Strathmore
UNIVERSITY

Modeling Own Source Revenue (OSR) of County Governments in Kenya

Mwaura Cynthia Wanjiru -084784

Submitted in partial fulfillment of the requirements for the Degree of Bachelor of Business Science in Finance at Strathmore University

Strathmore Institute of Mathematical Sciences
Strathmore University
Nairobi, Kenya

November, 2017

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Project contains no material previously published or written by another person except where due reference is made in the Research Project itself.

Cynthia Mwaura. [Name of Candidate]
.... [Signature]
14/12/2017. [Date]

This Research Project has been submitted for examination with my approval as the Supervisor.

Dr. Olukum [Name of Supervisor]
.... [Signature]
14/12/2017 [Date]

Strathmore Institute of Mathematical Sciences
Strathmore University

Abstract

Revenue forecasting is an essential part of budget making process in the public sector. This study considered the Time Series Modeling of Own Source Revenue of counties in Kenya. Data used was collected from the revenue collection system of one of the counties, in particular, daily revenue from July 2015 to June 2017 with the general objective of exploring the data and further establishing a suitable forecasting model which could be used to predict the amount of revenue to be collected in a certain specified period. Box and Jenkins method of time series analysis was used to analyse the series.

From the analysis, IMA (1,1) model was identified as a suitable model to forecast the own source revenue. The forecast generated holding other factors indicated that the revenue collected would remain within the same range as before and thus, the Commission of Revenue Allocation should continue allocating funds to counties as the county cannot fully rely on its own revenue for sustainability and economic development.

Table of Contents

DECLARATION ii

Abstract iii

List of Tables vi

List of Figures vii

Abbreviations viii

Chapter 1: Introduction 1

1.0 Background to the Study 1

1.1 Problem Statement 7

1.3 Aim of the Study 7

1.4 Research Objectives 7

1.5 Research Questions 7

1.6 Significance of the study 8

Chapter 2: Literature Review 9

2.0 Introduction 9

2.1 Theoretical Framework 9

2.1.1 Revenue 9

2.1.2 Revenue Forecasting 10

2.2 Empirical Evidence 12

2.2.1 Revenue Forecasting 13

2.3 Conceptual Framework 15

2.4 Research Gaps 16

Chapter 3: Research Methodology 17

3.0 Introduction 17

3.1 Research Design 17

3.2 Population 17
3.3 Sample 17
3.4 Data Collection 18
3.5 Data Analysis 18
3.6 Model Building for own source revenue 18
3.7 Other tests to be conducted 24
Chapter 4: Results and Discussion 25
4.1 Exploratory Data Analysis 25
4.2 Testing for stationarity in the series 29
4.3 ARIMA Model Identification 33
4.3.1 Parameter Estimation IMA (1,1) model 33
4.3.2 Diagnostic Check of Model Adequacy 34
4.3.3 The Forecasting 36
Chapter 5: Conclusion and Recommendation 37
5.1 Conclusion 37
5.2 Recommendations 37
References 39

List of Tables

Table 1: Identification of Models 23

Table 2: Descriptive Analysis of General Charges 29

Table 3: Estimation equation of IMA (1,1) Model 33

List of Figures

Figure 1: Commission on Revenue Allocation Parameter Weights 2

Figure 2: Own Source Revenue Trend...4

Figure 3: Total Revenue versus Expenditure... 5

Figure 4: The Conceptual Framework 15

Figure 5: A Scatterplot of the General Charges.. 25

Figure 6: A Time Series Plot of the General Charges 26

Figure 7: Histogram of General Charges 27

Figure 8: Contribution of Each Revenue Sub-Item 28

Figure 9: Quantile-Quantile plot for General Charges 28

Figure 10: Plot of the ACF and PACF of the original data 30

Figure 11: Time series Plot for Differenced General Charges 31

Figure 12: Histogram of differenced General Charges Series 32

Figure 13: Figure Plot of ACF and PACF of first order series............................. 32

Figure 14: Plot of Residuals.. 34

Figure 15: 4-Plot of Residuals 35

Figure 16: Forecast from ARIMA (0,1,1) Model 36

Abbreviations

ACF - Autocorrelation Coefficient Functions

AR - Auto Regressive

ARI - Autoregressive Integrated

ARIMA - Auto-Regressive Integrated Moving Average

ARMA - Autoregressive Moving Average

CRA - Commission of Revenue Allocation

FY - Financial Year

IMF - International Monetary Fund

IGR – Internally Generated Revenue

LIAFOMS - Local Authorities Integrated Financial Operation Management System

MA – Moving Average

OSR – Own Source Revenue

OLS - Ordinary Least Squares

PACF - Partial Autocorrelation Functions

Chapter 1: Introduction

1.0 Background to the Study

Own Source Revenue (OSR), which is also referred to as Internally Generated Revenue (IGR), is the revenue that county governments in Kenya generate within their jurisdictional area of unit. The adoption of the new Constitution of Kenya in 2010 saw the decentralization of the national government and the devolution of power and resources to 47 county governments. This means that the people living and conducting business within the county can easily access government services from their county government. According to the new constitution, counties should be able to fund themselves from collecting and managing their own source revenue. The constitution dictates that each county government shall have its own established revenue fund fully supported by money raised or received on behalf of the county government except money excluded by an Act of Parliament (The Constituion of Kenya , 2010).

County Governments in Kenya generate revenue from external and internal sources. Commission of Revenue Allocation (CRA) determines how revenue will be allocated between the national government and county governments and among county governments. This forms the external source of revenue for the county government. This sharable revenue is generated from grants from the state government and other institutions. However, county governments collect fees, rates, taxes and other sources from its community to form own source revenue (OSR) which is popularly referred to as internally generated revenue.

Before devolution, the national government collected revenue from receipts that the government would get. These included taxes, custom duties, foreign aid, capital revenues and revenue from state owned entities. However, with devolution came two tiers of government, national and county government.

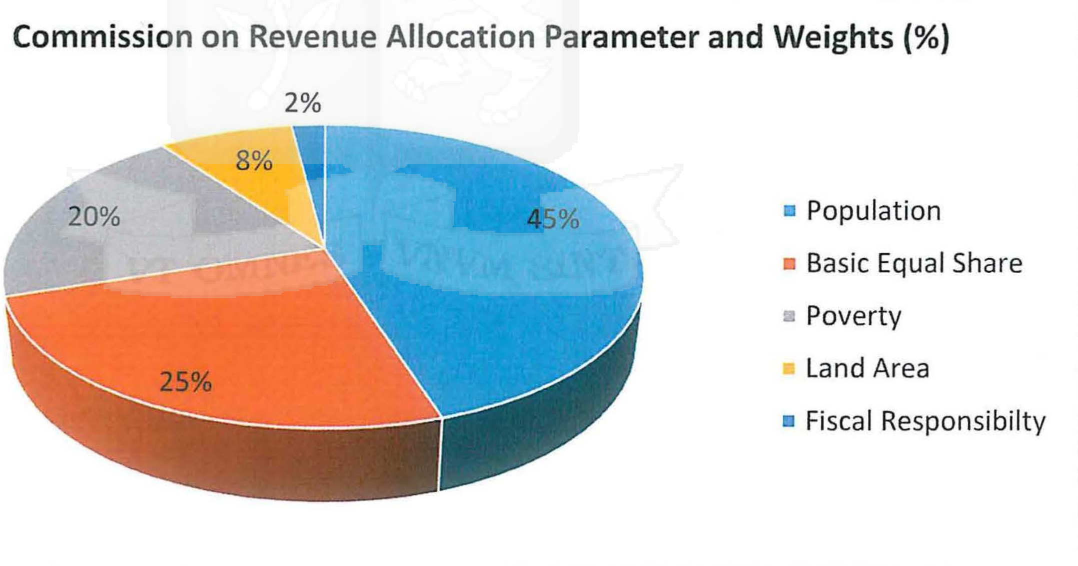
According to chapter 12 of the constitution, county governments' sources of revenue include taxes and single business permits as the core sources of revenue, property rates, licenses, fees, charges arising from delivery of services, interest and dividends, fines and fees, sales among others and grants and funds from the national government.

Counties may also borrow loans but it must be guaranteed by the national government and must be approved by the county government (The Constituion of Kenya , 2010). Counties also receive 15% from the national government to enable them to deliver as mandated. Marginalized counties receive an additional 0.5% from the equalization fund to cater for inequalities and marginalization among counties.

Revenue Allocation

Currently, the Commission of Revenue Authority allocates sharable revenue to counties based on the following parameters: - population 45%, basic equal share 25%, poverty 20%, land area 8% and fiscal responsibility 2%. These parameters are weighted based on expenditure trends over past years. Counties rely heavily on this allocation to fund their functions against what is required by the constitution, that is, counties are expected to fund their functions from own source revenue generated within the county (Commission on Revenue Allocation, 2015).

Figure 1: Commission on Revenue Allocation Parameter Weights



Source: Commission on Revenue Allocation (2015)

In Nigeria, since 1999, revenue allocation among states has been based on the following parameters: - equality of states 40%, population 30%, social development factor 10%, land mass 10% and internal revenue effort 10% (Salami, 2011).

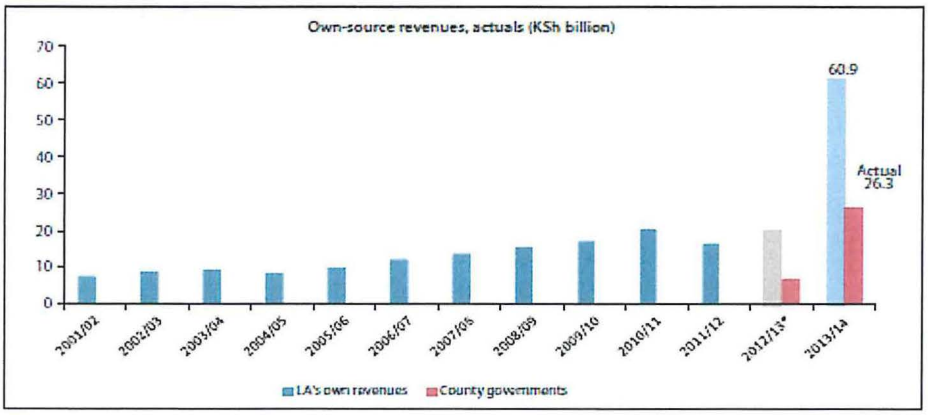
Revenue Collection

Revenue collection in most counties in Kenya is done through Local Authorities Integrated Financial Operation Management System (LAIFOMS) and manual receipting system (Oduor, Sevilla, & Wanyoike, 2016). This manual process exposed it to many challenges. Lack of adequate resources to collect revenue, employment of low skilled staff, fraud, corruption, misrepresentation of collected revenue and poor governance were and have been among the key limiting factors in the efficiency of the revenue collection system (Maina, 2016). The main drawbacks of the LIAFMOS system are lack of central data repository, limited checks and balances to mitigate fraud and limited comprehensive audit transactions (Oduor, Sevilla, & Wanyoike, 2016). These challenges have limited the ability of county governments to reach their revenue targets.

However, some county governments have adopted automated revenue collection systems which has transformed revenue collection and improved service delivery. These systems have removed poor collection, losses and pilferage.

The financial year 2013/14 was the first time county governments undertook planning under the devolved government framework (Commission on Revenue Allocation, 2014). County governments in Kenya raised Shs 33.9 billion in the financial year 2014/15 against a target of Shs 50.4 billion. This was an improvement of 28% from 2013/14 where counties managed to raise on Shs 26 billion. This improvement can be attributed to automation of revenue collection (Commission on Revenue Allocation, 2015).

Figure 2: Own Source Revenue Trend

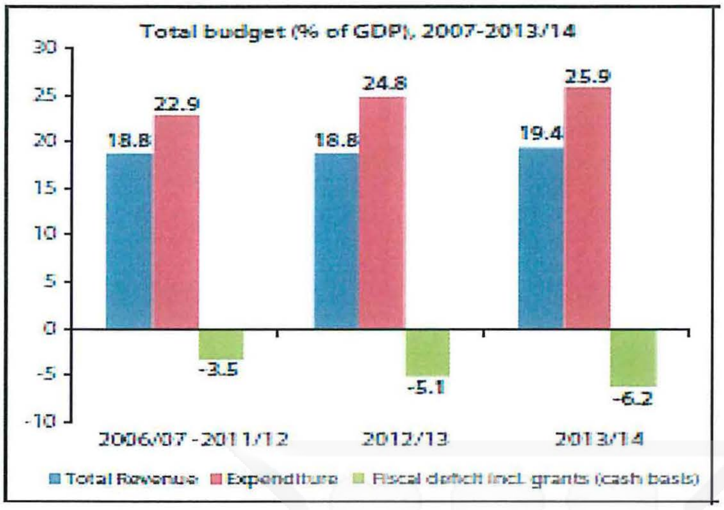


Source: World Bank Group (2014)

In addition, actual own source revenue (OSR) collection as a proportion of target in FY 2015/2016 dropped from 67.3% to 63.4%. It was also noted that OSR growth is growing significantly slower. In between FY 2013/14 and FY 2014/15, the counties recorded a growth of 28.9% but between FY 2014/15 and 2015/16, the growth in OSR was only 3.2%. This escalating negative misalignment between target and actual OSR collection highlight the difficulties counties face in preparing realistic forecasts (The National Treasury, 2017).

Government revenues are used to meet government expenditures. The main source of national government revenue is taxes. Whereas, for county governments, business permits and taxes are the main sources of revenue. If the government's expenditure exceeds the revenue collected, the government is forced to either borrow or raise taxes. All three options have a negative impact on the economy in the long run. Therefore, to maintain a balanced budget, government can either cut down on expenses or increase revenues. The former has proven to be quite difficult especially in trying to curtail recurrent expenditure as opposed to raising revenues. Under certain methodology, assumptions and constraints, revenue forecasts are done and are used in policy making to try and meet a balanced budget (Simba, 2012).

Figure 3: Total Revenue versus Expenditure



Source: World Bank Group

The figure above is obtained from a report done by World Bank indicating that expenditure exceeded the total revenue collected from the year 2007 to 2014.

Local governmnets in Nigeria generate internally generated revene through rates, taxes, fines and fees and other miscelanneous sources such as royalties and proceeds from commercial activities (Unyime, 2013). In the United States of America, states collect revenue from taxes from natural resources within the state, property taxes, sales and gross receipt taxes (Malm & Kant, 2013).

Budgeting Process

In Kenya, budgeting in county governments involves mobilizing sources of revenue and allocating the revenue to meet the needs of its people. At county level, budgeting begins with the integrated development planning process where the county plans and establishes its financial and economic priorities. Counties first receive the County Fiscal Strategy Paper, which provides guidelines on preparation of the budget which should be followed by all counties. This is followed by making estimates of the county government's revenue and expenditure. After this, the estimates are submitted to the County Assembly for approval before being handed over to the County Executive Committee for Finance (Oduor C. O., 2014).

Forecasting is a complex process in itself. More often than not, the predictions deviate from the actual revenue collected in a given period of time. States in the United States of America may use a number of analytical techniques while involving budgeters, policy makers, economists and finance officers to estimate individual state revenue forecasts. Sun and Lynch (2008), recognise that state governments in the US operate in fairly constrained environments where budget constraints and tax and expenditure limits place a premium on revenue and expenditure estimate accuracy. Generally, the process of forecasting involves many different types of quantitative and qualitative methods and the forecasts generated need different and multipe methods of analysis.

Internally generated revenue in Nigeria is restricted to moving-average based forecasts only (SPARC, 2016). This is because of limited state or local government level macroeconomic data. However, in forecasting Value Added Tax (VAT), moving average methodology or elasticity- based models are used.

Descriptions of budget preparation processes according to Kyobe and Danninger (2005), are generally not put down in formal documents and budgeting processes are often a mix of idiosyncratic budget practices and influences from legacy systems. Since revenue is undegenerated in counties, it is necessary to forecast revenue accurately so as to ensure that expenditures do not exceed revenues.

According to The National Treasury (2017), a recent survey done by the National Treasury of eight county governments found that revenue forecasting did not rely on macroeconomic assumptions or any credible methodology. Due to ambiguity within county governments' own source revenue forecasting process and failure in monitoring the budget within the year, revenue forecasts at county level are hardly updated.

1.1 Problem Statement

Despite the adaptation of automated revenue collection systems, revenue within local governments is still under-generated. This is because local governments have little knowledge and understanding of characteristics and behavior of own source revenue through modeling. This has led to counties failing to identify channels that generate the highest revenue that would have otherwise generated enough revenue to exceed expenses if managed well.

Kenya, with a population of over 40 million, has tremendous tax potential that can generate reliable and uninterrupted revenue to the county and national government. However, this can only be achieved through efficient and robust management of revenue channels that can generate reliable revenue enough to support recurrent and development expenditure. Budgeting and planning for county governments has been limited due to little knowledge and understanding of the trend, pattern and projections of own source revenue.

Therefore, this study sought to give a solution to modeling of own source revenue channels and forecasting through quantitative modeling of own resource revenue of counties in Kenya so as to be able to make realistic forecasts and make better decision on planning for the future through efficient management of revenue channels.

1.3 Aim of the Study

The purpose of this study was to assess the daily own source revenue and to try and identify the time series model of the own source revenue channels and forecast this revenue with a recommendation of efficient management of these revenue channels.

1.4 Research Objectives

1. To establish the trend and patterns of own source revenue in County Governments in Kenya.
2. To forecast own source revenue of counties in Kenya.

1.5 Research Questions

1. What trend and patterns does the own source revenue of County Governments in Kenya follow?
2. What is the forecast of own source revenue in County Governments in Kenya?

1.6 Significance of the study

This study is of significant importance to county government of Kenya as the model will assist in forecasting own source revenue for counties to enhance planning and budgeting processes. The forecast will assist the county governments in managing the revenue channels.

The model will assist the national government in allocation of revenue to the 47 County Governments. Depending on the forecast, it will enable the CRA to determine whether to increase or decrease allocation of revenue to counties of which the latter should be the case.

Chapter 2: Literature Review

2.0 Introduction

The objective of this chapter is to conceptualize the underpinnings upon which this study has been built on. It presents theories that have been used in revenue forecasting and considers few studies that have tested these theories. Their studies give insight on revenue forecasting models, their findings and methodology and how the same can be applied to the Kenyan County Government context.

2.1 Theoretical Framework

2.1.1 Revenue

Hamid (2008) defined revenue as the total income that accrues to a public or private organisation within a specified period of time. Government revenue can be generated from different sources such as taxes, borrowing, fines, fees etc. County governments' sources of revenue include taxes and single business permits as the core sources of revenue, property rates, licenses, fees, charges arising from delivery of services, interest and dividends, fines and fees, sales among others and grants and funds from the national government.

Hamilton (1994) stated that data collected over a period of time indexed by time form a time series. A time series can be decomposed into trends, patterns, that is, seasonal and cyclical, and lastly the residuals. Therefore, revenue collected by the county governments over a period of time can be referred to as a time series.

Montgomery et al ., (2008) stated that time series data can contain trends which may either be upward or downward. The trend models are classified into either linear, exponential or mixed (damped trend). In order to forecast, smoothing of this data is necessary. A linear trend means that the time series variables change by a constant amount in each period of time. Whereas, an exponential trend means that the time series changes by a constant percentage in each period of time.

Montgomery et al ., (2008) further stated that, a time series pattern can either be seasonal or cyclic. Seasonal patterns exists when the time series data is influenced by seasonal factors such as monthly, quarterly or daily and are related to the calendar. The season is always of a fixed and known period. For a period that is not fixed, a cyclic pattern exists when the data shows rises and falls and can be attributed to factors such as business cycles. The residuals form the random or systematic fluctuations.

Chatfield (2004) stated that a time plots gives a lot of information about the trend and the seasonal variations of the time series. The time plot also indicates the presence of outliers, that is, any data that is not consistent with the rest of the data.

2.1.2 Revenue Forecasting

Revenue forecasting techniques include qualitative and quantitative approaches. The qualitative approach imposes objectivity in the forecasting process by relying on expert judgment. According to Sun and Lynch (2008) this method is approached either through Delphi or expert judgment forecasting. A quantitative approach uses numerical data to generate forecasts. It relies on numerical data that is put through mathematical calculations under explicit assumptions to generate forecasts.

The methods for forecasting include general time series forecasting models, simulations, regression modeling and smoothing models (Montgomery, Jennings, & Kulahci, 2008). The commonly used techniques are, Unconditional Time Series Analysis and Conditional or Casual Model. Other techniques include but are not limited to Microeconomic Analysis, Structural Analysis and Micro-Simulation Analysis.

According to Jenkins et al . , (2000), macro-based models (conditional models) use aggregate data such as total consumption or imports in the economy to approach the problem from the perspective of the entire country's economy. They instead focused on micro-simulation techniques that estimate tax revenue where the unit of analysis was an individual, a household or a firm. They cited that the main advantage of using micro-simulation was that it enabled one to estimate the distributional effect of a given policy proposal on particular sectors.

Huizinga et al . , (2001) further stated that, macro models represent quantitative relationships among macroeconomic variables such as prices, employment, taxes, government expenditure, output, interest rates, exchange rates etc. These macro models are useful in forecasting and in policy analysis. These models use information from other series to explain behaviour of tax revenues (Simba, 2012).

Pescatori and Zaman (2011), recognized the use of Dynamic Stochastic General Equilibrium (DSGE) models for macroeconomic forecasts. The models follow a strict bottom-up approach, that is, from a micro to a macro level. The model begins by modeling household and firm behaviour from first principles. The aggregations derived from the microeconomic equations are then used to derive equations that relate macroeconomic variables such as output, consumption and investment.

However, these models are technically difficult to solve and analyse. Another challenge with these models is that they cannot easily absorb large array of high frequency data that is usually available to policy makers. This means that some variables are left out which may lead to serious misspecification. Nonetheless, these models provide forecasts that are very useful in policy making (Pescatori & Zaman, 2011).

Unconditional models forecast revenue by basing their forecast only on past revenue data and are not limited by other variables such as micro-economic and macro-economic variables. The forecast may be done as an unconditional prediction of the most likely outcome (Simba, 2012).

According to Golosov and King (2002), the method adopted in revenue forecasting is the major source of error in forecasting revenue as well as political and institutional factors.

Garret and Leatherman (2004) further stated that a quantitative approach is not necessarily more accurate than qualitative approach, as the nature of the revenue source dictates the best technique to use, whether quantitative or qualitative.

In addition, Sun and Lynch (2008) stated that a complex model for forecasting revenue does not necessarily imply accurate forecasts but rather a mix of both qualitative and quantitative approaches produces the best forecast. Therefore, a simple quantitative approach such as a simple trend analysis backed by qualitative expert judgments can be very useful in producing accurate forecasts.

The objectives of revenue forecasts include assessing the impact of revenue on policy making, assessing the impact of economic changes such as GDP, inflation, trade patterns on revenue, measuring the tax revenue potential of a department and the performance of the department in achieving its potential. Desirable characteristics of revenue forecast are accuracy and consistency with expenditure projections (Simba, 2012)

2.2 Empirical Evidence

Simiyu (2010), conducted a study on challenges affecting revenue collection and fund that county governments lose millions of shillings because tax officers or clerks accept bribes to reduce tax liability. This study backed what financial reports have shown concerning loss of billions of shillings through fraud between clerks and vendors. For this reason, domestic revenue mobilization faces many challenges resulting in inadequacy in financing county projects and in preparing realistic revenue forecasts. To curb these challenges, county governments have opted to adopt automated revenue collection systems to improve service delivery and accountability.

According to a report done by World Bank, for the financial year 2013/14, counties' own revenue collection was at a low of only 0.5% of the GDP compared to 1.2% of GDP annual target (World Bank Group, 2014). When looking across counties, actual own source revenue collected deviated significantly from the target. In overall, there was low level of revenue collection indicating that counties were inadequate in meeting service and equity targets.

International Monetary Fund (2014), found that in estimating revenue and expenditure, the forecast considers assumptions underlying growth and inflation such as regional and global trends, domestic conditions, international oil prices and fiscal and monetary policies. The revenue forecasted would then help in setting up performance targets.

2.2.1 Revenue Forecasting

Fullerton (1989), used a composite approach to model state government revenue in Idaho. He used a structural econometric model and a univariate ARIMA equation to generate baseline forecasts. He used constrained least squares analysis to determine the combination weights. His study revealed that the composite forecasts were more accurate than either econometric or ARIMA forecast and were more accurate than those previously published forecasts in the Idaho budget process from the year 1982 to 1985.

Howard (1990), further suggested a simple time-series models, that is, moving average model and exponential smoothing to forecast revenue of eight cities in Florida, USA within the fiscal year. He noted that limited expertise and data unavailability in local governments limited the use of these models. He further suggested that a degree of risk aversion may lead to revenue under-forecasting under the models tested in the study

Karingi and Ndung'u (2000), argued that there are major weakness in the two developed models for Kenya, MEPM AND MELT3, due to the fact that the models failed to recognize the importance of the supply-side factors in the Kenyan economy. This means that the forecast generated by the models disregarded the importance of structural rigidities in the economy in policy making. It was noted that the macroeconomic models were only useful in short-term forecasting and due to their weak representation of production in the real world, long-term forecasting was limited.

Gleday and Shukla (2001), evaluated revenue trends and forecasted revenue using trend analysis. Revenue collections from personal income tax, payroll tax and health contribution in Bhutan were evaluated, using exponential, linear and polynomial trends, three mathematical regression equations were developed to estimate revenue. To compare the accuracy of the model forecasts, Mean Absolute Percentage Error (MAPE) was used. He concluded that the most accurate projection model was the exponential model.

Using a sample of 34 countries from Africa, Asia, Middle East and Latin America, a study by IMF revealed that not all of the countries but a majority used macroeconomic variables as inputs in their revenue forecast (Kyobe & Danninger, 2005). The study revealed that only 13% of the sampled countries used formal econometric methods while 85% used a qualitative approach.

A survey done across states in the USA indicated that engaging multiple methods of forecasting, that is, a mix of quantitative and qualitative methods, realized better accuracy (Sun & Lynch, 2008). It was noted that highly complex analytical methods of forecasting do not necessarily result in greater accuracy. Forecasting remains as much of an art as it as a science as state governments continue to improve their accuracy in forecasting.

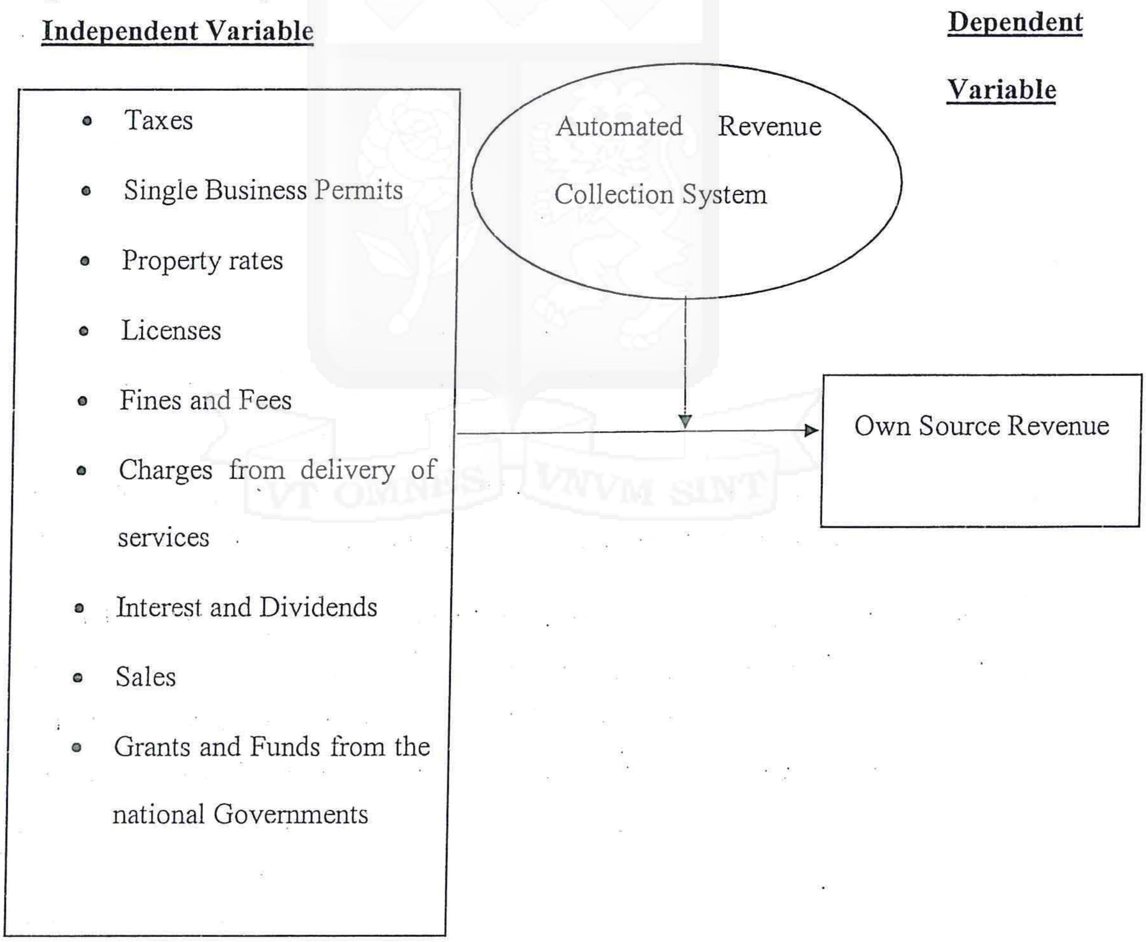
Sun and Lynch (2008), further revealed that states in the United States of America used a combination of simple trend analysis or linear regression along with concesus methods to estimate state revenues. Kansas, Utah and West Virginia states used simple trend analysis to frecats every tax revenue source whereas, Illinois, Minnesota, Mississippi and Ohio used simple trend analysis to forecast all revenue sources. Kentucky, Utah, New Mexico, Louisiana and Michigan used concencus forecasting for all the tax revenue sources. It was noted that the states reevaluated their estimates quartley rather than monthly but very few states indicated having a process of reevaluation in their revenue predictions.

Busari et al ., (2013), focused on the impact of Internally Generated Revenue (IGR) of local governments in Nigeria and their impact on the growth of Local Governments in Nigeria. They used pooled Ordinary Least Squares (OLS) regression technique to fit IGR of twenty local government areas of Lagos State in Nigeria. To validate their model, they used coefficient of determination (R^2), correlation coefficient (R), Durbin-Watson and F-statistic. Their result showed that the IGR had a huge impact on the growth of local governments.
Unyime (2013), came on to build on Busari's work by using Box and Jenkins method of time series analysis to forecast monthly internally generated revenue of Ikot Ekpene Local Government Area of Akwa Ibom State in Nigeria. From his analysis, ARIMA (2, 1, 2,) was identified as an appropriate model for the series and the Theil Coefficient indicated that the model was a perfect fit. According to SPARC (2016), internally generated revenue in Nigeria is forecasted using moving average-based models only. The revenue projection tool estimates the aggregate revenue collections based on historical moving average growth rates for an individual state or local government.

2.3 Conceptual Framework

The revenue for county governments in Kenya is the dependent variable of this study. The sources of revenue for the counties form the independent variables. These include, taxes and single business permits as the core sources of revenue, property rates, licenses, fees, charges arising from delivery of services, interest and dividends, fines and fees, sales among others and grants and funds from the national government. If these sources are high, then the revenue collected is high. How these sources behave over the years is used to predict how revenue will behave.

Figure 4: The Conceptual Framework



2.4 Research Gaps

A study conducted by IMF found that data and human capital constraints in low income countries led to the use of qualitative estimation techniques as the main methods for deriving revenue forecasts as econometric measures require a wealth of reliable and detailed data (Kyobe & Danninger, 2005).

Although literature has been reviewed on various techniques and methods used to forecast revenue for decentralized systems of governments, most of these studies are done in countries that have had devolved governments for a long period of time. Kenya is only at its adaptation stage and little research has been done concerning the prediction of own source revenue of county governments in Kenya.

Therefore, there existed a literature gap in modeling and forecasting of own source revenue of county governments in Kenya. This study therefore sought to fill this gap by using a quantitative unconditional approach in modeling and forecasting own source revenue for county governments. This will provide reliable forecasts that will assist in revenue management and policy making.

Chapter 3: Research Methodology

3.0 Introduction

This chapter addresses the methodology used to achieve the objectives set out in chapter one. It sets out the method, technique and procedures of the data collection as well as data analysis that was used to collect, analyse and interpret data which gives an overview of how the entire study was conducted to reach a final conclusion on the trend of own source revenue county governments and provide a forecast for county's own source revenue.

3.1 Research Design

The research was descriptive in nature as it aimed at portraying an accurate profile of the trend of county's own source revenue as well as finding the most suitable Time Series Model to forecast the OSR (Saunders, Lewis, & Thornbill, 2009).

The research was also quantitative in nature as it focused on numerical data and the output is numerical as well (O'Leary, 2010).

3.2 Population

The target population was the revenue channels for the 47 County Governments of Kenya. The research focused only on revenue channels that generated own source revenue for county governments in Kenya.

3.3 Sample

The sample size was drawn from daily own source revenue data of one of the county governments in Kenya with a focus of general charges given that there are over twenty revenue streams in the county finance act. Given the time frame of the existence of the County Governments from 2013 to date, the study of monthly or yearly revenue would provide inadequate results due to insufficient amount of data as required for optimum modeling under time series. However, the identity of the county cannot be disclosed due to confidentiality of the data.

A purposive sampling method was applied by choosing predefined own source revenue, that is, only revenue that can be collected on a daily basis was selected. The study was conducted with the purpose in mind of selecting the best suited data and target only daily data. This method was useful as it enabled the study to target only daily own source revenue for the county under consideration.

3.4 Data Collection

This study only used primary data. The primary data of the daily own source revenue of the county, was obtained from the county's revenue collection system, time varying from 2015 to 2017. Since the data was obtained from observations collected sequentially (daily) over time, the data was grouped as time series data.

3.5 Data Analysis

The research conducted a quantitative analysis as it is used graphs to explain the trend of the collected data as well as using the unconditional method of forecasting which required only the data of the own source revenue to forecast revenue independent of other factors otherwise called conditions.

3.6 Model Building for own source revenue

Box and Jenkins (1976) suggest the use of a multi-step building strategy in building a forecast model. The process involves model specification or identification, model fitting and model diagnostics which was used in the study.

To determine the order of the own source revenue, the following steps were applied.

Step 1: Examining the time series plot

The first steps involved plotting several time series plots over the period 2015 to 2017 and examining the plot to determine the components present in the time series. These are, the trend, the seasonal component and the outliers.

Tests

To model using time series data, the series has to be stationary to give good results and predictions. Time series are either stationary or non-stationary, a time series is stationary if the finite dimensional distributions are invariant under a time shift, that is, there's no systematic change in the mean and in the variance.

A time series is strictly stationary if the joint distribution of $Y_{t1}, Y_{t2}, Y_{t3}, \ldots, Y_{tn}$ is the same as the joint distribution $Y_{t1+h}, Y_{t2+h}, Y_{t3+h}, \ldots, Y_{tn+h}$ for all h ,where h is the distance between the observations.

The study used Autocorrelation Coefficient Functions (ACF) and Partial Autocorrelation Functions (PACF) graphs to test for stationarity. A decaying exponentially ACF implies that the data series is not stationary.

The data was not stationary; therefore, the data was differenced to achieve stationarity according to Box-Jenkins. The first difference which is stationary is:

$$y_t = y_t + y_{t-1} = \mu_t \tag{3.2}$$

Therefore, y is "integrated of order 1".

In other cases, there will be need to differentiate up to the d[th] difference to obtain stationarity.

Step 2: Examination of correlogram

A correlogram may show the trend and the seasonal components. A slowly damping correlogram indicates a slowly varying trend while a periodic fluctuating correlogram, indicates a periodic component. To remove any non-stationary components, the data was differenced at the appropriate time lags.

Several graphs of the own source revenue were drawn. The graphs were used to identify the trend of the data for the period mentioned.

Step 3: Determining the MA-order and the AR-order from the ACF and PACF

The behaviour of autocorrelation and partial correlation functions determines the model to be used. After achieving stationarity of the data, several models were used in the time series analysis. These include, Auto Regressive Model (AR), Moving Average Model (MA), Auto Regressive Moving Average Model (ARMA) and Auto Regressive Integrated Moving Average Model (ARIMA).

Auto Regressive Model (AR)

The general model for an Auto Regressive Model of order p [AR (p)] is given by;

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t \quad (3.3)$$

Where:

Y_T Is the daily own source revenue

$\emptyset_p$ Are the autoregressive parameters

Y_{t-i} Are prior observations

ε_t Is a purely random process that follows N (0, σ^2)

If p=1, $Y_t = \emptyset_t Y_{t-1} + \varepsilon_t$ (3.4)

And this is termed as Auto Regressive model for order one [AR (1)] with the properties

$$E(Y_t) = 0 \quad (3.5)$$

$$Var(Y_t) = \frac{\sigma^2}{1-\emptyset_i^2} \text{ If abs } (\phi_i) < 1 \quad (3.6)$$

If p = 2, $Y_t = \emptyset_t Y_{t-1} + \emptyset_2 Y_{t-2} + \varepsilon_t$ (3.7)

And this is termed as the Auto Regressive Model of order two [AR (2)]

Moving Average Model (MA)

The Moving Average of order p [MA (q)] is given by;

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3.8}$$

Where:

Y_t Is the series

θ_i Are Moving Average Parameters of the model.

ε_{t-i} Are the prior random shocks (prior white noise error terms)

If q=1, $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$

(3.9)

And this is termed as Moving Average Model of order one [MA (1)] with the properties:

$$E(Y_t) = \mu \tag{3.10}$$

$$\mathrm{Var}(Y_t) = \sigma^2 \sum_{i=0}^{q} \theta_i^2 \tag{3.11}$$

Auto Regressive Integrated Moving Average (ARIMA)

The general model for a non-seasonal ARIMA (p,d,q) model where:

- p is the number of autoregressive terms
- d is the number of non-seasonal differences needed for stationarity
- q is the number of lagged forecast errors in the prediction equation

Let y denote the d^{th} difference of Y which means,

If d=0: $y_t = Y_t$

$$(3.13)$$

If d=1: $y_t = Y_t - Y_{t-1}$

$$(3.14)$$

If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

$$(3.15)$$

This gives the general forecasting equation as:

$$\hat{y}_t = \mu + \emptyset_1 y_{t-1} + \cdots \emptyset_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

$$(3.16)$$

The most used model for forecasting a time series is the ARIMA model for which series are "fine-tuned" to make them stationary. The fine tuning involves transformations such as differencing and lagging the differenced series and/or the forecast errors to remove any autocorrelation from the forecast errors.

The autoregressive term in the ARIMA model is the lags of the differenced series while the lags of the forecast errors form the moving average term in the forecasting equation. To make the time series stationary, it is differenced and is said to be an integrated version of a stationary series.

Other special cases of ARIMA models include random walk and random-trend models, autoregressive models and exponential smoothing models.

Table 1: Identification of Models

Behaviour of theoretical ACF and PCF for stationary process		
Model	ACF	PACF
MA (q)	Finite. Cuts off after lag q	Infinite. Exponential decay and/or damped sine wave. Tails off
AR (p)	Infinite. Exponential decay and/or damped sine wave. Tails off	Finite. Cuts off after lag p
ARIMA(p,q)	Infinite. Exponential decay and/or damped sine waves after first p-q lags. Tails off.	Infinite. Exponential decay and/or damped sine waves after first p-q lags. Tails off.

Step 4: Model Estimation

To fit Box-Jenkins models, non-linear least squares and maximum likelihood estimation approaches are commonly used. In this study, the non-linear least squares approach was used. This model was used to generate the revenue forecast.

Step 5: Model Diagnostics

To diagnose models, the error term μ_t is assumed to follow a stationary unvaried process while the residuals should be white noise and have normal independent distributions from a fixed distribution with a constant mean and variance.

One can also use the Theil Coefficient to test the accuracy of the model.

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}(\hat{y}_{t+1} - y_{t+1})^2}{\sum_{t=1}^{n-1}\left(\frac{\hat{y}_{t+1} - y_{t+1}}{y_t}\right)^2}} \tag{3.17}$$

Where U lies between zero and one. If U is closer or equal to zero, then the estimates are adequate and the model is a perfect fit.

3.7 Other tests to be conducted

Testing for normality of the data

Moments around the mean are used to test for normality. These include the mean, the standard deviation, kurtosis and skewness. A data set is said to be normally distributed if at the basic level, the mean is equals to zero, the Standard deviation is one, kurtosis is three and skewness is zero (DeCarlo, 1997).

- Skewness is the third moment. It is defined as

$$S = \frac{E\left(Y - \mu_Y\right)^3}{\sigma_Y^3} \tag{3.18}$$

For the data set to be normally distributed, skewness is equal to zero, S=0

- Kurtosis is the fourth order moment and is defined as

$$K = \frac{E\left(Y - \mu_Y\right)^4}{\sigma_Y^4} \tag{3.19}$$

For the data set to be normally distributed, kurtosis is equal to 3, k=3

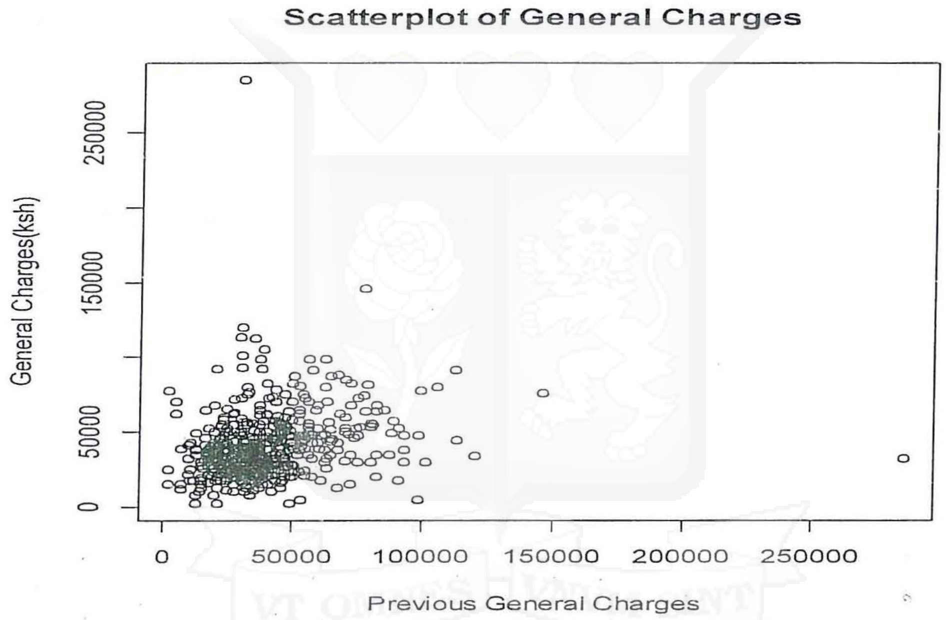
- QQ plots are also used to determine the normality and randomness of the data set. If the values fall along the straight line, then the data is said to be normally distributed otherwise it is not normally distributed.

Chapter 4: Results and Discussion

4.1 Exploratory Data Analysis

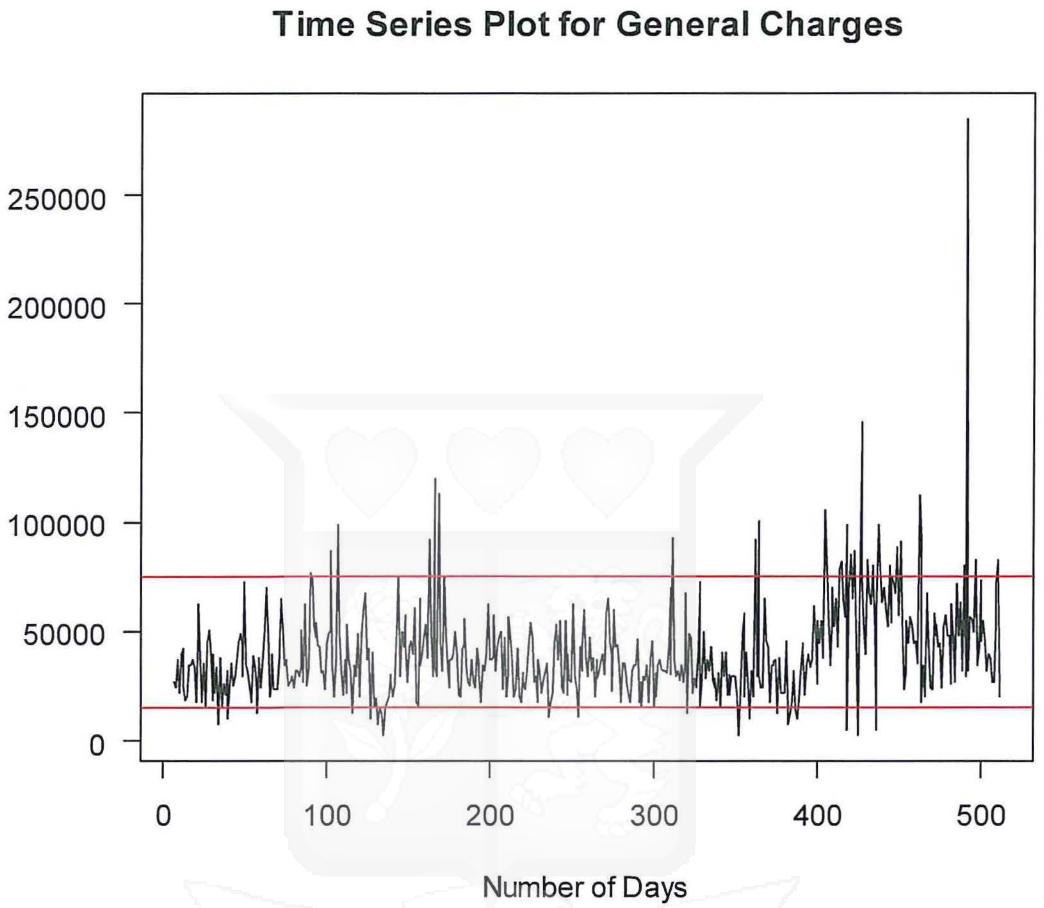
This section considered an exploratory data analysis of the series in order to identify the trend, the components of the series and the type of distribution that the data follows.

Figure 5: A Scatterplot of the General Charges



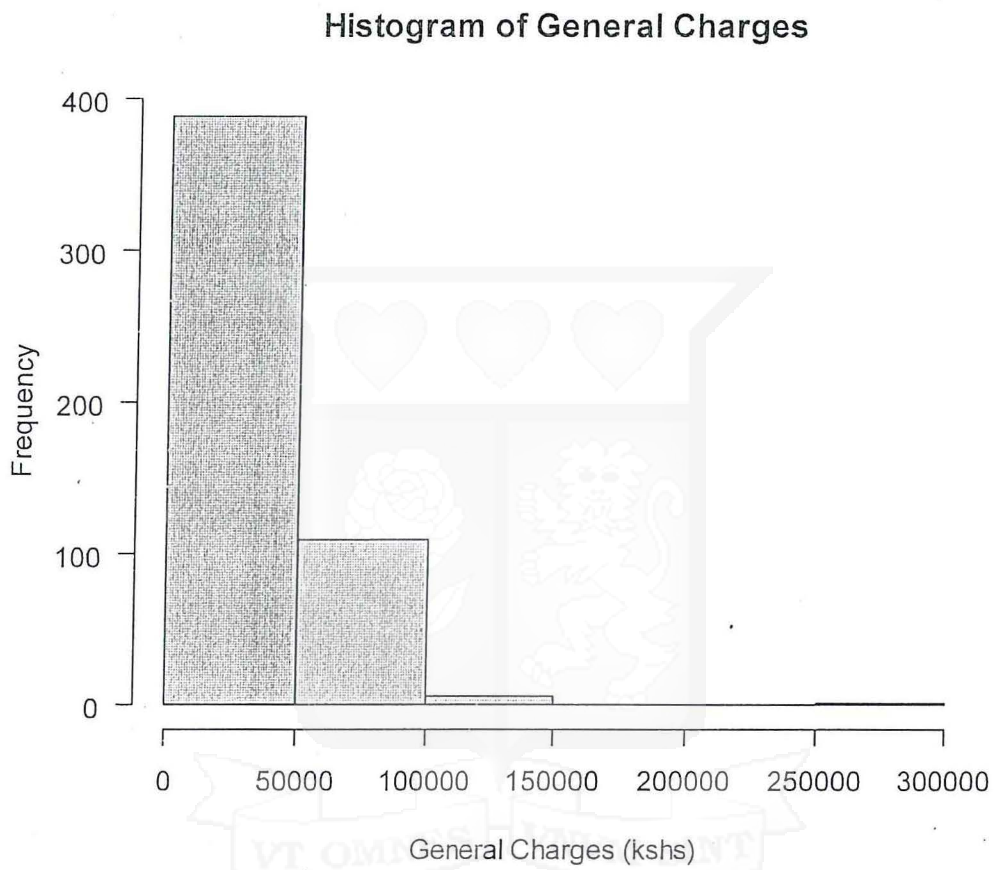
The scatterplot above in figure 5 showed that the data was concentrated in one region between the previous year's general charges and the current year's general charges. This showed that the data had little variation but instead remained within a certain range.

Figure 6: A Time Series Plot of the General Charges



The General Charges series is shown in the figure 6 above with data fluctuating within the range of about kshs 15,000 to kshs 70,000. The fluctuations are due to different revenue sub items being collected from day to day during the year. A large part of the data showed a regular pattern following a horizontal trend despite very few instances with spikes. The data exhibited no seasonal or cyclical pattern as the observations from day to day remained relatively within the same range. This indicates that if the data follows the same trend, then the forecast will be within the same range.

Figure 7: Histogram of General Charges



From the histogram above in figure 7, it was observed that most of the observations were located around ksh 50,000. This shows that the data is highly skewed to the left suggesting that the data does not follow a normal distribution.

Figure 8 below shows the contribution of four main revenue sub-items. Over the two-year period under analysis, items such as sale of budget copy, sale of project copy, perusal of county plan and hire of town chairs had no amount collected. Clearance certificates for rates and rents emerged as the main source of revenue for general charges followed closely by application for new plots or stalls. This indicates that the residents are geared towards opening businesses and renting out property.

Figure 8: Contribution of Each Revenue Sub-Item

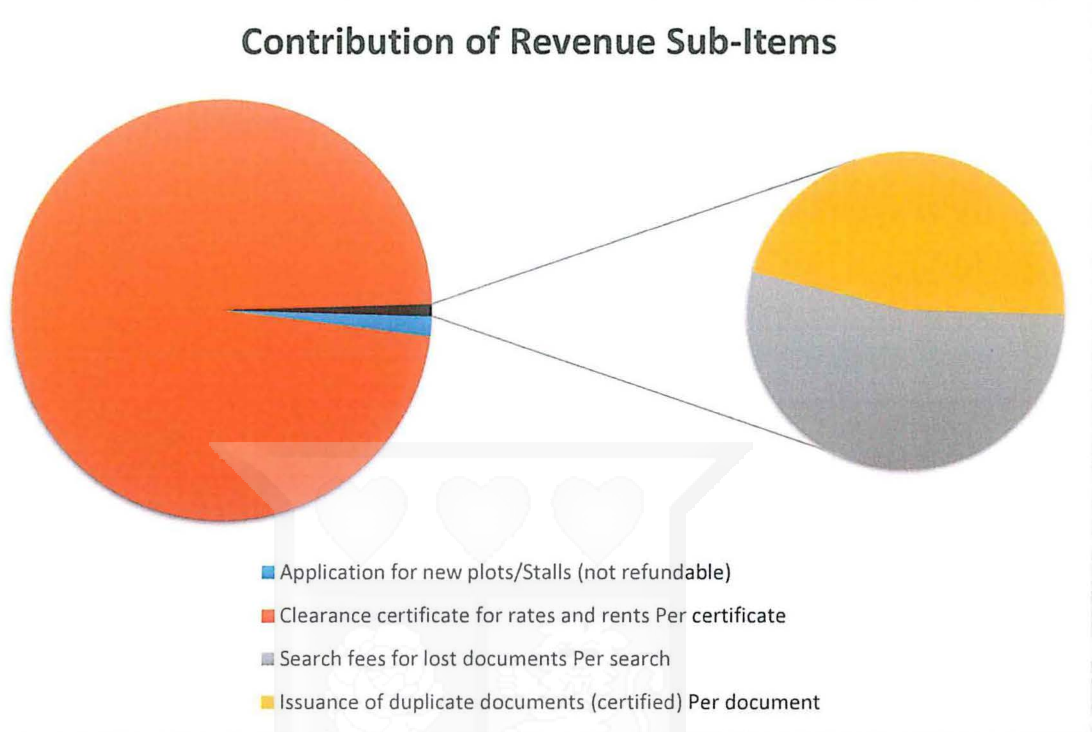
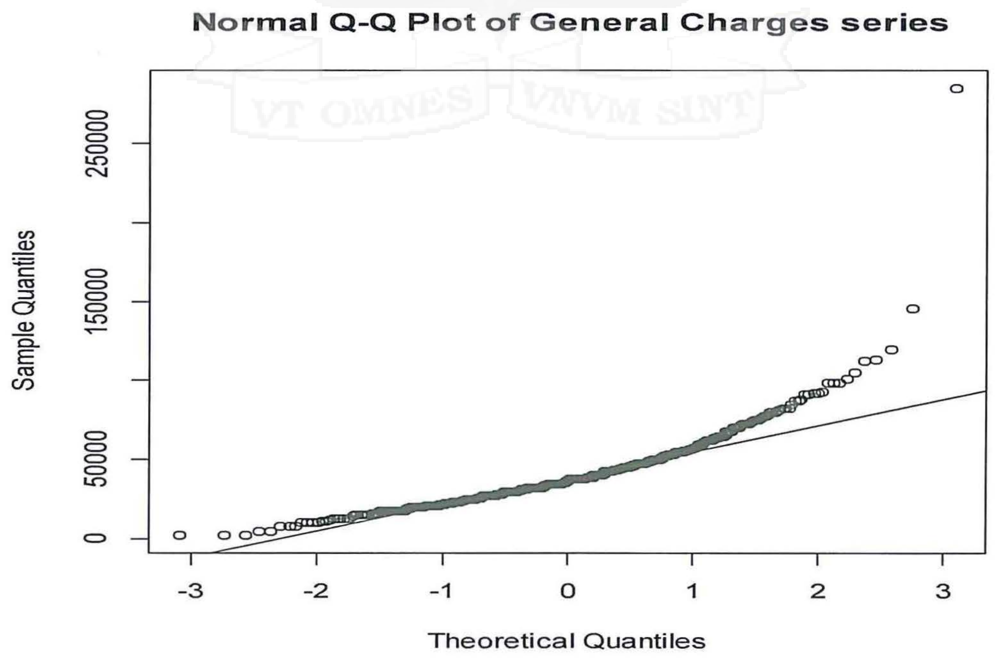


Figure 9: Quantile-Quantile plot for General Charges



The plot above in figure 9 shows the normal quantile-quantile plot for the general charges series. There is considerable curvature in the plot. The line passing through the second and third quartiles helps in pointing out the departure from a straight line in the plot.

Table 2: Descriptive Analysis of General Charges

Number of Observations	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Maximum
505	40,494	22.831.6	3.2321	26.346	2,500	285,000

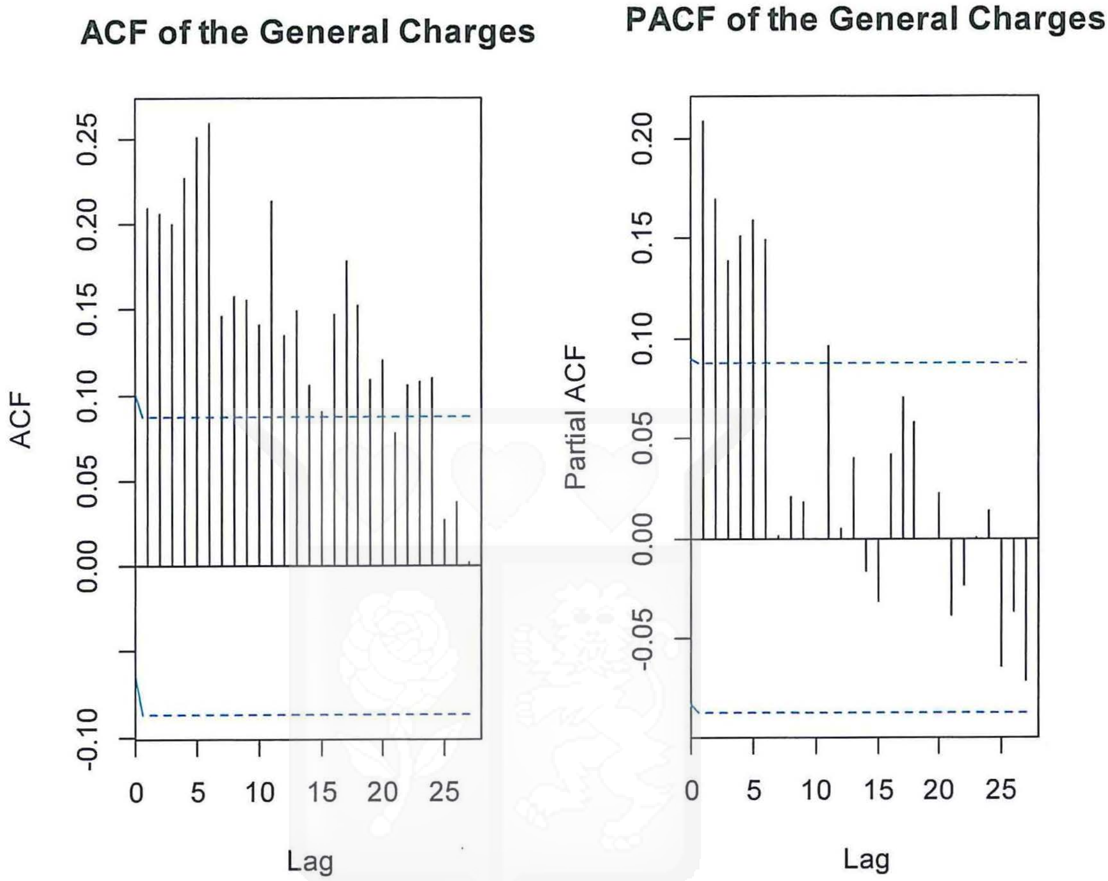
A data set is said to be normally distributed if the mean is equals to zero, the standard deviation is one, kurtosis is three and skewness is zero. Based on the observations, general charges are not normally distributed.

Kurtosis is greater than three. This means that the data set has heavier tails than a normal distribution and is referred to as a leptokurtic distribution. The skewness indicates the symmetry of the data set. In this case, the skewness is greater than zero indicating that the size of the right-hand tail is larger than the left-hand tail. This can be clearly observed through the histogram of the original data.

4.2 Testing for stationarity in the series

The stationarity of the general charges was tested through the autocorrelation function (ACF) test. ACF test is a unit root test. ARIMA models require a stationary working series in order to estimate and diagnose the model.

Figure 10: Plot of the ACF and PACF of the original data



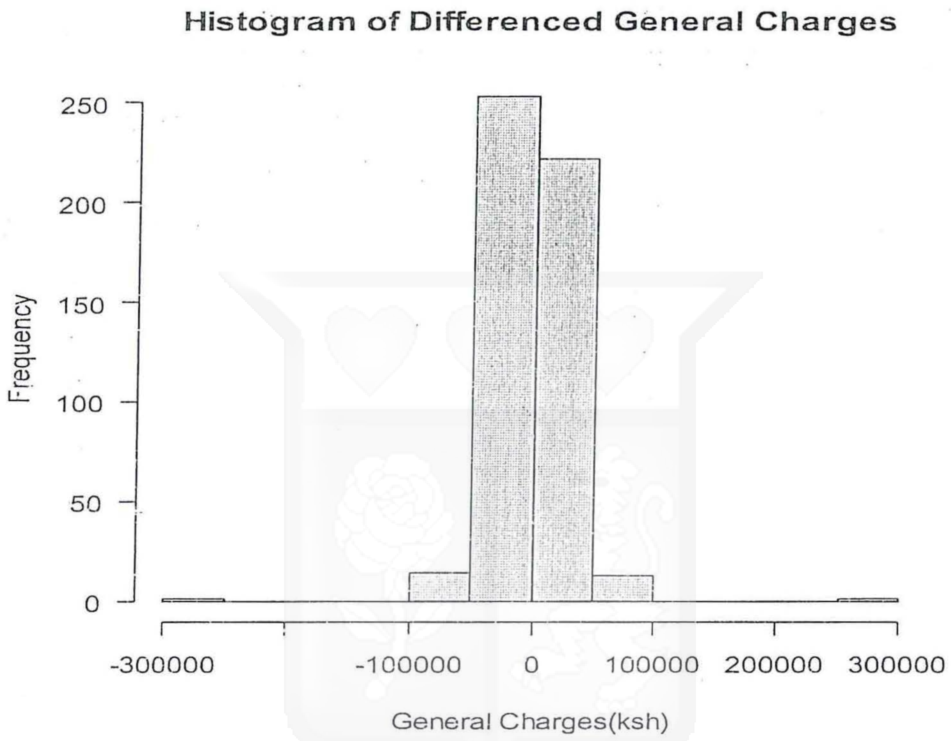
The autocorrelation function shown above in figure 10 indicated that the autocorrelations were strong and positive, starting high and decaying slowly thus indicating that the series was non-stationary and therefore, needed to be differenced. This suggested an ARIMA model.

In order to fit an ARIMA model the data needed to be stationary. Thus, the data was differenced and tested for stationarity.

Figure 11: Time series Plot for Differenced General Charges

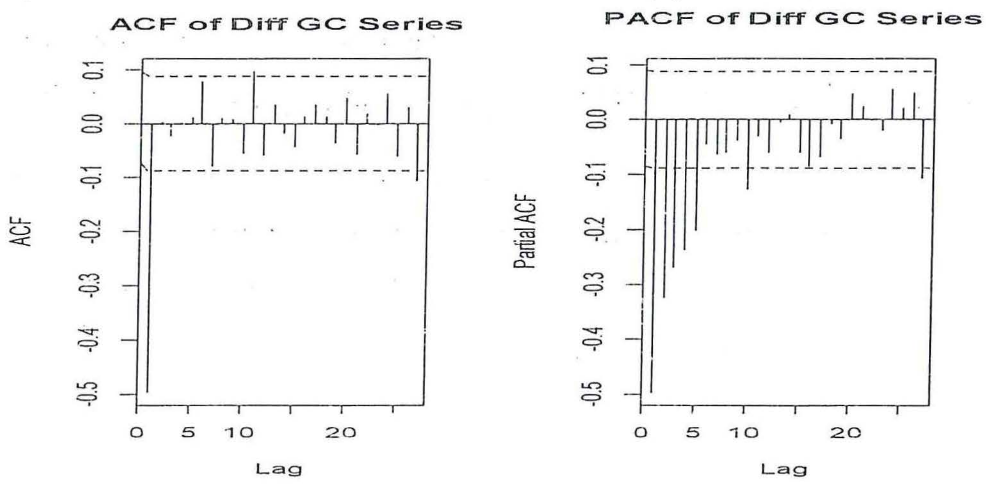
The first lagged difference from the original data time series shown above in figure 11, revealed a stationary series as the mean of the differenced data was around zero. The differenced data was less autocorrelated than the original data. However, there were some spikes in the figure representing high volatility periods.

Figure 12: Histogram of differenced General Charges Series



The histogram of the differenced series showed that the differenced series was near normal.

Figure 13: Figure Plot of ACF and PACF of first order series



From figure 13 above, the autocorrelations were distributed over the period and thus suggested stationarity. It further showed that only the autocorrelation at lag 1 was significant. Based on this autocorrelation plot, an MA (1) model was suggested for the differenced data. The partial autocorrelation of the differenced data suggested that the first, second, third, fourth and fifth lags were significant. This suggested an AR (5) model for the differenced data. Based on this plot, the ARIMA (0,1,1) model was examined in detail.

In the next step, Box-Jenkins method was used to find our model.

4.3 ARIMA Model Identification

A plot of the ACF for the differenced values of the series was done. The patterns of the ACF were used to determine the parameter values of p and q for the ARIMA model and showed that the autocorrelation at lag 1 was significant. Thus, the ARIMA model was set to be ARIMA (0,1,1). Since it has no autoregressive term, it is an integrated moving average, IMA (1,1).

4.3.1 Parameter Estimation IMA (1,1) model

Parameter estimations of the model are tabulated in the table below.

Table 3: Estimation equation of IMA (1,1) Model

Source	Parameter Estimate	Standard Error
Intercept	40.2566	97.6847
Ma1	-0.8987	0.0243
Estimate of Sigma Squared	451618496	
Log Likelihood	-5737.91	
AIC	11,479.83	

The general equation for an ARIMA model is given by

$$\hat{y}_t = \mu + \emptyset_1 y_{t-1} + \cdots \emptyset_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (4.1)$$

Since p=0 and q=1, from the summary of the ARIMA generated, the model equation was formed as: -

$$\hat{y}_t = \mu - 0.8987\varepsilon_{t-1} + 40.2566$$

(4.2)

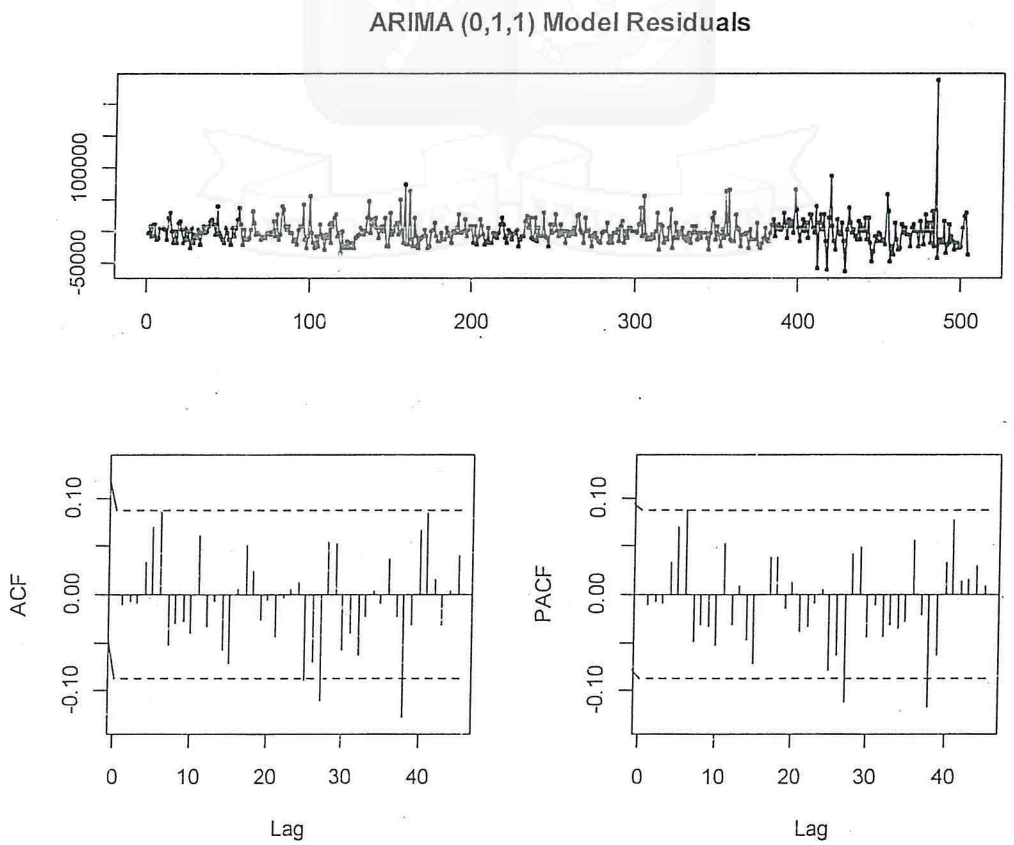
Equation 4.2 indicates that the current revenue estimate is a negative function of the previously collected revenue. This can be interpreted to mean that the estimated revenue will not be an increase of the previous revenue but instead a decrease of it.

4.3.2 Diagnostic Check of Model Adequacy

After fitting the model, the model was validated to check if it was appropriate using the non-linear least square fitting which uses residual analysis.

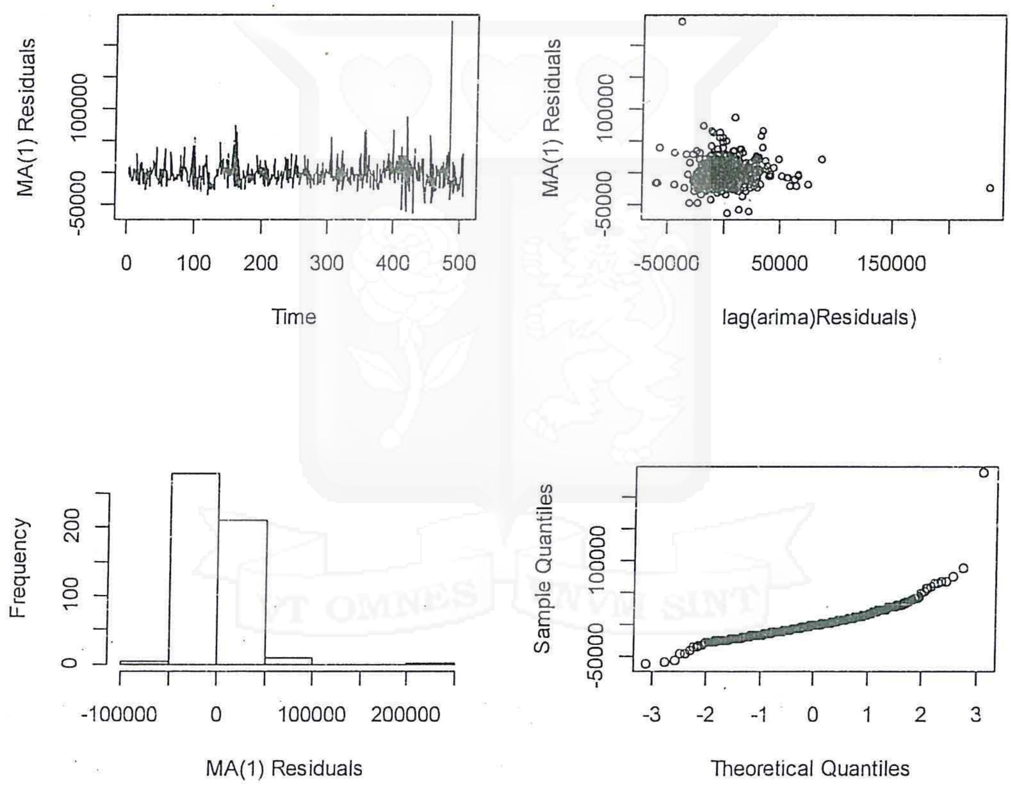
Figure 14 below illustrates a correlogram of residuals for IMA (1,1), it shows that the residuals are relatively small for both the ACF and PACF.

Figure 14: Plot of Residuals



The 4-plot is a convenient graphical technique for model validation as it tests the assumptions for the residuals on a single graph.

Figure 15: 4-Plot of Residuals



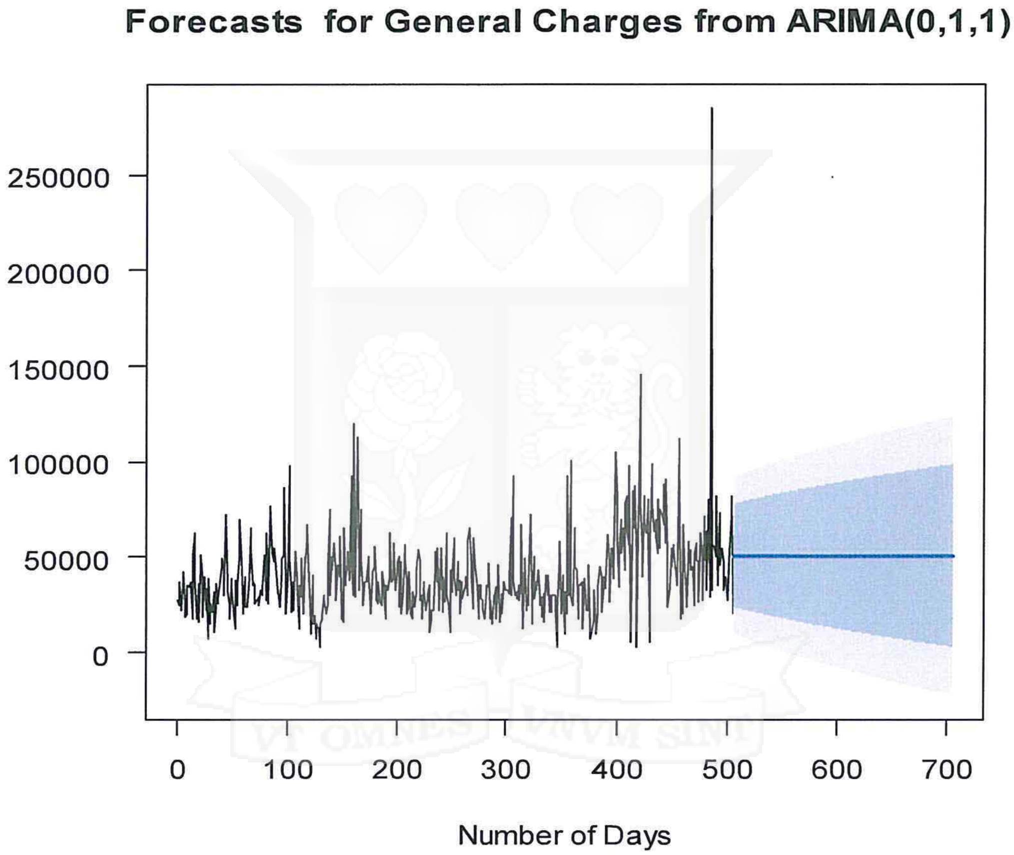
The run sequence plot shown in figure 15 above showed that the residuals did not violate the assumption of constant location and scale. The histogram and normal probability plot indicated that the normal distribution provided an adequate fit for this model.

The Box-Ljung test was also applied to the residuals. With a lag of fifty, the test's p-value was 0.06592 which is greater than 0.05. This means that the residuals are independent thus, the model was an appropriate model.

4.3.3 The Forecasting

In this step, the forecast of the general charges is done using IMA (1,1). Figure 16 below shows the general charges series together with the forecasts confidence bounds of 80% confidence limits in the darker blue region and 95% in lighter blue region.

Figure 16: Forecast from ARIMA (0,1,1) Model



The forecasts generated above for the general charges for the next 200 days were predicted to be within a range of ksh 5000 and ksh 100,000. As such, the revenue forecasted holding other factors constant indicates that the predicted revenue collected from general charges will remain within the same range unless other adjustments are put in place.

Chapter 5: Conclusion and Recommendation

5.1 Conclusion

This study used Time Series Analysis to model and forecast the daily own source revenue of a particular county in Kenya. Based on the results, the exploratory data analysis showed that the general charges exhibited a horizontal trend, that is, revenue collected from general charges remained relatively the same from day to day. The histogram revealed that the data was not normally distributed but was instead highly skewed to the left as the data was mainly concentrated at around kshs 50,000. Stationarity tests revealed that the data was non-stationary. This shows that the original data was not suitable for estimating and forecasting using any models and was thus differenced to obtain stationarity.

The results indicated that an IMA (1,1) model projection of the general charges provided a good forecast. The forecast indicated that the revenue would continue with the same trend with no increase in revenue collected. The IMA (1,1) that was identified in this paper was contrary to a similar research done by Unyime (2013), that found an ARIMA (2,1,2) as a suitable forecast model for modeling internally generated revenue of local governments in Nigeria.

In conclusion, the series shows that the revenue collected from general charges stagnated around the same range through the two years under analysis. This could be attributed to a slow adjustment from changing from a manual collection system to an automated revenue collection process.

5.2 Recommendations

The ARIMA models are useful in revenue forecasting and should be fully utilized since they are easy to apply. This research study solely looked into one revenue stream therefore, further research should be done on different revenue streams to identify if all revenue streams follow the same pattern or not. Further research should be done into identifying the optimal revenue channel having factored in certain constraints.

However, in order for these tasks to be achieved, the system should be able to generate data that can be easily analysed. That is, the system should generate the data as written in the county revenue act so as to enable ease of analysis.

This model provides only a quantitative baseline to measure other models. Thus, further research should be carried out in this area with a more complex model that incorporates qualitative aspects of revenue generation and collection in order to determine the correct forecast with a greater degree of certainty. Following the National Treasury's report, such a model can be updated regularly and fitted further with macroeconomic assumptions or any other credible methodology in order to eliminate ambiguity in own source revenue forecasting and budget making process and to allow for monitoring of the budget within the year.

The county under study has an automated revenue collection system. With this, it was expected that the county's revenue collection would increase due to automation. However, holding other factors constant, the data fails to display an upward trend in daily revenue collection in general charges. Therefore, measures need to be put in place to ensure that residents are incentivized to follow the required channels to pay for general services.

The forecast shows that the county will not be able to sustain itself through only its own source revenue. Therefore, the Commission for Revenue Allocation should maintain the current revenue allocation to counties as the forecast reveals that the collection of general charges will remain the same holding other factors constant.

References

Busari, G. A., Akinbo, R. O., & Adeboye, N. O. (2013). Impact of Internally Generated Revenue on the Growth of Local Governments on Nigeria. *Indian Journal of Research*, 267-268.

Chatfield, C. (2004). *The Analysis of Time Series.* Chapman & Hall.

Commission on Revenue Allocation. (2014). *Commission on Revenue Allocation Annual Report and Financial Statements 2013-2014.* Nairobi: CRA.

Commission on Revenue Allocation. (2015). *Recommendation on the Sharing of Revenue raised Nationally Between the National and County Governments for the Financial Year 2016/2017.* Nairobi: CRA.

DeCarlo, L. T. (1997). On the Meaning and Use of Kurtosis. *Psychological Methods Vol.2 , No. 3*, 292-307.

Fullerton, T. (1989). A Composite Approach to Forecasting State Government Revenues: Case of Idaho Sales Tax. *International Journal of Forecasting*, 373-380.

Garret, T. A., & Leatherman, J. C. (2004). *Selected Applications in Public Finance, Revenue Forecasting.* Manhattan.

Glenday, G., & Shukla, G. P. (2001). *Revenue Forecasting and Estimation, Techniques and Applications.* Durham: DCID, Duke Univeristy .

Golosov, M., & King, J. (2002). Tax Revenue Forecasts in IMF-Supported Programs. *IMF Working Paper.*

Hamilton, J. D. (1994). *Time Series Analysis.* Chichester: Princeton Univeristy Press.

Howard, F. (1990). Municipal Revenue Forecasting with Time Series Models: A Florida Case Study. *The American Review of Public Administration* , 45-59.

Huizinga, F., Fole, A. G., Karingi, S. N., & Ndung'u, N. S. (2001). *Theoretical Base for the Kenya Macro Model : The KIPPRA-Treasury Macro Model.* Nairobi: Kenya Institure for Public Research and Analysis.

International Monetary Fund. (2014). *Kenya, Fiscal Transparency Evaluation.* Washington D.C.: International Monetary Fund.

Jenkins, G., Kuo, C.-Y., & Shukla, G. (2000). *Tax Analysis and Revenue Forecasting.* Cambridge: Harvard Institute for International Development.

Karingi, S., & Ndung'u, N. (2000). *Macro Models of the Kenyan Economy: A Review.* Nairobi: Kenya Institute for Public Policy Research and Analysis.

Kyobe, A., & Danninger, S. (2005). Revenue Forecasting - How is it done? Results from a Survey of Low-Income Countries. *IMF Working Paper (WP/05/24).*

Maina, M. W. (2016). Factors Affecting Revenue Collection Efficiency in County Governments in Kenya. *International Journal of Recent Research in Commerce Economics and Management,* Vol. 3, Issue 2, pp 85-90.

Maim, E., & Kant, E. (2013). The Soruces of State and Local Tax Revenues. *Tax Foundation Fiscal Fact,* No. 354, pp 1-6.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting.* Hoboken: John Wiley & Sons, Inc. .

Oduor, C. O. (2014). *Handbook on County Planning, County Budgeting and Social Accountability.* Nairobi: Institute of Economic Affairs.

Oduor, C., Sevilla, J., & Wanyoike, T. (2016). *Automating revenue collections in Kenya: A case of Kiambu County.* Durban : IEEE.

O'Leary, Z. (2010). *The Essential Guide to Doing Your Research Project.* London: Sage .

Pescatori, A., & Zaman, S. (2011). Macroeconomic Models, Forecasting and Policymaking . *Economic Commentary,* 1-4.

Salami, A. (2011). Taxation, Revenue Allocation and Fiscal Federalism in Nigeria: Issues, Challenges and Policy Options. *Economic Annals,* Vol LVI, No. 189, pp 27-50.

Saunders, M., Lewis, P., & Thornbill, A. (2009). Research Methods for business students . Harlow: Pearson Education Limited.

Simba, J. N. (2012). *Revenue Forecasting: A Case Study of Import Duty.* Nairobi.

Simiyu, D. (2010). *Challenges Affecting Collection of Turnover Tax in Nairobi County.* Nairobi : Kenyatta University .

SPARC. (2016). *Revenue Projection Tool.* State Partnership for Accountability, Responsiveness and Capability.

Sun, J., & Lynch, T. D. (2008). *Government Budget Forecasting: Theory and Practice.* Boca Raton: CRC Press.

The Constituion of Kenya . (2010). Nairobi: National Council for Law Reporting.

The National Treasury. (2017). *2017 Budget Policy Statement.* Nairobi: Public Relations Office.

Unyime, P. U. (2013). Modeling Internally Generated Revenue (IGR) of Local Governments in Nigeria. *Mathematical Theory and Modeling, Vol.3, No.14* , 117-125.

World Bank Group. (2014). *Decision Time: Spend More or Spend Smart? Kenya Public Expenditure Review.* Nairobi: Macroeconomics & Fiscal Management, Trade & Competitiveness and, Governance Global Practices Africa Region.