BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

A TIME SERIES ANALYSIS OF MONYTHLY GENERATED REVENUE

A CASE STUDY OF CHIKOMBA RDC

BY

NOTHANDO RUMBIDZAISHE CHIREMBA

B1953761

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE BACHELOR OF SCIENCE HONOURS DIGREE IN STATISTICS AND FINANCIAL MATHEMATICS

SUPERVISED BY MR BASIRA

DECEMBER 2022

APPROVAL FORM

l Chiremba Nothando R, do hereby declare that this submission is my work apart from the references of other people's work which has been acknowledged .l do hereby declare that this work has neither been presented in whole or part of any digree at this university.

Student's Name

Nothando Chiremba	……….	……….
B1953761	Signature	Date

Certified by

Mr Basira	……….	….......
Supervisor	Signature	Date

Mrs Pagan'a	……….	……….
Chairperson	Signature	Date

DEDICATION

I dedicate this dissertation to my parents, Mr and Mrs Goni , my brothers and sisters who endured a long period of my absence during the course of my learning. Thank you for your undying support morally and financially. Above all glory and praise be to the Almighty God for the protection and guidance during my academic years.

ACKNOWLEDGEMENTS

Firstly l would like to thank God for the guidance, strength, time and knowledge. I am indebted to my family for their tireless effort in helping me in my academic endeavours. Most gratitude goes to my supervisor Mr K Basira for his patience, encouragement, constructive comments, suggestions, corrections and remarks which were quite beneficial to the learner. Without his constant guidance, l would not have been able to finish this dissertation.

I would want to express my gratitude to the SFM department staff at large for fostering an environment that allowed me to finish my studies and for providing me with the support that I needed to excel in my studies. I would want to convey my gratitude to my classmates for the time we shared together.

ABSTRACT

A time series is a collection of data points that are typically measured at regular intervals of time. Examples can be found in a wide range of disciplines, from economics to engineering, and time series analysis techniques are a significant component of statistics. Time series analysis refers to techniques for deriving significant properties from time series data and predicting future values. A class of linear models called Autoregressive Integrated Moving Average (ARIMA) models, also known as Seasonal Autoregressive Moving Average (SARIMA) methods, can describe both stationary and nonstationary time data.

In ARIMA models, autocorrelation patterns play a significant role. In order to ascertain whether there is an ideal frame, this study will examine the application of the ARIMA and SARIMA approach to income earned, specifically sampling at various time intervals. The primary focus of the study was on the analysis of Chikomba Rural District Council's revenue generation using time series from the period of January 2013 through December 2021 with the intention of creating the best fit time series model to forecast revenue collection for the following three years. It aimed to highlight the difficulties Chikomba Rural District Council had making money.

The Chikomba Rural District Council's Finance Department provided secondary data to the researcher. The information given was for the overall revenue made each month during the given time period.

E-views 12 SV (64) and Excel were used by the researcher to analyze the data as part of the descriptive research design. The time series model with the best fit for predicting the monthly income earned over the next three years was ARIMA (4, 1, 4). The observed outcomes demonstrated that revenue fluctuated between increases and decreases at a steady rate. Additionally, researchers should think about conducting a qualitative research design employing crucial data from interviews with staff members and stakeholders to find the causes of the observed pattern.

Time series analysis, revenue, ARIMA, SARIMA, modeling, statistics, and stationary

TABLE OF CONTENTS

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

CRDC-Chikomba Rural District Council

LGA'S-Local Government Authorities

MA-Moving Average

AR-Auto-Regressive

SARIMA-Seasonal Auto-Regressive Intergreted Moving Average

ARIMA-Auto-Regressive Intergrated Moving Average

ARMA-Auto-Regressive Moving Average

RMSE-Root Mean Squared Error

MPE-Mean Percentage Error

VAT-Value Added Tax

MOF-Ministry Of Finance

NRB-Nepal Rastra Bank

AIC-Akaike Information Criterion

GDP-Gross Domestic Product

BIC-Bayesian Information Criterion

ACF-Auto-correlation Function

PACF-Partial –Autocorrelation Function

ADF-Argumented Dickery FullerTest

KPSS-Kwiatkowski-Phillips-Schmidt-Shin

CHAPTER 1

1.0 Introduction

This chapter provides a broad summary of the research. The researcher goes over the background of the monthly revenue generated at the Chikomba Rural District Council. In this chapter, the problem statement, study objectives, research questions, purpose of the study, study scope, and study significance are all developed. This study focuses on the CRDC's revenue production process and how that process has changed through time, both in terms of yearly growth and decline. The goal is to establish a pattern in the management of revenue generation and determine whether there has been a consistent rise in revenue over time.

1.1 Background of the study

The Local Governmental Finance Act No. 9 of 1982 gives Local Government Authorities (LGAs) the authority to raise revenue from a variety of sources in order to carry out the duties and exercise the powers granted by Act No. 8 of 1982 while also making sure that the money is used appropriately. This implies that LGAs must collect revenue from the sources mentioned in the Act. Various sources of income have been used since the LGAs were first established, but none of them has ever been able to generate enough income to cover all of the budget's necessary expenses for the LGA's activities.

Local authorities are municipal councils, town councils, and/or local bodies created in accordance with the Urban Councils Act [Chapter 29:15] or Rural District Councils Act [Chapter 29:13] with the general responsibility of regulating specific Council areas (Chiri, 2015,2017). One very important factor in the formation of local council is the ability of local government to produce revenue internally. Delivering services in line with the national goal of enhancing public service delivery is their primary goal (GOT, 2000).The process through which a local government attempts to recoup unpaid debts from the public is often referred to as revenue collection (Cross, 2019). Taxes, service fees, levies, fines or penalties, as well as income from using state services, are only a few possible sources of income.

The term "revenue" refers to government earnings. It is crucial because it greatly affects how much money will be available for spending. Without revenue, a budget and corresponding spending cannot exist. Taxes, levies, service fees, and licenses are typically the main sources of funding for public expenditures. Different local authorities have varying degrees of success in securing the necessary financing from the municipality for the services they are in charge of providing. Municipalities therefore receive a lower or larger portion of their funding from taxation collected at the federal level, known as the National Revenue Fund. The methods used to generate revenue also affect equality and the growth of the economy.

A social contract between the local people and the local government is developed as a result of the local people paying taxes and other service fees to the local authority and having the right to demand responsibility for the how, where, and why their taxes are spent. Approximately 75% of councils, according to Slack (2016), are experiencing weak incomes, infrastructure backlogs, and declining populations. He went on to say that as a result of these revenue constraints, councils are unable to maintain their current basic infrastructure and offer the people the services they require.

Due to public aversion toward paying taxes and the District's lack of control over conditional and donor funding, CRDC was faced with severe issues with poor revenue collection. The amount of money being taken in from sources like trade licenses, market dues, fines, and fees was also drastically falling, making it impossible to use it to successfully offer the necessary services to the general people. The District's numerous deficit financial budgets from previous years provide evidence of this, demonstrating that the projected expenses exceed the projected revenues. The researcher used time series analysis in Chikomba RDC to look into the monthly generated revenue in light of these current events.

1.2 Statement of the problem

One of the biggest issues facing local governments in Zimbabwe has been the availability of funding. Some of these issues include the State Government's refusal to pay Local Government 10% of its revenue, theft by Local Government tax collectors, and misappropriation of council finances by the Chairman of Local Government. The researcher felt it was necessary to undertake a time series study of the monthly generated revenue in Chikomba Rural District Council in light of the aforementioned issues.

1.2 Research objectives

The researcher was guided by the following objectives in carrying out the research

- To specify time series models.
- To identify the models.
- To estimate the models.
- To diagnose the models.
- To forecast using the models.

1.4 Research questions

In order to achieve the objectives of the study, the following research question was posed for the appropriate answers:-

- What are the time series models that can be used to analyse monthly generated revenue?

1.5 Assumptions

The researcher will make the following assumptions to ensure the ongoing validity of this study:-

• All information sources will generate reliable information and be free from bias.

• The data gathered will be accurate and trustworthy.

• Relevant authorities will be accommodating and cooperative in granting information access.

• The sample that the researcher employed will accurately reflect the entire population in terms of the results.

1.6 Significance of study

- This knowledge will enable management of councils in Zimbabwe to put in place effective and adequate controls so that council will be able to collect revenue efficiently which in turn will strengthen capacity of the council to discharge its prime responsibility that mandate its establishment.

- Moreover this study serves as a requirement of partial fulfillment of the requirement for the award of BSCSFM of Bindura University therefore this study is of great significance as it guarantees award of the degree to the researcher.

- The study would be used as empirical literature review with scholars interested in this field under the study to acquire more skills in research methodology and data analysis.

- The residents of various districts in the country may get to appreciate their role in making service delivery easy for their respective councils as they provide the required financial resources to make service delivery possible.

- The Management of CRDC and other councils in Zimbabwe may benefit from the research if solutions proposed by the researcher are adopted or put into use in their day to day running of their institutions.

- The study was beneficial to policy makers of CRDC and other local government in finding a solution to the problems of fluctuations in the local revenue collections and appraises their performance.

1.7 Limitations of the study

The study was conducted at CRDC, the following are some of the limitations that the researcher encountered during the study:-

- Given the amount of time required to conduct an appropriate and valuable investigation, the researcher's biggest constraint was time. In reality, the deadline for submitting the researcher's article severely limited the amount of time that could be used to analyze the research problem and track change over time. To get around this, the researcher covered some of the issues that were preventing the completion of this investigation throughout the weekends and late-night hours.
- Worthy mentioning is the bias of individual respondents towards certain schools of thought as well as the unlikelihood of respondents to give information which they thought to be confidential.
- The fund to support the study was a problem.

1.8 Deliminations of the study

According to the explanation provided below, the researcher divided the study's focus into four categories: geographical, temporal, subject, and the size and duration of the problem's solution.

- Geographic focus: The study was conducted in Zimbabwe, specifically at the Chikomba Rural District Council, due to the ease of access to information and the availability of the most recent audited final accounts.
- Time frame: The research was carried out over a nine-year period, from January 2013 to December 2021.

- Study objective: To analyze the monthly revenue generated by Chikomba Rural District Council.

1.9 Defination of terms

- According to Adenugba and Chike (2013), revenue refers to the sources of funds needed by the government to pay its operations. These monies come from a variety of sources, including taxes, loans, penalties, fees, and statutory requirements, among others. The fact that revenues are inflows is a crucial aspect. The business receives something in return for giving customers products and services.

- Revenue collection, which covers revenue collection, client management, debt and credit management, and registration and management of the poor. The fact that a significant number of municipal customers are impoverished and hence unable to pay for services must be taken into account when developing financial plans and business strategies (USAID, 2007).

- Local government refers to a political subdivision established inside a state to carry out the duties and responsibilities bestowed by legislative or constitutional laws. Like other governmental units, local governments have a set territory, a population, an organization, the right to engage in, and the ability to carry out, public activities. In any political system, local government is the smallest administrative division. It is the lowest level of government in a contemporary state organization and is legally independent, having the ability to generate its own income and carry out tasks allocated to it by the constitution.

- A time series is a collection of data points that appear in a particular order over a certain amount of time. Cross sectional data, which records a moment in time, can be compared to this.

- Time series models are a vibrant area of research with the goal of meticulously compiling and thoroughly analyzing the historical observations of a time series in order to create a model that accurately captures the series' underlying structure. The series for which forecasts are needed are then generated using this approach.

10. Summary

The study's guidelines are provided in this introductory chapter, which also explains what the study is about. This chapter served as an introduction to the entire study. This chapter provided a detailed summary of the research questions, objectives, and problem statement. The focus of the following chapter will be a review of the literature on time series analysis-based revenue generating.

CHAPTER 2

2.0 Introduction

Various literature evaluations and earlier research papers pertaining to revenue collection in Local Authorities are highlighted in this chapter. This chapter lists a number of contributions from numerous authors and experts on revenue collection that assisted in identifying any gaps in the mentioned literature. The chapter highlighted and examined their opinions, suggestions, and critically evaluated and interpreted contributions to the topic under consideration. In order to conduct a trustworthy investigation for this chapter, both previous and current information was used.

2:1Theoretical framework

2.1.0 Time series analysis

A time series is a collection of ordered data, according to Robinson (2020). The ordering normally refers to time, but other orderings, like over-space, etc., could be envisioned. The statistical techniques for assessing and modeling an ordered sequence of observations are the subject of time series analysis. Time series analysis is used to identify trends in statistical data over a regular time period (G,2013). A time series' prior observations are meticulously gathered and thoroughly examined with the goal of creating a model that accurately captures the series' innate structure. Generally speaking, time series attempt to comprehend the underlying context of the pertinent data points through the forecasting of future values from recorded past values.

There are stationary and non-stationary time series. However, stationary time series are the focus of time series theory. A time series is deemed stationary if it's mean and variance remain constant, and vice versa. When it comes to time series analysis, it's roots may be found in the use of time series models in the study of revenue collection. Since then, a number of models have been constructed to explain the revenue collection process, and more are continually being generated. Many statistical or data-based models have been used to simulate different revenue collecting tactics.

In addition to MA, AR, SARIMA, and ARIMA models, there are other time series models as well, although SARIMA and ARIMA models are the main emphasis of this study. The majority of the time, these models are employed to forecast the future and fit historical data. All statistical forecasting techniques are extrapolative in nature, i.e., they entail predicting future relationships or patterns based on past patterns.

The ability of a method or model to predict future outcomes will always be restricted, and the accuracy of future forecasts will also depend on how well-versed a person is in the data or variable of interest. As data availability grows as a result of improved technology and data collection efficiency, quantitative methods are currently widely used. Beyond the known variables that may reduce predictive power, there are unknowable events that may also inject uncontrollable flaws into future forecasts. The future may bring unanticipated changes because it is not constant over time, even when we anticipate some previous occurrences to continue (Hyndman et al., 1998). This is evidence in favor of short-term forecasts as opposed to long-term ones.

Despite the possibility of certain people knowing you, time series models frequently apply the maxim "no one knows you better than you." Shumway and Stoffer (2006) interpret this to suggest that time series approaches presuppose that explanatory variables be recorded in the historical occurrences of the variable of interest. As actual realizations for the variable under study become available, it is crucial to update the model because doing so could improve the results that can be predicted in the future. The anticipation for the future will, however, be off if the historical data is, by chance, erroneous.

In the instance of the CRDC, inaccurate data may have been used as a result of human error in recording financial flows or transactions as well as ignorance. Future projections will benefit from data cleaning and verification before building time series models. When the finalized figures or the adjusted/corrected figures for the previous month were available around the 15th of the following month, internal corrections were made to the monthly data used in this study. Included in this is the allocation of unallocated revenues to the appropriate revenue streams.

In order to spot trends and be able to predict or make forecasts for the future, historical data must be readily available. Chikomba Rural District Council provided the secondary information that was gathered from January 17, 2013, through December 31, 2021 consisting of monthly revenue generated. A trustworthy and more realistic model for industrial exploration will be developed by the researcher with the aid of this vast sample of data.

2.1.1 Arima model

ARIMA models are one of the most popular and widely used statistical method for time series forecasting. Three basic time series models are used to create ARIMA models.

a) Autoregressive (AR)

b) Moving Average (MA)

c) Autoregressive Moving Average (ARMA)

The MA model treats the time series' current value as a linear function of both its present and past residual values. In the ARMA model the AR and MA models are merged taking into account both historical values and residuals. The time series necessary for AR, MA, and ARMA models are stationary processes, which means that the series' mean and covariance does not change over time. The Auto-Regressive Integrated Moving Average (ARIMA) model is a traditional machine learning paradigm that varies from deep learning techniques.

This is a linear regression model that excels at forecasting time series, making it the perfect choice for this study. The ARIMA model is distinct from other machine learning models in that it only generates predictions based on historical values of the target variable and excludes exogenous variables as features. As a result, it is a model with a strong statistical and mathematical foundation.

AR: Auto-regressive

In this model, the target variable is autoregressively regressed on its historical data.

The lagged data of the target are used as the X variable.

$Y = B0 + B1 * Ylag1 + B2 * Ylag2 + \cdots + Bn * Ylagn$

As seen by the aforementioned equation, the current value Y is a linear function of the previous n values. B values are the regression beta values that are fitted to the model during training.

The equation below demonstrates how this equation might be modified to predict the future

$$Y_{forward1} = B0 + B1 * Y + B2 * Y_{lag1} + B3 * Y_{lag3} + \cdots Bn * Y_{lag\ (n-1)}$$

I: Integrated means that the data are subjected to a separate equation that applies a differencing step:

$Y\ foward\ 1 - Y = B0 + B1 * (Y - Ylag1) + B2 * (Ylag1 - Ylag2\ ) + \cdots$

The above equation shows that the relationship between the future value of Y and its past values is linear. This is done to make the mean-variance stationary for the Y values during time series forecasting.

MA: Moving Average

A moving average's mathematical formula

$Y = B0 + B1 * Elag1 + B2 * Elag2 + \cdots + Bn * Elagn$

The random residual discrepancies between the model and the target variable are represented by E in the moving average model. Therefore, E is the difference between the exact value and the model's estimated value. When all these factors are considered, the ARIMA model provides a good baseline model due to its simple design and absence of exogenous variables.

Although the creation of ARIMA is a difficult process, it can be boiled down to four steps:

- Identification of the ARIMA (p, d, q) structure
- Estimating the coefficients of the formulation
- Fitting test on the estimated formulation
- Predicting future results using historical data

The ARIMA model is a three parameter stochastic process (p, d, and q) where p represents the Auto-Regressive AR (p) process, d stands for the integration (which is required for the transformation into a stationary stochastic process) and q is for the Moving Average MA (q) process.

2.1.2 Sarima model

The SARIMA model, sometimes referred to as the seasonal autoregressive integrated moving average model, is an extension of the ARIMA model that enables the direct modeling of the seasonal component of the series. For stochastic model data with a seasonal data pattern, it is a time series forecasting technique.

According to Box and Jenkins (1976) and (R.H. & David S, 2006), the SARIMA model involves four steps, including:

<u>The model identification phase</u>

This phase involves identifying the variables to analyze and confirm the stationarity of the time series and choosing the most pertinent auto-regression and moving average combination.

<u>The model estimation phase</u>

It evaluates the models found in the preceding stage and chooses the most effective one.

The model validation phase

In this stage, the chosen model's accuracy is evaluated, and potential improvements are also determined..

The model forecasting phase

It forecasts the series' upcoming data, which are provided with a confidence interval.

The theoretical framework of the forecasting process is shown below:

- Defining the forecasting problem
- Gathering information
- Preliminary analysis
- Choosing a forecasting method
- Using and evaluating methods

2:2 Empirical literature

Finance, tourism, and transportation are just a few of the industries that have conducted studies on the modeling of data series using time series models (Pelinescu et al. (2010), Koirala (2012), Brojba (2010), Jayesekara and Passty (2009), and Slobodnitsky and Drucker (2008)). These models' accuracy in predicting both in-sample and out-of-sample values has made them excellent tools for modeling univariate data. Over the past few years, there has been an upsurge in the usage of the ARIMA and SARIMA models to analyze state revenue, with many writers advocating their use.

2:2:1 Review studies on Arima models

In order to aid authorities in developing effective plans and controlling local income and expenditure with the help of an effective strategic management tool, Pelinescu et al. (2010) examined the Romanian local budget. This developed as a result of the local authorities' difficulty in forecasting future revenue when creating their annual budgets. The Holt-Winters multiplicative and additive models were used in the authors' study to forecast total local revenue and own revenue of local authorities using historical data from the first quarter of 2000 through the third quarter of 2010. The Holt-Winters equations were created and run using the E-views software, and a model was chosen that minimized the Root Mean Squared Error (RMSE). Because it is user-friendly and produces reliable forecasts, the study suggested using Holt-Winters models as a tool for multi-annual budget forecasting.

In a comparable study for Romania, Brojba (2010) modeled the entire budget revenue using ARIMA models on data from monthly earnings for the years 2007 to 2008 (the period of the economic crisis). The data contained or displayed trend and seasonality, which allowed the ARIMA models to successfully capture data movement during the financial crisis. The study found that as the fitted values were reasonably close to the actuals, ARIMA models can be used to define goals and forecast future improvements. The short-term projections are the most accurate, but the model is not without flaws because the parameters are sensitive to sample size (Brojba, 2010).

The ARIMA model was also employed by Chatagny and Soguel (2009) to predict tax receipts for all 28 cantons or districts. The primary goal of the study was to provide evidence that employing univariate time series models can reduce forecast bias. Data on tax collections for the years 1944 to 2006 were acquired from the districts together with actual forecasts that were seen. In order to account for some districts' failure to record historical data in certain years, the time series data were split into two samples (1944 to 2006 and 1976 to 2006). The mean percentage error was used to categorize the over, under, and zero error per canton or district to evaluate the ARIMA model's performance against the observed projections for the two sample periods.

The findings of the mean percentage error showed that the ARIMA models and observed projections under expected tax revenue had mean percentage errors (MPE) close to zero in the two sample periods. Using straightforward univariate models, the study found that bias from the observed forecast can still be reduced (ARIMA). Because bias can be caused by a variety of circumstances, the ARIMA model's primary drawback is its inability to explain its origin. But because the models don't require explanatory variables and can therefore be used for forecasting without incurring any costs for data collection, they were thought to be useful.

By combining direct and indirect tax revenue variables such as VAT, Income Tax, Corporate Tax, Tax on Oil Products, and Other Tax Revenues, Silvestrini et al. (2008) estimated annual budget deficits in France using monthly data from government revenue and expenditure, spanning from January 1996 to December 2004. On the expenditure side, the relevant variables included interest on debt, salaries and pensions, operating expenses, interventions, civil capital expenditures, and military expenditures. The goal was to create a statistical univariate model that would be able to predict when the government deficit will close or grow, so that the government could be advised on policy decisions and deficit regulation.

All revenue and expenditure variables were modelled using seasonal ARIMA models using the reduced sample ending in December 2001. Data for 2004–2005 were saved to compare model predictions with actuals (model validation). Then, if fresh data became available, the monthly projections were adjusted. The monthly predictions were also added together to create two other forecasts: i monthly cumulative forecasts for the two validation years, and (ii) temporary aggregated annual ARIMA models, which were built for yearly projections one year in advance. The 2004 and 2005 annual realizations were compared to the monthly cumulative forecasts, the annual projections as a whole, and the official French forecast (traditional forecasts) (actuals). Comparing the temporary aggregated yearly ARIMA forecasts to the conventional forecasts, it was found that they were fairly accurate.

2:2:2 Review studies on Sarima models

The Ministry of Finance (MOF) and the Nepal Rastra Bank are the two principal organizations in Nepal that undertake revenue estimates (NRB). However, the techniques employed to anticipate Nepalese revenues were ineffective in capturing revenue flows, and there was a dearth of a well-documented system for doing so Koirala (2012). For the financial years 2012/13 and 2013/14 (Nepal's financial years start in August and end in July), Koirala estimated and forecasted Nepal's total revenue using monthly data from Nepal Rastra Bank from August 1997 to August

2012. Additionally, five methods—the Holts, Winters, Decomposition, Seasonal ARIMA, and Growth methods—were developed to help with this activity.

Winters (with seasonal component) and Seasonal ARIMA were determined to best reflect Nepal's total income out of all the techniques, as seen by their decreased Mean Percentage Error (MPE) and MAPE in comparison to the other three methods. The two techniques were suggested because they decreased forecasting mistakes in Nepal's overall income estimate.

The Box-Jenkins approach is frequently employed in a variety of disciplines, including the study of natural meteorological phenomena like rainfall. The average rainfall in Tamilnadu, India, was modeled using SARIMA (SARIMA(0,1,1)(0,1,1)12) by Nirmal and Sundaram (2010). The Indian Institute of Tropical Methodology (IITM), located in Pune, India, provided the study with sample data spanning the years 1871 to 2006. The mean absolute percentage inaccuracy was the criterion for performance evaluation (MAPE). According to the study's findings, SARIMA models are effective time series models for predicting Tamilnadu's monthly rainfall.

Similar research was conducted in Dhaka, Bangladesh, using seasonal ARIMA to model and predict rainfall. The purpose of the study was to help water management organizations prioritize and control water demand. The RMSE and AIC criteria were used to determine which model of the actual rainfall was the greatest match for the monthly rainfall data from 1981 to June 2010. In order to adequately describe or depict the rainfall time series data for the chosen sample, the ARIMA(0,0,1)(0,1,1)12 model was found. The closeness of the fitted values to the real values was verified using model adequacy procedures. In order to forecast two years (July 2010 to June 2012), Mahsin et al. applied a seasonal ARIMA model (2012).

Every country's economic development depends in large part on tourism; as a result, the GDP of a nation increases as more tourists visit that nation. When predicting the future flow of tourists, time series analysis is crucial. Singh (2013) developed SARIMA ARIMA models with seasonal effects to forecast the number of visitors from abroad that will visit Bhutan, India. Because there was no literature on modeling tourism arrival for Bhutan, this study represented the first attempt to apply the SARIMA to simulate tourist arrivals. A number of SARIMA models were developed using data on monthly international visitor arrivals from January 1983 to December 2012. In addition to the R2, RMSE, MAPE, and BIC statistics, the model with white noise residuals, ARIMA(0,1,1)(1,1,1), was chosen and used to produce monthly forecasts for 2013 and

2014. The study came to the conclusion that the SARIMA model's forecast might offer useful data for Bhutan's tourism arrivals.

3:0 Summary

Every nation's economy benefits from making the most accurate revenue projections possible since doing so results in a more equitable distribution of future budgets. It is clear from the literature analysis up top that time series models have proven to be effective at forecasting sales. However, it is necessary to have a recorded quantitative data sample of the same historical course. Because the precision of these approaches decreases over time, they are more accurate for short-term forecasts (two to three years).More in-depth knowledge of the variables of interest, a well-defined model, and the consideration of statistics like root mean squared error, mean absolute percentage error, quadratic lost function, and many others are required in order to obtain or generate long-term forecasts that are more accurate. In addition, it is important to monitor the derived forecasts and take them into account for future revisions as needed.

CHAPTER 3: RESEARCH METHODOLOGY

3:1 Introduction.

The term "research methodology" refers to the theoretical examination of the philosophies and methods used in a certain field of knowledge. It outlines the precise steps taken to locate, pick, and evaluate data on a certain subject. In order to determine a study's validity and reliability, the researcher must critically evaluate it. This chapter focuses on an examination of the fundamental terms and ideas of time series analysis, including the background conditions, presumptions, and procedures involved in the analysis and use of autoregressive integrated moving averages (ARIMA) and seasonal autoregressive integrated moving averages (SARIMA). In this chapter, the research methodologies are described. It consists of the study design, data sources, research methodologies, data gathering strategies, and data analysis procedures.

3.2 Research type

The research is a quantitative one. Secondary data was collected, analyzed and used to develop an ARIMA and a SARIMA to forecast monthly generated revenue.

3.3 Research design

The general method for conducting research, known as "research design," includes data collection, interpretation, analysis, and logical discussion (Creswell, 2014).This is a broad technique employed by the researcher to logically and coherently combine several study components. Since it offers responses to the research questions, it serves as the study's primary tool. The researcher will employ a quantitative study approach because it involves less subjective judgment and is more reliable.

3.4 Research instrument

Any equipment used to gather and evaluate data is referred to as a research instrument. E-Views was used to evaluate the monthly revenue time series data. In order to do the research, the internet was also employed to obtain publications and theories.

3.5 Data collection

The Department of Finance at the Chikomba Rural District Council provided the researcher with the data. The information was secondary, meaning it wasn't acquired by the researcher from the primary source but rather was aggregated and assembled by other organizations and was prepared for analysis. Press releases, annual reports, published audited financial statements, public documents, and corporate profiles are some important sources of secondary data that can be used to find the answers to research queries. The researcher opt to use secondary data because it is difficult to obtain primary data related to this topic. The information was gathered throughout a nine-month period starting in January 2013 and ending in December 2021.

3.6 Data cleaning

Not all times will the data be in a format that is ideal for developing models. Data cleaning seeks to organize unorganized data. A tidy dataset, according to Wickham (2014), can be characterized by three characteristics:

- Each observation comprises a row,
- a variable from the column and
- an observational unit produces a table.

The data are shown in a tabular format and cover the entire period from January 1, 2013, to December 31, 2022. After the data has been pre-processed, each observation is represented by a row, and each variable by a column.

The monthly produced revenues and the date are the two variables in the dataset. The accuracy and presence of duplicates in the data were verified.

3:7 Data presentation and analysis

3:7:1 Time series analysis

An ordered succession of observations can be analyzed and modelled using statistical techniques in time series analysis. Time series analysis is the understanding of formulas that break down a series into individual parts and atoms, allowing for the detection of underlying trends, which then leads to the generation of estimates and projections. By extrapolating future values from recorded previous values, time series analysis fundamentally aims to comprehend the underlying context of the pertinent data points. These time series models include MA, AR, ARIMA, and SARIMA, among others, although the SARIMA and ARIMA models serve as the primary focus of this study.

3:7:2 Lag

A period between two points or observations is defined as lag. This is a type of backward lagging, for instance, lag 1 is between Yt and Yt-1, lag 2 is between Yt and Yt-2, lag 3 is between Yt and Yt-3, and so on. Additionally, time series can be Yt and Yt+1 lag ahead. The value of the prior observation, Yt-1, is dependent on the observation at the current time, Yt.

3:7:3 Differencing

Other time-series do not require differencing and over-differenced series might result in subpar estimate values. This idea is used to make the series stationary, to d-trend, and to control autocorrelations. In the formula for differencing, the value of an earlier observation is subtracted from the value of a later observation. However, for this series, there are potential changes in both the mean and the dispersion with time. There is an increase in variability as the mean edges upward. By differencing once or twice for a dynamic mean, the trend is eliminated. To achieve a stationary process in the case of changing variability, logarithmic transformation is helpful.

A non-stationary mean is thought to be more easily made stationary by varying the scores. The number of times the scores must diverge in order for the process to become stationary is known as the parameter d. The model is stationary and has no trend if d=0. D=1 and linear trends are eliminated once the series has been differentiated once. Once the difference has been divided, both the linear and quadratic trends are eliminated, giving d=2. D values of 1 or 2 are typically sufficient to render the mean stationary for non-stationary series.

3.7.4 Stationary and Non-Stationary Series

In contrast to non-stationary series, which contain systematic tendencies like linear, quadratic, and so forth, stationary series move around a constant mean level and can be either systematically increasing or decreasing over time. In time series analysis, when the raw data are frequently changed to stationary, the stationary property is a useful tool. Economic data, for instance, are frequently seasonal or reliant on a non-stationary price level. Poor forecasting is the outcome of using non-stationary time series, which create incorrect and fake data.

In these cases, converting time series data to stationary form is a possible approach. A non-stationary mean can be made stationary more easily by varying the scores. The number of times the scores must be differentiable in order to render the process stationary is known as d. The model is already stationary and does not have a trend when d=0. The linear trend is eliminated if d = 1, which denotes a single change in the series. After the difference is divided, d=2 and both the linear and quadratic trends are eliminated. The mean can typically be made stationary for non-stationary series with d values of 1 or 2.

3.8 Components of time series

When choosing the best modeling and forecasting method, it is crucial to take into account the types of data patterns that can be seen in the time-series graphs of the time plots. When it comes to patterns in time series data, the following elements are the most often occurring sources of variation:

- Trend (T)
- Cyclical (C)
- Seasonal (S)
- Horizontal (H)
- Irregular (I

- **Trend (T)**

It happens when the data show a long-term increase or reduction. When discussing a trend in a cyclical context, the term "long-term movement" is typically used. The long-term movement in a time series without calendar-related and irregular influences that serves as a display of the underlying level is referred to as a trend, according to the African Society of Statistics (ASS, 2012). It comes about as a result of factors like population growth, price inflation, and broader economic changes. To make a model with two trend terms (d=2) stationary, it must be differenced twice. Depending on the circumstance, the first difference eliminates a linear trend, the second difference eliminates a quadratic trend, and so forth.

- **Cyclical(**C)

When the data exhibits cyclical rises and cyclical dips that are not associated with a specific period, a cyclical pattern is present (Markakis, Wheelwright, & Hyndman, 2004). (1998). Data is said to be in a cycle when there are rises and dips that are not periodic. These fluctuations are generally linked to the business cycle and are typically brought on by the state of the economy.

- **Seasonal (S)**

When a time series is impacted by seasonal elements like the time of year or the day of the week and recurs on a regular periodic basis, it is said to have a seasonal pattern. Seasonality in a time series refers to a recurring pattern of changes across S time periods, where S is the average number of time periods between each pattern. A seasonal first-order autoregressive model would employ Xt-12 for prediction given monthly data and S = 12. Seasonality always has a set and recognizable periodicity. Seasonal patterns are any repeatable, predictable patterns that last for an entire year.

- **Horizontal (H)**
  A series is said to be "stationary" in its mean when data values oscillate horizontally about it, according to Makridakis, Wheelwright, and Hyndman's (1998) interpretation.

- **Irregular (H)**

The irregular component of a time series is an unpredictable component that remains after the seasonal and trend components have been approximated and eliminated.

3.9 Identification of Arima and Sarima parameters

3.9.1 Arima Model Parameters

- **Autoregressive component**
  The term "autoregressive" (abbreviated as "AR") refers to a parameter that is regressive. It is assumed that there is no auto-correlation in the time series if p=0.There is still one lag in the series auto-correlation for p=1.The series is stationary for d=0, therefore we do not have to compute the difference for it. When d=1, the series is said to be non-stationary; consequently, the difference must be obtained first in order to make the series stationary. Two differences have been made in the series when d=2.

- **Moving average component, MA**

  In ARIMA, a moving average of q=1 signifies an error term and an auto-correlation with one lag. Q stands for the moving average parameter. To determine if the series and associated error terms are auto-correlated, the W-D test, ACF, and PACF are helpful.

3.9.2 Sarima model parameters

SARIMA = p, d, q, P, D, Q, S, where

- seasonal P specifies the number of autoregressive terms; lags of the stationary series, given the set of universal parameters.
- D and seasonal D indicates that a sequence of stationaries needs to be differentiated.
- q represents the number of moving average terms, along with seasonal Q.
- S stands for the length of the data.

Descriptive statistics, which provide the mean, maximum, minimum, and standard deviation among others will be used to interpret the data. The Serial LM Test which looks at residual autocorrelations for model construction will be used by the researcher in the analysis.

.

3.10 Model identification

Before data can be fitted to a time series model, it must be stationary. The mean, variance, and covariance of stationary data are constant and solely depend on the lag k. We shall first determine whether the data are stationary before determining the model. The Augmented Dickey-Fuller (ADF) test will be used for this. The researcher will utilize the differencing method up to the data are stationary if the data are not stationary. The researcher will recognize the model using the ACF and PACF graphs when the data is steady. By issuing the command auto. arima, the software will automatically determine the optimal model. The ARIMA model, which is Hyndman and

Khandakar, is then discovered by combining the unit root test, minimization of Akaike Information Criteria with a correlation (AICc), and the maximum likelihood function.

3.11 Diagnostic checking

The data's suitability for time series analysis will be evaluated using the model adequacy of both AR and MA models. To examine the behavior of the data, ACF and PACF will also be added. Montgomery et al. (2016) assert that the residuals ought to behave like a white noise process. Since there are no trends visible, a rectangular form indicates that the model is acceptable. If the model is acceptable, there should be no discernible structure in the residual sample autocorrelation function.The researcher will make use of the Hyndman and Khandakar ARIMA Chart, which displays the methods employed to analyze the data using E-views.

3.12 Summary

The research design, methodology, and data collection process were all covered in this chapter. The chapter continued by outlining the data analysis steps the researcher will do in the following chapter. The analysis and display of data are covered in the following chapter.

CHAPTER 4

4:0 Introduction

The outcomes of the data analysis are reported in this chapter. The chapter discusses the procedures the researcher used to analyze the data. Additionally, it offers the outcomes of the researcher's investigation into the ARIMA and SARIMA models. The results were done using Excel and E Views software. Prior to fitting the data into the model, stationarity was tested. The researcher employed quantitative analysis for analyzing the data. The presentation of descriptive statistics is based on secondary data gathered from a sample population in Chikomba RDC.

4:1 Data presentation

After collecting secondary data from Chikomba Rural Distrisct, it was uploaded into E views software. Data was checked whether it was time series data or not. The software detected that data was not time series and it was converted to time series.

4.1.1 Descriptive statistics

The descriptive statistics were calculated and below are the results that were obtained from the software

	Income generated
mean	1802.049
median	1790.280
maximum	4050.650
minimum	280.7800

Standard deviation	803.2562
skewness	0.266346
kurtosis	3.121869
Jarque-Bera	1.343755
probability	0.510749
sum	194621.3
Sum sq deviation	69038599
observations	108

Table 4.1: Descriptive statistics for monthly generated revenue

Table 4:1 above displays descriptive statistics for monthly revenue collected between January 2013 and December 2021. The sample consisted of 108 observations in total. According to the table above, the revenue collected ranged from 280.7800 to 4050.650, with 280.7800 serving as the minimum and 4050.650 serving as the maximum. The skewness of the data is 0.266346 and the average generated revenue has a positive value of 1802.049 indicating that the average monthly income has increased. The probability of the data being normally distributed is 0.510749, which is greater than the 0.05 level of significance. A favorable value of 1.343755 for Jarque-Bera can also be used to support this.

4.1.2 Time series plot of monthly generated revenue

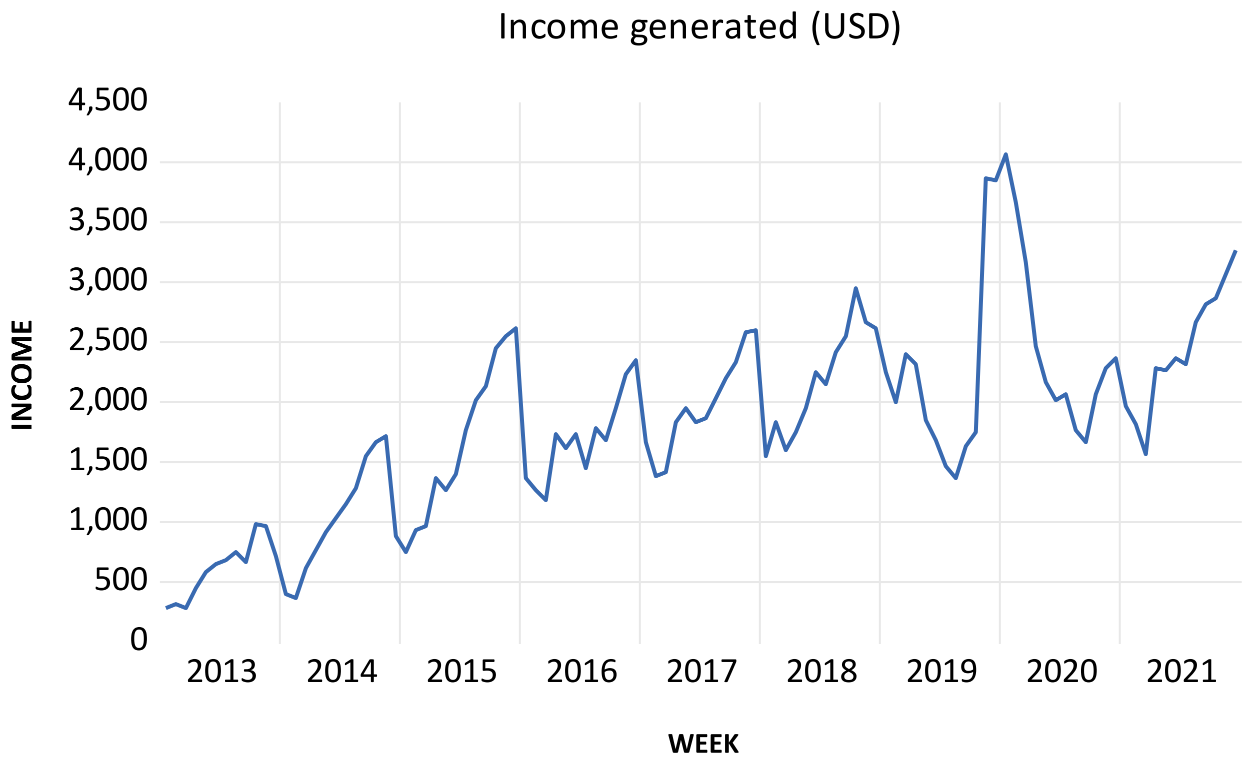


Figure 4.1: Time series plot of monthly generated revenue

The data pattern was not stationary, necessitating differencing; as a result, the researcher was able to produce a stationary time series plot of revenue collection at first differencing, as shown below. Time series plot of the revenue collected for the period January 2013 up to December 2021 was completed as illustrated above.

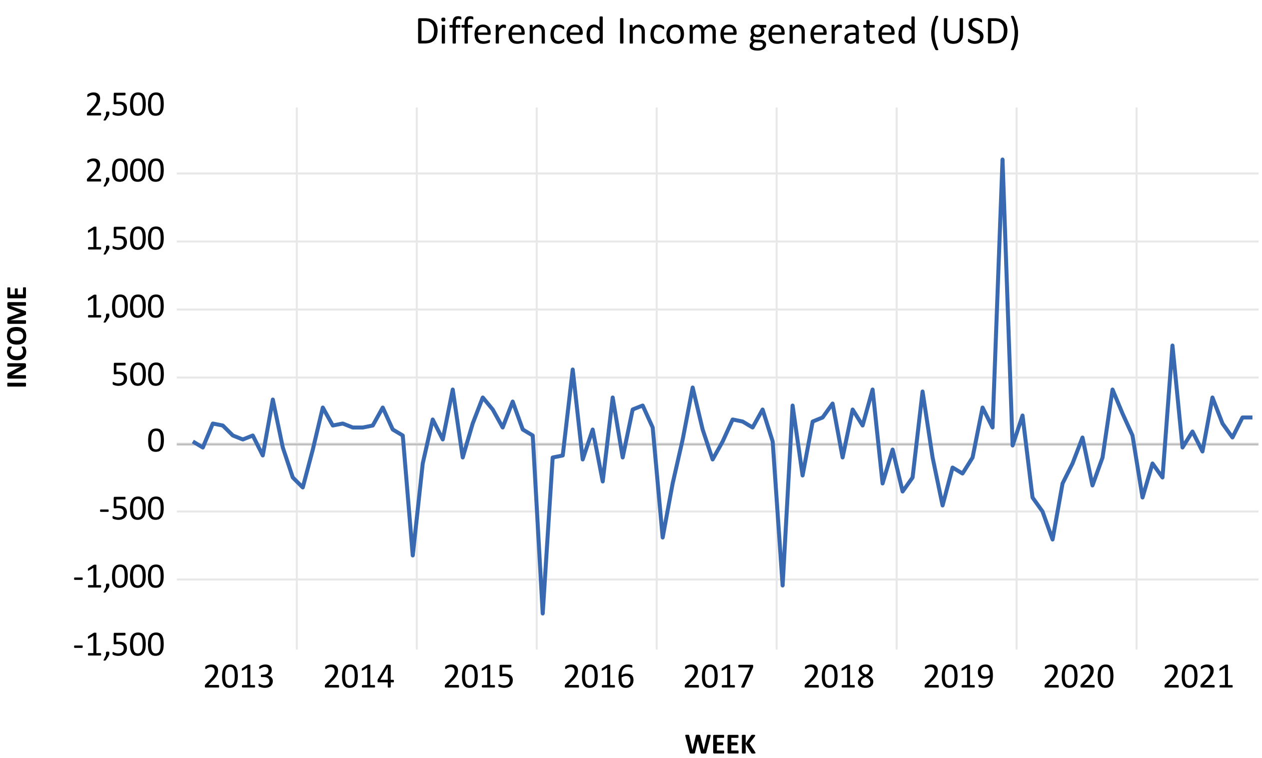


Figure 4.2: Time series plot of differenced income generated.

4.2 Stationarity

To test for stationarity before running the statistical test the researcher used ACF and PACF graphs as illustrated below. The data showed that there was stationarity at 1st differencing as shown in Figure 4.2.1 below.

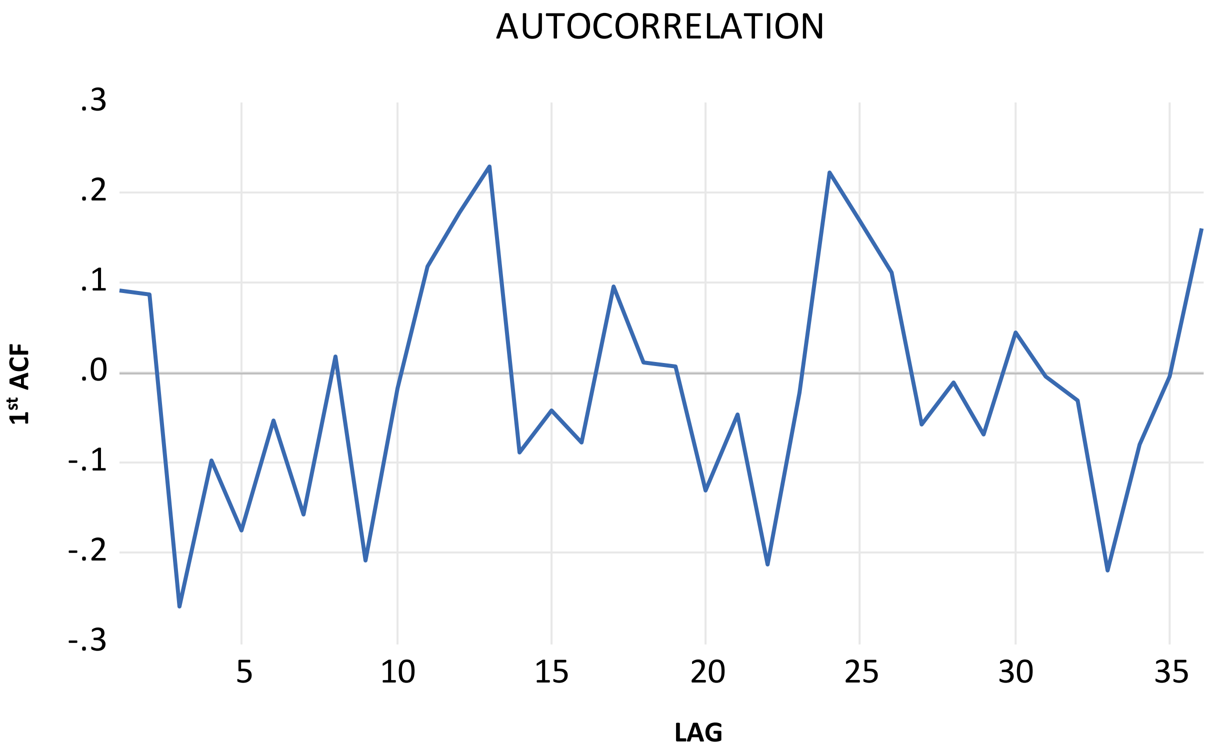


Figure 4.3

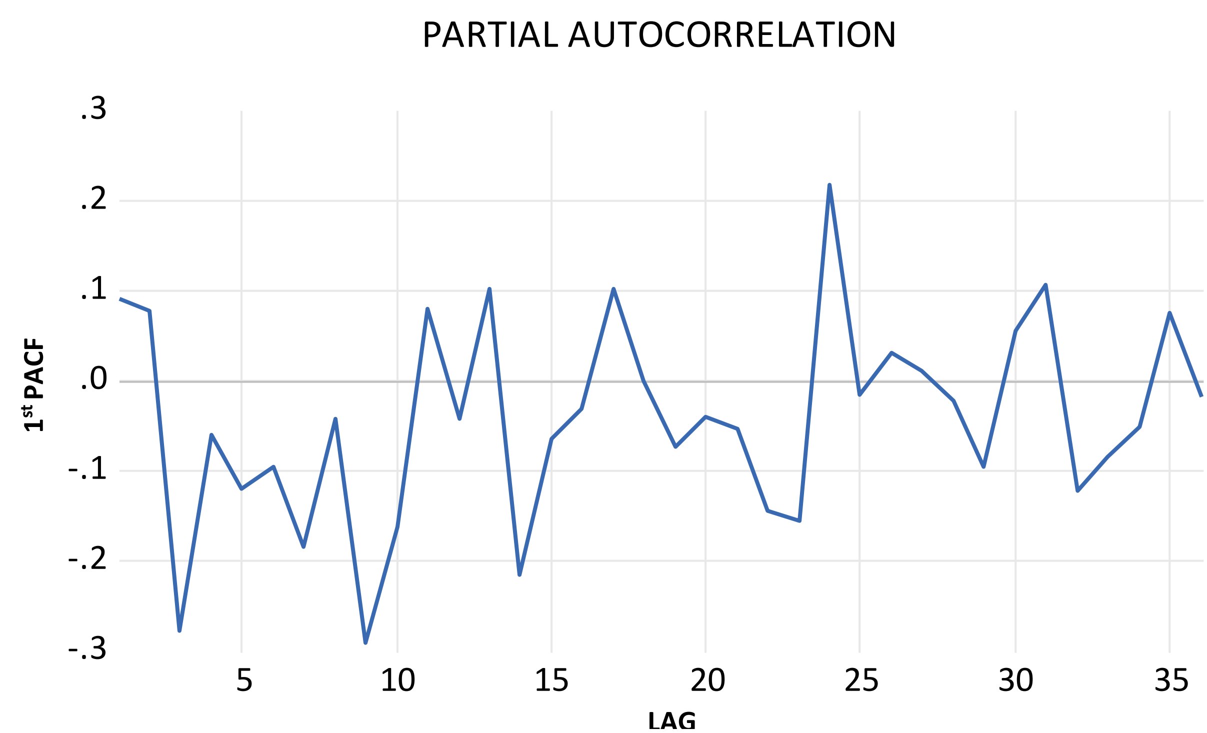


Figure 4.4

The plot showed that there was a constant increase and decrease in revenue collected. Therefore the time series data shows a stationary trend since there is constant variation.

4.3 Tests of stationarity

Stationarity is the main component in performing ARIMA. The researcher should ensure that the data is stationary before model building. In this case the researcher used KPSS Test to test for trend stationarity and an ADF test to test for stationarity.

4.3.1 KPSS test results

	Intercept	**Intercept and trend**
Level	**1.073116***	**0.132985**
1st difference	**0.082527**	**0.081086**

Table 4.2

KPSS test results depicted that revenue generated is stationary at level for intercept with a t statistic value of 1.07316 * which is greater than the critical t values as illustrated in appendix page.

4.3.2 ADF unit root test results

	Intercept	**Intercept and trend**
Level	**-2.469516***	**-5.340184**
1st difference	**-9.311660**	**-9.266514**

Table 4.3

Ho: There exists a unit root (non-stationary)

H_1: Unit root does not exist (stationary)

Table above shows the unit root test/stationarity test to check whether income generated was stationary at level or after 1st differencing for both intercept and trend plus intercept. Testing the above hypothesis at 5% level of significance and taking absolute values of the t statistic the outcomes were as follows:

- 5.340184>3.453179 at level ,trend and intercept
- 9.311660>2.888932 at 1st difference , intercept
- 9.266514>3.452764 at 1st difference , trend and intercept

Since t statistic is greater than the t critical as illustrated above, we reject H0 and conclude that income generated is stationary. See appendix for critical values at 5%.

4.4 Model identification

At this stage the researcher will identify the best fit time series model for revenue collection. The researcher will identify the Autoregressive and Moving Average terms which suits the data. The results obtained after running the command showed that ARIMA (4, 1, 4) is the best model for the data as shown in the table below.

MODEL	**LOGL**	**AIC**
ARIMA(4,1,4)	28.946385	-0.354138
ARIMA(4,1,3)	26.999718	-0.342834
ARIMA(3,1,4)	23.550518	-0.336443
ARIMA(2,1,2)	23.493629	-0.328047
ARIMA(1,1,3)	23.340538	-0.326984
ARIMA(0,1,4)	23.340538	-0.324122

ARIMA(0,1,3)	21.630223	-0.310845
ARIMA(3,1,2)	23.612266	-0.310510
ARIMA(1,1,4)	23.494022	-0.308299
ARIMA(2,1,3)	23.493971	-0.308298
ARIMA(4,1,1)	22.904929	-0.297288
ARIMA(3,1,3)	23.631614	-0.292180
ARIMA(4,1,2)	23.628872	-0.292128
ARIMA(2,1,4)	23.502220	-0.229761
ARIMA(3,1,0)	20.177414	-0.283690
ARIMA(2,1,1)	20.170076	-0.283553
ARIMA(3,1,1)	20.181424	-0.265073
ARIMA(4,1,0)	20.179756	-0.265042
ARIMA(1,1,2)	18.771400	-0.257409
ARIMA(0,1,0)	14.437587	-0.232478
ARIMA(1,1,0)	15.365527	-0.231131
ARIMA(0,1,1)	15.330419	-0.230475
ARIMA(0,1,2)	15.678994	-0.218299
ARIMA(2,1,0)	15.397086	-0.213030
ARIMA(1,1,1)	5.371400	-0.212550

Table 4.4

4.4.1 Parameter estimation

The parameters were estimated by e views and the results obtained are shown in the table below.

Arima(1,1,1)	
Coefficients :AR(1)	-0.746517
:MA(1)	0.796993
S. E :AR(1)	0.121051
:MA(1)	594.674

Sigma SQ	0.032346
Log likelyhood	28.94640
AIC	-0.354138
BIC	-0.104341
HQ	-0.252874

Table 4.5

Results of Parameter Estimation

There is no additional differencing in the model as shown in the table above. This is because the lower the AIC value the better the fit of the model and a negative BIC as illustrated in the table above indicates the preferred model. This means that the model shows that it has AR (p=1), order of differencing (d=1), MA (q=1) and standard error (s.e).

4.5 Diagnostic checking

It is at this point that the researcher checked if all the assumptions of ARIMA are fulfilled, that is stationarity, normality

4.5.1Model residual results

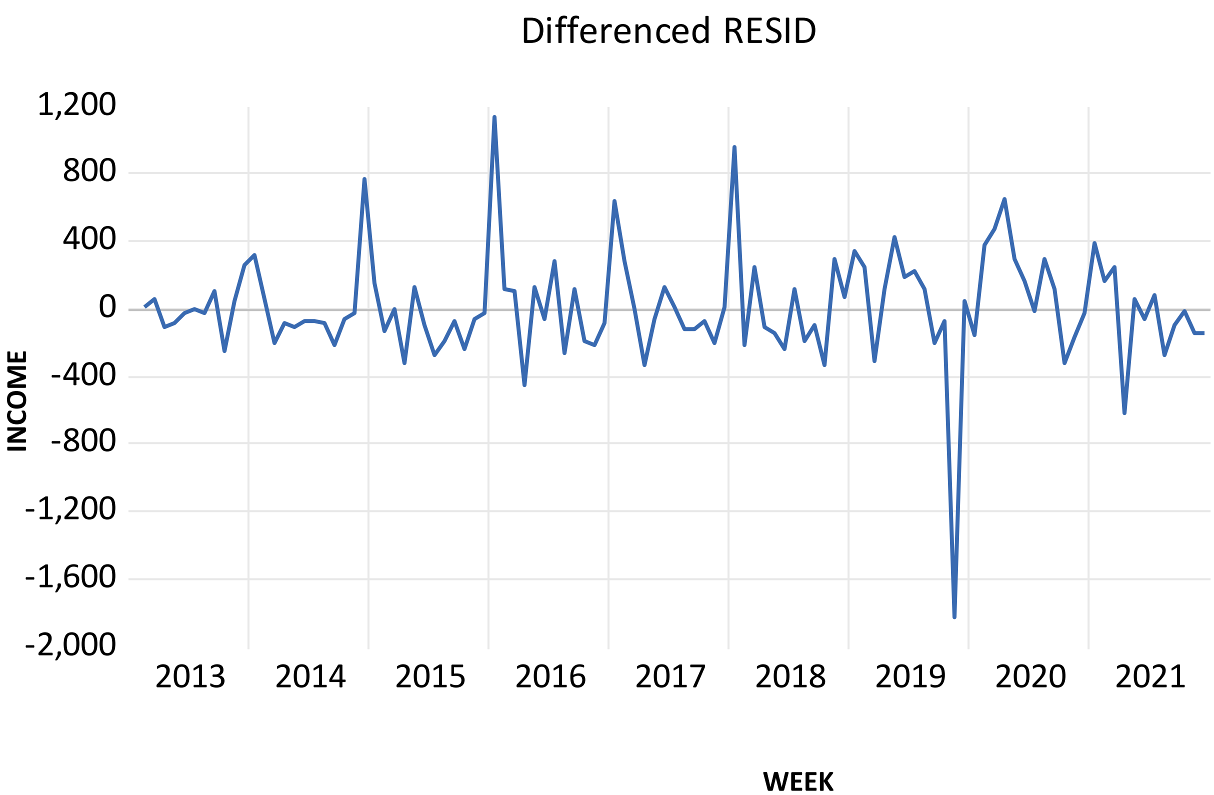


Figure 4.5

The graph in the figure above clearly shows a white noise structure where by the residuals deviated around mean zero and a constant variation.

4.5.2 Normality

Below is the histogram of residuals showing the normal curve obtained from the data after running the commands using E-views. The figure below is bell shaped showing normality of residuals and therefore the selected model satisfies the normality assumption.

Histogram residuals

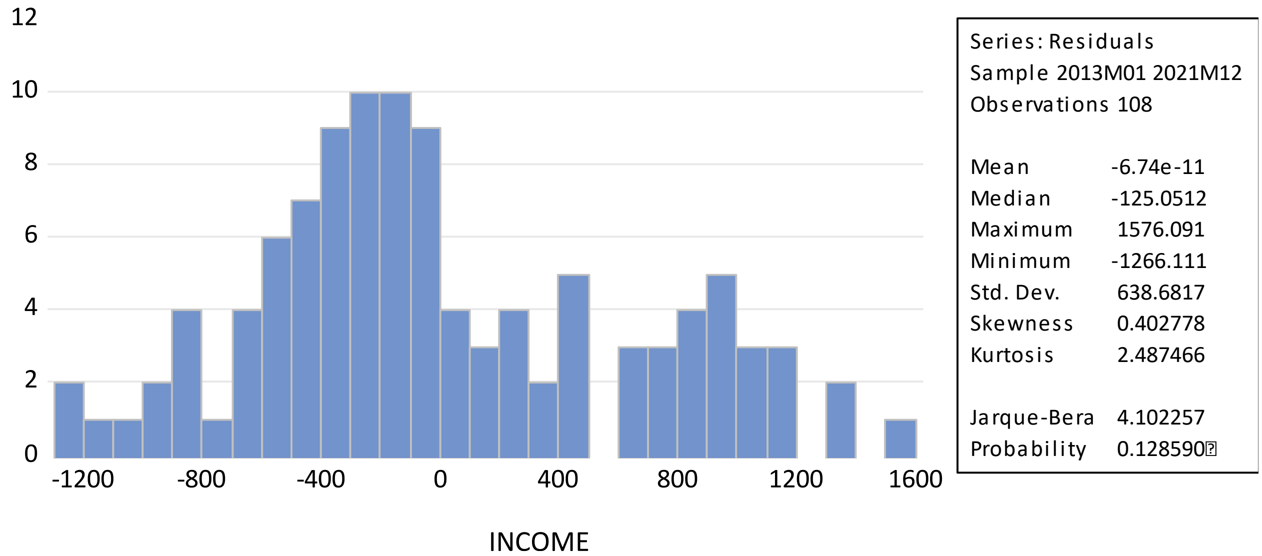


Figure 4.6

4.5.3 Serial Correlation LM Test

The researcher used Breusch- Godfrey test to test for serial correlations of residuals using the following hypothesis:

H0: There are no serial correlation of the time series

H1: There are serial correlations of the time series

F-statistic	176.0839	Prob. F(2,104)	0.000
Obs*R-squared	83.37749	Prob.Chi-Square(2)	0.000

Table 4.6

Since all the p values are less than 5% significance level it means that the result is significant hence we reject the null hypothesis and conclude that there are serial correlations in the data.

4.6 Testing for seasonality

A real quick way of checking for seasonality in a data set is by looking at the correlogram for that data and using e views it is pretty easier. The researcher checked the correlogram (appendix page) of monthly generated income which showed autocorrelation with no kind of identifiable pattern, ups and downs, highs and lows or changes that indicates some kind of pattern so there seems to be no seasonality in monthly revenue generated. Therefore the SARIMA model cannot be used on the data set.

4.7 Forecasting

The researcher forecasted future revenue collection of Chikomba RDC as shown below

Forecasts from ARIMA (4, 1, 4)

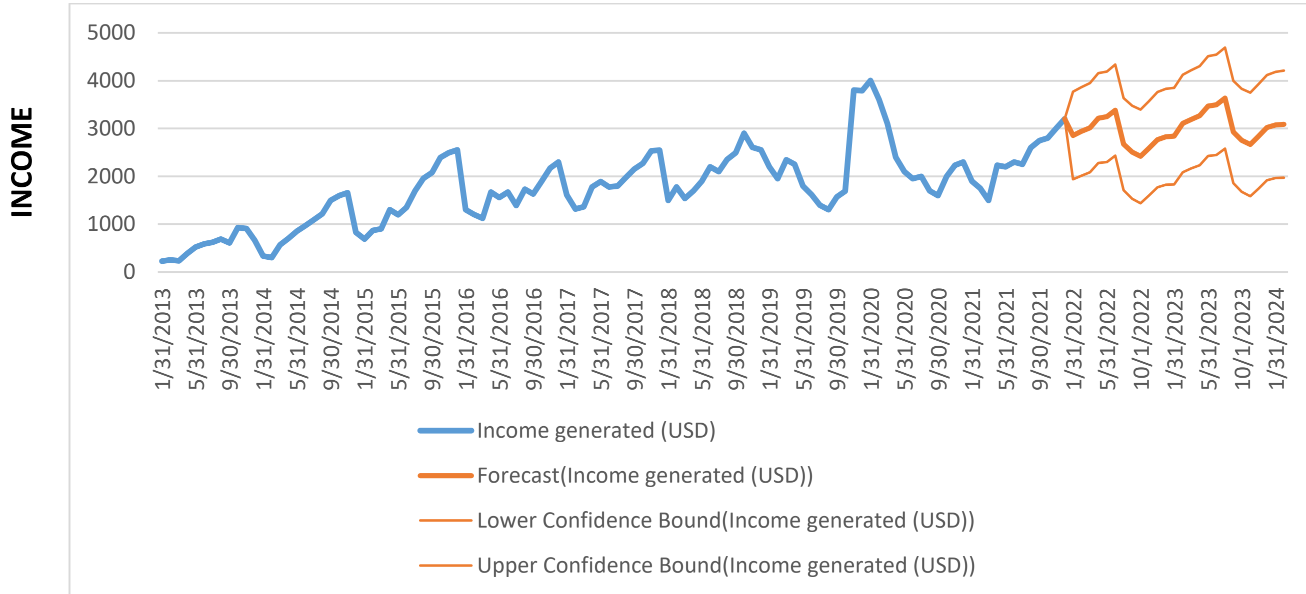


Figure 4.7

The above forecast shows the predicted revenue collection with a constant trend. It can be clearly shown from the forecast that the revenue collection will slightly increase and decrease at a constant rate.

4.8 Summary

This chapter showed data presentation together with its analysis which enabled the researcher to come up with the best fit time series model for the data which was done through model diagnostic checking. This also helped to forecast revenue collection of the Chikomba RDC for the next 3 years.

CHAPTER 5

5.0 Introduction

This chapter summarizes the research’s findings and suggestions.

5.1 Summary of the findings

Through the use of time series analysis, the study examined income generation in Chikomba RDC. The nine-year period from January 2013 to December 2021, when the research was conducted using monthly income. Chikomba RDC, Department of Finance, was the source of the information. Because the study's interests were centered on the trends in revenue collection, the research concentrated on the time series analysis of monthly income. The researcher had to locate the time series model that predicted revenue collection for the following three years with the best fit in order to get the best outcomes.

The ARIMA and SARIMA models were chosen to forecast the revenue trends over the following 36 months in accordance with the theoretical framework. In order to forecast revenue collection, the researcher was able to develop the best time series model. The best time series model was ARIMA (4, 1, 4), which was effective and trustworthy in producing precise revenue projections. The researcher utilized a correlogram to test for seasonality in the data set and found that there was none. As a result the SARIMA model cannot be used to predict revenue collection in the Chikomba RDC. A review of the time series analysis literature was also carried out while important empirical investigations were taken into consideration. Excel and the E-views SV(x64) 12 package were used for forecasting.

The researcher tested the stationarity of numerous graphs and stationarized the data to create an appropriate model for the data. The projected statistics showed a consistent trend in income from

January 2013 through December 2021. Due to a lack of data to predict future trends, this study primarily focused on Chikomba RDC's revenue collection, ignoring the town council.

5:2 Conclusion

According to the study's findings, the ARIMA (4, 1, 4) time series produced a steady trend in forecasted data, making it the best matched time series for predicting revenue collection in the Chikomba RDC.

5:3 Recommendations

In light of the research's findings, the researcher advises other academics to carry out the forecasting process using several techniques in order to develop the time series model that fits the data the best. Researchers can also conduct additional research on the town council's revenue collection, which was neglected in this study. The best time series model for predicting income collection for the following n years is ARIMA (4, 1, 4) and Chikomba RDC can utilize it to inform its strategic planning. In addition, users ought to backup their information using digital copies and update the software they plan to use to the best or most recent version.

The government should provide subsidies to Chikomba Rural District Council so that citizens' access to services can be enhanced. Time series models can be used to predict revenue collection with a fair degree of accuracy, however some researchers may need to include exogenous variables in the models to increase accuracy. Government spending and the income tax rate are two exogenous variables that might be tested.

REFERENCES

1. Adenugba A.A and Chike F.O.(2013),The Effect of Internally revenue generation on Infrastructural development, study of Lagos State Internal Revenue. *Journal Educational and Social Reasearch* 3(2) 6
2. Box, G. E. ., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. John Wiley and Sons, Inc
3. Box, G., & Jenkins, G.(1976). The Box-Jenkins Method. In Time Series *Analysis* (pp. 470-484)
4. Brojba L.C., Dumitru C.G., and Belciug A.V., 2010. On the use of some optimal strategies of fiscal administration during economic crisis. Romania: *Romanian Journal of Economic Forecasting,*
5. Chatagny F., and Soguel N.C., 2009. Tax Revenue Forecasting in the Swiss Cantons: A Time Series Analysis. Switzerland. IDHEA
6. Chiri, M. (2015,2017). *Local Authorities Report.* Harare: Auditor General.
7. Creswell, John W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Thousand Oaks: SAGE Publications
8. Cross, V. (2019, July 19). *Smallbusiness*. Retrieved July 19, 2019, from Chron: https://smallbusiness.chron.com/strategic-plans-boost-revenue-collection-14627.html
9. G, O. (2013). Electricity Use in Namibia. *Developing Algorithms to Encourage More*.
10. Hyndman, R J, & Athanasopoulos, G. (2014), Forecasting: Principles and practice. OTexts IBM. (2013). IBM SPSS Forecasting 22
11. Hyndman, R, & Khandakar, Y.(2008). Automatic Time Series Forecasting: The forecast package for R. *Journal of Statistical Software*, 1-22. https://doi.org http://article.sapub.org/10.5923.j.statistics.20140406.06.html

12. Hyndman R.J., Makridakis S., and Wheelwright S.C., 1998. Forecasting methods and application:Third edition USA: John Wiley and Sons.
13. Jayesekara N., and Passty B.W., 2009. City of Cincinnati Income Tax Revenue: A Forecast Based on Economic Data and Causal Inference. *University of Cincinnati.*
14. Kahneman, D. T. (2011). How Predictive Analysis and Choice Architecture Can Improve Student Success. *Research and Practise Assessment, 9*, 61-69.
15. Koirala T.P., 2012.Government Revenue Forecasting in Nepal. Nepal. Nepal Rastra Bank
16. Mahsin M.D., Akhter Y., and Begum M., 2012. Modeling rainfall in Dhaka division of Bangladesh using time series analysis . Dhaka, Bangladesh. *Journal of Mathematical Modeling and Application,*1:67−73.
17. Montgomery, D.C., Jennings, C, ., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting* (Second Edi). John Wiley& Sons
18. Nirmal M., and Sundaram S.M., 2010. A seasonal ARIMA model for forecasting monthly rainfall in Tamilnadu. Tamilnadu, India. *National Journal of advances in building sciences and mathematics,* 1.
19. Pelinescu E., Anton L.V., Ionescu R., and Tasca R., 2010. The analysis of local budgets and their importance in the fight against the economic crisis effects. Romania. *Romanian Journal of Economic Forecasting*,1.
20. R. Salgotra,2020, Time series analysis and forecast of the Covid 19 pandemic in India
21. S., & David S, S. (2006). *Time Series Analysis And Explanatory Data Analysis.*
22. Shumway R.H., and Stoffer D.S, 2006. Time Series Analysis and Its Applications With R Examples:Third edition.USA:Springer
23. Silvestrini A., Salto M., Moulin L., and Veredas D., 2008. Monitoring and forecasting annual public deficit every month: the case of France. France. *Empirical Economics,*34:493−534.
24. Singh E.H., 2013. Forecasting Tourist Inflow in Bhutan using Seasonal ARIMA Bhutan,India. *International Journal of Science and Research*, 2:2319−7064
25. Slack, D. E. (2016). *IMPROVING LOCAL GOVERNMENT REVENUE IN NSW. WHAT ARE THE OPTIONS.* Toronto: Institute of Municipal Finance and Governance

APPENDIX

Descriptive statistics

	INCOME_G...
Mean	1802.04888...
Median	1790.28
Maximum	4050.65
Minimum	280.78
Std. Dev.	803.256215...
Skewness	0.26634577...
Kurtosis	3.12186858...
Jarque-Bera	1.34375507...
Probability	0.51074872...
Sum	194621.28
Sum Sq. Dev.	69038598.5...
Observations	108

Correlogram

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.089	0.089	0.8801	0.348
		2	0.085	0.077	1.6790	0.432
		3	-0.261	-0.279	9.3425	0.025
		4	-0.098	-0.060	10.425	0.034
		5	-0.176	-0.122	13.955	0.016
		6	-0.055	-0.097	14.300	0.026
		7	-0.159	-0.185	17.230	0.016
		8	0.016	-0.043	17.261	0.027
		9	-0.209	-0.293	22.443	0.008
		10	-0.019	-0.164	22.484	0.013
		11	0.116	0.078	24.117	0.012
		12	0.176	-0.043	27.911	0.006
		13	0.228	0.102	34.384	0.001
		14	-0.090	-0.216	35.396	0.001
		15	-0.043	-0.066	35.629	0.002
		16	-0.078	-0.033	36.407	0.003
		17	0.094	0.102	37.560	0.003
		18	0.009	0.000	37.571	0.004
		19	0.006	-0.074	37.575	0.007
		20	-0.133	-0.041	39.935	0.005
		21	-0.048	-0.054	40.245	0.007
		22	-0.214	-0.146	46.548	0.002
		23	-0.023	-0.156	46.624	0.002
		24	0.221	0.216	53.480	0.000
		25	0.167	-0.016	57.427	0.000
		26	0.109	0.030	59.145	0.000
		27	-0.058	0.009	59.644	0.000
		28	-0.012	-0.023	59.666	0.000
		29	-0.071	-0.097	60.410	0.001
		30	0.043	0.054	60.687	0.001
		31	-0.005	0.106	60.691	0.001
		32	-0.032	-0.123	60.848	0.002
		33	-0.221	-0.085	68.572	0.000
		34	-0.082	-0.053	69.654	0.000
		35	-0.005	0.075	69.658	0.000
		36	0.158	-0.018	73.742	0.000

Unit Root Tests

1. KPSS TEST

- **Level, Intercept**

Null Hypothesis: INCOME_GENERATED__USD_ is stationary
Exogenous: Constant
Bandwidth: 8 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		1.073116
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	639246.3
HAC corrected variance (Bartlett kernel)	3575014.

KPSS Test Equation
Dependent Variable: INCOME_GENERATED__USD_
Method: Least Squares
Date: 11/20/22 Time: 16:15
Sample (adjusted): 1/31/2013 12/31/2021
Included observations: 108 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1802.049	77.29337	23.31441	0.0000

R-squared	0.000000	Mean dependent var	1802.049
Adjusted R-squared	0.000000	S.D. dependent var	803.2562
S.E. of regression	803.2562	Akaike info criterion	16.22444
Sum squared resid	69038599	Schwarz criterion	16.24928
Log likelihood	-875.1198	Hannan-Quinn criter.	16.23451
Durbin-Watson stat	0.207657		

- **Level , Trend and Intercept**

Null Hypothesis: INCOME_GENERATED__USD_ is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 5 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.132985
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	286923.0
HAC corrected variance (Bartlett kernel)	869360.8

KPSS Test Equation
Dependent Variable: INCOME_GENERATED__USD_
Method: Least Squares
Date: 11/20/22 Time: 16:18
Sample (adjusted): 1/31/2013 12/31/2021
Included observations: 108 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	783.4342	103.3358	7.581442	0.0000
@TREND("1/31/2013")	19.03953	1.668842	11.40883	0.0000

R-squared	0.551154	Mean dependent var	1802.049
Adjusted R-squared	0.546920	S.D. dependent var	803.2562
S.E. of regression	540.6816	Akaike info criterion	15.44188
Sum squared resid	30987684	Schwarz criterion	15.49155
Log likelihood	-831.8617	Hannan-Quinn criter.	15.46202
F-statistic	130.1613	Durbin-Watson stat	0.460249
Prob(F-statistic)	0.000000		

- **1st diffrerence , Intercept**

Null Hypothesis: D(INCOME_GENERATED__USD_) is stationary
Exogenous: Constant
Bandwidth: 11 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.082527
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	133214.3
HAC corrected variance (Bartlett kernel)	41384.81

KPSS Test Equation
Dependent Variable: D(INCOME_GENERATED__USD_)
Method: Least Squares
Date: 11/20/22 Time: 16:23
Sample (adjusted): 2/28/2013 12/31/2021
Included observations: 107 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	27.75421	35.45051	0.782900	0.4354

R-squared	0.000000	Mean dependent var	27.75421
Adjusted R-squared	0.000000	S.D. dependent var	366.7029
S.E. of regression	366.7029	Akaike info criterion	14.65628
Sum squared resid	14253931	Schwarz criterion	14.68126
Log likelihood	-783.1111	Hannan-Quinn criter.	14.66641
Durbin-Watson stat	1.819042		

- **1st difference ,Trend and Intercept**

Null Hypothesis: D(INCOME_GENERATED__USD_) is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 11 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.081086
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	133212.9
HAC corrected variance (Bartlett kernel)	41268.07

KPSS Test Equation
Dependent Variable: D(INCOME_GENERATED__USD_)
Method: Least Squares
Date: 11/20/22 Time: 16:26
Sample (adjusted): 2/28/2013 12/31/2021
Included observations: 107 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29.83492	71.73974	0.415877	0.6783
@TREND("1/31/2013")	-0.038532	1.153199	-0.033413	0.9734

R-squared	0.000011	Mean dependent var	27.75421
Adjusted R-squared	-0.009513	S.D. dependent var	366.7029
S.E. of regression	368.4430	Akaike info criterion	14.67496
Sum squared resid	14253779	Schwarz criterion	14.72492
Log likelihood	-783.1106	Hannan-Quinn criter.	14.69522
F-statistic	0.001116	Durbin-Watson stat	1.819063
Prob(F-statistic)	0.973409		

2. ADF TEST

- **Level , Intercept**

Null Hypothesis: INCOME_GENERATED__USD_ has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.469516	0.1258
Test critical values:	1% level	-3.492523	
	5% level	-2.888669	
	10% level	-2.581313	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(INCOME_GENERATED__USD_)
Method: Least Squares
Date: 11/20/22 Time: 16:28
Sample (adjusted): 2/28/2013 12/31/2021
Included observations: 107 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INCOME_GENERATED__USD_(-1)	-0.108129	0.043786	-2.469516	0.0151
C	221.1449	85.62538	2.582703	0.0112

R-squared	0.054893	Mean dependent var	27.75421
Adjusted R-squared	0.045892	S.D. dependent var	366.7029
S.E. of regression	358.1898	Akaike info criterion	14.61852
Sum squared resid	13471493	Schwarz criterion	14.66848
Log likelihood	-780.0907	Hannan-Quinn criter.	14.63877
F-statistic	6.098507	Durbin-Watson stat	1.729302
Prob(F-statistic)	0.015141		

- **Level , Trend and Intercept**

Null Hypothesis: INCOME_GENERATED__USD_ has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.340184	0.0001
Test critical values:	1% level	-4.047795	
	5% level	-3.453179	
	10% level	-3.152153	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(INCOME_GENERATED__USD_)
Method: Least Squares
Date: 11/20/22 Time: 16:31
Sample (adjusted): 4/30/2013 12/31/2021
Included observations: 105 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INCOME_GENERATED__USD_(-1)	-0.370647	0.069407	-5.340184	0.0000
D(INCOME_GENERATED__USD_(-1))	0.247970	0.093294	2.657954	0.0092
D(INCOME_GENERATED__USD_(-2))	0.277675	0.095735	2.900440	0.0046
C	314.8786	86.05345	3.659106	0.0004
@TREND("1/31/2013")	6.792170	1.664697	4.080124	0.0001

R-squared	0.232812	Mean dependent var	28.26371
Adjusted R-squared	0.202125	S.D. dependent var	370.1772
S.E. of regression	330.6566	Akaike info criterion	14.48649
Sum squared resid	10933376	Schwarz criterion	14.61286
Log likelihood	-755.5405	Hannan-Quinn criter.	14.53770
F-statistic	7.586558	Durbin-Watson stat	1.953435
Prob(F-statistic)	0.000022		

- **1st difference, Intercept**

Null Hypothesis: D(INCOME_GENERATED__USD_) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.311660	0.0000
Test critical values:	1% level	-3.493129	
	5% level	-2.888932	
	10% level	-2.581453	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(INCOME_GENERATED__USD_,2)
Method: Least Squares
Date: 11/20/22 Time: 16:34
Sample (adjusted): 3/31/2013 12/31/2021
Included observations: 106 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INCOME_GENERATED__USD_(-1))	-0.910375	0.097767	-9.311660	0.0000
C	25.43106	35.90475	0.708292	0.4803

R-squared	0.454661	Mean dependent var	1.645189
Adjusted R-squared	0.449417	S.D. dependent var	496.9259
S.E. of regression	368.7253	Akaike info criterion	14.67667
Sum squared resid	14139672	Schwarz criterion	14.72692
Log likelihood	-775.8634	Hannan-Quinn criter.	14.69704
F-statistic	86.70701	Durbin-Watson stat	2.011999
Prob(F-statistic)	0.000000		

- **1st difference, Trend and Intercept**

Null Hypothesis: D(INCOME_GENERATED__USD_) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.266514	0.0000
Test critical values:	1% level	-4.046925	
	5% level	-3.452764	
	10% level	-3.151911	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(INCOME_GENERATED__USD_,2)
Method: Least Squares
Date: 11/20/22 Time: 16:35
Sample (adjusted): 3/31/2013 12/31/2021
Included observations: 106 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INCOME_GENERATED__USD_(-1))	-0.910402	0.098246	-9.266514	0.0000
C	27.00202	73.58273	0.366961	0.7144
@TREND("1/31/2013")	-0.028812	1.176190	-0.024496	0.9805

R-squared	0.454664	Mean dependent var	1.645189
Adjusted R-squared	0.444075	S.D. dependent var	496.9259
S.E. of regression	370.5099	Akaike info criterion	14.69553
Sum squared resid	14139589	Schwarz criterion	14.77091
Log likelihood	-775.8631	Hannan-Quinn criter.	14.72608
F-statistic	42.93719	Durbin-Watson stat	2.011954
Prob(F-statistic)	0.000000		

Model Selection

- **Summary**

Automatic ARIMA Forecasting
Selected dependent variable: DLOG(INCOME_GENERATED...
Date: 11/20/22 Time: 16:38
Sample: 1/31/2013 1/12/2022
Included observations: 107
Forecast length: 0
Model maximums: (4,4)2(0,0)
Regressors: C

Number of estimated ARMA models: 25
Number of non-converged estimations: 0
Selected ARMA model: (4,4)(0,0)
AIC value: -0.354138035818

- **Equation output**

Dependent Variable: DLOG(INCOME_GENERATED__USD_)
Method: ARMA Maximum Likelihood (BFGS)
Date: 11/20/22 Time: 16:41
Sample: 2/28/2013 12/31/2021
Included observations: 107
Convergence not achieved after 500 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.017904	0.008247	2.170946	0.0324
AR(1)	-0.177232	0.143119	-1.238351	0.2186
AR(2)	1.314230	0.126755	10.36831	0.0000
AR(3)	-0.089625	0.115167	-0.778222	0.4383
AR(4)	-0.746517	0.121051	-6.166952	0.0000
MA(1)	0.196847	8.017739	0.024551	0.9805
MA(2)	-1.512128	569.1276	-0.002657	0.9979
MA(3)	-0.192956	68.03871	-0.002836	0.9977
MA(4)	0.796993	594.6714	0.001340	0.9989
SIGMASQ	0.032346	1.588354	0.020365	0.9838

R-squared	0.276401	Mean dependent var	0.022888
Adjusted R-squared	0.209263	S.D. dependent var	0.212423
S.E. of regression	0.188894	Akaike info criterion	-0.354138
Sum squared resid	3.461053	Schwarz criterion	-0.104341
Log likelihood	28.94640	Hannan-Quinn criter.	-0.252874
F-statistic	4.116897	Durbin-Watson stat	1.815974
Prob(F-statistic)	0.000165		

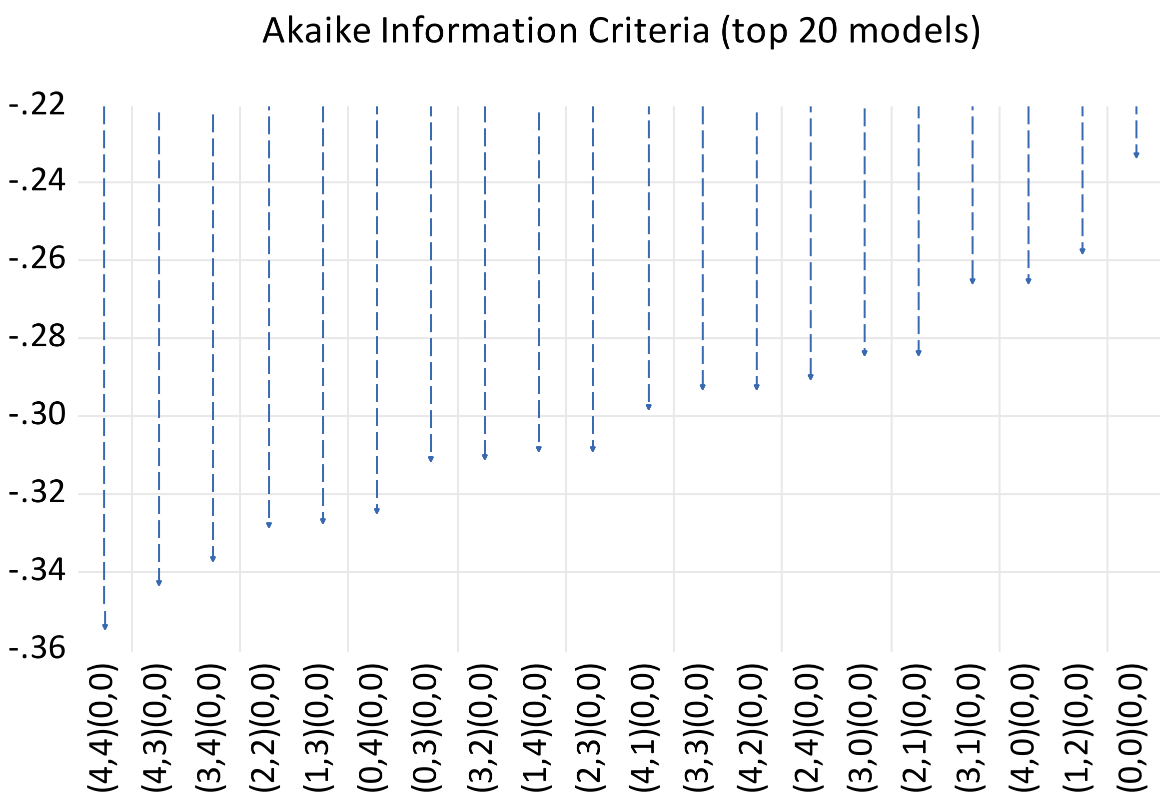
Inverted AR Roots	.83+.40i	.83-.40i	-.92+.20i	-.92-.20i
Inverted MA Roots	.86+.23i	.86-.23i	-.96+.28i	-.96-.28i

- **Arma Criteria Table**

Model Selection Criteria Table
Dependent Variable: DLOG(INCOME_GENERATED__USD_)
Date: 11/20/22 Time: 16:42
Sample: 1/31/2013 1/12/2022
Included observations: 107

Model	LogL	AIC*	BIC	HQ
(4,4)(0,0)	28.946385	-0.354138	-0.104341	-0.252874
(4,3)(0,0)	27.341639	-0.342834	-0.118017	-0.251696
(3,4)(0,0)	26.999718	-0.336443	-0.111626	-0.245305
(2,2)(0,0)	23.550518	-0.328047	-0.178169	-0.267288
(1,3)(0,0)	23.493629	-0.326984	-0.177105	-0.266225
(0,4)(0,0)	23.340538	-0.324122	-0.174244	-0.263364
(0,3)(0,0)	21.630223	-0.310845	-0.185947	-0.260213
(3,2)(0,0)	23.612266	-0.310510	-0.135652	-0.239625
(1,4)(0,0)	23.494022	-0.308299	-0.133442	-0.237414
(2,3)(0,0)	23.493941	-0.308298	-0.133440	-0.237413
(4,1)(0,0)	22.904929	-0.297288	-0.122430	-0.226403
(3,3)(0,0)	23.631614	-0.292180	-0.092342	-0.211168
(4,2)(0,0)	23.628872	-0.292128	-0.092291	-0.211117
(2,4)(0,0)	23.502220	-0.289761	-0.089923	-0.208750
(3,0)(0,0)	20.177414	-0.283690	-0.158791	-0.233058
(2,1)(0,0)	20.170076	-0.283553	-0.158654	-0.232921
(3,1)(0,0)	20.181424	-0.265073	-0.115195	-0.204315
(4,0)(0,0)	20.179756	-0.265042	-0.115164	-0.204284
(1,2)(0,0)	18.771400	-0.257409	-0.132511	-0.206777
(0,0)(0,0)	14.437587	-0.232478	-0.182519	-0.212225
(1,0)(0,0)	15.365527	-0.231131	-0.156192	-0.200752
(0,1)(0,0)	15.330419	-0.230475	-0.155536	-0.200096
(0,2)(0,0)	15.678994	-0.218299	-0.118380	-0.177793
(2,0)(0,0)	15.397086	-0.213030	-0.113111	-0.172524
(1,1)(0,0)	15.371400	-0.212550	-0.112631	-0.172044

- **Arma Criteria Graph**



- **Breusch-Godfrey Serial Correlation LM Test**

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	176.0839	Prob. F(2,104)	0.0000
Obs*R-squared	83.37749	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 11/20/22 Time: 16:51
Sample: 1/31/2013 12/31/2021
Included observations: 108
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INCOME_GENERATED__USD_	-0.060237	0.038959	-1.546185	0.1251
C	110.8190	76.39976	1.450514	0.1499
RESID(-1)	0.986418	0.097534	10.11354	0.0000
RESID(-2)	-0.121068	0.099872	-1.212223	0.2282

R-squared	0.772014	Mean dependent var	-6.74E-11
Adjusted R-squared	0.765437	S.D. dependent var	638.6817
S.E. of regression	309.3243	Akaike info criterion	14.34299
Sum squared resid	9950878.	Schwarz criterion	14.44233
Log likelihood	-770.5215	Hannan-Quinn criter.	14.38327
F-statistic	117.3893	Durbin-Watson stat	1.888503
Prob(F-statistic)	0.000000		

- **Seasonality Test**

Date: 11/20/22 Time: 16:55
Sample (adjusted): 1/31/2013 12/31/2021
Included observations: 108 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.864	0.864	82.924	0.000
		2	0.714	-0.131	140.00	0.000
		3	0.548	-0.145	173.99	0.000
		4	0.440	0.133	196.11	0.000
		5	0.357	0.007	210.85	0.000
		6	0.319	0.077	222.72	0.000
		7	0.291	0.013	232.69	0.000
		8	0.299	0.118	243.28	0.000
		9	0.300	0.003	254.06	0.000
		10	0.366	0.288	270.30	0.000
		11	0.430	0.087	292.94	0.000
		12	0.461	-0.096	319.21	0.000
		13	0.439	-0.039	343.27	0.000
		14	0.369	-0.112	360.51	0.000
		15	0.328	0.191	374.29	0.000
		16	0.307	0.050	386.47	0.000
		17	0.303	0.019	398.45	0.000
		18	0.275	-0.150	408.41	0.000
		19	0.248	0.031	416.61	0.000
		20	0.219	0.043	423.07	0.000
		21	0.216	0.016	429.47	0.000
		22	0.211	-0.058	435.61	0.000
		23	0.239	0.022	443.61	0.000
		24	0.244	0.018	452.04	0.000
		25	0.203	-0.170	457.96	0.000
		26	0.131	-0.061	460.43	0.000
		27	0.080	-0.018	461.36	0.000
		28	0.051	-0.011	461.75	0.000
		29	0.029	-0.074	461.88	0.000
		30	0.022	0.083	461.95	0.000
		31	0.009	-0.075	461.96	0.000
		32	-0.002	-0.054	461.96	0.000
		33	-0.013	-0.008	461.99	0.000
		34	0.026	0.140	462.10	0.000
		35	0.093	0.154	463.50	0.000
		36	0.156	0.000	467.53	0.000