ISSN (online) 2664-9535 (print) 2664-9527 https://atcr.kra.go.ke

Use of Time Series Models on Forecasting of Value Added Tax Revenue in Kenya Revenue Authority

Muthuri Evans Kithure[1]

[1] Kenya Revenue Authority, Nairobi, Kenya

Received 30 June 2021
Accepted for publication 14 July 2021
Published 18 August 2021

Abstract

Taxation is one of the means by which governments finance their expenditure by imposing charges on citizens and corporate entities. Tax revenue forecasting plays a central role in annual budget formulation. It provides policy makers and fiscal planners with the data needed to guide borrowing, use accumulated reserves, or specify monetary measures to balance the budget. Therefore, it is necessary for a government to forecast the revenue it collects for planning purposes. Kenya Revenue Authority (KRA), is an agency of the government of Kenya that is responsible for the assessment, collection and accounting for of all revenues that are due to government, in accordance with the laws of Kenya. The main objective of this study was to fit time series models in the data series of revenue collections and establish their effectiveness as far as revenue forecasting is concerned. The study used the monthly Value Added Tax (VAT) collections data from the financial year 2009/2010 to 2015/2016 with the general objective of exploring patterns in the data such as trend, seasonal components, cycles among others and further establish a suitable forecasting model which can be used to predict the amount of VAT revenue to be collected in a certain specified period. The first step was to check if the series was stationary by using Dickey Fuller Test, thereafter transformation by differencing if the series was not stationary. The order of the models was tentatively chosen by analyzing the ACF and PACF plots of the data. This resulted to AR(3) model, MA(1) model, ARMA(3,1) model and ARIMA(3,1,1) model which were fitted to the data series. In order to select the best potential model, different statistics were used like BIC, AIC, AICc, and forecast accuracy measures like ME, MAE and MPE. ARIMA(3,1,1) was selected as the best model compared to the other models. Diagnostic check was made to test for correlation and normal distribution of the residuals using Box-Ljung test, Q-Q plots and Shapiro wilk test and the results showed normal distribution of the residuals with no correlations for all the models. Using the models, forecast of VAT revenue collection in the financial year 2016/2017 was made and the forecasted values compared with the revenue collections that were made in the respective financial year. ARIMA (3,1,1) produced better forecasted values as compared to the other models. Therefore ARIMA (3,1,1) was chosen as the best effective model to fit the data series and forecast the VAT revenue collections.
Keywords: time series model, VAT

ISSN (online) 2664-9535 (print) 2664-9527 https://atcr.kra.go.ke

1. Background

Taxation is the largest source of government revenue in Kenya. Kenya Revenue Authority (KRA) is the agency charged with the responsibility of collecting revenue on behalf of the Government of Kenya. Revenue collected is used to finance public expenditure for goods and services consumed by the public

Both tax analysis and revenue forecasting are very vital in providing directions as to enhancing tax revenue, improving equity and efficiency of taxes, promoting investments and consequently economic growth with a view of increasing national income. In addition, it is important in monitoring processes and budget planning. It defines the resource envelope forming the basis for effective medium-term and long-term planning (ICPAK, 2016).

Tax revenue forecasting plays a central role in annual budget formulation. It provides policy makers and fiscal planners with the data needed to guide borrowing, use accumulated reserves, or specify monetary measures to balance the budget. It also informs about what fiscal actions are sustainable and hence how to balance fiscal policy to address the problems in the balance of payments and hence foreign debt. Accurate forecasting of revenues and expenditures is important for avoiding both underfunding and excessive funding of the government, and related consequences of associated surpluses or deficits (Annex3, 2015).

Revenue forecasting is an essential part of budgeting in the public sector and, hence, it is necessary for a government to forecast the revenue it collects for planning purposes. Budgetary uncertainties have led to increased reliance on economic and revenue forecasting by state governments in recent years. Because of the magnitude of the fiscal problems facing many states, forecasting has assumed a more central role in the policy making process. As a result, revenue forecasts are closely examined and accuracy is essential for planning purposes. To improve accuracy, there is need for analysts to assemble as much information about their respective state economies as possible, including formal and informal consideration of alternative forecasts (ICPAK, 2016).

Glenn P. et al. (2000) noted that tax analysis and forecasting of revenues are of critical importance to governments in ensuring stability in tax and expenditure policies. To augment timely and effective analysis of the revenue aspects of the fiscal policy, governments have increasingly turned toward in-house tax policy units rather than relying on tax experts from outside. These tax policy units have been increasingly called upon to analyze the impact of tax policies on the economy and to estimate the revenue implications of tax measures, with the ultimate objective of ensuring a healthy fiscal situation within the economy. Tax policy units also help ensure that tax systems are efficient, fair, and simple to understand and comply with. Such systems help to create an economic environment that is conducive to greater social justice.

In Kenya, the revenue forecast is produced by a single agency, which is the Ministry of Finance (MoF) in conjunction with the Kenya Institute for Public Policy Research and Analysis (KIPPRA). The MoF formulates the government budget by undertaking revenue projections and at the same time prepares the monthly/annual revenue targets for relevant revenue collection agencies in the country.

Against this background this study sought to provide an analysis of the VAT revenue collected by KRA. Exploratory data analysis was performed on the VAT series data and a model was established which can accurately forecast the revenue in future for a given period.

VAT Revenue

VAT is an indirect tax levied on the consumption of goods and services. It was introduced in Kenya in January 1990 to replace Sales Tax, which had been in operation since 1973.

It is charged at each stage of production and distribution chain up to the retail stage. Therefore, it is a tax on the difference between what a producer pays for inputs and what the producer charges for finished/final goods and services. VAT on imported goods and services is payable at the point of entry by the importer and this tax is collected by the Commissioner of Custom Services Department

The VAT standard rates are 0% and 16%. The former applies to supply of goods and services which are known as zero rated supplies while the latter applies to supply of goods and services which are neither exempt or zero rated. Zero rating is a term that is used in the VAT law to refer to taxable supplies of goods and services that are subject to tax at the rate of 0%. Therefore, a taxpayer dealing with these supplies does not charge the customers any output tax. However, he is entitled to recover input tax charged on these supplies. Exempt supplies are business transactions on which the VAT is not chargeable. This means that, goods and services under this category are not taxable. Persons dealing with such supplies are not required to register for VAT. However, if a trader deals with both exempt and taxable supplies, the trader should be registered for VAT which is eligible on taxable supplies. Therefore, any persons dealing exclusively in exempt supplies cannot claim input tax on them. Both exempt and zero-rated goods and services are listed in the first and second schedule of the VAT act respectively (KRA, 2015)

After charging/collecting the VAT, registered persons then remit it to the Authority every month. Registered persons only act as VAT agents in collecting and paying the tax since it is borne by the final consumer of goods and services. The due date of filing VAT return and making payment is 20th of every month

Forecasting of VAT

On average, VAT accounts, for 28% of total tax revenue collected in every financial year (Mutua, 2012). Being a tax on consumption, the revenue collected from it fluctuates in different months of each financial year.

Table 1.1: VAT Collections

Table 1.1 shows that VAT had a positive trend from the financial year 2012/2013 to financial year 2014/2015. However, the trend reversed with a negative growth up to the financial year 2016/2017. These variations can be attributed to various economic factors in the country.

Components in the VAT data series such as trend, seasonal components, cyclical components and random components can be identified by carrying out some statistical analysis. Time series modeling relies on these major components as far as forecasting is concerned. Time series forecasting of revenues assumes that patterns in the historical data of a data series can be used to project future revenues. Therefore, accurate forecasts of VAT revenue can be done using time series models. This is important for various reasons such as budget planning and decision making.

1.2 Statement of the problem

The purpose of this study was to establish the effectiveness of time series models on forecasting of VAT revenue collected in Kenya. This was with an objective of making reliable forecasting of revenue which is important for avoiding both underfunding and excessive funding of the government, and related consequences of associated surpluses or deficits.

A number of revenue forecasting models, which have been developed by various researchers, are mostly conditional models e.g. macroeconomic models that are developed conditionally on the accuracy of macroeconomic variables that are used as a basis for the prediction. A lot of different types of data are needed to construct such models. Conversely, the unconditional models only use revenue series data and its forecasts do not depend on forecasts of any other economic variable.

A good example of unconditional models that have been used in forecasting is the time series models. These models include exponential smoothing model, autoregressive model, moving average model and autoregressive moving average model among others. These models have widely been used and they produce different results depending on the nature of data. Therefore, it is not clear which of these models produces the best results as far as revenue forecasting is concerned.

At the moment there is substantial data available on revenues, which can be used in revenue projections. This research study made use of times series models on VAT revenue data in establishing a suitable forecasting model. The unconditional method can be done faster and more easily in the sense that it only uses revenue series data and its forecasts does not depend on forecasts of any other economic variable, such as GDP or consumption. This study will act as a baseline study and possibly if further explored can be a convenient way of forecasting revenue in the country, since you don’t have to worry about the accuracy or significance of the other variables to be used.

1.3 Objectives of the study

The general objective of the study is to establish the effectiveness of time series models on forecasting of VAT revenue in KRA

1.4 Specific objectives of the study

To establish the applicability of AR model on forecasting of VAT revenue

To establish the applicability of MA model as far as VAT revenue projection is concerned

To establish the applicability of ARMA model on forecasting of VAT revenue

To establish the applicability of ARIMA model on forecasting of VAT revenue

1.5 *Research Hypothesis*

How effective is the AR model in forecasting of VAT revenue?

How efficient is the MA model in forecasting of VAT revenue?

How effective is the ARMA model in forecasting of VAT revenue?

How efficient is the ARIMA model in forecasting of VAT revenue?

1.6 *Justification of the study*

There has been a debate on how realistic Kenya’s budgeting framework is. One of the reasons that has ensued is that KRA has on several occasions not been able to meet the revenue targets of various financial years. This notwithstanding, government’s ambition to roll out a significant number of infrastructural developments has been poised to compound the budget financing challenges.

These concerns should lead our policy makers to ponder on whether our economic statistical forecasts are right. Given that Kenya’s revenue portfolio is highly driven by tax revenues, I pose to ask in this study whether we have established an efficient revenue forecasting method for accurate revenue projections of various tax heads.

This study sought to establish the trend and behavior of VAT revenue collections by exploring the monthly revenue collection data. This was with an aim of establishing a suitable model for accurate revenue projections by making use of time series models. The established model will assist in revenue forecasting of the VAT revenue and other tax heads thereby enhancing the planning and budgeting processes. The model will also assist in the setting of collection targets.

Regular forecasting of VAT revenue will reveal the prospects of attaining the revenue targets of various periods. Again, it will help in measuring tax revenue potential of a country as well as measuring performance of revenue departments. In addition, it will help the authority to monitor its progress and make sure it's still on track in fulfilling its core mandate of adequate revenue collection. Again, this project will be of great importance in enhancing my skills and abilities in modeling time series data.

1.6 Scope of the study

This study was limited to revenue collections of the previous seven financial years. This was a period from the financial year 2009/2010 to 2015/2016. Since VAT revenue is collected on a monthly basis, this gave a total of 84 observations which were adequate enough to make accurate revenue projections using time series models

2. Literature Review

The purpose of literature review is to get an insight into similar studies on revenue forecasting models, their findings and methodology used and how the same can be applied to the Kenyan context.

2.2 Theoretical Review

Sequence or time-lag theory is the most important theory of business forecasting. It is based on the behavior of different businesses which show similar movements occurring successively but not simultaneously. There is time-lag between different movements, for example, expenditure on advertisement may not at once lead to increase in sales. As such, this method considers time lag based on the theory of lead-lag relationship which holds good in most cases. The series that usually change earlier serve as forecast for other related series. However, the accuracy of forecasts under this method depends upon the accuracy with which time lag is estimated (Dhaval, 2012).

Action and reaction theory is based on the Newton's 'Third Law of Motion', which states that for every action there is an equal and opposite reaction. In business, this law implies that if there is depression in a particular field of business, there is bound to be boom in it sooner or later. It reminds us of the business, cycle which has four phases, i.e. prosperity, decline, depression and prosperity. This theory regards a certain level of business activity as normal and the forecaster has to estimate the normal level carefully. According to this theory, if the price of commodity goes beyond the normal level, it must come down also below the normal level because of the increased production and supply of that commodity (Dhaval, 2012).

Theory of economic rhythm propounds that the economic phenomena behave in a rhythmic manner and cycles of nearly the same intensity and duration tend to recur. According to this theory, the available historical data have to be analyzed into their components such as trend, seasonal, cyclical and irregular variations. The secular trend obtained from the historical data is projected a number of years into the future on a graph or with the help of mathematical trend equations. If the phenomena is cyclical in behavior, the trend should be adjusted for cyclical movements. When the forecast for a year is to be split into months or quarters then the forecaster should adjust the projected figures for seasonal variations also with the help of seasonal indices. It becomes difficult to predict irregular variations and hence, rhythm method should be used along with other methods to avoid inaccuracy in forecasts. However, it must be remembered that business cycles may not be strictly periodic and the very assumptions of this theory may not be true as history may not repeat (Homeworkhelp, 2012).

Specific historical analogy theory assumes that history repeats itself. It simply implies that whatever happened in the past under a set of circumstances is likely to happen in future under the same set of conditions. Thus, a forecaster has to analyze the past data to select such period whose conditions are similar to the period of forecasting. Further, while predicting for the future, some adjustments may be made for the special circumstances which prevail at the time of making the forecasts (Dhaval, 2012).

Cross-section analysis theory is based on the knowledge and interpretation of current forces rather than projection of past trends. The theory assumes that no two cycles are alike. By the like causes always produce like results. All the factors bearing upon a given situation are assembled and relying upon the knowledge of economic processes. The forecaster concludes whether the situation is favorable or not, immediate recognition is given to the fact that business conditions are shaped by simultaneous inflationary and deflationary forces. Predominance of inflationary forces results in booms, whereas predominance of deflationary forces leads to depression. The forecaster who utilizes this method enumerates stable forces and a third which sets forth deflationary forces on the basis of judgment (Homeworkhelp, 2012).

2.3 Empirical Review

Thomas M. (2000) examined the effectiveness of composite forecasting of sales tax revenues in Idaho. An econometric model and a univariate time series model provided base line projections. The composite forecasts were found to outperform both base line forecasts. In the composite forecasting ordinary least squares was used to estimate the parameters, except in cases where autocorrelation correction was necessary. For the latter case, an autoregressive moving average exogenous (ARMAX) nonlinear least squares correction technique was used.

According to Joselito (2005) there are two ways in which revenue forecasting is normally practiced. It can be calculated as an unconditional prediction of the most likely outcome or it may be performed conditionally on the accuracy of macroeconomic variables that are used as a basis for the prediction. In a study by the IMF, using a sample of 34 countries from Africa, Asia, Latin America and the Middle East, they showed that although not all of the countries relied on macroeconomic forecasts as inputs to the revenue forecast, majority still do. While 85% of the sampled countries use subjective assessment and basic extrapolation techniques as their main forecasting methodology, only about 13% use formal econometric methods.

Graham Glendy (2001) evaluated the revenue collections from personal income tax (PIT), payroll tax, and health contribution in Bhutan in the period 1993/4 to 2005/6. Revenue trends were evaluated and revenue for up to fiscal year 2007-2008 was forecasted using trend analysis model. Three mathematical regression equations to estimate the revenue were developed using the exponential, linear, and polynomial trends. The Mean Absolute Percentage Error (MAPE) was used to compare the forecasting accuracy between models, and hence concluded that the exponential model produced the most accurate projections.

Joselito.A. (2005) forecasted revenue for Bureau of Customs (BoC) of the Philippines, which collects import duties from oil and non-oil, imports, and also levies VAT and Excise taxes from the import of commodities. The parameters estimates were obtained through ordinary least square (OLS) method of regression. The study concluded that the amount of import duties collected is a function of the tariff rates and the value of dutiable imports.

A.Kyobe (2005) in his study of revenue forecasting in low-income countries found out that little research has been carried out on the determinants of revenue forecasting practices. One possible explanation he came up with is that a systematic and comparative analysis requires a wealth of institutional knowledge. On the other hand, descriptions of budget preparation processes are generally not put down in formal documents, and country practices are often a mix of idiosyncratic budget practices and influences from legacy systems.

Favero (2005) assessed the possibility of producing unbiased forecasts for fiscal variables in the Euro area by comparing a set of procedures that rely on different information sets and econometric techniques. In particular, he considered autoregressive moving average models, Vector auto-regressions, small-scale semi-structural models at the national and Euro area level, institutional forecasts (Organization for Economic Co-operation and Development), and pooling. He ranked models on the basis of their forecasting performance using the mean square and mean absolute error criteria at different horizons. His study concluded that simple time-series methods were able to deliver unbiased forecasts, or slightly upward-biased forecast for the debt–GDP dynamics.

Krol (2010) compared alternative time-series models to forecast state tax revenues. He used forecast accuracy to compare to a benchmark random walk forecast. Quarterly data for California was used to forecast total tax revenue along with its three largest components, sales, income, and corporate tax revenue. His study found out that for one and four-quarter-ahead forecasts from 2004 to 2009, Bayesian vector auto-regressions generally forecasted best based on root mean squared errors compared to standard vector auto-regressions or a random walk model. He proposed that similar to the macroeconomic forecasting experience, Bayesian vector auto-regressions should be considered as a useful and cost effective revenue forecasting model for state governments.

Corvalo (2010) took a study on the possibility of improving the monthly forecasts of the Value Added Tax on Merchandise and Services (ICMS in Portuguese) collected by the State of Santa Catarina, Brazil. Dynamic regression was used based on the concepts of co-integration and error correction utilizing the general to specific approach suggested by the London School of Economics (LSE). Different data series were selected and analyzed for the final model industry profit, consumption of electric energy and other energy sources, and cement, and business consultations to the Credit Service Protection Agency (SPC). In the process of the choice of the variables, Granger's tests of causality and the analysis of long-run equations were used. The results obtained were very satisfactory for forecasts both inside and outside the sample period, indicating that the use of this model by the Budget Department of the State of Santa Catarina will provide more suitable values for the decision making process and improvement in budget planning.

Streimikiene (2018) took a study to forecast the tax revenue of Pakistan for the fical year 2016–17 using three diffrent time series techniques and also to analyse the impact of indirect taxes on the working class. The study analysed the effiency of three diffrent time series models mainly the Autoregressive model (AR with seasonal dummies), Autoregressive Integrated Moving Average model (ARIMA), and the Vector Autoregression (VAR) model. The data used for this paper was from July 1985 to December 2016 (monthly) and focused on forecasting for 2017. Components of tax revenues such as direct tax, sales tax, federal excise duty and customs duties were used for forecasting of total tax revenue. The results of the study revealed that among these models the ARIMA model gives better-forecasted values for the total tax revenues of Pakistan.

2.3.1 Tax revenues in Kenya

In his study Simple (2012) used the monthly customs data, in particular the Import duty from January 2003 to December 2010 with the general objective of exploring the data and establishing a suitable forecasting model which can be used to predict the amount of import duty to be collected in a certain specified period. The exploration data analysis revealed that the import duty had a positive trend, a strong positive correlation over time and that an IMA (1,1) model was established as a suitable model to forecast the tax.

James Murunga (2016) took a study on tax effort and determinants of tax ratios in Kenya. Analysis of time series data running from 1980 to 2015 was done using ordinary least squares regression making use of per capital GDP, share of service sector in GDP, share of external debt in GDP, share of agriculture in GDP, share of exports in GDP and share of imports in GDP as the explanatory variables. The study found Kenya's tax effort to be less than unity meaning the tax capacity was not fully utilized which implies that the country has a potential of raising more tax. The study recommended the need for political will, efficient legal system and consistency in the implementation of tax policy

Conceptual Framework

Through use of data on previous monthly collections, patterns in the historical data were analyzed and time series models were used to forecast the future amount of revenue that can be collected in a given period. These models responded differently depending on the nature of data and their outputs varied. The relationship between the models and the future collection can be illustrated in a conceptual framework as shown

Figure 2.1: Conceptual Framework

Autoregressive (AR) Model

An AR model is a linear predictive modelling technique which predicts future behavior based on past behavior. It's used for forecasting when there is some correlation between values in a time series and the values that precede and succeed them. The model makes use of past data to model the behavior, hence the name autoregressive. The process is basically a linear regression of the data in the current series against one or more past values in the same series.

Moving Average (MA) Model

This model is similar to an AR model. However, rather than using past values of the forecast variable in a regression, a MA model uses past forecast errors in a regression-like model to forecast the future values. Therefore, these forecast errors are used as explanatory variables.

Autoregressive Moving Average (ARMA) Model

ARMA model refers to a forecasting model or process in which both auto-regression analysis and moving average methods are applied to a well-behaved time series data. ARMA assumes that the time series is stationary, fluctuates more or less uniformly around a time-invariant mean.

Autoregressive Integrated Moving Average (ARIMA) Model

ARIMA models establish a powerful class of models which can be applied to many real time series. It's a forecasting technique that projects the future values of a series based entirely on its own inertia. The model is composed of different terms, an autoregressive term (AR), a moving-average term (MA), and an integration term (I) that accounts for the non-stationarity of the time series. Its main application is in the area of short term forecasting requiring at least 40 historical data points. It works best when your data exhibits a stable or consistent pattern over time with a minimum number of outliers.

Critique of Existing Literature

Most studies have been carried about forecasting using various techniques. Some of these techniques include qualitative methods, time series methods and causal methods. A lot of studies about revenue forecasting in different countries have been done. One of the studies has been done by Streimikiene (2018). He took a study to forecast the tax revenue of Pakistan for the fical year 2016–17 using three diffrent time series techniques. The study analysed the efficiency of three diffrent time series models mainly the Autoregressive model (AR with seasonal dummies), Autoregressive Integrated Moving Average model (ARIMA), and the Vector Autoregression (VAR) model. The results of the study revealed that among these models the ARIMA model gives better-forecasted values for the total tax revenues of Pakistan. Different forecasting methods have produced good forecasts depending on the nature of the data series used. Time series modelling being widely used method of forecasting, its application has been limited as far as revenue forecasting is concerned especially in Kenya. Therefore, this study sought to explore the efficacy of time series methods in forecasting of tax revenue in KRA.

Table 2: Literature review summary

Forecasting provides relevant and reliable information about the past and present events and the likely future events. This is necessary for sound planning. Forecasts are possible only when a history of data exists. Previous studies about forecasting have been carried out using different techniques yielding different results. Steps involved in forecasting process are usually problem definition, gathering of information, preliminary analysis of data, choosing and fitting of models and later using and evaluating a forecasting model.

Research Gap

Tax revenue forecasting remains an important issue in many developing countries. Though it has not been studied a lot in the literature compared to the issues of GDP forecasting, its effects are of utmost importance in developing countries. Several researchers conducted researches on this field and concluded that under or over-prediction of tax revenues in government budgets persisting over a period of years has emerged as a problem both in developed and developing countries. For instance, Kenya being one of the developing countries has little research done about revenue forecasting. This has resulted to over-prediction of tax revenues in most instances which is evidenced by the fact that KRA has been unable to meet the revenue targets in different periods of most of the financial years. This research sought to provide a solution to this problem by exploring the data series on revenue collection, to analyze it and develop a model for accurate projection.

Time series modelling is one of the methods that has been applied as far is forecasting is concerned. Most of the studies have showed the models produce very good forecasts especially if the data contains components such as trend, cycles and seasonal components. Little studies have been done on their effects in tax revenue forecasting. This study therefore explored on how time series modelling can be applied easily and effectively in tax revenue forecasting in KRA

3. Research Methodology

This chapter on research methodology presents the research design adopted by the study, target population, sample frame, sample size and sampling technique, data collection instruments, data collection procedure, data analysis and data presentation

Research Design

The research design can be viewed as the overall strategy chosen to integrate the different components of the study in a coherent and logical way, thereby, ensuring that the research problem has been addressed effectively. It constitutes the blueprint for the collection, measurement, and analysis of data. Since VAT revenue collection for the previous periods is known, descriptive design was used to help determine how much of VAT revenue can be collected in future.

Target population

The research made use of previous revenue collections in Kenya. Therefore, the unit of analysis will comprise of the previous monthly revenue collected by KRA in the respective financial years.

Sampling Frame

Since the population under study is revenue collected in Kenya by KRA, the sampling frame will therefore consisted of the monthly VAT revenue collections of the previous financial years. Since VAT is a tax on consumption which depends on business activities, it was considered because its data series would produce time series components that are depended upon when doing forecasting

Sample Size

The sample size was the previous monthly VAT revenue collections for a period of 7 financial years from 2009/2010 to 2015/2016. This was a total of 84 observations. Since data is obtained from observations collected sequentially (monthly) over time, it was grouped as time series data.

Research instruments

The research instruments to be used were KRA records. These will be records of the previous revenue collections of the respective financial years.

Data Collection Procedure

Secondary data was administered throughout the research making use of KRA data of previous monthly VAT revenue collections of the respective financial years for a period of 10 years. That data was obtained from KRA database through the data analysis unit in the Domestic Taxes Department.

Data Analysis

Several statistical analysis packages have been developed which can be used to analyze any time series data. In order to model and forecast data on VAT revenue collections, R software was used. The software are chosen because of its wide application and acceptance in the field of Statistics and Econometrics.

Analytical Models

The four main time series models were used as the analytical models. These include;

Autoregressive (AR) models

In an autoregressive model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregressive indicates that it is a regression of the variable against itself. Thus a time series model x_t is an autoregressive model of order p, AR(p) if

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t$$

$$x_t = \alpha_0 + \sum_{i=1}^{p} \alpha_{t-i} \ \varepsilon_t$$

Where

x_t is the actual value

ε_t is random error (or random shock) at time period t

α (i=1,2,…p) are model parameters

α_0 is a constant

These expressions state that the estimated value of x at time = t is determined by the immediately previous value of x (i.e. at time = t-1) multiplied by a measure α of the extent to which the values for all pairs of values at time periods lag 1 apart are correlated plus a residual error term ε at time t

Moving average (MA) models

A MA model is similar to an AR model, except that instead of being a linear combination of past time series values, it is a linear combination of the past white noise terms i.e it uses past error terms as explanatory variables (ε_t)

A times series model x_t is a MA of order q, MA (q) if

$$x_t = [\![\varepsilon_t + \beta]\!]_1 \ \varepsilon_{t-1} + [\![\varepsilon_t + \beta]\!]_2 \ \varepsilon_{t-2} + \cdots + [\![\varepsilon_t + \beta]\!]_q \ \varepsilon_{t-q}$$

Where

β(i=1,2,…q) are model parameters

ε_t is random error (or random shock) at time period t

The random shocks are assumed to be a sequence of independent and identically distributed random variables with zero mean and a constant variance σ^2

Autoregressive Moving Average (ARMA) Model

AR and MA models can be effectively combined together to form a general and useful class of time series models, known as the ARMA models. A time series model $[\![x]\!]_t$ is an ARMA model of order p,q, ARMA (p,q), if

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t [\![+\beta]\!]_1 \ \varepsilon_{t-1} + [\![\varepsilon_t + \beta]\!]_2 \ \varepsilon_{t-2} + \cdots + [\![\varepsilon_t + \beta]\!]_q \ \varepsilon_{t-q}$$

Where

x_t is the actual value

ε_t's are random errors (or random shocks) at time period t

α (i=1,2,…p) and β(i=1,2,…q) are model parameters

α_0 is a constant

By setting $p \neq 0$ and $q = 0$ we recover the AR (p) model. Similarly, if we set $p = 0$ and $q \neq 0$ we recover the MA (q) model

ARMA model can be used to model a time series with fewer parameters overall compared to either an MA or an AR model by themselves. The estimated value at time t is expressed as the sum of q MA terms that represent the average variation of random variation over q previous periods plus the sum of p AR terms that compute the current value of x as the weighted sum of the p most recent values.

Autoregressive Integrated Moving Average (ARIMA) Model

An ARMA model assumes that the time series data is stationary i.e it has a constant mean, which is not always the case. Most datasets are not always stationary as they usually contain trends and periodicity. These datasets need to be made stationary before applying an ARIMA model which is done through differencing typically once, twice or three times, until the series is at least approximately stationary - exhibiting no obvious trends or periodicities. As with the MA and AR processes, the differencing process is also described by the order of differencing. This gives rise to an ARIMA model

A time series model x_t is an ARIMA model of order p,d,q, ARIMA (p,d,q), if

$$x_t' = \alpha_0 + \alpha_1 x_{t-1}' + \alpha_2 x_{t-2}' + \cdots + \alpha_p x_{t-p}' + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q}$$

Where

x_t' is the differenced series (it may have been differenced more than once)

ε_t's are random errors (or random shocks) at time period t

α (i=1,2,…p) and β(i=1,2,…q) are model parameters

α_0 is a constant

4. Research Findings and Discussion

This chapter is concerned with the analysis of data, where the main objective of the study was to establish the efficacy of time series models as far as revenue forecasting is concerned and thereafter build a model, which can describe historical data of the data series. This is also a model which has forecasting power so that it can be used to forecast future revenue collections both in short term and long term for the VAT obligation. The analysis was carried out in different steps so as to achieve both the main objective and specific objectives. Different tests were also applied to access the accuracy of the results.

The monthly VAT revenue collections that were used for this research are expressed in millions of Kenyan shillings. They covered the period from July 2009 to June 2017 making a total of 96 observations. Therefore, the unit of time is a month, the data from July 2009 to June 2015 was used for model fitting while the one from July 2016 to June 2017 was used for accessing the forecasting power and accuracy. The data was obtained from KRA database.

Figure 4.1: Monthly VAT Collections

From the figure 4.1, the plot shows a general increment of the VAT revenue collections for a period of 7 financial years. That long-term increase of data represents a trend which also explains that the mean of the data is not constant. Since the mean of the VAT data is not constant, the data series is therefore not stationary. In order to perform time series modelling on the data, it had to be transformed to become stationary.

4.2 Descriptive Statistics

A test was done on the data which produced the following results of the descriptive statistics

Table 4.1: Descriptive Statistics of VAT

From the Table 4.1, the VAT distribution is skewed to the right since the value of skewness is positive. This is also confirmed by the mean being greater than the median. As the kurtosis is negative, the distribution is relatively flat and therefore the data is not normally distributed.

4.3 Stationary Test

If the series exhibits originally long-term increase or any statistical properties varying over time, it must be transformed into stationary series. This is to make it compatible with different time series models. The aim of this test is to check the presence or absence of unit root, where the null hypothesis is stated by default as the presence of unit root and the alternative hypothesis as absence of a unit root

H0: Presence of unit root (Not stationary)

H1: Absence of unit root (Stationary)

By using Augmented Dickey Fuller (ADF), the test gave a p-value of 0.495. As the p-value is greater than 5% significance level, we fail to reject the null hypothesis and hence conclude that the data series is not stationary.

Stationarity can also be checked by inspecting the behavior of the plot of ACF and PACF.

Figure 4.2: ACF of Monthly VAT Collections

Figure 4.2 shows a slow descent of the ACF which indicates also a serial correlation of data in different lags. These properties also indicate that the data series is not stationary. We therefore need to transform it to a stationary series

4.4 Differencing

Differencing a time series means computing the difference between consecutive observations to make it stationary. It was used to make the VAT data series stationary.

Figure 4.3: Differenced VAT Series

After the first differencing of the original series, from the Figure 4.3 we can see that there was no systematic increase or decrease of series, both the mean and the variance became constant over time. Again, the test from the Dickey Fuller gave the p-value of less than 0.01. As the p-value is less than 5% significance level, we reject the null hypothesis and confirm the stationarity of the series after the first difference.

4.5 Model Identification

Autoregressive (AR) model

In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. An AR (autoregressive) model is usually used to model a time series which shows longer term dependencies between successive observations. Plotting and examining the graph of partial correlogram of the stationary time series enabled us to identify the order of the model

Figure 4.4: PACF of differenced VAT

The partial correlogram in figure 4.4 shows that the partial autocorrelations at lags 1, 2 and 3 exceed the significance bounds, are negative, and are slowly decreasing in magnitude with increasing lag (lag 1: -0.350, lag 2: -0.274, lag 3: -0.228). The partial autocorrelations tail off to zero after lag 3. This means that we will have an autoregressive model of order 3 (p=3) i.e AR(3). This model can also be written as ARIMA(3,0,0)

Moving Average (MA) model

A MA model is similar to an AR model, except that instead of being a linear combination of past time series values, it is a linear combination of the past white noise terms. A MA model is usually used to model a time series that shows short-term dependencies between successive observations.

Plotting and examining the graph of correlogram of the stationary time series enabled us to identify the order of the model

Figure 4.5: ACF of differenced VAT

We see from the correlogram of figure 4.5 that the autocorrelation at lag 1 (-0.355) exceeds the significance bounds, but all other autocorrelations do not seem to exceed the significance bounds. This means that our moving average will be of order 1 (q=1) i.e MA (1). This model can also be written as ARIMA(0,0,1)

Autoregressive Moving Average (ARMA) model

AR and MA models can be effectively combined together to form a general and useful class of time series models, known as the ARMA models. Having initially identified the orders of the AR an MA models, our ARMA model will therefore be ARMA (3,1). It can also be written as ARIMA(3,0,1)

Autoregressive Integrated Moving Average (ARIMA) model

This model is similar to ARMA where it also combines both the AR and MA models respectively. In addition, it has the order of differencing which is used to describe the differencing process of making a non-stationary series become stationary. Since our data series was differenced once to make it stationary, the order will be 1. Therefore, our ARIMA model will be ARIMA (3,1,1).

4.6 Model selection

Having identified the order of our models, selection of the best model was based on the values of AIC (Akaike Information Criteria). After running the tests, the models produced the following results

Table 4.2: Model Selection with BIC, AIC and AICc

The best model is normally the one with the least value of AICc. Also, the principle of model parsimony is usually applied to select the model with the least parameters as possible. From the table above, the ARIMA (3,1,1) is the one with the least values of AICc. Since the purpose of this study was to model and forecast the future value of VAT revenue, we cannot only validate the model on the basis of AICc only. Even though this model fits the data well, it's not a guarantee that it will produce the best forecast. Therefore, forecasting accuracy must be performed as well. Accuracy test on the models was done and it produced the following results

Table 4.3: Accuracy test of the models

The potential model selected is the one that minimizes the forecasting errors. From the table above, it is also clear that ARIMA (3,1,1) had the minimum errors as compared to the other models. Before confirming the selected model, diagnostic checking must be performed to check whether the errors are uncorrelated (behave like white noise), and normally distributed.

4.7 Diagnostic Checking

As suggested by Box and Jenkins, after selecting a potential model, diagnostic checking is needed to check whether the selected model fits the data well. The main assumption is that residuals from the fitted model are expected to be randomly independent and identically distributed following the normal distribution, in short, they must behave like a white noise. Box-Ljung test and correlogram plot were used to check whether the residuals are correlated while for normality checking, normal Q-Q plot and Shapiro test were used.

From the Box-Ljung test, the p-values for the models were 0.3902 for ARIMA(3,0,0), 0.6788 for ARIMA(3,0,0), 0.9471 for ARIMA(3,0,1) and 0.9182 for ARIMA(3,1,1).

The hypothesis is usually stated as

H0: The data is independently distributed (no correlation)

H1: The data is not independently distributed (serial correlation)

With the p-values being greater than 0.05, we fail to reject the null hypothesis and conclude that the residuals of all the models are independent with no correlation

Correlogram plot of all the residuals of the models gave results which were almost similar

Figure 4.6: ACF plot of Residuals

Figure 4.6 above, we can see on the ACF plot that there was no any significant spike, all the others are within the boundaries. Therefore, we can confirm that there was no correlation in the residuals which means there was no information left in the residuals to be used in fitting the models.

From Shapiro-Test for normality the p-values for the models were 0.0648 for ARIMA(3,0,0), 0.0587 for ARIMA(3,0,0), 0.0621 for ARIMA(3,0,1) and 0.0552 for ARIMA(3,1,1). The null hypothesis suggests the normality of residuals, and the alternative hypothesis suggests lack of normality in residuals. Therefore with 5% confidence level we fail to reject the null hypothesis and assume that the residuals for all the models are normally distributed.

Again, the Q-Q plots for all the residuals of the models produced the same plots showing most of the points having fallen in the straight line as shown in figure 4.7 below

Figure 4.7: Q-Q Plot of residuals

We can conclude that the residuals of all the models are normally distributed and they are all potential models to be used for forecasting.

4.8 Parameter Estimation

Parameters of each model were estimated by likelihood estimation and the results produced are shown in the table below

Parameter	ARIMA (0,0,1)	ARIMA (3,0,0)	ARIMA (3,0,1)	ARIMA (3,1,1)
〖ar〗_1	0.6222	0.5249	1.08588	0.0277
〖ar〗_2	0	0.1496	-0.20656	-0.1393
〖ar〗_3	0	0.2311	0.10288	-0.0948
〖ma〗_1	0	0	-0.64655	-0.5885

Table 4.4: Parameter Estimation

4.9 Forecasting

Each of the model was used to make forecast for the next financial year and the results were compared with the collections that were made in the respective financial year

Autoregressive (AR) model (ARIMA (3,0,0))

The first forecast was made using AR model. This model produced results which varied in different months across the financial year as shown in the table below

Table 4.5: Forecasting using AR model

From the table above, comparison was made between the forecasted values and the collections by computing the variance. The projected values obtained were under forecasted on several months of the financial year. This is because the forecast values were less than the amount collected in the respective months. The average variance was also obtained so that it could be used to make comparison with the other models.

Moving Average (MA) model (ARIMA (0,0,1))

Forecast was made using the MA model and the results were as shown in the table below

Table 4.6: Forecasting using MA model

Using AR model, the forecast values produced were constant for the entire period of the financial year except for the month of July. Computing the variance produced values which were negative since they were less than the collections. Therefore, all the projected values were under forecasted making the model unreliable for forecasting.

Autoregressive Moving Average (ARMA(3,0,1)) model

Forecasting using this model produced results which were almost similar to those of the AR model. The forecasted values varied in different months across the financial year also.

Table 4.7: Forecasting using ARMA model

From the table above, we realize that the model made projections which were under forecasted on several months. This is because the forecast values were less than the amount collected in the respective months. The variance was also computed and the average figure obtained so that it can be used to make comparison with the other models.

Autoregressive Integrated Moving Average (ARIMA(3,1,1)) model

The last forecast was done using ARIMA model. The forecasted values yielded varied across the months in the financial year as shown in the table below

Table 4.8: Forecasting using ARIMA model

The model produced projections whose values were only under forecasted in the month of July and February. During the other periods, the projected values were slightly higher than the actual collection. Looking at the targets that were set, we realize that the values were overestimated in several months like November, January, March, April and June. The value of the average variance that was computed was higher for the targets than for the projected values. Therefore, this model did better revenue projection than the one used for projecting the targets. Again, comparing the value of the average variance of this model with the other model, this model produced the least value making it the best suitable and effective model for revenue forecasting.

Plotting the forecast graph produced the following graph

Figure 4.8: Forecast Plot of VAT

Figure 4.8 shows a continuity of the positive trend of the VAT revenue collections. Therefore, more revenue is anticipated to be collected in the next financial year as compared to the previous financial years

5. Summary, Conclusions and Recommendations

This is the final chapter of the study. It presents summary of findings, discussions, and relevant conclusions, study recommendations and suggestions for further investigations.

5.2 Summary

The main objective of this study was to establish the efficacy of time series models in forecasting of revenue collection. This was by narrowing down to VAT obligation with an aim of selecting the best and accurate model which posses' high power of predictability (forecasting power). This model would be used for making accurate and realizable revenue projection for the obligation.

Time series forecasting of revenues assumes that patterns in the historical data of a data series can be used to project future revenues. VAT data series for a period of 7 financial years was explored and found to contain components such as trend, seasons and random changes. Time series forecasting relies on these concepts to make forecast. Since the data was found to contain some trend, the mean was not constant and therefore the data was not stationary. There was need to transform and make it stationary by differencing it

The models that were fitted on the data are AR model, MA model, ARMA model and ARIMA model.

Autoregressive (AR) model

AR model was first to be considered. Plotting and examining the graph of partial correlogram of the differenced series showed that the AR model should be of order 3 i.e AR(3). This is equivalent to ARIMA(3,0,0). After running both the best fit model and accuracy measure tests, this model underperformed compared to other models. Forecasting using the model produced values which were under forecasted on several months making it unsuitable to forecast revenue collections of VAT.

Moving Average (MA) model

Later MA model was also fitted on the differenced series. In order to determine its order, a graph of collelogram was plotted and examined and it gave order 1 i.e MA(1). This can also be written as ARIMA(0,0,1). Again, both the best fit model and accuracy measure tests were done on this model too. Just like the AR model, this one also underperformed compared to other models. The forecasted values produced by the model were constant for all the months in the financial year which was unrealistic. Therefore, this model was not the best to use.

Autoregressive Moving Average (ARMA) model

ARMA model being a combination of both AR and MA model yielded ARMA (3,1) which is the same as ARIMA(3,0,1). This one was underperformed compared to other models that were fitted after running the best fit model test and the accuracy measure test. Forecasting using the model produced similar results as the AR model where the values were under forecasted in several months of the financial year. This also made the model unsuitable for forecasting of revenue collection.

Autoregressive Integrated Moving Average (ARIMA) model

ARIMA model was the last to be fitted. It is similar to ARMA model with a differencing order. After differencing the series once, it became stationary. Therefore our ARIMA model became ARIMA(3,1,1).

Best fit model test and accuracy measure test showed that the model was better the former models. Forecasting using the models produced better results than the other models. This is because the values were under forecasted only in two months. A comparison between the projected values and the targets showed that the projected values were more realizable and easily achievable in most of the months. Therefore, this model was found to be the best to fit the VAT data series and make VAT forecasts as well.

Using maximum likelihood, parameters of the model were estimated and the resulting ARIMA(3,1,1) using backward shift operator was as follows

$$(1-\emptyset_1 \beta-\emptyset_2 \beta^2-\emptyset_3)(1-\beta) y_t=(1+\theta_1 \beta)\epsilon_t$$

$$(1-1.08588\beta+0.20656\beta^2-0.10288\beta^3)(1-\beta) y_t=(1-0.64655\beta)\varepsilon_t$$

5.3 Conclusions

The main objective of this study was to establish the effectiveness of time series models on forecasting of VAT revenue in KRA. The first specific objective sought to look at the AR model and the conclusion made was that it was ineffective as far as revenue forecasting is concerned. This is because it made projections which were under forecasted most of the time. The conclusion of the second objective was that it was ineffective too. This is due to the reason that it projected values which were constant and under forecasted in all the months. Looking at the third objective which took note of the ARMA model, the results were similar to that of an AR model making it ineffective also. Our fourth objective yielded better results than the others using the ARIMA model. The projected values of this model were under forecasted only on two occasions. A comparison made between the projected values and the actual collections showed this model produced the least value of variance compared to the other models. Therefore, this study concluded that the ARIMA model was the best effective model for making VAT revenue projections.

5.4 Recommendations

Tax revenue forecasting plays a central role in annual budget formulation. Time series forecasting models have proven useful in many applications, but have not been extensively used especially in tax revenue forecasting. Since the results of this study established that the ARIMA model is the best effective model for projecting VAT revenue, I would recommend its adoption by KRA and other governments agencies for forecasting of revenues and expenditures as well. This is important for avoiding both underfunding and excessive funding of the government, and related consequences of associated surpluses or deficits. I would also recommend the future researchers to examine the effectiveness of other forecasting methods in tax revenue forecasting.

6. References

[1.] Annette, K. (2005). Revenue Forecasting- How is it done? IMF Working Paper (WP/05/24).

[2.] Annex3. (2015). REVENUE AND EXPENDITURE FORECASTING TECHNIQUES FOR A PER SPENDING.

[3.] Corvalo, E. D. (2010). Forecasting the collection of the state value added tax (ICMS) in Santa Catarina. Brazilian Journal of Operations & Production Management.

[4.] Dhaval. (2012). Business Forecasting: Business Management Ideas. Retrieved from Business Management Ideas: http://www.businessmanagementideas.com

[5.] Ehrhart, E. &. (2010). Is VAT stabilizing?

[6.] Favero, C. A. (2005). Modelling and forecasting fiscal variables for the Euro Area. Oxford Bulletin of Economics and Statistics.

[7.] Glenday, G. (2001). Revenue Forecasting and Estimation, Techniques and Applications. Durham: Duke University.

[8.] Homeworkhelp, T. (2012). Home: Statistics Homework Help:Business Forecasting: Forecasting Theories. Retrieved from The Homeworkhelp: http://thehomeworkhelp.co.uk

[9.] ICPAK. (2016). KENYA'S REVENUE ANALYSIS 2010-2015. Nairobi, Kenya.

[10.] James Murunga, M. M. (2016). Tax Efforts and Determination of Tax Ratios in Kenya. European Journal of Economics, Law and Politics, Vol3 No 2.

[11.] Jenkins, G. P. (2000). Tax analysis and revenue forecasting. . Harvard: Harvard Institute for International Development.

[12.] Joselito. (2005). Forecasting Revenues from the BIR and BOC. Towards the Improvement of Cash Programming in the Philippines.

[13.] KRA. (2015). VALUE ADDED TAX ACT. Nairobi: National Council for Law.

[14.] Krol, R. (2010). Forecasting State Tax Revenue: A Bayesian Vector Autoregression Approach. California State University, Northridge.

[15.] Mosomi, N. A. (2015). Determinants of Tax Efficiency Perceptions by Domestic Taxpayers in Kenya. Nairobi, Kenya.

[16.] Murithi, M. a. (2003). Tax Elasticity and Bouyancy in Kenya. Nairobi, Kenya: University of Nairobi.

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

[18.] Otieno, S. (2015). Comparison of lasso and mle. East african journal of health sciences, 20.

[19.] Samwel, K. C. (2012). Elasticity and Bouyancy of Tax Components and Systems in Kenya. Nairob, Kenya.

[20.] Simba, J. N. (2012). Revenue Forecasting: A Case of Import Duty.

[21.] Streimikiene, D. a. (2018). Forecasting tax revenues using time series techniques--a case of Pakistan. Economic Research-Ekonomska Istra.

[22.] Wawire, N. (2000). Tax buoyancy and income-elasticity of Kenya's tax system. Nairobi, Kenya: University of Nairobi.

[23.] Wawire, N. H. (2011). DETERMINANTS OF VALUE ADDED TAX REVENUE IN KENYA. Nairobi, Kenya: Kenyatta University.

Annex

Table 1.1: VAT Collections

FINANCIAL YEAR	VAT COLLECTION (Kshs MILLIONS)
2012/2013	101,446
2013/2014	124,144
2014/2015	143,922
2015/2016	137,951
2016/2017	134,490

Source: KRA Data (2012-2017)

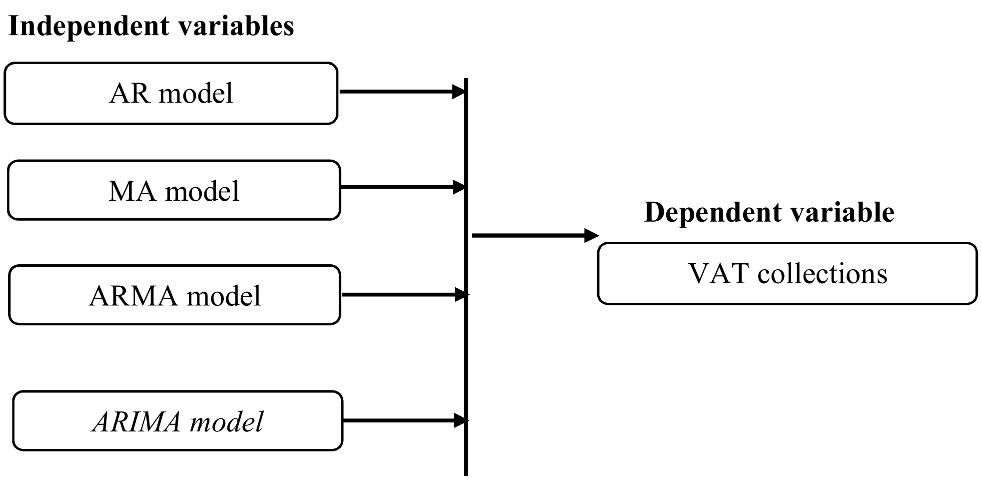

Figure 2.1: Conceptual Framework

Table 2: Literature review summary

Year	**Researcher**	**Title of Study**	**Findings**
2005	Annette Kyobe	Revenue Forecasting- How is it done?	Little research has been carried out on the determinants of revenue forecasting practices. One possible explanation is because a systematic and comparative analysis requires a wealth of institutional knowledge
2010	Corvalo, Eder Daniel and Samoryl, Robert Wayne and Brasil, Gutemberg Hespanha	Forecasting the collection of the state value added tax (ICMS) in Santa Catarina	Use of dynamic regression produced results which were very satisfactory for forecasts both inside and outside the sample period
2010	Krol, Robert	Forecasting State Tax Revenue: A Bayesian Vector Autoregression Approach	Bayesian vector auto-regressions generally forecasted best based on root mean squared errors compared to standard vector auto-regressions or a random walk model

2012	Simba, Jeliah Nyanduko	Revenue Forecasting: A Case of Import Duty	Import duty had a positive trend, a strong positive correlation over time and an IMA(1,1) model was established as a suitable model to forecast the tax.
2018	Streimikiene, Dalia and Raheem Ahmed, Rizwan and Vveinhardt, Jolita and Ghauri, Saghir Pervaiz and Zahid, Sarwar	Forecasting tax revenues using time series techniques--a case of Pakistan	ARIMA model gives better-forecasted values for the total tax revenues of Pakistan compared with other models like AR and Vector Autoregression (VAR)

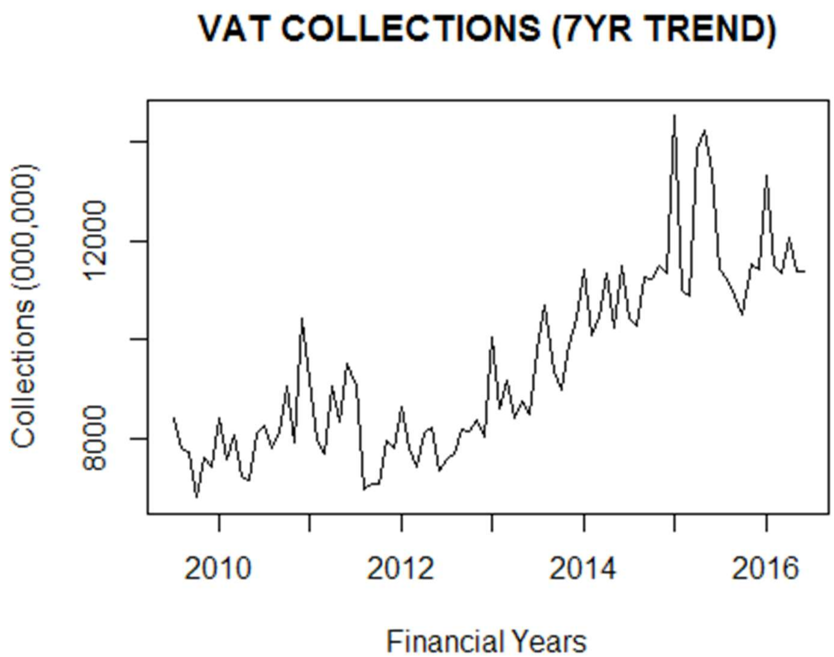

Figure 4.1: Monthly VAT Collections

Table 4.1: Descriptive Statistics of VAT

Statistics	Value
Mean	9,480.79
Median	9,043
Minimum	6,789
Maximum	1,4526
Standard Deviation	1,859.7
Skewness	0.67
Kurtosis	-0.34
Observations	84

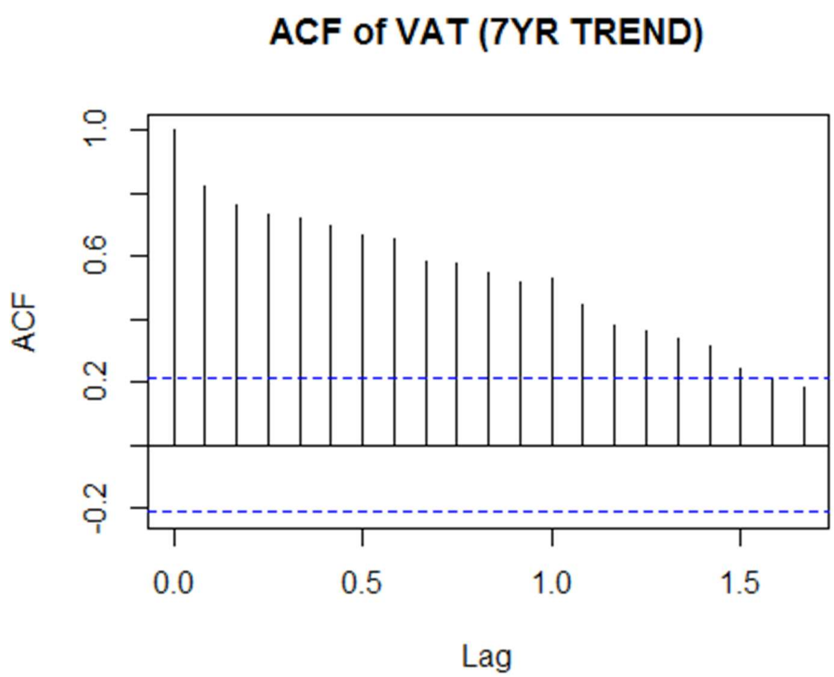

Figure 4.2: ACF of Monthly VAT Collections

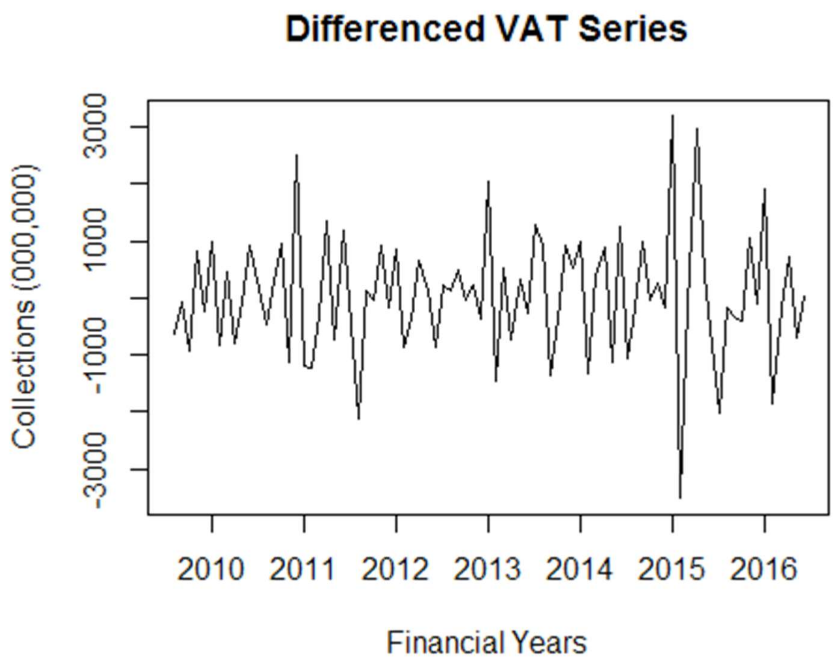

Figure 2.3: Differenced VAT Series

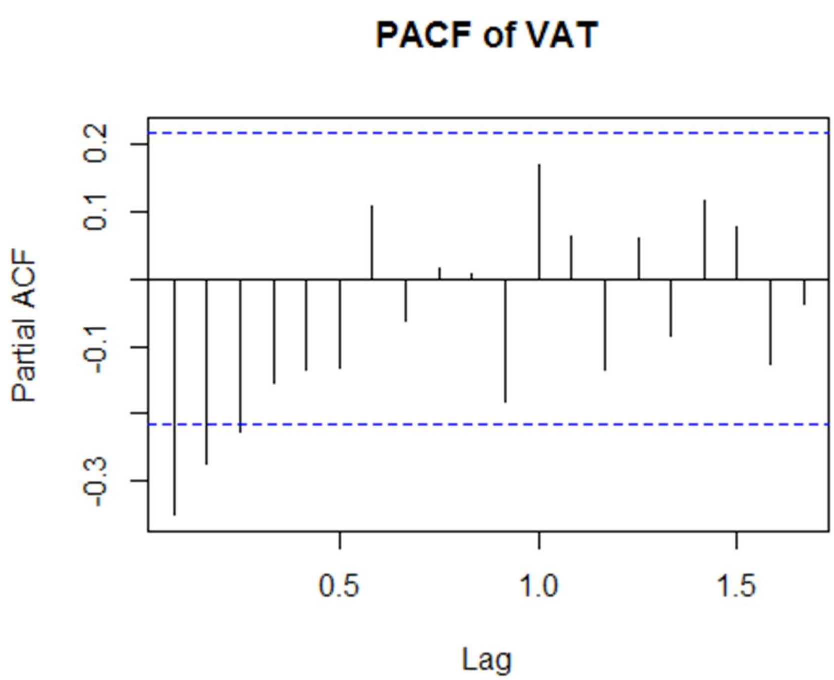

Figure 4.4: PACF of differenced VAT

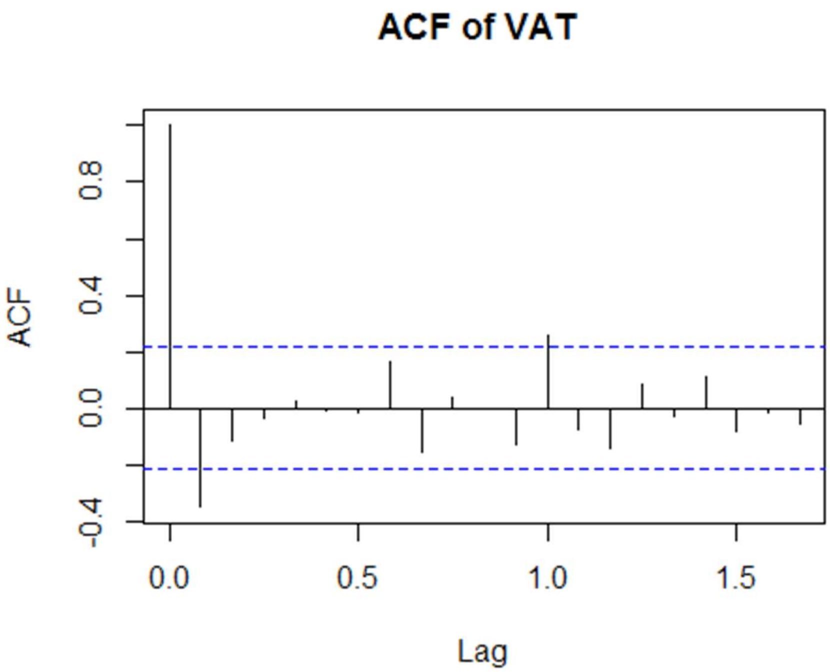

Figure 4.5: ACF of differenced VAT

Table 3.2: Model Selection with BIC, AIC and AICc

Model	AIC	BIC	AICc
ARIMA (0,0,1)	1459.54	1466.83	1459.84
ARIMA (3,0,0)	1403.86	1416.02	1404.63
ARIMA (3,0,1)	1400.69	1415.28	1401.79
ARIMA (3,1,1)	**1381.08**	**1393.17**	**1381.86**

Table 4.3: Accuracy test of the models

Model	RMSE	MAE	MAPE	MASE
ARIMA (0,0,1)	1380.46	1159.59	12.3277	0.927489
ARIMA (3,0,0)	961.956	744.326	7.73274	0.595341
ARIMA (3,0,1)	930.793	704.462	7.26226	0.563457
ARIMA (3,1,1)	**926.077**	**691.128**	**7.06233**	**0.5527918**

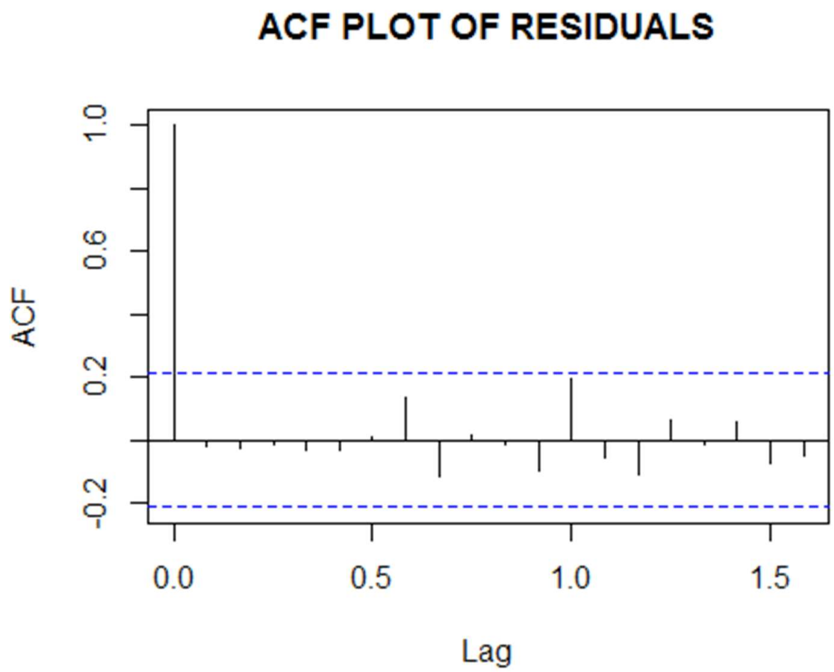

Figure 4.6: ACF plot of Residuals

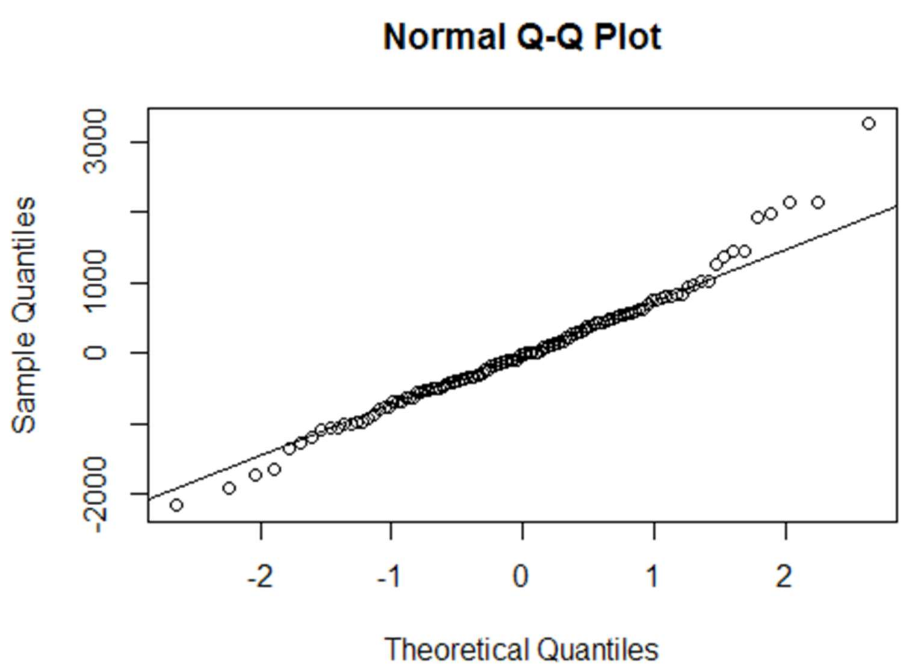

Figure 4.7: Q-Q Plot of residuals

Table 4.4: Parameter Estimation

Table 4.5: Forecasting using AR model

Parameter	ARIMA (0,0,1)	ARIMA (3,0,0)	ARIMA (3,0,1)	ARIMA (3,1,1)
ar_1	0.6222	0.5249	1.08588	0.0277
ar_2	0	0.1496	-0.20656	-0.1393
ar_3	0	0.2311	0.10288	-0.0948
ma_1	0	0	-0.64655	-0.5885

TIME	FORECAST	COLLECTION	VARIANCE
Jul-16	11,368	12,836	(1,468)
Aug-16	11,202	11,595	(393)
Sep-16	11,117	10,797	320
Oct-16	11,042	10,520	522
Nov-16	10,952	10,642	310
Dec-16	10,874	11,191	(317)
Jan-17	10,802	11,582	(780)
Feb-17	10,732	12,252	(1,520)
Mar-17	10,666	10,279	387
Apr-17	10,605	11,424	(819)
May-17	10,546	11,263	(717)
Jun-17	10,491	10,108	383
	AVERAGE VARIANCE		(341)

Table 4.6: Forecasting using MA model

TIME	FORECAST	COLLECTION	VARIANCE
Jul-16	10,282.77	12,836	(2,553)
Aug-16	9,489.18	11,595	(2,106)
Sep-16	9,489.18	10,797	(1,308)
Oct-16	9,489.18	10,520	(1,031)
Nov-16	9,489.18	10,642	(1,153)

Dec-16	9,489.18	11,191	(1,702)
Jan-17	9,489.18	11,582	(2,093)
Feb-17	9,489.18	12,252	(2,763)
Mar-17	9,489.18	10,279	(790)
Apr-17	9,489.18	11,424	(1,935)
May-17	9,489.18	11,263	(1,774)
Jun-17	9,489.18	10,108	(619)
	AVERAGE VARIANCE		(1,652)

Table 4.7: Forecasting using ARMA model

TIME	FORECAST	COLLECTION	VARIANCE
Jul-16	11,479	12,836	(1,357)
Aug-16	11,453	11,595	(142)
Sep-16	11,654	10,797	611
Oct-16	11,455	10,520	855
Nov-16	11,480	10,642	702
Dec-16	11,402	11,191	123
Jan-17	11,284	11,582	(298)
Feb-17	11,254	12,252	(998)
Mar-17	11,225	10,279	946
Apr-17	11,196	11,424	(228)
May-17	11,168	11,263	(95)
Jun-17	11,356	10,108	1,033
	AVERAGE VARIANCE		160

Table 4.8: Forecasting using ARIMA model

TIME	FORECAST	COLLECTION	FORECAST VARIANCE	TARGETS	TARGETS VARIANCE
Jul-16	11,986	12,836	**(850)**	11,910	**(926)**
Aug-16	11,617	11,595	**22**	11,587	**(8)**

Sep-16	10,982	10,797	**185**	11,152	**355**
Oct-16	11,024	10,520	**504**	10,632	**112**
Nov-16	11,122	10,642	**480**	11,884	**1,242**
Dec-16	11,571	11,191	**380**	11,905	**714**
Jan-17	11,586	11,582	**2**	14,377	**2,795**
Feb-17	11,573	12,252	**(679)**	12,063	**(189)**
Mar-17	10,642	10,279	**363**	11,659	**1,380**
Apr-17	11,572	11,424	**148**	12,786	**1,362**
May-17	11,324	11,263	**61**	11,669	**406**
Jun-17	11,022	10,108	**914**	11,155	**1,047**
	AVERAGE VARIANCE		**127**		**691**

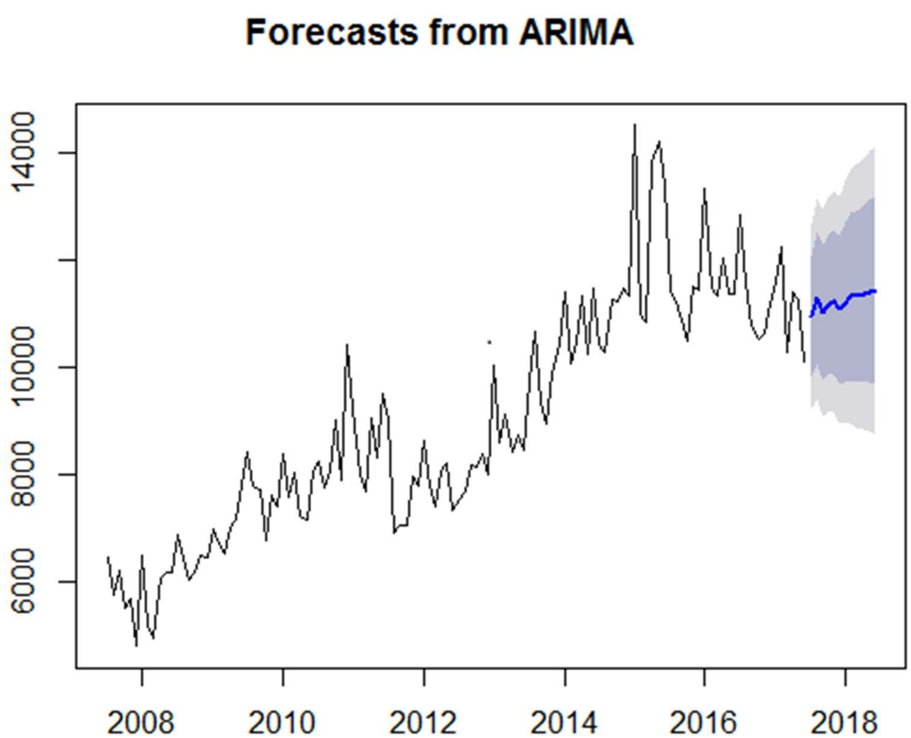

Figure 4.8: Forecast Plot of VAT