of their direction, and is given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{t} |y_i - x_i| \tag{3.19}$$

Whereby MAE denotes the Mean Absolute Error; x_i is true value; y_i is Prediction; $|y_i - x_i|$ are absolute errors; and n is the total number of data points

3.4.3.3. Root Mean Square Error

The root mean square error (RMSE) measures differences between predicted values by a model or an estimator and observed values. Th RMSE is the standard deviation of the residuals (prediction errors) and measures how residuals are spread out i.e., it indicates the concentration of the data around the line of best fit and is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}} \tag{3.20}$$

Whereby $RMSE$ is Root Mean Square Error; x_i denote the actual observation; $\hat{x}_i$ is estimated value; and N indicates the number of data points.

3.4.3.4. Mean Percentage Error

The mean percentage error (MPE) is a measure of the magnitude of errors. Smaller values imply closeness to the accepted or real value. MPE is given by:

$$MPE = \frac{100\%}{n}\sqrt{\sum_{t=1}^{n}\frac{a_t - f_t}{a_t}} \tag{3.21}$$

Where MPE is Mean Percentage Error; f_t denotes the forecast; a_t denotes the actual value of quantity being forecasted; and n indicates the number of different times for which the variable is forecasted.

3.4.3.5. Mean Absolute Percentage Error

The mean absolute percentage error (MAPE) measures the mean of the absolute percentage errors of forecasts. It is given by:

$$MAPE = \frac{100\%}{n}\sqrt{\sum_{t=1}^{n} | \frac{a_t - f_t}{a_t} |} \tag{3.22}$$

Where $MAPE$ is Mean Absolute Percentage Error; f_t denotes the forecast; a_t denote the actual value of quantity being forecasted; and n indicates the number of different times for which the variable is forecasted.

3.4.3.6. Mean Absolute Scaled Error

The mean absolute scaled error (MASE) measures the accuracy of forecasts and determines the effectiveness of forecasts generated through an algorithm by comparing the predictions with the output of a naïve forecasting approach. It is given by:

$$MASE = \frac{|e_t|}{\frac{1}{n-1} x \sum_{t=2}^{n} |y_t - y_{t-1}|} \qquad \{y_t\}\ is\ non\ seasonal \tag{3.23}$$

$$MASE = \frac{|e_t|}{\frac{1}{n-M} x \sum_{t=M+1}^{n} |y_t - y_{t-M}|} \qquad \{y_t\}\ is\ seasonal \tag{3.24}$$

Whereby $MASE$ is Mean Absolute Squared Error; $\{y_t\}$ denotes the actual observation in time series; and e_t indicates the forecast error for a given period.

3.4.3.7. Residual standard deviation

The residual standard deviation (RSD) measures the standard deviation of the residual values, i.e., the difference between a set of observed and predicted values. It is given by:

$$RSD = \sqrt{\frac{\sum_{i=1}^{n}(Y - \hat{Y})^2}{n-2}} \tag{3.25}$$

Where RSD denotes the residual standard deviation; $\hat{Y}$ is the estimated value; Y is the observed value; and $(Y - \hat{Y})$ is the residual; and n indicates the number of data points in the population.

3.4.4. Residual diagnostics

The "residuals" are is remainders after fitting a model which is the difference between the observed and the corresponding fitted values in majority of the models. It is given by:

$$e_t = y_t - \hat{y}_t \tag{3.26}$$

Residuals are vital in examining adequacy of the model in capturing the information in the data. Consequently, good forecasting model yields residuals which are uncorrelated; and have zero mean. Modification of forecasting method may be undertaken to give better forecasts if either these properties is not satisfied (Hyndman & Athanasopoulos, 2018).

In addition to these properties, it is essential for residuals to be normally distributed and have constant variance for ease of calculation of prediction intervals (Hyndman & Athanasopoulos, 2018).

3.5. Forecasting

Forecasting aims at predicting one or several periods ahead. This assumes that the variable's past pattern and behaviour will be portrayed in future. Nevertheless, likelihood of obtaining forecast errors are unavoidable as the forecast period is extended.

Suppose $y_1, y_2, \ldots, y_k$ are observed values in time series, in forecasting, we estimate the future values such as y_{t+h} made at time $t = k$ for h steps ahead.

Before seasonal naïve model with deterministic trend, let's consider forecast from mean method, naïve method and seasonal naïve methods:

3.5.1. Mean method

The mean method postulates that forecasts of all future values equals the average (or "mean") of the historical data. Suppose historical data is given by $y_1, \ldots, y_t$, the forecasts from the mean method are defined by:

$$\hat{y}_{t+h|t} = \bar{y} = (y_1 + \cdots + y_t)/t \quad (3.27)$$

3.5.2. Naïve method

The naïve method sets all forecasts to be equal to the value of the last observation. The naïve forecasts are defined by:

$$\hat{y}_{t+h|t} = y_t \quad (3.28)$$

3.5.3. Seasonal naïve method

The seasonal naïve method sets each forecast to be equal to the last observed value from the same period of the year. It is defined by:

$$\hat{y}_{t+h|t} = y_{t+h-m(k+1)} \quad (3.29)$$

Whereby m denotes the seasonal period; and k is the integer part of $(h-1)/m$ (i.e., the number of complete years in the forecast period prior to time). For instance, the forecast for all future July values is equal to the last observed July value, in case of the monthly data. The forecast of all future Quarter 1 values is equal to the last observed Quarter 1 value, for the quarterly data. Similar rules apply for other seasonal periods.

3.5.4. Seasonal naïve method with deterministic trend

Let's consider the seasonal naïve model with deterministic trend in equation (3.3) given by:

$$Y_t = c + Y_{t-s} + \beta t + a_t$$

We define forecast for time $t+h$ by:

$$\hat{y}_{t+h|t} = c + h\beta t + y_{t+h-s(k+1)} \quad (3.30)$$

whereby c is the estimate of the level of the time series at time t; βt indicates the estimate of the trend (slope) of the series at time t; s indicates the frequency of seasonality; and k is the integer part of $(h-1)/m$ (i.e., the number of complete years in the forecast period prior to time ($t+h$). Here, the forecast function is trending and the h-step-ahead forecast is equal to the drift plus h times the last estimated trend value and seasonal naïve forecast.

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter presents analysis and discussion of the study results. The monthly tax revenue data from the Kenya Revenue Authority spanning July 2000 to June 2016 comprising of 192 monthly data series was used in this study (see **Appendix 1**). The R version 4.2.0 statistical software was used for analysis.

4.1. Exploratory data analysis

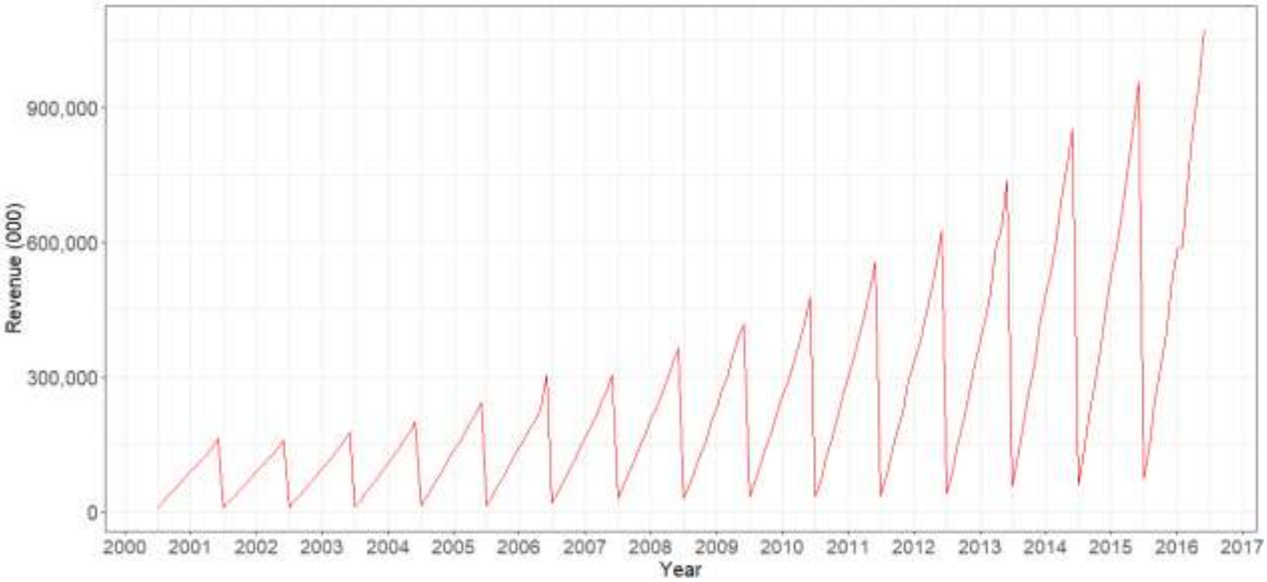

Figure 4.1: Cumulative monthly tax revenue in Kenya from July 2000 to June 2016

Exploratory data analysis was done by plotting graph of original time series tax revenue data to portray overall characteristics of the tax revenue data as illustrated Figure 4.1. The time plot deduces existence of trend in the data due to systematic change in the plot. There is evidence of increasing linear trend component whereby each monthly tax revenue data of the year is increasing with time, hence upward trend. The observations also have seasonality (ups and downs on the regular intervals). The tax revenue data is therefore non-stationary.

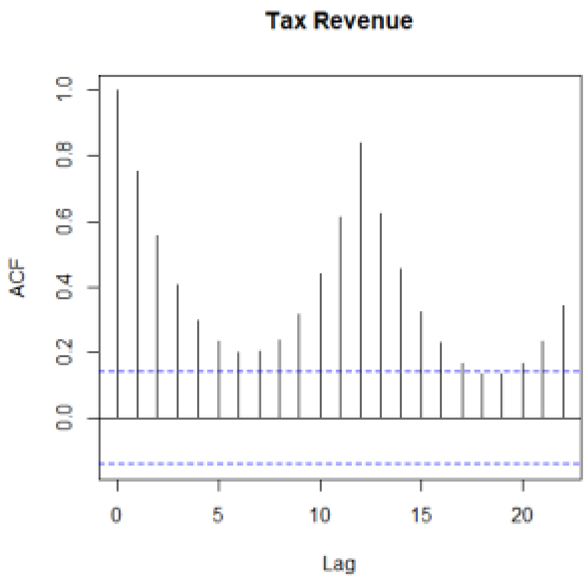

Figure 4.2: ACF for the cumulative monthly tax revenue in Kenya

The sample ACF plot decreases in attenuating sine wave pattern as indicated in Figure 4.2 which implies that trend and seasonal component exist in the time series. The trend is depicted by the slow decline in the ACF as the lags increase while “scalloped” shape indicates seasonality in the data.

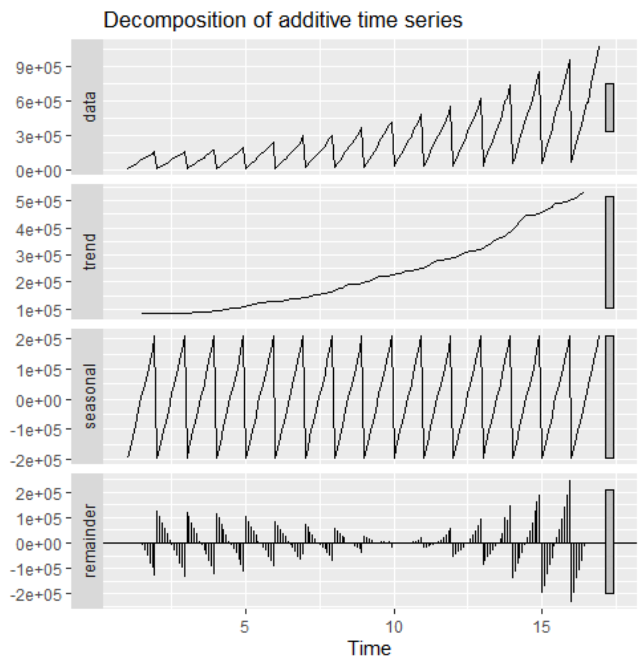

Figure 4.3: Decomposition of tax revenue in Kenya

The decomposed tax revenue in Kenya indicated in Figure 4.3 confirms existence of trend, seasonal and random effects components in the tax revenue data. The first panel shows the original data series. The second panel plots the trend component and we see a clear trend pattern. The third panel shows the very strong seasonal component in the tax revenue data. The fourth panel shows the remainder component or random effects, which is left after removing the trend and seasonal components. The non-stationary tax revenue data was applied in order to estimate trend and seasonal components. Differencing the data to make it stationary was not undertaken since it would have removed trend component.

4.2. Modelling and Forecasting

4.2.1. Model Estimation

The seasonal naïve model with deterministic trend in equation (3.3) is given by:

$$Y_t = c + Y_{t-s} + \beta t + a_t$$

The estimate of β in equation (3.4) is given by:

$$\hat{\beta} = \frac{\sum_{t=1}^{T} tY_t - c\sum_{t=1}^{T} t - \sum_{t=1}^{T} tY_{t-s}}{\sum_{t=1}^{T} t^2}$$

Based on the tax revenue data, the values for c (mean) and $\hat{\beta}$ are 2540.63 and -1645.34 respectively. The length of seasonal cycle s = 12. Replacing c, $\hat{\beta}$ and s in equation (3.3), we obtain the estimated equation as:

$$\widehat{Y_t} = 2540.63 + Y_{t-12} - 1645.34t \tag{3.31}$$

4.2.2. Model Validation

We employed in-sample validation to the tax revenue data to compare forecasting performance of seasonal naïve model with deterministic trend with benchmark models (mean model, naïve model and seasonal naïve model). The experimental data was divided into the training data and the testing data. The training data comprised of 156 data points from July 2000 to June 2014 while the testing data comprised of 24 data points from July 2014 to June 2016.

We first established the baseline forecasts using mean, naïve and seasonal naïve methods and compared with seasonal naïve with deterministic trend forecast:

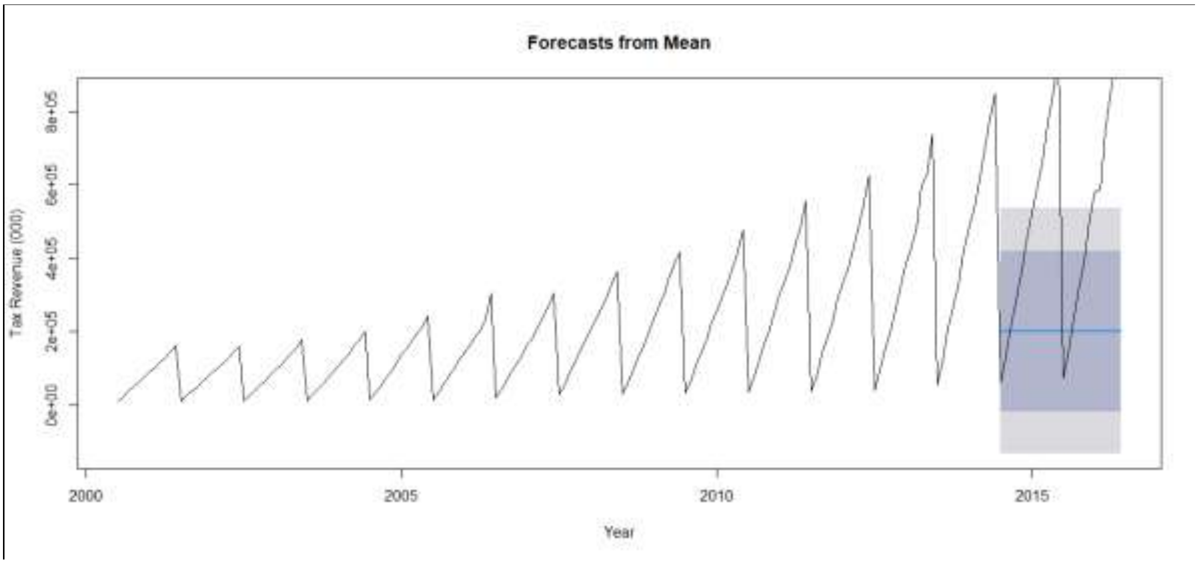

Figure 4.4 Tax revenue, mean forecasts and confidence intervals

The tax revenue forecasts for the two years (July 2014 -June 2016) using the mean method is indicated in Figure 4.4. The 80% confidence limits and 95% confidence limits of the forecast estimates are indicated by dark colour and lighter colour, respectively. Based on forecasts, although mean forecast has low intervals it failed to consider the high seasonality in the data. The comparisons of mean forecasts and actual data depict huge deviation since the model fails to consider seasonality of the tax revenue data. The forecasts assume that the future tax revenue is same as the mean of the historical data. i.e. the monthly tax revenue forecasts from July 2014 to June 2016 is equal to the average of the tax revenue from July 2000 to June 2014. Consequently, the model may not be suitable for forecasting tax revenue data exhibiting seasonality.

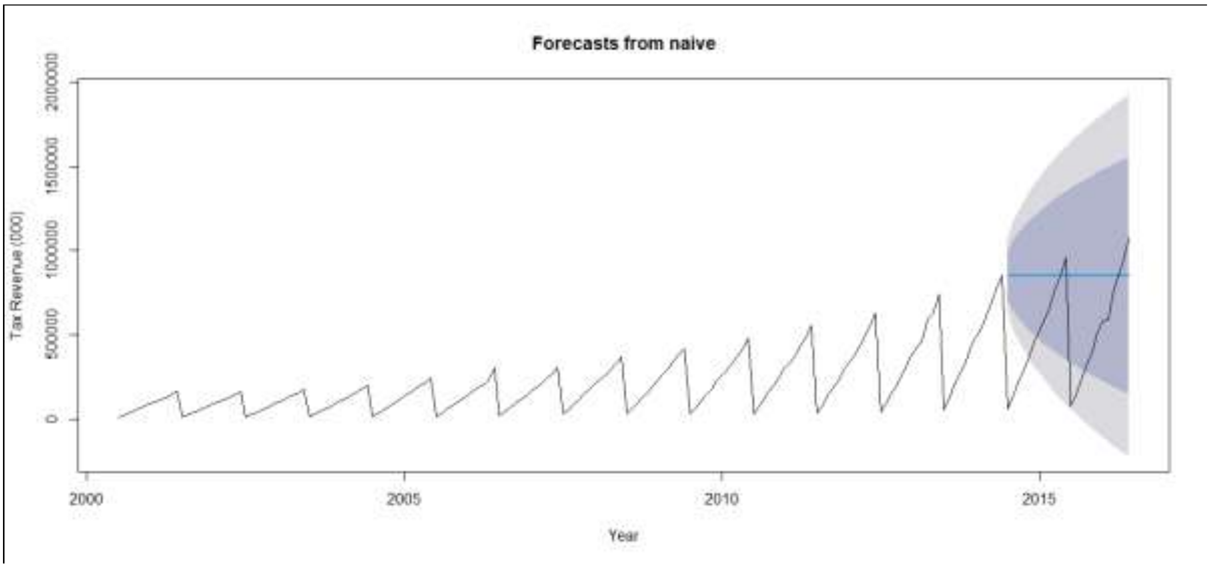

Figure 4.5: Tax revenue, naive forecasts and confidence intervals

The tax revenue forecasts for the two years (July 2014 -June 2016) using naïve method is portrayed in Figure 4.5. The 80% confidence limits and 95% confidence limits of the forecast estimates are indicated by dark colour and lighter colour, respectively. Based on forecasts, naïve model has very high prediction intervals which makes it a worse model for tax revenue data. The comparisons of forecasts and actual data depict huge deviation since the model fails to consider seasonality of the data. The model sets forecasts to be same as the value of the last observation of the historical data. i.e. the monthly tax revenue forecasts from July 2014 to June 2016 is equal to the tax revenue of June 2014. Consequently, the model may not be suitable for forecasting tax revenue data exhibiting seasonality.

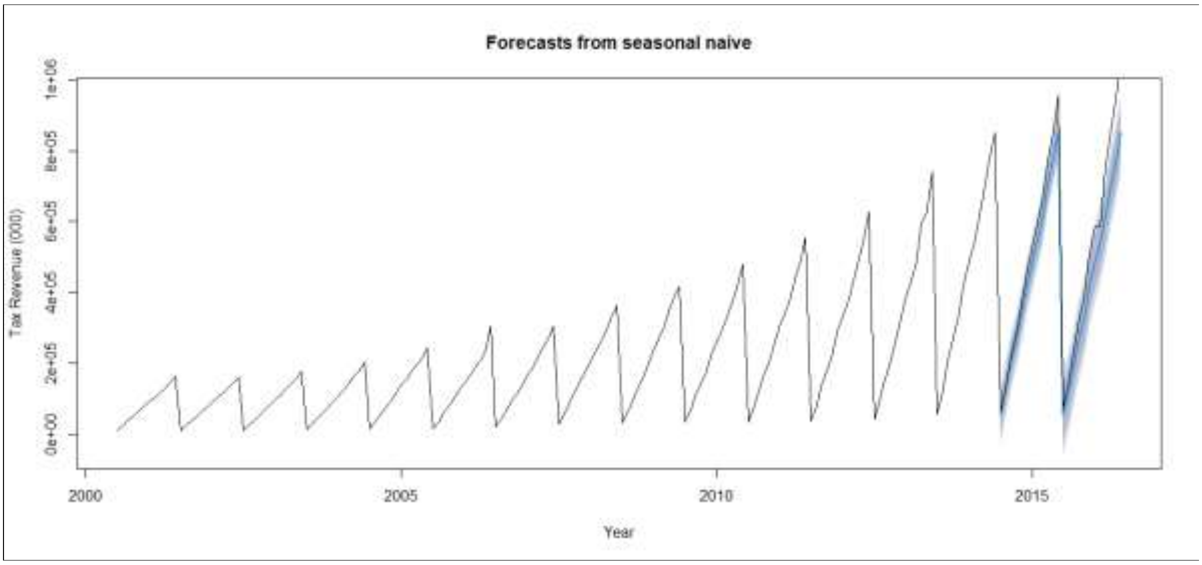

Figure 4.6: Tax revenue, seasonal naïve forecasts and confidence intervals

The tax revenue forecasts for the two years (July 2014-June 2016) using seasonal naïve method is indicated in Figure 4.6. The 80% confidence limits and 95% confidence limits of the forecast estimates are indicated by dark colour and lighter colour, respectively. Based on forecasts, seasonal naïve forecasts take cognisance of seasonality as opposed to mean and naïve forecasts. The comparisons of forecasts and actual data depict small variation since the model considers seasonality of the data. The model forecasts are same as the last observed value from the same period last year. i.e. the monthly tax revenue forecast for June 2014 is same as monthly tax revenue for June 2015 and June 2016; monthly tax for August 2013 is same as August 2014, etc. Consequently, the model may not be suitable for forecasting tax revenue data exhibiting both seasonality and trend as it fails to consider the trend component in the data. In addition, the seasonal naïve model exhibit memory loss as the forecast values depend on the previous realization.

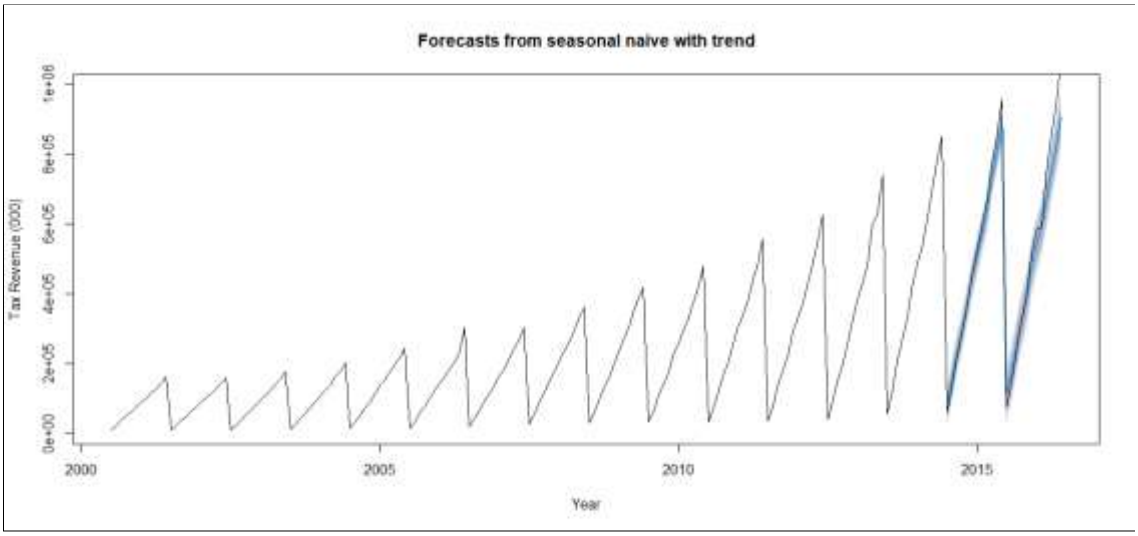

Figure 4.7: Tax revenue, seasonal naive forecasts with trend and confidence intervals

The tax revenue forecasts for the two years (July 2014-June 2016) using seasonal naïve method with trend is indicated in Figure 4.7. The 80% confidence limits and 95% confidence limits of the forecast estimates are indicated by dark colour and lighter colour, respectively. Based on forecasts, seasonal naïve forecasts with trend take cognisance of both seasonality and trend in the data as opposed to seasonal naïve forecasts. The comparisons of forecasts and actual data depict small variation since the model considers trend and seasonality in the data. The model adds trend component to the seasonal naïve model. i.e. the monthly tax revenue forecast for June 2013 is same as monthly tax revenue for June 2014 plus trend component. Consequently, the model is most suitable for forecasting tax revenue data exhibiting both seasonality and trend components. The majority actual tax revenue values lie within 95 per cent confidence interval as indicated in **Appendix 2**.

The experimental results indicated that seasonal naïve method with trend has higher forecasting accuracy than mean, naïve and seasonal methods, which is shown in Table 4.1. below:

Table *4.1: Forecast Metrics Comparison*

Method	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Residual SD
Mean	3.749229e-11	170691.1	131544.8	-150.7847	179.5697	4.701385	0.738251	171201.6355
Naïve	5010.81	111886.7	57363.98	-82.60474	120.9506	2.050177	-0.1072557	112205.1692
Seasonal naïve	25891.11	38917.47	25987.41	11.00796	11.16718	0.9287847	0.7828172	40385.64101
Seasonal naïve with deterministic trend	**4.889521**	**28160.02**	**20606.97**	-19.13876	23.93777	**0.7364889**	0.7696307	**29317.15218**

The forecasting accuracy was measured by RMSE, MAE and Residual Standard Deviation as metrics to compare different models. Looking at forecast from all the models in table 4.1, the forecast from seasonal naïve model with deterministic trend is showing a better performance (lower RMSE, MAE & smaller residual standard deviation) hence superior forecasting model compared to mean, naïve and seasonal naïve models.

The AIC, AICc and BIC were applied to check the goodness of fit of the suggested models as indicated in Table 4.2 below:

Table *4.2: AIC, AICc and BIC for four suggested models*

Method	AIC	AICc	BIC
Mean	4528.76	4528.83	4535.01
Naïve	4359.75	4359.78	4362.87
Seasonal naïve	3753.86	3753.89	3756.91
Seasonal naïve with deterministic trend	**3654.92**	**3654.99**	**3661.01**

The AIC, AICc and BIC values for mean model, naïve model, seasonal naïve model, and seasonal naïve model with deterministic trend is illustrated in Table 4.2. The model diagnostics indicate that seasonal naïve model with deterministic trend has the smallest AIC, AICc and BIC values in comparison with the other competing models. This implies that the fitted values using seasonal naïve with deterministic trend model are very close to the actual values. Therefore, the model is appropriate for forecasting monthly tax revenue in Kenya.

4.2.3. Residual analysis

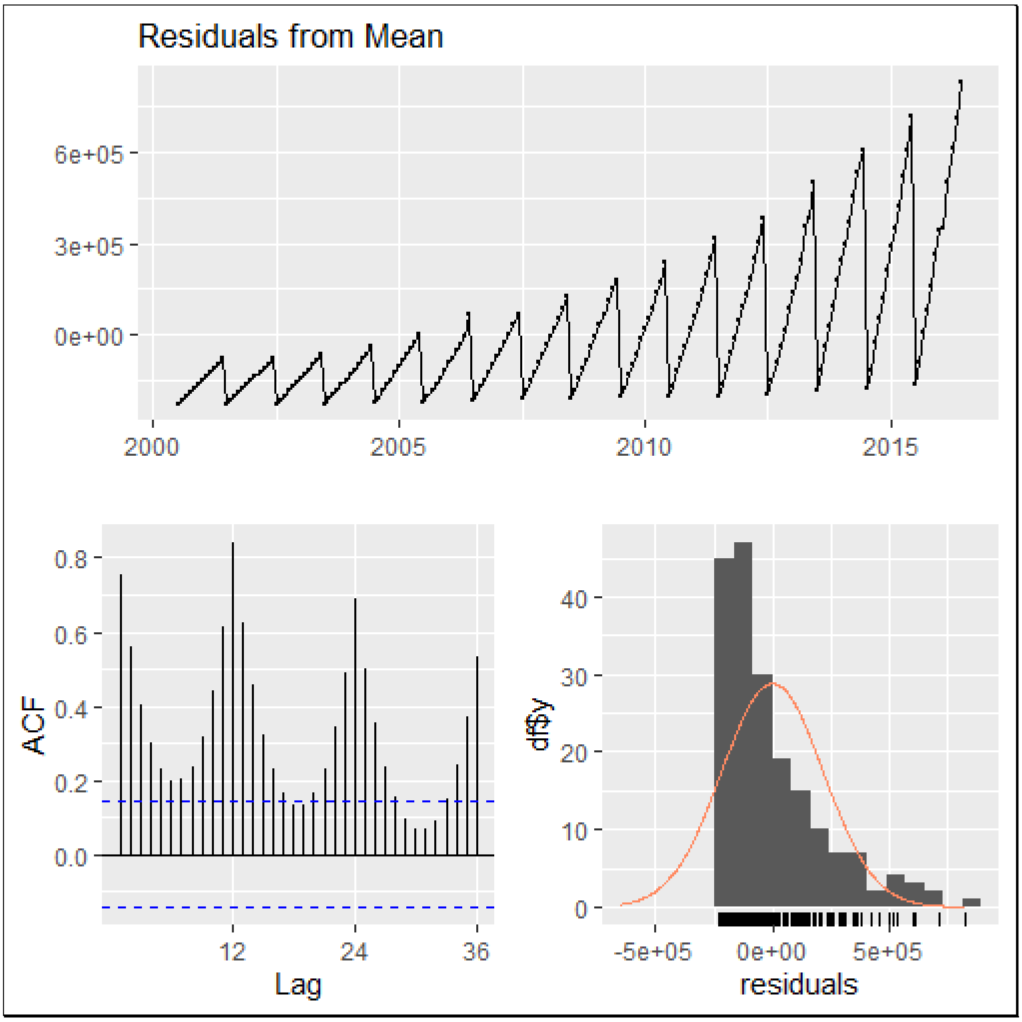

Figure 4.8: Residuals from fitted mean model for tax revenue data

Figure 4.8 shows residuals of fitted mean model of tax revenue. The graph shows that residuals and ACF from the mean method replicate the tax revenue data structure. The residuals are not normally distributed as shown in the histogram i.e., they are right tailed. Therefore, the computed prediction intervals, assuming a normal distribution may not be accurate hence forecasts may probably not be good.

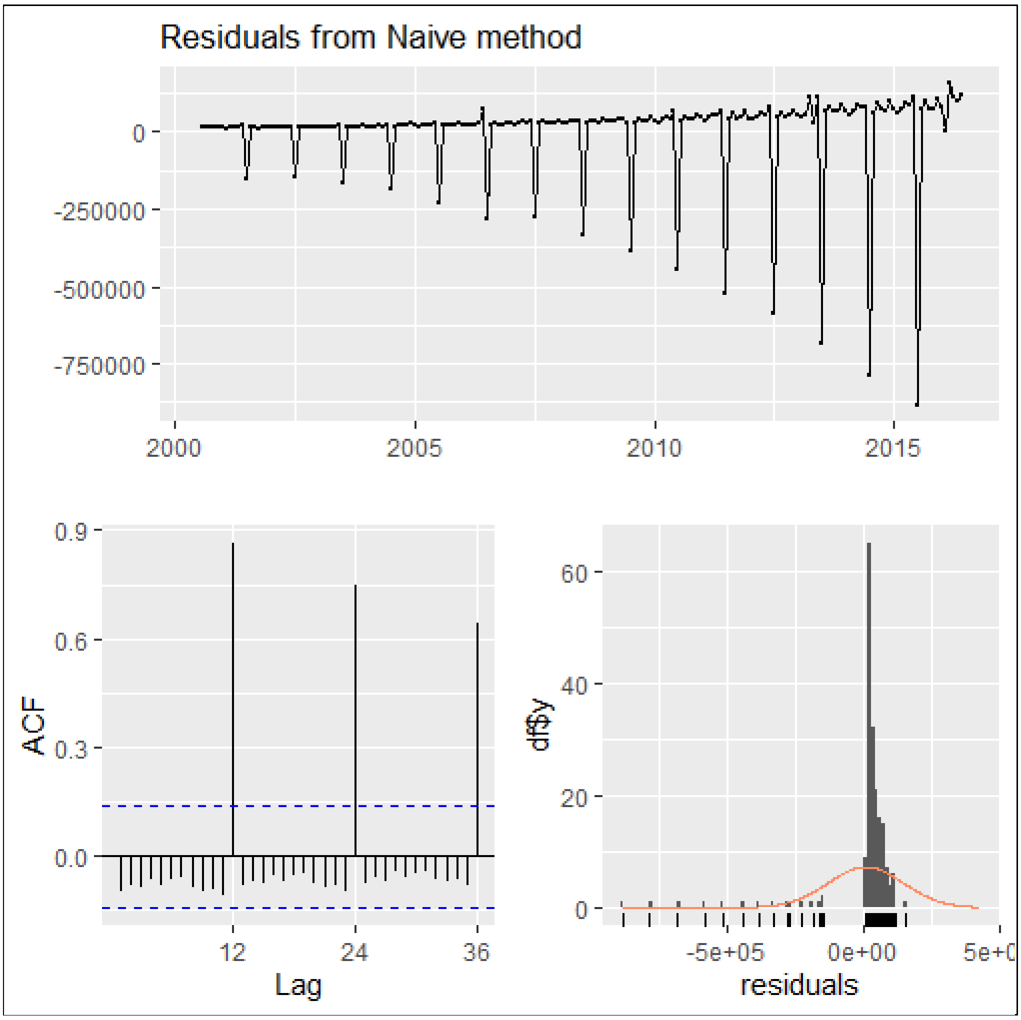

Figure 4.9: Residuals from fitted naïve model for tax revenue data

Figure 4.9 shows residuals of fitted naïve model of tax revenue. The graph shows that residuals from the naïve method are stationary on mean. However, the variation in the plot as depicted by continuous increase towards the right of the graph indicates that residuals are non-stationary on variance. The ACF of residuals indicate spikes at lags of multiples from 12 (i.e. 12, 24, 36, etc.) outside insignificant zone for ACF plot hence residuals are not random. The residuals are not normally distributed as shown in the histogram, just as residuals from the mean method i.e., they are right tailed. Therefore, the forecasts may be inaccurate.

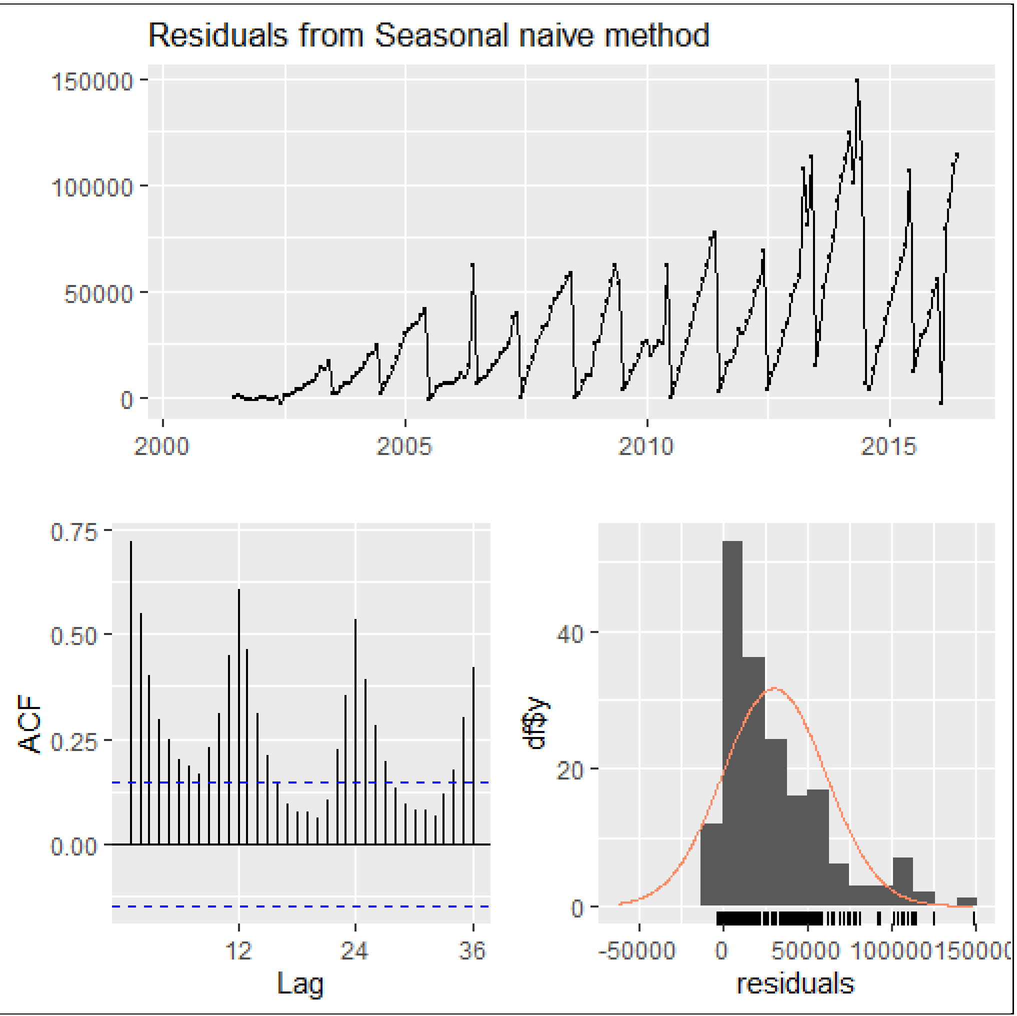

Figure 4.10: Residuals from fitted seasonal naïve model for tax revenue data

Figure 4.10 shows residuals of fitted seasonal naïve model of tax revenue which produced similar residuals when trend component was incorporated. The graph shows that residuals from the seasonal naïve method are not stationary due to the presence of the trend component in the tax revenue data. The ACF of residuals indicate presence of both seasonal and trend components as evidenced by attenuating sine wave patterns. The residuals tend to follow normal distribution as indicated by the histogram hence forecasts from this method may probably yield better results compared to mean and naïve models.

The Ljung-Box test values for the model residuals is indicated below:

Table *4.3: Ljung-Box test statistic values for suggested models*

Method	**Ljung-Box Statistic (Q*)**	**Total lags used**	**P value**
Mean	670.67	24	< 2.2e-16
Naïve	266.75	24	< 2.2e-16
Seasonal naïve	680.63	24	< 2.2e-16
Seasonal naïve with deterministic trend	486.93	24	< 2.2e-16

Table 4.3 indicates the Ljung-Box test statistic (Q*) values at a lag of 24 for four suggested models. The p-values are relatively small (p-value< 2.2e-16) and less than 0.05. We therefore conclude that the residuals are not independently distributed hence exhibit serial correlation.

4.2.4. Forecasting tax revenue

4.2.4.1. Seasonal Naïve forecast with deterministic trend

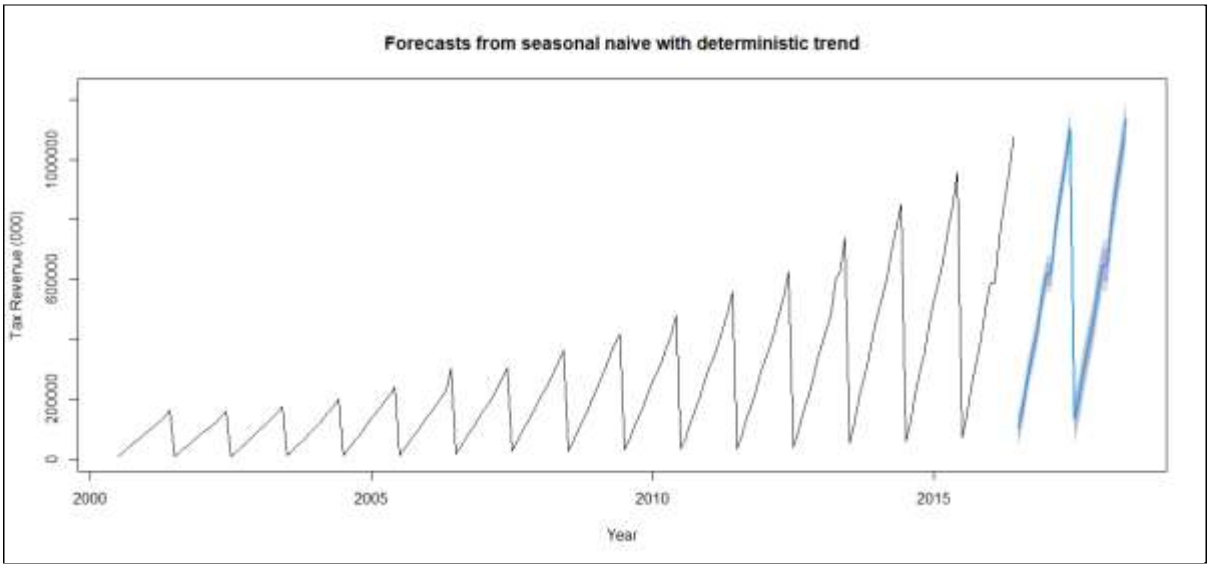

Figure 4.11: Tax revenue, seasonal naïve forecasts with deterministic trend and confidence intervals

The tax revenue forecasts for the two years (July 2016 -June 2018) using the seasonal naïve with trend method is indicated in Figure 4.11 (see **Appendix 3** for two-year revenue forecast figures). The 80% confidence limits and 95% confidence limits of the forecast estimates are indicated by dark colour and lighter colour, respectively. According to the forecasts, we can deduce that cumulative monthly tax revenue in Kenya will continue to grow overtime as evidenced by increasing linear trend.

4.2.4.2. Forecasting trend

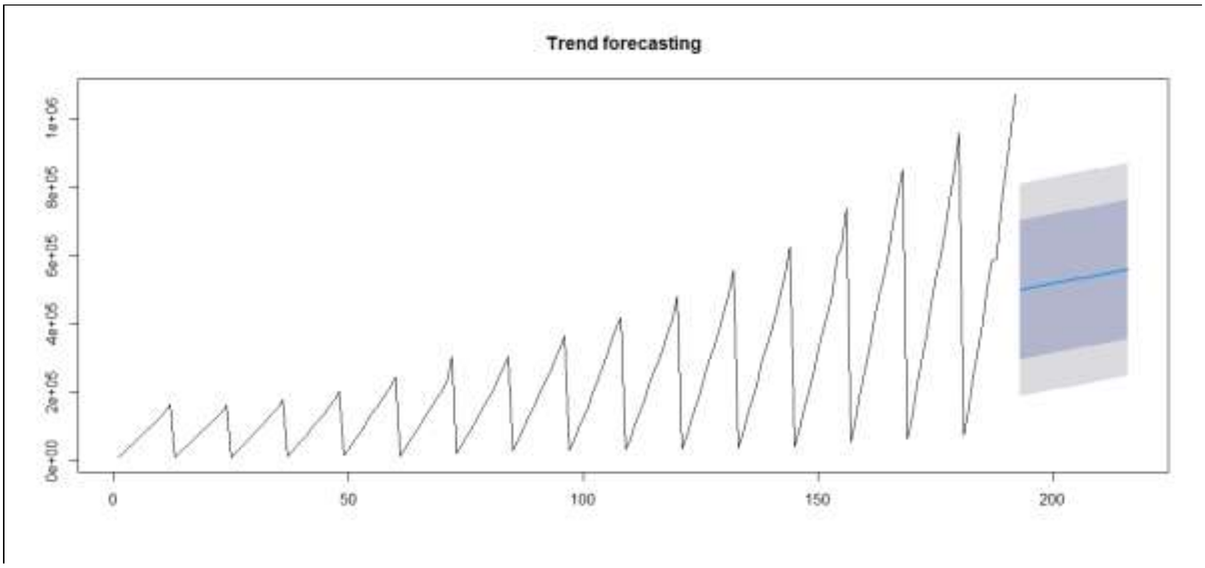

Figure 4.12: Tax revenue, trend forecasts and confidence intervals

The tax revenue trend forecasts for the two years (July 2016 -June 2018) is indicated in Figure 4.12 (see **Appendix 4** for two-year revenue trend forecast figures). The 80% confidence limits and 95% confidence limits of the forecast estimates are indicated by dark colour and lighter colour, respectively. The increasing linear trend depicts growth in cumulative monthly tax revenue.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

The seasonal naïve model with deterministic trend was the best model for forecasting tax revenue data since it recognizes both seasonal and trend components in the data. This is because the developed model returned the least forecast error compared to seasonal naïve model by 99.98% (for ME), 27.64% (for RSME), 20.70% (for MAE) and 20.70% (for MASE). Further, the study established that cumulative monthly tax revenue will grow over time.

5.2. Recommendations

This study recommends application of seasonal naïve model with deterministic trend for forecasting monthly tax revenue in Kenya. Modelling tax revenue causality considering other variables which account for seasonality such as inflation, exchange rates, public expenditure and public debt should be explored as future area of study.

REFERENCES

[1] Anderson, D. R., Burnham, K. P., & White, G. C. (1998). Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies. *Journal of Applied Statistics*, 25(2), 263-282.

[2] Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

[3] Batóg, B., & Batóg, J. (2021). *Regional Government Revenue Forecasting: Risk Factors of Investment Financing.* Risks, 9(12), 210.

[4] Belinga, V., Benedek, D., Mooij, R. d., & Norregaard, J. (2014). Tax Buoyancy in OECD countries *IMF working paper*.

[5] Bolnick, B. and Haughton, J. (1998). Tax policy in Sub-Saharan Africa: Examining the role of excise taxation. *African Economic Policy, Discussion Paper No. 2,* Harvard Institute for International Development, Cambridge, MA.

[6] Botrić, V., & Vizek, M. (2012*)*. Forecasting fiscal revenues in a transition country: the case of Croatia. *Zagreb International Review of Economics & Business, 15(1), 23-36.*

[7] Brian, S. E & David, C. H. (2005). *Least Squares Estimation.* https://stat.ethz.ch/~geer/bsa199_o.pdf. Accessed on 16.04.2017

[8] Chimilila, C. (2017). Forecasting tax revenue and its volatility in Tanzania. *African Journal of Economic Review, 5(1), 84-109.*

[9] Frechtling, D.C. (2001). *Forecasting Tourism Demand: Methods and Strategies*. Woburn: Butterworth Heinemann, Elseivier group.

[10] Glenn, Jenkins, Kuo, Chun-Yan & Shukla, Gangadhar. (2000). Tax Analysis and Revenue Forecasting. *JDI Executive Programs, Development Discussion Papers*

[11] Government of Kenya (GoK). (2009). *Income Tax Act. Chapter 472 of the Laws of Kenya*. Nairobi: Government Printers.

[12] Government of Kenya GOK. (1986*). "Sessional Paper No 1 of 1986 on Economic Management for Renewed Growth"*, Government Printer, Nairobi.

[13] Hurvich, C.., & Tsai, C (1989). *Regression and Time Series Model Selection in Small Sample*. *Biometrika*, 76, 297– 307.

[14] Hyman, D. N. (1987). *Public Finance: A Contemporary Application of Theory of Policy (2nd ed.)*. Chicago: The Dryden Press.

[15] Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice,* 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 24.10.2021

[16] Institute of Economic Affairs. (2012). *Tax Incentives and Exemption Regime in Kenya: Is It Working?* Issue No. 30

[17] International Monetary Fund (2011). *Revenue Mobilization in Developing in Developing Countries.* Retrieved from https://www.imf.org/external/np/pp/eng/2011/030811.pdf. Accessed on 16.04.2017

[18] Kenya Institute for Public Policy Research and Analysis (2016). *Kenya Economic Report 2016: Fiscal Decentralization in support of Devolution*. Nairobi: KIPPRA

[19] Kenya Institute for Public Policy Research and Analysis (2020). *Excise Taxation in Kenya: A Situation Analysis*. Nairobi: KIPPRA

[20] Kenya National Bureau of Statistics (2016). *Economic Survey*. Nairobi: KNBS

[21] KIPPRA. (2004a). "*Tax Compliance Study*", Tax Policy Unit, Macroeconomics Division, Kenya Institute of Public Policy Research and Analysis, Nairobi, Kenya.

[22] KIPPRA. (2004b). *"Tax Reform Experience and the Reform Agenda for Kenya", Policy Brief No 3*, Kenya Institute of Public Policy Research and Analysis, Nairobi.

[23] KRA (2021). *Value Added Tax (VAT).* https://kra.go.ke/individual/filing-paying/types-of-taxes/value-added-tax. Accessed on 24.10.2021

[24] Li-Xia, L., Yi-Qi, Z., & Liu, X. Y. (2011). *Tax forecasting theory and model based on SVM optimized by PSO. Expert Systems with Applications*, *38*(1), 116-120.

[25] Lu, S., Cai, Z. J., & Zhang, X. B. (2009, August). *Application of GA-SVM time series prediction in tax forecasting.* In 2009 2nd IEEE International Conference on Computer Science and Information Technology (pp. 34-36). IEEE.

[26] Luchko, M., Drozd, I., Plutytska, K., Ruska, R., & Vovk, I. (2021). Analysis and modelling of value added tax revenues on imports: Some issues of application in Ukraine. *International Journal of Production Management and Engineering, 9(1), 37-46.*

[27] Mamula, M. (2015). Modelling and forecasting international tourism demand-evaluation of forecasting performance. *International Journal of Business Administration, 6(3), 102.*

[28] Molapo, M. A., Olaomi, J. O., & Ama, N. O. (2019). Bayesian vector auto-regression method as an alternative technique for forecasting South African tax revenue. *Southern African Business Review, 23(1).*

[29] Moyi, E. D and Muriithi M.K. (2003). "Tax Reforms and Revenue Mobilization in Kenya", *Research Paper No 131, African Economic Research Consortium (AERC*), Nairobi, Kenya.

[30] Moyi, E., & Ronge, E., (2006). *Taxation and Tax Modernization in Kenya: A Diagnosis of Performance and Options for Further Reforms*, Institute of Economic Affairs.

[31] Musgrave, R., & Musgrave, P. (1989). *Public finance in theory and practice* (5th ed.). New York: McGraw-Hill Book Co.

[32] Mutua, J. M. 2012. *A Citizen's Handbook on Taxation in Kenya*: Institute of Economic Affairs, Nairobi, Kenya.

[33] Nandi, B. K., Chaudhury, M., & Hasan, G. Q. (2014). Univariate time series forecasting: A study of monthly tax revenue of Bangladesh. *East West Journal of Business and Social Studies, 1.*

[34] Nau, R. (2017). *Seasonal random trend model.* https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/411seart.htm. Accessed on 16.04.2017

[35] Nau, R. (2017). *Seasonal random walk model.* https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/411searw.htm. Accessed on 16.04.2017

[36] Nau, R. (2017). *Working with seasonal data*. https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/411l397.htm. Accessed on 16.04.2017

[37] Nazmi, N., & Leuthold, J. H. (1988). Forecasting economic time series that require a power transformation: Case of state tax receipts. *Journal of Forecasting*, 7(3), 173-184.

[38] Ochieng, O. (2015, May 21). Kenya loses over Sh600bn every year in tax evasion. *Daily Nation,* Retrieved from https://www.nation.co.ke/business/Kenya-loses-over-Sh600bn-every-year-in-tax-evasion/996-2724818-5jswbtz/index.html. Accessed on 16.04.2017

[39] Osoro, N. E. (1993). *Revenue productivity implications of tax reform in Tanzania*. Nairobi; African Economic Research Consortium

[40] PWC (2021). *Import (Customs) duty*. https://taxsummaries.pwc.com/kenya/corporate/other-taxes. Accessed on 24.10.2021

[41] Sakamoto, Y., Ishinguro, M., & Kitagawa, G. (1986). *Akaike Information Criterion Statistics.* D. Reidel Publishing Company.

[42] Schwarz, G. (1978). *Estimating the Dimension of a Model.* Annals of Statistics, 6(2), 461–464.

[43] Shahnazarian, H., Solberger, M., & Spånberg, E. (2017). Forecasting and analysing Corporate tax revenues in sweden using Bayesian VAR Models. *Finnish economic papers, 28(1).*

[44] Streimikiene, D., Rizwan Raheem, A., Vveinhardt, J., Pervaiz Ghauri, S., & Zahid, S. (2018). Forecasting tax revenues using time series techniques–a case of Pakistan. *Economic research-Ekonomska istraživanja, 31(1), 722-754.*

[45] Ticona, W., Figueiredo, K., & Vellasco, M. (2017, August). Hybrid model based on genetic algorithms and neural networks to forecast tax collection: application using endogenous and exogenous variables. *In 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON) (pp. 1-4). IEEE.*

[46] Todaro, P. M. & Smith, C. S. (2003). *Economic development*. Eighth Edition. Addison Wesley

[47] Tsay, R. S. (2010). *Analysis of Financial Time Series*, 3rd ed., Wiley, Hoboken, NJ.

[48] Wagacha, M. (1999). Fiscal policies and their macroeconomic implications for Kenya. In P. Kimuyu, M. Wagacha and O. Abagi, eds., *Kenya's Strategic Policies for the 21st Century: Macroeconomic and Sectoral Choices.* Nairobi: Institute of Policy Analysis and Research (IPAR).

[49] Wawire, N. H. W. (2003). Trends in Kenya's tax ratios and tax effort indices, and their implication for future tax reforms. *Egerton Journal, 4, 256-279*

[50] Wawire, N. H. W. (2017). Determinants of Value Added Tax Revenue in Kenya. *Journal of Economics Library, 4(3), 322-344.*

[51] World Bank (1990). Argentina: *Tax Policy for Stabilization and Economic Recovery.* Country Study. Washington, D.C.: The World Bank.

Appendix 1: Cumulative tax revenue in Kenya from July 2000 to June 2016

Year/Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000							9,998	22,346	37,475	49,681	63,682	76,969
2001	90,241	101,786	115,517	128,085	143,685	163,171	10,026	23,660	37,230	48,761	62,550	76,215
2002	90,212	102,231	114,716	127,584	143,466	160,394	10,807	24,206	38,981	52,478	66,299	81,948
2003	96,782	110,021	124,775	141,737	156,917	176,999	12,716	26,151	43,578	58,616	72,868	91,661
2004	108,080	123,372	140,719	161,746	177,828	201,544	14,865	32,475	53,463	72,842	91,464	116,617
2005	138,298	155,059	174,862	196,846	216,775	242,927	14,154	33,861	58,129	78,428	97,976	123,436
2006	145,000	163,386	186,205	206,676	231,776	305,040	20,787	42,231	67,924	90,428	112,970	140,221
2007	165,473	185,723	211,973	243,952	271,767	305,040	29,179	55,911	87,194	116,814	146,096	174,197
2008	207,445	232,305	260,903	295,750	328,450	363,621	29,452	57,410	94,667	126,818	156,692	199,197
2009	233,418	271,011	305,986	350,074	390,213	417,354	33,410	63,724	106,704	141,955	176,478	224,451
2010	259,767	290,690	329,879	376,403	415,380	479,568	33,845	70,044	119,470	162,648	203,025	257,979
2011	302,869	339,789	385,369	438,905	489,735	557,171	36,753	78,451	135,905	179,392	225,147	289,675
2012	332,835	375,669	426,272	488,853	544,007	626,445	40,395	89,569	151,062	200,992	255,823	324,709
2013	380,793	428,830	483,594	596,623	625,524	739,894	55,113	121,100	202,790	267,076	330,930	416,966
2014	484,350	540,830	608,312	697,686	774,383	851,804	61,395	124,714	216,264	290,720	357,398	453,687
2015	528,793	591,738	666,851	762,506	845,831	958,186	73,656	146,102	245,689	321,369	397,113	504,095
2016	584,481	589,149	745,709	855,143	955,465	1,072,417						

Appendix 2: Actual, forecasts and confidence intervals from Seasonal naïve model with deterministic trend

			95% Confidence Limit	
Period	**Actual (Ksh)**	**Forecast (Ksh)**	**Lower**	**Upper**
Jul-14	61,395.00	82,989.31	25,528.75	140,449.90
Aug-14	124,714.00	148,976.31	91,515.75	206,436.90
Sep-14	216,264.00	230,666.31	173,205.75	288,126.90
Oct-14	290,720.00	294,952.31	237,491.75	352,412.90
Nov-14	357,398.00	358,806.31	301,345.75	416,266.90
Dec-14	453,687.00	444,842.31	387,381.75	502,302.90
Jan-15	528,793.00	512,226.31	454,765.75	569,686.90
Feb-15	591,738.00	568,706.31	511,245.75	626,166.90
Mar-15	666,851.00	636,188.31	578,727.75	693,648.90
Apr-15	762,506.00	725,562.31	668,101.75	783,022.90
May-15	845,831.00	802,259.31	744,798.75	859,719.90
Jun-15	958,186.00	879,680.31	822,219.75	937,140.90
Jul-15	73,656.00	110,865.62	29,604.11	192,127.10
Aug-15	146,102.00	176,852.62	95,591.11	258,114.10
Sep-15	245,689.00	258,542.62	177,281.11	339,804.10
Oct-15	321,369.00	322,828.62	241,567.11	404,090.10
Nov-15	397,113.00	386,682.62	305,421.11	467,944.10
Dec-15	504,095.00	472,718.62	391,457.11	553,980.10
Jan-16	584,481.00	540,102.62	458,841.11	621,364.10
Feb-16	589,149.00	596,582.62	515,321.11	677,844.10
Mar-16	745,709.00	664,064.62	582,803.11	745,326.10
Apr-16	855,143.00	753,438.62	672,177.11	834,700.10
May-16	955,465.00	830,135.62	748,874.11	911,397.10
Jun-16	1,072,417.00	907,556.62	826,295.11	988,818.10

Appendix 3: Seasonal naïve forecasts with deterministic trend for tax revenue in Kenya from July 2016 to June 2018

		80% Confidence Limit		95% Confidence Limit	
Month	**Forecast (Ksh.)**	**Lower**	**Upper**	**Lower**	**Upper**
Jul-16	104,143.50	64,869.80	143,417.20	44,079.55	164,207.50
Aug-16	176,589.50	137,315.80	215,863.20	116,525.55	236,653.50
Sep-16	276,176.50	236,902.80	315,450.20	216,112.55	336,240.50
Oct-16	351,856.50	312,582.80	391,130.20	291,792.55	411,920.50
Nov-16	427,600.50	388,326.80	466,874.20	367,536.55	487,664.50
Dec-16	534,582.50	495,308.80	573,856.20	474,518.55	594,646.50
Jan-17	614,968.50	575,694.80	654,242.20	554,904.55	675,032.50
Feb-17	619,636.50	580,362.80	658,910.20	559,572.55	679,700.50
Mar-17	776,196.50	736,922.80	815,470.20	716,132.55	836,260.50
Apr-17	885,630.50	846,356.80	924,904.20	825,566.55	945,694.50
May-17	985,952.50	946,678.80	1,025,226.20	925,888.55	1,046,016.50
Jun-17	1,102,904.50	1,063,630.80	1,142,178.20	1,042,840.55	1,162,968.50
Jul-17	134,631.00	79,089.61	190,172.40	49,687.76	219,574.30
Aug-17	207,077.00	151,535.61	262,618.40	122,133.76	292,020.30
Sep-17	306,664.00	251,122.61	362,205.40	221,720.76	391,607.30
Oct-17	382,344.00	326,802.61	437,885.40	297,400.76	467,287.30
Nov-17	458,088.00	402,546.61	513,629.40	373,144.76	543,031.30
Dec-17	565,070.00	509,528.61	620,611.40	480,126.76	650,013.30
Jan-18	645,456.00	589,914.61	700,997.40	560,512.76	730,399.30
Feb-18	650,124.00	594,582.61	705,665.40	565,180.76	735,067.30
Mar-18	806,684.00	751,142.61	862,225.40	721,740.76	891,627.30
Apr-18	916,118.00	860,576.61	971,659.40	831,174.76	1,001,061.30
May-18	1,016,440.00	960,898.61	1,071,981.40	931,496.76	1,101,383.30
Jun-18	1,133,392.00	1,077,850.61	1,188,933.40	1,048,448.76	1,218,335.30

Appendix 4: Trend forecasts for tax revenue in Kenya from July 2016 to June 2018

		80% Confidence Limit		95% Confidence Limit	
Month	**Forecast (Ksh.)**	**Lower**	**Upper**	**Lower**	**Upper**
Jul-16	499,247.00	295,921.30	702,572.80	188,287.10	810,207.00
Aug-16	501,924.00	298,598.30	705,249.80	190,964.10	812,884.00
Sep-16	504,601.00	301,275.30	707,926.80	193,641.10	815,561.00
Oct-16	507,278.00	303,952.30	710,603.80	196,318.10	818,238.00
Nov-16	509,955.00	306,629.30	713,280.80	198,995.10	820,915.00
Dec-16	512,632.00	309,306.30	715,957.80	201,672.10	823,592.00
Jan-17	515,309.00	311,983.30	718,634.80	204,349.10	826,269.00
Feb-17	517,986.00	314,660.30	721,311.80	207,026.10	828,946.00
Mar-17	520,663.00	317,337.30	723,988.80	209,703.10	831,623.00
Apr-17	523,340.00	320,014.30	726,665.80	212,380.10	834,300.00
May-17	526,017.00	322,691.30	729,342.80	215,057.10	836,977.00
Jun-17	528,694.00	325,368.30	732,019.80	217,734.10	839,654.00
Jul-17	531,371.00	328,045.30	734,696.80	220,411.10	842,331.00
Aug-17	534,048.00	330,722.30	737,373.80	223,088.10	845,008.00
Sep-17	536,725.00	333,399.30	740,050.80	225,765.10	847,685.00
Oct-17	539,402.00	336,076.30	742,727.80	228,442.10	850,362.00
Nov-17	542,079.00	338,753.30	745,404.80	231,119.10	853,039.00
Dec-17	544,756.00	341,430.30	748,081.80	233,796.10	855,716.00
Jan-18	547,433.00	344,107.20	750,758.80	236,473.10	858,393.00
Feb-18	550,110.00	346,784.20	753,435.80	239,150.10	861,070.00
Mar-18	552,787.00	349,461.20	756,112.80	241,827.10	863,747.00
Apr-18	555,464.00	352,138.20	758,789.80	244,504.10	866,424.00
May-18	558,141.00	354,815.20	761,466.80	247,181.10	869,101.00
Jun-18	560,818.00	357,492.20	764,143.80	249,858.10	871,778.00