UNIVERSITY OF NAIROBI

SCHOOL OF MATHEMATICS

Revenue Forecasting: A Case of Import Duty.

This research project is submitted to the School of Mathematics of the University of Nairobi in partial fulfillment of the requirement for the degree of Masters of Science in Social Statistics.

By

Jeliah Nyanduko Simba

©July 2012

DECLARATION

This project as presented in this report is my original work and has not been presented for any other university award

Student:

JELIAH NYANDUKO SIMBA **Reg. No.: I56/78050/09**

Signature: ____________________ Date: 30 / 07 /2012

This project has been submitted as a partial fulfillment of the requirements for Masters of Science in Social Statistics of the University of Nairobi with our approval as the university supervisors:

Mr. John Ndiritu

Signature: ____________________ Date: 30 / 07 /2012

Dr. Ivivi Mwaniki

Signature: ____________________ Date: 30 / 07 /2012

DEDICATION

This research is dedicated to my loving family; my husband; Ondieki; Children; Enock; Millicent; Hillary and Arnold for their support and encouragement during the entire period of the project.

ACKNOWLEDGEMENT

I owe my gratitude's to the Customs department and the Kenya Revenue Authority as a whole for providing and allowing me to use their data in the forecasting of the Import Duty Revenue. I will not forget to thank my supervisors John M. Ndiritu and Dr. Ivivi Mwaniki for their sincere and able guidance, direction and support during preparation and completion of this project. Lastly, to all those who made this project possible, thank you and God bless you all.

ABSTRACT

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. This study uses the monthly customs data, in particular the Import duty from January 2003 to December 2010 with the general objective of exploring the data and further establish 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 has an upward/positive trend, has a strong positive correlation over time and that an IMA (1, 1) model was established as a suitable model to forecast the tax.

TABLE OF CONTENTS

LIST OF FIGURES

TABLES

ACRONYMNS

ACF-Autocorrelation Coefficient Functions

ADF-Augmented Dickey Fuller

AIC-Akaike Information Criterion

AICC-Akaike Information Criterion Corrected

AR-Auto Regressive

ARI - Autoregressive Integrated

ARIMA-Auto-Regressive Integrated Moving Average

ARMA-Autoregressive Moving Average

ARMAX-Auto-Regressive Moving Average Exogenous

BIC-Bayesian Information Criterion

BIR-Bureau of internal Revenue

BOC-Bureau of Customs

CCCN-Customs Cooperative Council Nomenclature

CIF-Cost Insurance & Freight

EAC-East African Community

GDP-Gross Domestic Product

IMA-Integrated Moving Average

COMESA- Common Market for Eastern and Southern Africa

EAC-East African Community

GDP-Gross Domestic Product

IGAD-Inter-Governmental Authority on Development

IMA-Integrated Moving Average

IMF-International Monetary Fund

KIPPRA-Kenya Institute for Public Policy Reserve

KRA-Kenya Revenue Authority

KSHs-Kenya Shillings

LCL-Lower class Limit

MA-Moving Average

MAPE-Mean Absolute Percentage error

M-Million(s)

MoF-Ministry of Finance

OECD-Organization for Economic Cooperation and Development

OLS-Ordinary Least Squares

PACF-Partial Autocorrelation Functions

PIT-Personal Income Tax

UCL-Upper Class Limit

UK-United Kingdom

USA-United States of America

VAT-Value Added Tax

CHAPTER ONE

1.0 INTRODUCTION

1.1.0 Background information

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.

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.

Thiess B. et al (2009) established that revenue forecasting in some selected leading OECD countries e.g. USA, UK, Austria and Germany, which have well established forecasting institutions are prone to forecast errors. On average the countries forecast had an underestimation of 4.4%, and an overestimation of 5.2%. Furthermore, IMF study (Kyobe and Danninger, 2005) established that institutions relevant for revenue forecasting in most low-income countries are underdeveloped and thus score low on the quality of revenue forecasting practices if judged by basic standards. Only 12.9% of the low-income countries use econometric techniques. This is because such techniques require a wealth of reliable and relatively detailed data which most of the countries lack. Very few low-income countries have formal rules and regulations dictating the forecasting process, agencies responsibilities are not well defined. It is thus necessary for any government to forecast the revenue it collects using well-laid standard procedures for planning purposes.

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 paper will provide an analysis of the international trade taxes, in particular the import duty, which is collected by customs within the Kenya Revenue Authority (KRA). Exploratory data analysis will be performed on the import duty series data and a model be established which can possibly forecast the import duty tax.

1.1.1Trade Taxes

Trade Taxes are basically taxes on oil and non-oil imports. Non -oil trade taxes comprise import duty, excise duty, VAT on dry imports and the import declaration fees, while Oil trade taxes comprise import duty, excise duty and VAT on petroleum products imports. The importance of trade taxes to the overall total government tax collection has been declining due to trading blocks and liberalization. Nevertheless, trade taxes still account for about 35% of the total revenue in Kenya. This dominant role of trade taxes, as a major revenue source for the government can be attributed to the following factors: First, trade taxes are easy to collect as they are levied and administered at specific border points throughout the country. Second, trade taxes provide a useful economic tool for government in achieving economic goals. For example import substitution policies are potentially achieved when taxes are properly levied on imported goods. Third, trade taxes have an influence on the foreign exchange market in the economy and also contribute to savings and earnings in foreign currencies. Fourth, trade taxes are imposed as part of trade protection of the local industries by acting as anti-dumping or countervailing duties against unfair trade practices of foreign suppiier. The import declaration fee is entered before the import duty is paid for on any goods (imports).

The taxation of trade taxes has undergone several changes the main ones being:

- In 1988, the number of tariff categories was reduced from the wide-raging 25 categories to 17 categories. Eight of the tariffs were abolished namely 55%, 65%, 75%, 90%, 95%, 110%, 125%, and 170%.
- In 1989, the classification of traded goods and commodities was changed from Customs Cooperative Council Nomenclature (CCCN) to the Harmonized system. The Harmonized system allowed for a uniform classification of goods in international trade. Duty categories were further reduced from 17 to 12.
- In 1991, rationalization of the import duty rate structure continued by reducing the number of duty rates from 15 to 11. The highest duty rate of 100 was lowered to 70%. As part of the import liberalization policy aimed at lowering the cost of manufacturing and raising the competitiveness of local goods, duty rate on raw materials, intermediate goods and spare parts was further reduced by 5% point.
- In 1992, in keeping with the policy of replacing quantitative restrictions by tariffs, which was intended to lower high duty rates; top duty rate of 70% was reduced to 60% and 60% rates to 50%. Tariff bands were reduced from 11 to 9 by increasing 2% rate to 10%.
- In 1994, the number of tariff bands were reduced from 9 to 6 namely 0%, 10%, 20% 25%, 30% 40%.
- In 1995, the number of tariff bands was reduced from 6 to 5 namely 0%, 5%, 10%, 25% and 40%.
- In 1996, the top rate of 40% was lowered to 35% and 10% rate lowered and merged with 5% band. Some items were shifted to 15% band.
- In 1997, the tariff structure was rationalized down to three positive duty rates namely, 5%, 15% and 25%.

- In 1998, the programme of rationalizing duty rates was halted. The Pre-shipment companies were required to issue a clean report of finding on second hand vehicles imported into the country.
- In 1999, measures were put in place to protect domestic businesses e.g. raising duty on agricultural, livestock, horticultural products and imposing suspended duty on barley and malt.
- In 2000, duty on imports from COMESA countries was reduced to zero.
- In 2001, import duty rates on raw materials and inputs were reduced from 2.55 to 0% to curb dumping of raw materials from COMESA. To harmonize tariff structures within COMESA and EAC the top tariff band was lowered to 35%.
- In 2002, all duties on raw materials were abolished to lower the cost of production for manufacturers.

Currently, the import duty rate on dry goods are 0%, 10%, 25%, 35%, 40%, 45%, 50%, 60%, 75% and 100%. Excise duty rate are 0%, 5%, 7%, 10%, 20%, 50%, 54%, 65%, 120% and 130%. The standard VAT rate on imports is 16%.

Goods imported to Kenya are divided into different sectors which are: 1) Agriculture sector (1-24); 2) Chemical Sector (25-40); 3) Petroleum Sector (27)[1]; 4) Leather, Paper & Textile Sector - Articles of skin, leather, wood, paper , and paper products(41-49); 5) Articles of stone, glass, precious metals(68-71), Machinery (80-84) and steel sector(72-83); 6) Optical & Medical Apparatus, clocks and watches & misc. manufactured articles(90 - 97), Vehicles and transport Equipment(86-89).

1.1.2Import Duty

Import duty or tariff or custom duty is a tax imposed on imported merchandise that serves to raise the domestic price above the landed international price level by the margin of the tax. Import duty is a predominant source of revenue in Kenya, preferred because they are relatively stable and their collection is more administratively feasible than other taxes. The collection cost of import duty is about 5 to 11 per cent according to World Bank estimates of 2004, which are relatively low, compared to other taxes.

Economic, demographic and institutional changes affect the performance of the import duty. The ever-growing population in Kenya implies pressure for public expenditure and increased demand for imports. The rising use of e-commerce in business transactions has contributed to the increase of import duty due to demand for more technology. Recent trade liberalization and regional trade blocks such as COMESA, IGAD and EAC are expected to continue to lower the dependence of government revenues on import duty. The international trade sector can be important for the government, since trade tax revenue is directly related to the structure of the industry and size of the trading sector in the economy.

1.1.3Determinants of revenue forecasting

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.

However, in the case of the import duty tax, studies that have been conducted have used different macro variables as determinants of the import duty tax. The determinants used include; Foreign exchange earnings, Relative prices, Real income, Employment, Interest rates, Real exchange rate, Growth in the imports and services, Lagged value of import duties, Import Value etc. The variables significantly determine the behavior of total imports in the reference period.

1.1.4 Revenue forecasting

Governments need funds to finance the budget expenditures. Taxes are the major source of government revenues. If expenditures exceed revenues, governments resort to deficit financing through borrowing or raising taxes. Either case can have negative consequences on the economy over time. To maintain a balanced budget, governments can either curtail their expenditures and investments or increase revenues. Experience has shown that pruning of expenditures especially if they are recurrent is difficult to achieve as in an increase in taxes. There are several factors involved in the preparation of revenue forecasts of a tax system. All forecasts are done according to certain underlying methodology and common assumptions.

Generally, the revenue forecasting and tax analysis/ revenue estimation techniques may include:

(i) Qualitative Approach: - forecasts rely on expert judgment; some governments rely on few experts to be responsible for tax analysis and revenue forecasting.

(ii) Quantitative Approach: - these methods rely on numerical data, with explicit assumptions and procedures used to generate forecasts. The techniques commonly used are; 'Unconditional' Time Series Analysis, 'conditional'/Causal Models, microeconomic analysis, structural (National Account/Input-Output) Analysis, Micro-Simulation Analysis etc.

'Unconditional' Time Series Analysis includes trends and growth factor analysis. This approaches are said to be 'unconditional', because they base their forecast only on past revenue data. While, 'conditional'/Causal Models use information from other series (usually tax bases) to explain the behavior of tax revenues. (Andrew F.,1994).

In macroeconomic models, revenue forecasts depend on macroeconomic performance (real economic growth, inflation, interest rates etc), structure of the economy, structure of the tax system (tax base, rates, exemptions, incentives etc) and administrative capacity and taxpayer compliance. At the same time, the macroeconomic forecasts are also affected by fiscal policy (especially debt policy). Revenue forecasts and macroeconomic forecasts must be made on consistent policy. Macro economic forecasts are done by the Macro working group. Unrealistic estimation of macro variables could affect the quality of revenue forecast.

In revenue forecasting, sophistication does not necessarily mean more accurate. In fact, an experienced finance officer can often "guess" what is likely to happen with a great deal of accuracy. In general, forecasters use a variety of techniques, recognizing that some perform better than others depending on the nature of the revenue source (Thomas A, 2004).

Revenue forecasting has the following objectives:

(i) Assessing revenue impact of policy changes: This includes assessing revenue impact of changes in tax policies, non-tax policies and structural changes.

(ii) Appraising revenue impacts of economic changes such as GDP growth, inflation, changing trade patterns and so forth

(iii) Measuring tax capacity and tax effort: Measuring the tax revenue potential of a country and the degree to which it has been achieved.

(iv) Measuring performance of the revenue department.

The following are the desirable characteristics of revenue forecasts

(i) Accuracy: Revenue forecasts need to be as accurate as possible. The closer forecasts can come to actual revenue, the better spending decisions the Government can make.

(ii) Consistency with expenditure projections: If government agencies are aware beforehand of how much resources will be available to them, then planning will be better though out and therefore more accurate and credible.

1.2 Research Problem

An IMF study (Stephen et.al 2005), found out that data and human capital constraints in a large number of low-income countries led to the use of qualitative estimation techniques as the main methods for deriving budget revenue forecasts, since econometric techniques require a wealth of reliable and relatively detailed data.

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 uses revenue series data and its forecasts do not depend on forecasts of any other economic variable.

At the moment there is substantial data available on revenues, which can be used in forecasting. This research study will use unconditional models on the import duty 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 do 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 broad objective of this paper is to explore the monthly data for the import duty and try to establish a suitable mathematical equation to forecast the tax heads. Specifically, the study seeks to;

- To explore the monthly import duty data from January 2003 to December 2011
- To establish the trend and behavior of the import duty over the period under study.
- To find a suitable ARIMA model to forecast the import duty revenue

1.4 Purpose of the Study

The study endeavored to shed light on the trends and patterns of the Import Duty and also in establishing a suitable ARIMA forecasting model. The established model will assist in revenue forecasting of the import duty and thereby enhancing the planning and budgeting processes of the import duty. The model will also assist in the setting of targets.

1.5 Organization of the Paper

The rest of the paper will be organized as follows: Chapter two which is on the literature review looks into similar studies on the import duty forecasting models; chapter three outlines the methodology adopted in the study while chapter four presents the data analysis findings and conclusions and recommendations of the study.

CHAPTER TWO

2.0 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.

According to Joselito (2005) there are two ways in which revenue forecasting is normally practiced. First, the forecast may be calculated as an unconditional prediction of the most likely outcome. Secondly, a forecast 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 show that although not all of the countries rely 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.

2.1 Macroeconomic Models (Conditional Models)

Huizunga et al (2001) defines macro models as a representation of the quantitative relationships among macroeconomic variables such as employment, output, prices, government expenditure, taxes, interest rates etc. Macro models are useful in forecasting and in analysis of policy.

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. The collection of taxes, both VAT and excise, is a similar performance of applying a tax rate to the value of the commodity.

Using the ordinary least squares (OLS), Smal (1995) forecasted the total customs duties in South Africa. In the model, the tax base for the customs duties was specified as the value of merchandise imports, excluding oil imports. Customs duties are payable within one month of the arrival of the goods in South Africa and the tax base was consequently specified without a lag structure. A dummy variable was included in the customs duty component to be able, if necessary, to manipulate the effective tax rate because the rate is the estimated coefficient of the tax base.

Graham G, (2002) proposes three approaches to forecast customs duties. These are the imposition of tariff, the macro-base approach and the Micro-Simulation Approach. The imposition of a tariff approach takes cognizance of the fact that tariff revenue is a function of quantity imported, tariff rate, prices in domestic currency and the price elasticity for the goods. In this scenario it is assumed that the exchange rate has no variation when tariffs are imposed. For *ad valorem* excise duty the value of imports as opposed to quantity of imports is used.

The macro base approach is based on GDP growth rate and the change in the price of imports relative to the GDP deflator. Thus custom duties are a function of growth rate of import values/ quantities, elasticity of imports with respect to GDP, elasticity of tariff revenue with respect to import values. The key to applying the above is to estimate elasticity of imports with respect to GDP, elasticity of imports with respect to relative prices to GDP deflator, and elasticity of tariff revenue with respect to the import values which can then be estimated using regression analysis.

The micro simulation approach uses a trade tax calculator which is a micro-simulation model that uses large amount of actual raw data collected at the entry points such as the value of commodity I in a certain fiscal year, quantity of imports of commodity I, world prices of commodity I in a certain year, nominal exchange rate, real growth of import value of commodity I, percentage change in the import price index, tariff rates and the exemption rules among others.

Chun-Yan Kao et. al in import duties forecasting used a methodology which involved three step process. First, the elasticity of demand for imports is obtained by regressing real imports (i.e., imports in nominal dollars divided by the import price index) with real GDP and relative prices (i.e. the import price index divided by the GDP deflator) in log form. The second step is to estimate the elasticity of tariff revenues with respect to imports. Regressing, real tariff revenues (i.e., nominal import duties divided by the GDP deflator) with respect to real imports in log terms. Thirdly, the forecast for import tariff for the following year is calculated by; a multiplication of tariff revenue of the previous year, one plus percentage change in Imports and tariff revenue elasticity with respect to imports.

2.2 Unconditional Time Series Analysis

These approaches are said to be 'unconditional', because they base their forecasts only on past revenue data.

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.

Roshan .B (2000), estimated the import duties in Ghana by multiplying the effective tax rate with the Cost Insurance & Freight (CIF) import value. The effective rate of import duties is defined as the ratio of the total import duties to the total import value measured by the CIF value. The effective tax rate is measured by the summation of import duties collected across all imported goods to the total import value measured in CIF prices, including tariff exempt and zero- rated goods.

Rosham et.al also projected the total annual CIF import values. To obtain values of the upcoming year, they first estimate imports for each month of the coming year and then sum over the 12-month estimates to get the annual forecast. The monthly estimate is obtained by a simple extrapolation or from a simple regression analysis of using the same calendar month's import value against time. Thus once the effective tax rate and the total import value are known, their product can calculate the amount of import duties.

Thomas M. (1989) examined the effectiveness of composite forecasting of sales tax revenues in Idaho. An econometric model and a univariate time series model provide base line projections. The combined forecasts are found to outperform both base line forecasts. In the composite forecasting Ordinary least squares is used to estimate the parameters, except in cases where autocorrelation correction is necessary. For the latter case, an autoregressive moving average exogenous (ARMAX) nonlinear least squares correction technique is used. The structural parameters and the ARMA coefficients are jointly estimated.

2.3 Determinants of Revenue forecasting practices.

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.

Festus O. (1999) carried out a study to examine the determinants and major components of aggregate imports and its major components in the period 1953 to 1993 in Nigeria. The estimated equations rest on the stock adjustment import exchange model that has its roots in the balance of payments theory and in the consumer theory of demand as in the traditional import demand function. Quantitative estimates, based on integration and error correction specification, indicate that, foreign exchange earnings, relative prices and real income all significantly determine the behavior of total imports in the reference period. The results of the disaggregated imports also reveal the importance of foreign exchange. Thus, it is concluded if the Nigerian government wishes to increase imports, it is essential to implement economic policies that will enhance foreign exchange availability.

Abhijit S. G. (2007) examines the main determinants of Tax Revenue performance in Developing Countries. The results confirm that structural factors such as per capita GDP, agriculture share in GDP, trade openness and foreign aid significantly affect revenue performance of an economy. Other factors include corruption, political stability, share of direct and indirect taxes etc. The paper also makes use of a revenue performance index, and finds that while several Sub Saharan African countries are performing well above their potential, some Latin American economies fall short of their revenue potential.

CHAPTER THREE

3.0 METHODOLOGY

3.1 Introduction

This study utilizes the 'unconditional' method of forecasting where only the data of the import duty will be used to estimate revenue independently.

3.2 Data

The data used in this study, is the secondary data of the monthly Import Duty obtained from Custom Services Department, time- varying from January 2003 to December 2010. Thus, as this data is obtained from observations collected sequentially (monthly) over time can be grouped as time series data.

3.3.0 Statistical software

To analyze the data the study will use excel and R statistical software. The software's are chosen because of their wide application and acceptance in the field of Statistics and Econometrics. SPSS will also be used.

3.3.1 Introduction to time series

A graph of the import duty will be drawn. This will give the trend for the period mentioned. I'll then give a description of the results based on the occurrences in Kenya in the period mentioned. 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 there's no systematic change in the mean i.e. no trend or there's no systematic change in the Variance. A time series is strictly stationary if the joint distribution of $Y_{t_1}, Y_{t_2}, Y_{t_3}, \cdots, Y_{t_n}$ is the same as the joint distribution of $Y_{t_1+h}, Y_{t_2+h}, Y_{t_3+h}, \cdots, Y_{t_n+h}$ and h being the distance between the observations. In other words shifting the time origin by an amount h has no effect on the joint distribution, which must only depend on the interval between $t_1, t_2, ..., t_n$. That means that for instance if h=1, the distribution of X_t must be the same for all values of t i.e. the mean $\mu_t = \mu$ and $\sigma_t^2 = \sigma_t$ both constants don't depend on t.

To test for stationarity we will use the Autocorrelation coefficient functions (ACF) and Partial Autocorrelation functions (PACF) graphs. If the ACF is decaying exponentially, it implies that the data series is not stationary. However this will be formally tested by using the Augmented Dickey-Fuller (ADF) test for stationarity . ADF tests do determine whether a time series is stationary or, specifically, whether the null hypothesis of a unit root can be rejected.

Test statistic is basically a t-statistic

$$t^* = \frac{\hat{\gamma}}{s.e(\hat{\gamma})} \tag{3.1}$$

Hypothesis tested is

H_0: $\theta = 0$ (the data is not stationary (unit root))

H_1: $\theta \neq 0$ (the data is stationary (no unit root))

Decision rule:

If t^* > ADF critical value, we do not reject null hypothesis, i.e., unit root exists.

If t^* < ADF critical value, we reject null hypothesis, i.e., unit root does not exist.

This will call for differencing if it is found that $P - \text{value} > \alpha = 0.05$ then we conclude H_0 that the data is not stationary.

- Differencing non-stationary series: Box-jenkins recommends differencing non-stationarity series one or more times to achieve stationarity. Doing so produces an ARIMA model, with "I" standing for "integrated". But its first difference

$$y_t = y_t + y_{t-1} = \mu_t \qquad (3.2)$$

is stationary, so y is "integrated of order 1"

- Logarithm transformation is also a useful method in certain circumstances. We frequently encounter series where increased dispersion seems to be associated with higher levels of the series – the higher the level of the series, the more variation there is around that level and conversely. These transformations will make the data stationary.

There are several models that are employed in time series analysis, namely: Auto Regressive (AR), Moving Average (MA), Auto Regressive Moving Average (ARMA), and Auto Regressive Integrated Moving Average (ARIMA) among others. We'll discuss them briefly in the following subsections.

3.3.2 Autoregressive model (AR) model

The notation AR (p) refers to the autoregressive model of order p. The AR (p) model is written

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2}, \dots \alpha_p X_{t-p} + \varepsilon_t \qquad (3.3)$$

Where ε_t~N (0, σ^2) and i.i.d., α_i are the parameters of the model

Some constraints are necessary on the values of the parameters of this model in order that the model remains stationary. For example, processes in the AR (1) model with $|\varphi_1| \geq 1$ are not stationary.

The simplest of the AR process is when p=1 i.e. AR (1)

The properties of AR (1) process:

$$E(Y_t)=0 \qquad (3.4)$$

$$Var(Y_t)=\frac{\sigma^2}{1-\theta_1^2} \quad if \; abs(\theta_1)<1 \qquad (3.5)$$

$$Corr(Y_t,Y_{t-k})=\rho_k \quad if \quad abs(\theta_1)<1, k>0 \qquad (3.6)$$

3.3.3 Moving Average (MA) model

The notation MA (q) refers to the moving average model of order q:

$$Y_t=\beta_0\varepsilon_t+\beta_1\varepsilon_{t-1}+..+\beta_q\varepsilon_{t-q}+\varepsilon_t \qquad (3.7)$$

Where β_j's are the parameters of the model, μ is the expectation of Y_t (often assumed to equal 0), ε_t are white noise error terms and $Var(Y_t)=\sigma^2\sum_j\beta_j^2$

3.3.4 Autoregressive Moving Average model

The notation ARMA (p,q) refers to the model with p autoregressive terms and q moving average terms. This model contains the $AR(p)$ and $MA(q)$ models,

$$Y_t=\alpha_1Y_{t-1}+\alpha_2Y_{t-2}+\cdots+\alpha_pY_{t-p}+\beta_1\varepsilon_{t-1}+\beta_2\varepsilon_{t-2}+\cdots+\beta_q\varepsilon_{t-q}+\varepsilon_t \qquad (3.8)$$

Where $\varepsilon_t \sim N(0,1)$, $p,q\geq 0$ are integers and p,q are the order of the model

3.3.5 Auto-Regressive Integrated Moving Average (ARIMA)

ARIMA models are, in theory, the most general class of models for forecasting a time series, which can be made stationary by transformations such as differencing and logging. In fact, the easiest way to think of ARIMA models is as fine-tuned versions of random-walk and random-trend models: the fine-tuning consists of adding lags of the differenced series and/or lags of the forecast errors to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors.

Lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of non-seasonal differences, and
- **q** is the number of lagged forecast errors in the prediction equation.

For an $ARIMA(p,d,q)$

is a non-stationary ARMA (p,q). $\Delta^d Y_t = W_t$

$$W_t = \phi_1(W_{t-1}) + \phi_2(W_{t-2}) + \cdots + \phi_p(W_{t-p}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_q e_{t-q} \qquad (3.9)$$

For $d = 1$

Or in other terms,

$$Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \cdots + \phi_p(Y_{t-p} - Y_{t-p-1}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \qquad (3.10)$$

If the process contains no autoregressive terms, we call it an integrated moving average and abbreviate the name to IMA (d,q), for example the simple IMA (1,1) model which satisfactorily represents numerous time series, especially those arising in economics and business. If no moving average terms are present, we denote the model as ARI (p,d).

The IMA (1, 1) Model

The simple IMA (1,1) model satisfactorily represents numerous time series, especially those arising in economics and business. In difference equation form, the model is

$$y_t = y_{t-1} + e_t - \theta e_{t-1} \qquad (3.11)$$

To write y_t as a function of present and past noise values we have;

$$y_t = e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + \dots + (1-\theta)e_{-m} - \theta e_{-m-1} \quad (3.12)$$

Thus in contrast to stationary ARMA models, the weights on the white noise terms do not die out as we go into the past since we are assuming that $-m \angle 1$ and $0 \angle 1$ we may think of y_t as mostly an equally weighted accumulation of a large number of white noise values.

For IMA (1,1) model with constant terms we have

$$y_t = e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + \dots + (1-\theta)e_{-m} - \theta e_{-m-1} + (t+m+1)\theta_0 \quad (3.13)$$

This shows an added linear deterministic time trend $(t+m+1)\theta_0$ with slope θ_0. An equivalent representation of the process would then be

$$y_t = y_t' + \beta_0 + \beta_1 t \quad (3.14)$$

Where, y_t' is an IMA (1, 1) series with $E(\Delta y_t') = 0$ and $E(\Delta y_t) = \beta_1$.

To identify the appropriate ARIMA model for a time series, we begin by identifying the order(s) of differencing needing to stationarize the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating. If you stop at this point and predict that the differenced series is constant, you have merely fitted a random walk or random trend model. However, the best random walk or random trend model may still have auto-correlated errors, suggesting that additional factors of some kind are needed in the prediction equation.

3.4 Model building for the import duty

In exploring the import duty series data, several time series plots will be plotted using the R software, to determine their trends and patterns. A description of the results based on the occurrences from the series data will be given. To model using time series data, the series has to be stationary to give robust results and predictions.

We will develop a multi-step model-building strategy espoused so well by Box and Jenkins (1976). There are three main steps in the process, that is; Model specification (or identification), Model fitting, and Model diagnostics.

To determine the order of the import duty, I'll apply the steps below;

Step 1: Examination of time-series plot

The first step is to produce a time-series plot of the import duty over the period January 2003 and December 2010 and examine the plot to identify the present components of the time series i.e. trend component, seasonal components, and outliers. Tests will be carried out in this stage to determine if the series is stationary using the Autocorrelation function (ACF) and the Augmented Dickey Fuller (ADF) test. The components observed will be removed through differencing.

Step 2: Examination of correlogram

Trend and seasonal components may show up in a correlogram. A slowly damping correlogram is indicative of a slowly varying trend component. A periodic fluctuating correlogram is indicative of a periodic component. Taking the difference at appropriate time lags may remove those non-stationary components.

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

The data is AR (p) if: ACF will decline steadily, or follow a damped cycle and PACF will cut off suddenly after **p** lags.

The data is MA (q) if; ACF will cut off suddenly after q lags and PACF will decline steadily, or follow a damped cycle.

Table 3.0: characteristics of ACF and PACF

Behavior of theoretical ACF and PACF for stationary process		
Model	***ACF***	***PACF***
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR(p)	Exponential decay and/or damped sinusoid	Cuts off after lag p
ARIMA(p,q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

Step 4: Determining the orders using AIC, AICC or BIC. Each method has its own merit. A practically relevant question is when to use what, although a general answer to this question is inconceivable. The choice should depend on the nature and the aim of the data analysis. Empirical experience suggests that AIC is a good starting point. AIC is defined as

$$AIC = -2 * \log - likelihood + 2p \tag{3.15}$$

An alternative widely used criterion is the Bayesian Information Criterion (BIC) which essentially replaces the term $2p$ in the AIC with the expression $p + p \ln N$.

Hence BIC is

$$BIC = -2 * \log - likelihood + p(1 + \ln N) \tag{3.16}$$

and the corrected AIC i.e. AICC

$$AICC = AIC + \frac{2(k+1)(k+2)}{n-k-2} \tag{3.17}$$

Where n is the sample is size and k is the total number of parameters excluding the noise variance.

Step 5. Model estimation: the main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation. Maximum likelihood estimation is generally the preferred technique.

Step 6. Model diagnostics; model diagnostics for Box-Jenkins models is similar to model validation for non-linear least squares fitting. That is, the error term $\mathbf{u}_t$ is assumed to follow the assumptions for a stationary unvaried process. The residuals should be white noise (or independent when their distributions are normal) drawings from a fixed distribution with a constant mean and variance.

3.5 Other tests to be conducted

Testing for Normality of the data

Some of the stylized facts are determined by measuring the third and fourth moment's i.e. skewness and kurtosis respectively.

- Skewness is the ratio of the third order moment and is defined as

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

❖ For normality assumption, $S = 0$, and kurtosis is a ratio of the fourth order moment, which is assumed to exist to the squared second-order moment and is represented as

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

And under normality assumption, $k = 0$

❖ QQ plots: It tests for normality and the randomness of the data, we use the QQ plots are used. Normality can be checked more carefully by plotting the so-called normal scores or QQ plot. Such a plot displays the quartiles of the data versus the theoretical quartiles of a normal distribution. With normally distributed data, the QQ plot looks approximately like a straight line and hence the values are expected to fall along or close to the line otherwise the data isn't.

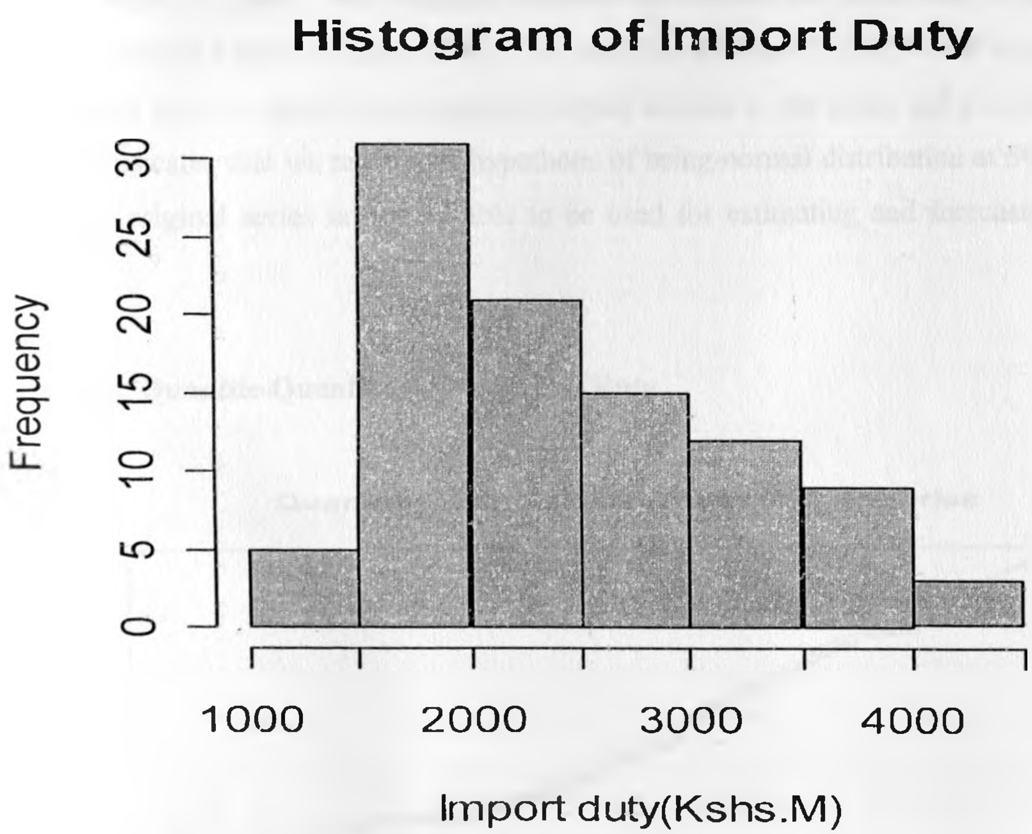
❖ Shapiro-Wilk'stest:It essentially calculates the correlation between the residuals and the corresponding normal quartiles. The lower this correlation, the more evidence we have against normality alternatively, if $p-value > \alpha$ we reject the null hypothesis.

In the next chapter I'll analyze the data based on the methodology.

The import duty series is shown in the figure above; the import duty collection fluctuates in the range of about 1300M to 4500M. The fluctuations might be due to the import tariffs, which change from financial year to another. The upward trend might be due to various reforms going on in the department to better the revenue collections. The series has positive growth overtime.

The import duty series does not show a very regular pattern thus seasonal variations might be absent, though from 2007 each financial year start with low values and end up with higher values in the that particular year.

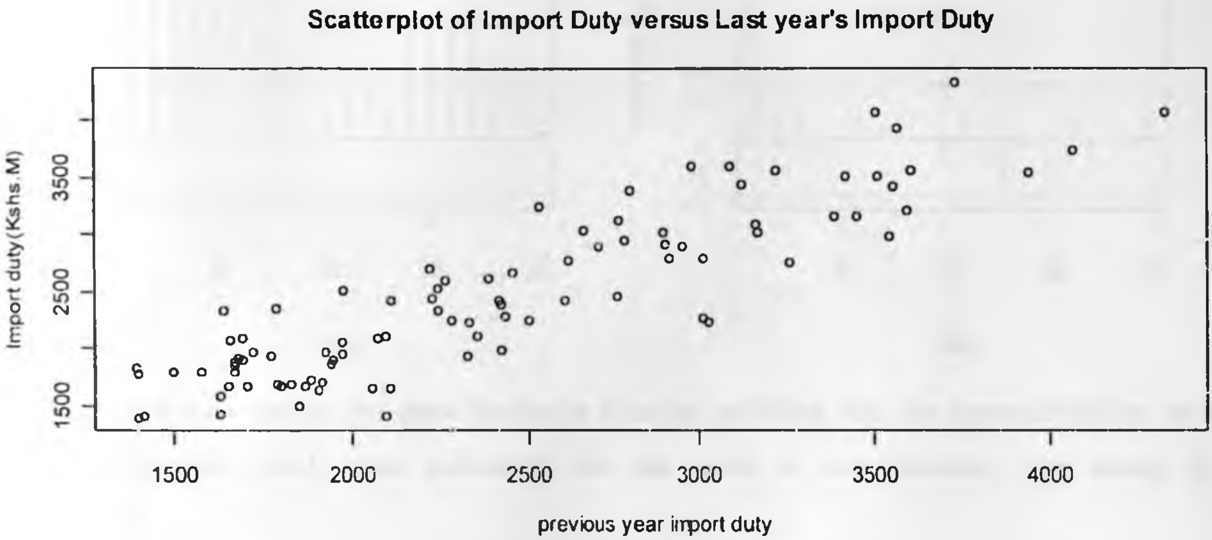
Figure 4.2: Histogram of the Import Duty



From the histogram above, it can be observed that a great number of observations are located around kshs. 2000M.

The plot above displays the normal quantile –quantile plot for the import duty series. There is a considerable curvature in the plot. A line passing through the first and third normal quartiles helps point out the departure from a straight line in the plot.

Figure 4.4: A Scatterplot of the Import Duty

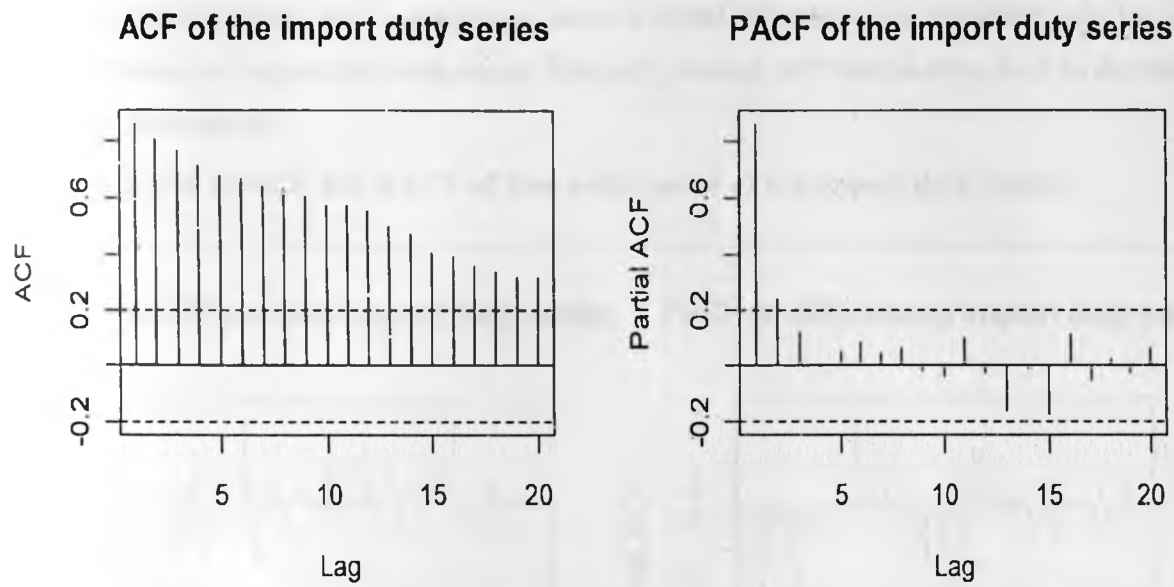


The scatterplot above shows some linear trends, there is a correlation between the previous year's import duty collection and the current year's collection.

4.3 Testing for non-Stationarity in the series

ARIMA models require a stationary working series prior to estimation and diagnostic checking. As can be seen in figure 1, the import duty data has a positive nonlinear trend in the mean of the series. The unit roots test will help us determine the stationarity of a series. Thus the stationarity of the monthly import duty data is checked by autocorrelation function (ACF) and the augmented dickey-fuller test (ADF) test.

Figure 4.5: plot of the ACF and PACF of the original data



In the import duty series, the autocorrelation function indicates that the autocorrelations start high and decline slowly thus indicating that the series is non-stationary and should be differenced.

Table 4.2: ADF test for the original import duty

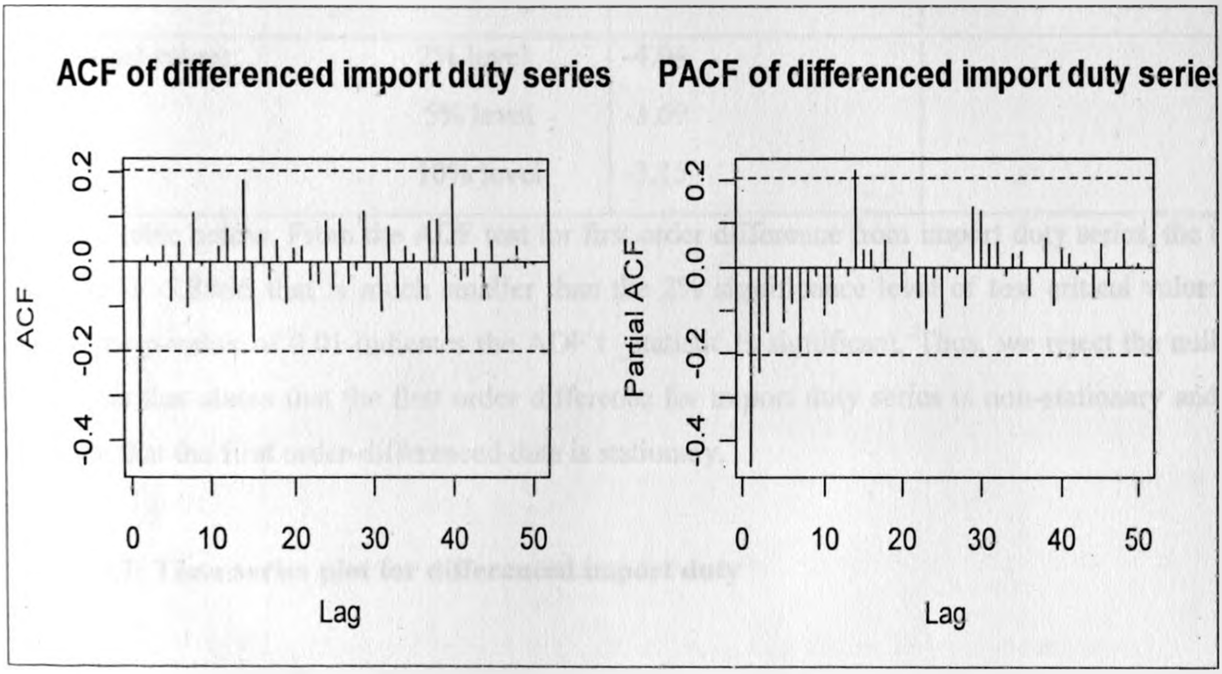
Null Hypothesis: import duty revenue data is non stationary Alternative Hypothesis: The data is stationary			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.3196	0.4444
Test critical values:	2% level	-4.04	
	5% level	-3.69	
	10% level	-3.15	

Likewise, the ADF test for import duty revenue series indicates that the ADF test statistic is −2.3196 which is greater than test critical values of −4.04, −3.69 and −3.15 at 2%, 5% and 10%

significance levels. The P- value of 0.444 strongly disagrees that the series is stationary. Thus, we do not reject the null hypothesis of being non-stationary.

The process is continued until a stationary series is found. We check the stationarity for the first order difference of import duty data series. Similarly, we use ADF test and the ACF to determine the series' stationarity.

Figure 4.6: plot of ACF and PACF of first order series of the import duty series.



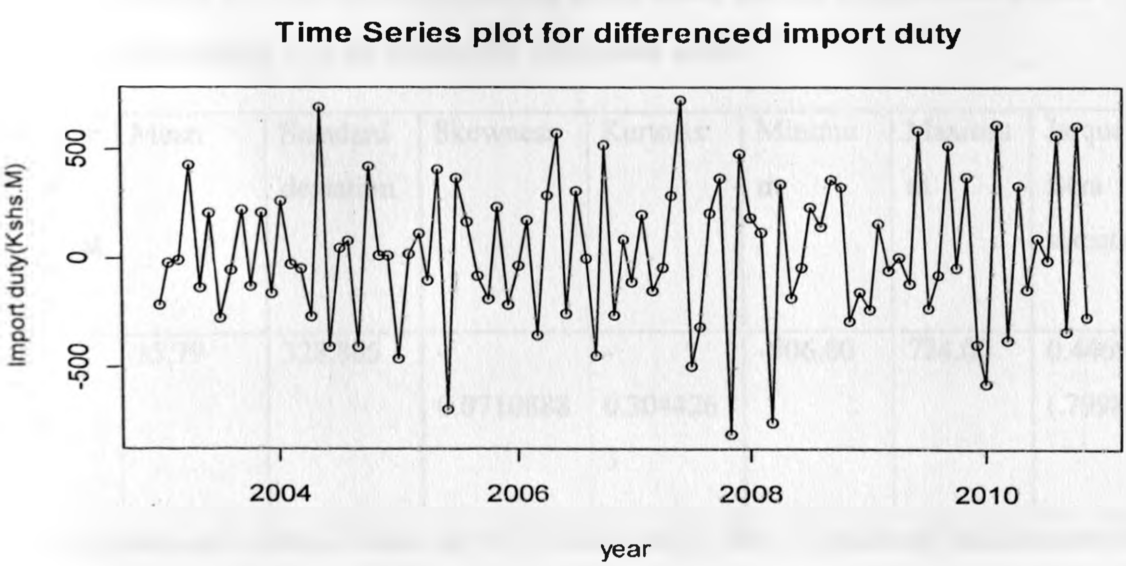
From figure 6 above, the autocorrelations are distributed over the period and hence suggests stationarity

Table 4.3: ADF test for first order import duty revenue

Null Hypothesis: import duty revenue data is non stationary Alternative Hypothesis: The data is stationary			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.8866	0.01
Test critical values:	2% level	-4.04	
	5% level	-3.69	
	10% level	-3.15	

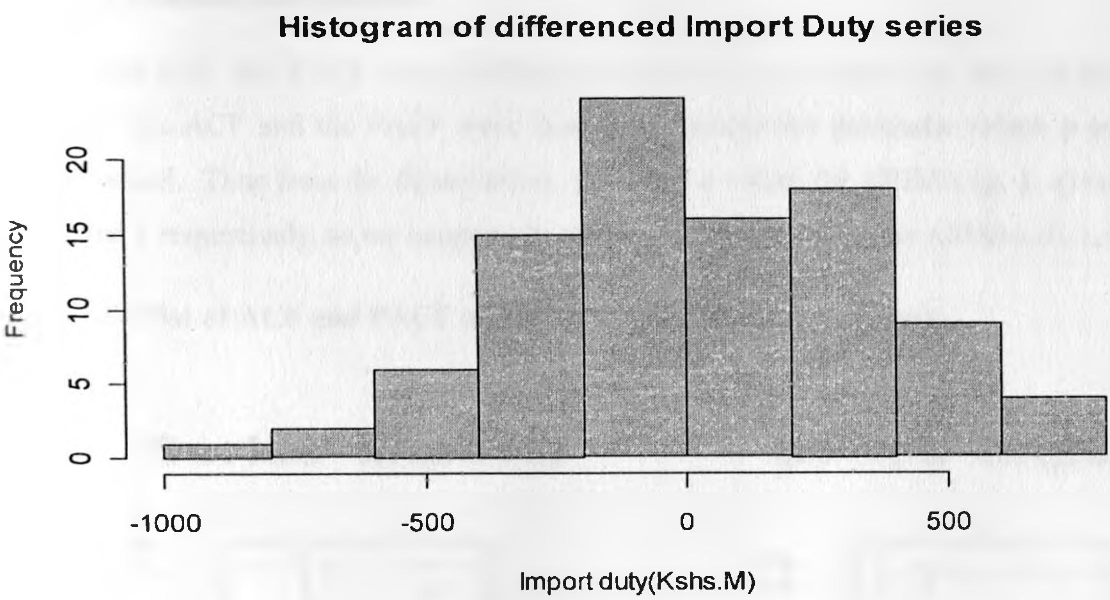
From the table below, From the ADF test for first order difference from import duty series, the t _statistic is -5.8866 that is much smaller than the 2% significance level of test critical value. Hence, the p-value of 0.01 indicates the ADF t _statistic is significant. Thus, we reject the null hypothesis that states that the first order difference for import duty series is non-stationary and conclude that the first order-differenced data is stationary.

Figure 4.7: Time series plot for differenced import duty



First lagged difference from the original data time series, which is a stationary series is shown in figure above. From the figure it can be seen that the difference with respect to the first order for import duty series is stationary because most of the import duty values are located around mean of zero. However, there are some spikes in the figure, representing high volatility periods.

Figure 4.8: Histogram of differenced import duty series



The histogram of the differenced import duty series shows that the series is near normal

Table 4.4: Normality test for first order difference series

Number of observations	Mean	Standard deviation	Skewness	Kurtosis	Minimum	Maximum	Jarque-Bera statistic
96	25.79	328.805	-0.0710888	-0.3044263	-806.00	724.00	0.4469 (.7998)

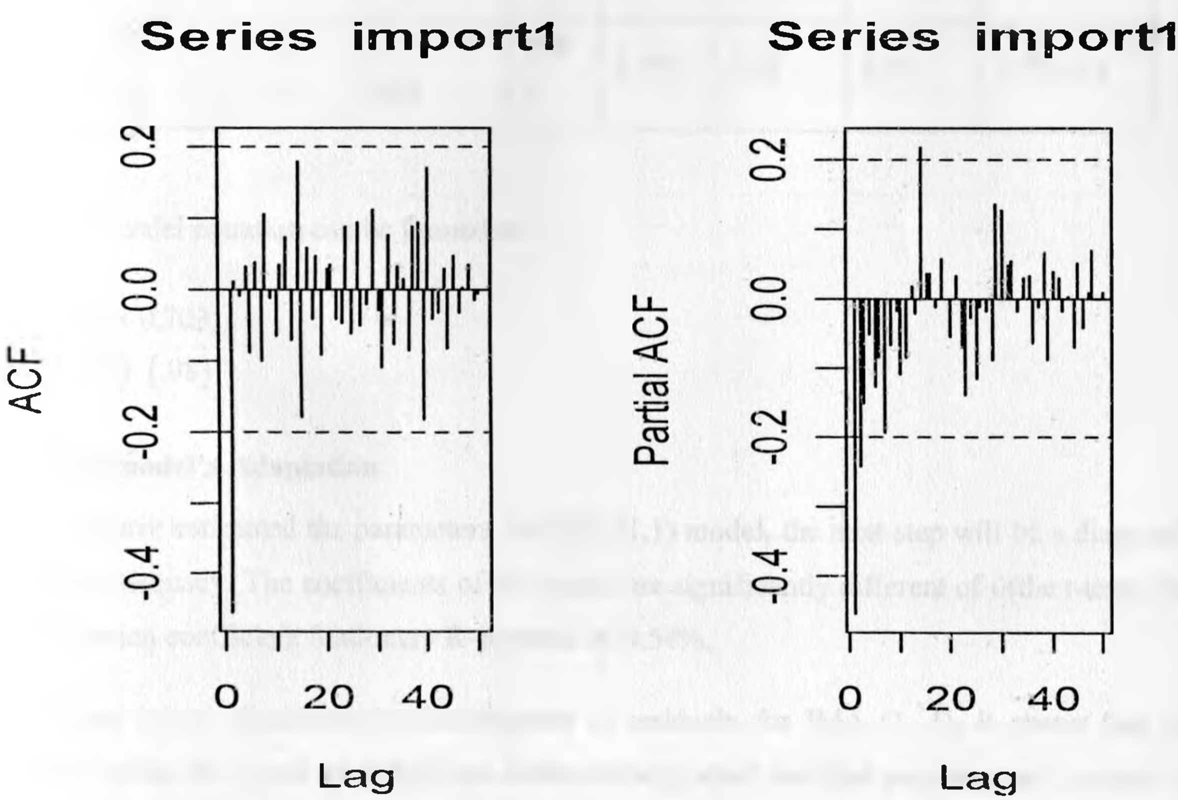
The skewness and kurtosis values are -0.0710888 and 0-.3044 respectively which shows that the distribution is slightly asymmetric and highly platykurtic.

With the information obtained above, we have the stationary series after one lagged difference from the original import duty series. In the next step, we will use the first order difference for import duty series to find our models using Box-Jenkins.

4.4 ARIMA Model Identification

A plot of the ACF and PACF for the differenced values of the import duty data was done. The patterns of the ACF and the PACF were done to determine the parameter values p and q for ARIMA model. Thus from the figure below, the p and q values for ARIMA (p, 1, q) model are set at 0 and 1 respectively, so we temporarily set our ARIMA model to be ARIMA (0, 1, 1).

Figure 4.9: Plot of ACF and PACF of differenced series of Import Duty



Since the process depicted does not contain the autoregressive terms, we call it an integrated moving average and abbreviate the name to IMA (d, q), thus IMA (1, 1).

4.4.1 Parameter estimation IMA (1, 1) Model

Parameters estimation of the model is as tabulated in the table below;

Table 4.5: Estimation equation of IMA (1, 1) Model

ARIMA Model Parameters							
				Estimate	SE	t	Sig.
import duty-Model_1	import duty	Constant		26.32	8.76	3.00	0.00343
		Difference		1			
		MA	Lag 1	0.703	0.08	9.00	2.72E-14

Thus, the model equation can be formed as

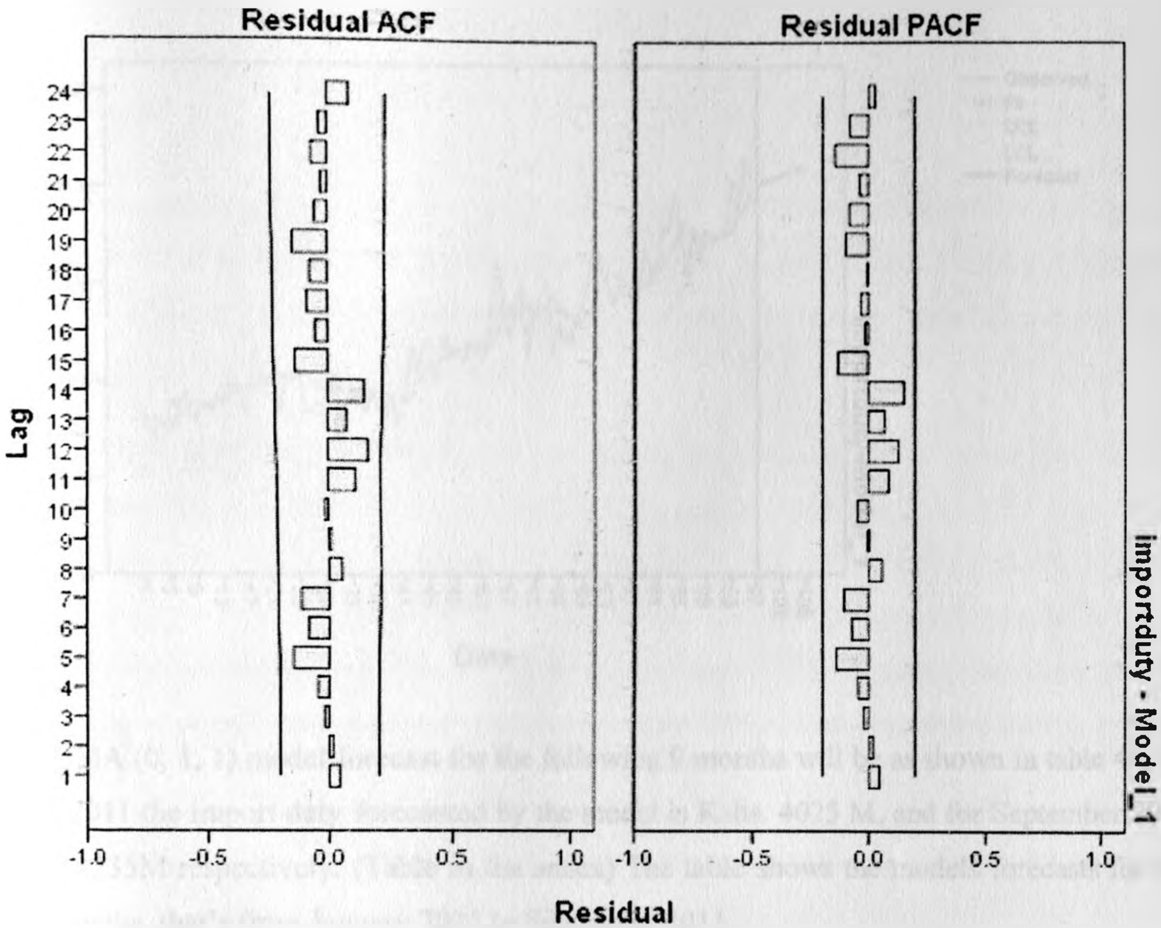
$$\hat{Y}_t = 26.32 + 0.703_{t-1}$$
$$(8.76) \quad (.08)$$

4.4.2 The model's Adaptation

After we have estimated the parameters for IMA (1,1) model, the next step will be a diagnostic of model adequacy. The coefficients of the model are significantly different of 0(the t-test). The determination coefficient Stationary R-squared is 29.54%.

The figure below illustrates the correlogram of residuals for IMA (1, 1), it shows that the residuals of the ACF and the PACF are both relatively small and that any term isn't exterior to the confidence intervals and the Q-statistic has a critical probability near to 1 (appendix 1: table4).

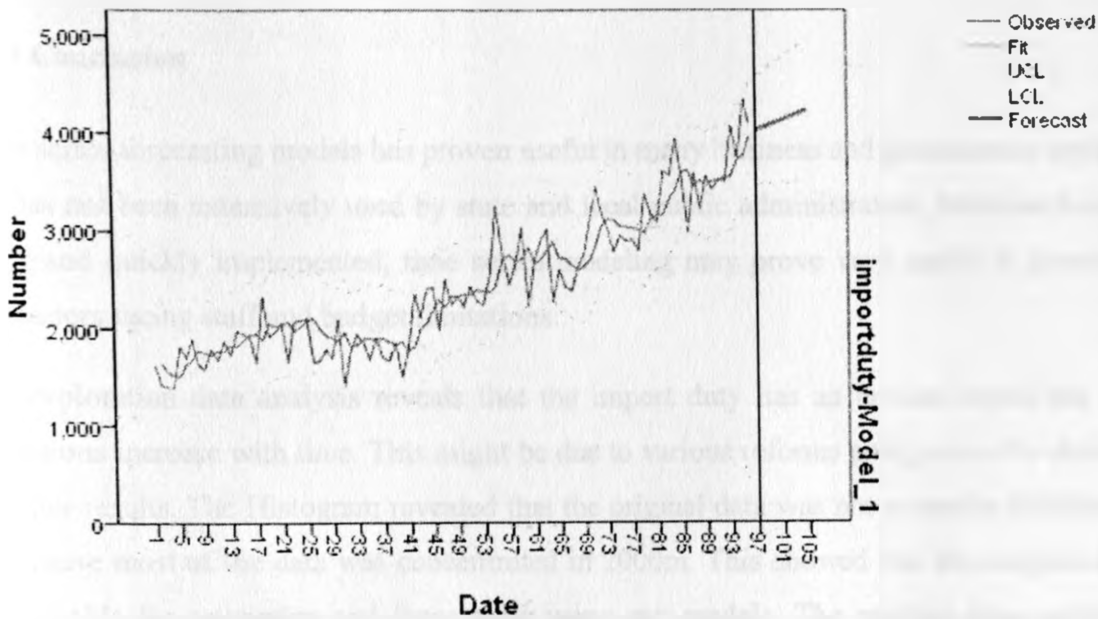
Figure 4.10: Correlogram of the residuals



4.4.3 The forecasting

In this step the forecast of the import duty using IMA (1, 1) Model is done. The figure below displays the import duty series together with the forecasts with the upper and lower 95% prediction limits for those forecast.

Figure 4.11: Forecasts from ARIMA (0, 1, 1)



The ARIMA (0, 1, 1) model forecast for the following 9 months will be as shown in table 4.6. In January 2011 the import duty forecasted by the model is Kshs. 4025 M, and for September 2011 is Kshs. 4235M respectively. (Table in the annex) The table shows the models forecasts for the next 9 months, that's from January 2011 to September 2011.

Table 4.6: Forecast of January 2011 to September 2011

Forecast										
Model		97	98	99	100	101	102	103	104	105
import duty-Model_1	Forecast	4025	4051	4077	4104	4130	4156	4183	4209	4235
	UCL	4575	4625	4675	4724	4771	4819	4865	4911	4957
	LCL	3475	3477	3480	3484	3489	3494	3500	3507	3514

CHAPTER FIVE

5.0 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

Time series forecasting models has proven useful in many business and government applications, but has not been extensively used by state and local public administrators. Because it is easy to apply and quickly implemented, time series modeling may prove very useful to governments' forecasters facing staff and budget limitations.

The exploration data analysis reveals that the import duty has an upward trend that is, their collections increase with time. This might be due to various reforms going on in the departments to better results. The Histogram revealed that the original data was not normally distributed; this is because most of the data was concentrated in 2000m. This showed that the original data was not suitable for estimating and forecasting using any models. The analysis also indicates that there is a strong positive correlation over time indicating that a linear regression model might be appropriate.

ARIMA models require a stationary working series prior to estimation and diagnostic checking. Since the original data was not stationary, it had to be differenced. A plot of ACF and PACF for differenced values of the import duty data was done. The p and q values for ARIMA (p,1,q) Model was set as 0 and 1 respectively. The results reported here indicate that an IMA (1, 1) model projection of the import duty provides good forecasts.

This model provides a baseline to measure other models and hence further research should be carried out in this area with a more complex model to determine the correct forecast with a greater degree of accuracy.

5.2 Recommendations

The ARIMA models are useful in revenue forecasting and should be fully utilized since they are easy to apply. Good practice of data keeping should be applied to enable such analysis to be done.

6.0 REFERENCES

1. Andrew F. Siegel, (1994), The Students Companion for use with Practical Business Statistics2nd edition

2. Annette Kyobe and Stephan Danninger (2005), Revenue Forecasting- How is it done? Results from a survey of Low-Income Countries, *IMF Working Paper (WP/05/24)*

3. Festus O. Egwaikhide (1999), Determinants of imports in Nigeria: A dynamic specification, *African Economic Research Consortium Research Paper 91*
4. Glenn P. Jenkins, Chun-Yan Kuo ,Gangadhar P. Shukla, (2000) Tax Analysis and Revenue Forecasting -- Issues and Techniques ,*Harvard Institute for International Development Harvard University*

5. Graham Glenday, Gangadhar P. Shukla, RabinoSagana (2001), Revenue Forecasting andEstimation, Techniques and Applications, *DCID, Duke University, Durham*
6. Huizinga F, et al (2001). *The Theoretical Base for Kenya Macro – Model*: The KIPPRATreasury Macro Model. KIPPRA, Nairobi.
7. Jonathan D. Cryer, Kung-Sik Chan, (2008) *Time Series Analysis with Applications in R*, 2nd Ed.
8. Joselito R. A. (2005). Forecasting Revenues from the BIR and BOC: Towards the Improvement of Cash Programming in the Philippines,*Department of Budget and Management, Republic of the Philippines*
9. NIST, *Exploratory Data Analysis.*http://www.itl.nist.gov/div898/handbook/eda/eda.htm.
10. RoshanBajrachaya and chun-Yan Kuo (2000) Tax Revenue Estimation in Ghana;*InternationalInstitute for Advanced Studies inc.*
11. Smal, M, M. 1995: The framework of the Public Sector Equations in the South Africa Reserve Bank's Econometric Model,*Occasional Paper No. 7.*
12. Stephen few,(2007) Visualizing Change: An Innovation in Time-series Analysis. *Perceptual Edge.*
13. Thiess Buettner, Bjoern K., (2009) Revenue Forecasting Practices: Differences across countries and Consequences for Forecasting Performance; CESIFO working paper No. 2628 category 1: Public Finance.

14. Thomas A. Garrett, John C. Leatherman, (2004), An Introduction to State and Local Public Finance; 5th Edition, Part 2 - Selected Applications in Public Finance, section (iv),*Revenue Forecasting.*

15. Thomas M. Fullerton Jr.,(1989), A composite approach to forecasting state government revenues: Case study of the Idaho sales tax,*International Journal of Forecasting 5.*

Table 7.3: Model statistics

Model	Model Fit statistics					
	Stationary R-Squared	RMSE	MAPE	MaxAPE	MaxAE	Normalized BIC
import duty-Model_1	0.295417	277.4769	8.993971	40.07139	808.7711	11.34735

Ljung-Box Q(18)		
Statistics	DF	Sig.
15.64817	17	0.548899

Table 7.4: The Correlogram of Residuals

	AC	PAC	Q-stat	Prob
1	0.0492	0.049235	0.6292	
2	0.0236	0.021268	0.6297	
3	-0.0196	-0.02184	0.7089	
4	-0.0510	-0.04969	0.7214	
5	-0.1479	-0.14291	0.7552	0.385
6	-0.0823	-0.06933	0.7587	0.685
7	-0.1066	-0.0995	0.7594	0.859
8	0.0560	0.059653	0.7599	0.944
9	0.0136	-0.00333	0.7974	0.957
10	-0.0116	-0.04613	0.8384	0.991
11	0.1067	0.083426	0.8655	0.997
12	0.1563	0.128196	1.7869	0.987
13	0.0707	0.065995	1.787	0.994
14	0.1448	0.149813	1.8626	0.997
15	-0.1274	-0.12561	1.8896	0.999
16	-0.0471	-0.01159	1.9128	1.0000
17	-0.0904	-0.03773	1.9875	1.0000
18	-0.0679	-0.00344	2.0483	1.0000
19	-0.1398	-0.09828	2.1214	1.0000
20	-0.0448	-0.07908	2.1935	1.0000
21	-0.0187	-0.03213	2.3163	1.0000
22	-0.0662	-0.14801	2.3449	1.0000
23	-0.0337	-0.07981	2.3749	1.0000
24	0.0910	0.034	2.4072	1.0000

7.2 Appendix 2: Programme codes

Using R.

```
##TIME SERIES ANALYSIS##
import=read.csv("importduty.csv", header=T)
import
attach(import)

#Libraries required for this analysis
library(fracdiff),library(tseries),library(moments),library(forecast),library(stats)
library(fSeries),library(graphics),library(MASS),library(FinTS),library(fArma)
library(TSA),library(fEcofin),library(quadprog)

## Month,Import.duty
##Explolatory data analysis, of original data
import2=ts(import,frequency=12,start=2003)
plot(import2[,-1],main="Time Series plot for import duty"
,type="o",xlab="year", ylab="Import duty(Kshs.M)")
win.graph(width=4.875, height=3,pointsize=8)
hist(import[,2],xlab="Import duty(Kshs.M)",main="Histogram of Import Duty",col="5")
qqnorm(import[,2],main="Quantile-Quantile plot of Import Duty series");qqline(import[,2])
shapiro.test(import[,2])
summary(import[,2])
kurtosis(import[,2])
skewness(import[,2])
sd(import[,2])
jarque.bera.test(import[,2])
win.graph(width=3, height=3,pointsize=8)
plot(y=(import[,2]),x=zlag(import[,2]),xlab="previous year import duty", ylab="Import
duty(Kshs.M)"
,main="Scatterplot of Import Duty versus Last year's Import Duty")
```

```
###checking for stationarity of imports
acf(import[,2],20,main="ACF of the import duty series")
pacf(import[,2],20,main="PACF of the import duty series")
##Augumented Dickey-Fuller test
adf.test(import[,2])
##first order difference for import duty
import1=diff(import[,2])
par(mfrow=c(1,2))
acf(import1,50,main="ACF of differenced import duty series")
pacf(import1,50,main="PACF of differenced import duty series")
plot(import1, type="l", xlab="time", ylab="Differenced import duty"
, main="Time Series plot of differenced import duty")
hist(import1,xlab="Import duty(Kshs.M)"
,main="Histogram of differenced Import Duty series",col="5")
summary(import1)
kurtosis(import1)
skewness(import1)
sd(import1)
adf.test(import1)
jarque.bera.test(import1)
Box.test(import1, type="Ljung")
###Model parameter estimation by differenced data
###and ARIMA model selection
arima=auto.arima(import[,2])
###forecasting using ARIMA model
fcast=forecast(arima2,10)
plot(fcast,ylab="Import duty",type="o",xlab="time")
##model diagnistics
tsdiag(arima2)
#######evaluation of the forecast##########
```

```
summary(fcast)
#diagnistic tests for residuals
auto0=arima(import[,2])
auto0
summary(auto0)
tsdiag(auto0,400)
plot(rstandard(auto0), main="Standardised residuals", type="l")
acf(auto0$residuals,20, main="ACF Standardised Residuals")
pacf(auto0$residuals,20, main="PACF Standardised Residuals")
```

% Model creation in spss

```
GET
 FILE='C:\Documents and Settings\Happiness\Desktop\late-7\import.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
PREDICT THRU 105.
* Time Series Modeler.
TSMODEL
  /MODELSUMMARY  PRINT=[MODELFIT RESIDACF RESIDPACF] PLOT=[ SRSQUAR
E RSQUARE]
  /MODELSTATISTICS  DISPLAY=YES MODELFIT=[ SRSQUARE RMSE MAPE MAE M
AXAPE MAXAE NORMBIC]
  /MODELDETAILS  PRINT=[ PARAMETERS FORECASTS]
  /SERIESPLOT OBSERVED FORECAST FORECASTCI FITCI
  /OUTPUTFILTER DISPLAY=ALLMODELS
  /AUXILIARY  CILEVEL=95 MAXACFLAGS=24
  /MISSING USERMISSING=EXCLUDE
  /MODEL DEPENDENT=importduty INDEPENDENT=month
    PREFIX='Model'
  /ARIMA AR=[0]  DIFF=1  MA=[1]
    TRANSFORM=NONE  CONSTANT=YES
  /AUTOOUTLIER DETECT=OFF.
```

Appendix 3: Data used

year	month	import duty
2003	January	1630
2003	February	1417
2003	March	1399
2003	April	1395
2003	May	1823
2003	June	1690
2003	July	1899
2003	August	1628
2003	September	1575
2003	October	1794
2003	November	1666
2003	December	1877
2004	January	1716
2004	February	1975
2004	March	1947
2004	April	1902
2004	May	1640
2004	June	2331
2004	July	1924
2004	August	1975
2004	September	2055
2004	October	1654
2004	November	2074
2004	December	2092
2005	January	2104
2005	February	1648
2005	March	1668

2005	April	1783
2005	May	1688
2005	June	2093
2005	July	1401
2005	August	1768
2005	September	1938
2005	October	1860
2005	November	1676
2005	December	1910
2006	January	1700
2006	February	1669
2006	March	1846
2006	April	1497
2006	May	1782
2006	June	2357
2006	July	2109
2006	August	2417
2006	September	2421
2006	October	1980
2006	November	2503
2006	December	2246
2007	January	2333
2007	February	2231
2007	March	2432
2007	April	2286
2007	May	2247
2007	June	2533
2007	July	3257
2007	August	2763
2007	September	2454

2007	October	2664
2007	November	3030
2007	December	2224
2008	January	2705
2008	February	2894
2008	March	3013
2008	April	2265
2008	May	2605
2008	June	2425
2008	July	2387
2008	August	2620
2008	September	2766
2008	October	3126
2008	November	3451
2008	December	3167
2009	January	3014
2009	February	2786
2009	March	2950
2009	April	2899
2009	May	2910
2009	June	2800
2009	July	3386
2009	August	3161
2009	September	3088
2009	October	3606
2009	November	3568
2009	December	3939
2010	January	3549
2010	February	2978
2010	march	3594

2010	April	3221
2010	May	3557
2010	June	3421
2010	July	3514
2010	August	3506
2010	September	4070
2010	October	3737
2010	November	4344
2010	December	4080