BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARMENT OF STATISTICS AND MATHEMATICS

Time Series Analysis On Non-Performing Loans Of Peoples' Own Savings Bank (2022-2023)

BY
CHIKETAH MONALISA T
B201276B

A DISSERTATION SUBMITTED TO BINDURA UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE BACHELOR OF SCIENCE *HONOURS DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS*

SURPERVISOR: MS P HLUPO

JUNE 2024

APPROVAL FORM

The undersigned certify that they have read and recommended to the Bindura University of Science Education for acceptance of a dissertation entitled “TIME SERIES ANALYSIS ON NON-PERFORMING LOANS OF POSB” Submitted by CHIKETAH MONALISA T, Registration Number B201276B in partial fulfillment of the requirements for the Bachelor of Science Honor’s degree in Statistics and Financial Mathematics.

CHIKETAH MONALISA T………… ……………..10/06/2024……………..

B201276B Signature Date

Certified by

Ms P Hlupo…………… ……………12/06/2024……………

SUPERVISOR Signature Date

Certified by

MRM. MAGODORA … ………………………………………

Chairman of Department Signature Date

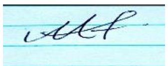
………………………………………………………………………………………………

DECLARATION FORM

I Chiketah Monalisa T hereby declare that this submission is my own work apart from the references of other people's work which has duly been acknowledged. I hereby declare that, this work has neither been presented in whole nor in part for any degree at this university or elsewhere.

Author: Chiketah Monalisa T

Registration Number: B201276B

Signature: 

Date: 10 June 2024

DEDICATION

This dissertation is devoted to my father Mr. J Chiketah and my late mom Mrs. S Chiketah.

ACKNOWLEDGEMENTS

First and foremost l would like to give thanks to the almighty for gift of life and for helping me embrace this journey. I would also like to give gratitude to my supervisor, Ms. P Hlupo for her academic input, for her unwavering academic guidance, which significantly contributed to the successful completion of this project. Additionally, l would like to express my gratitude to all staff members of the department of Statistics and Financial Mathematics for the moral support throughout the duration of the course. I am indebted to my friend Brandon Holandi, Alex Manjoro and Denzel Gombarago for their encouragement and companionship. Special gratitude is reserved for my family, whose enduring patience, financial assistance and technical expertise have been invaluable throughout my life.

ABSTRACT

N on-Performing loans threatens the stability and profitability of financial institutions. This research utilize a complete time series analysis on NPLs of POSB, which aims to recognize trends, patterns and potential causes of NPLs over the period 2022 to 2023. This research employs ARIMA model forecasting models and trend analysis. The adjusted R^2, Sigma volatility, Akaike information criterion and Bayesian information criterion analytical tools were used to evaluate the reliability of the model. The diagnostic analysis indicated that the ARIMA (3,2,1) model is the most applicable for forecasting NPLs at POSB as determined by the AIC. The researcher would recommend financial institutions like POSB to implement proactive measures to control NPL risk, enhance loan portfolio management and maintain financial health. The results of this research add to the established knowledge base in financial risk management and offer valuable insights in for policy makers, regulators and practitioners within POSB. Additionally, the methodology and insights displayed can be adapted and applied to similar studies in other financial institutions, adopting a more robust understanding of NPL trends so that they can decide on appropriate risk control strategy in the light of forecast made.

Keywords: AR, ARIMA, ADF, Non-Performing Loans, Forecasting, MA

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

NPL NON-PERFORMING LOAN

POSB PEOPLES OWN SAVINGS BANK

ARIMA AUTORESSIVE INTEGRATED MOVING AVERAGE

AR AUTOREGRESSIVE MODEL

MA MOVING AVERAGE MODEL

ARMA AUTOREGRESSIVE MOVING AVERAGE

SAMA SEASONAL AUTOREGRESSIVE MOVING AVERAGE

ACF AUTOCORRELATION FUNCTION

PACF PARTIAL AUTOCORRELATION FUNCTION

ADF AUGMENTED DICKEY-FULLER

CHAPTER 1: INTRODUCTION

1.0 Introduction

An increasing percentage of non-performing loans in the banking industry could affect the financial stability, delay the flow of funds from savers and borrowers and potentially reduce investment, impacting long-term economic growth. Therefore, it is important to point out the factors behind non-performing loans and understand their influence on future non-performing loan levels. As for POSB, NPLs could be higher because of the perceived risk which could result from some client's features and adverse business conditions that the bank could be exposed to. Hull (2007) explains that one of the basic formations of every organization, most importantly a banker is to understand the portfolio of risk it faces currently and the risks it plans to take in the future. The performance of the financial institutions of Zimbabwe in the previous years was constantly poor due to the fact that most of the loans were non-performing and the loan defaulters were not a small number.

The introductory chapter of this dissertation provides an overview of the context and purpose of research. It explains that the study aims to investigate the time series analysis on NPLs in Zimbabwe, with a specific focus on the People's Own Savings Bank (POSB). The introductory part of the paper states the reasons why the study is conducted in this particular area and explains the purpose and significance of the present research findings for NPLs in POSB, Zimbabwe. "Statement of the problem" is the next sentence, specifically arguing that the influence of these techniques on credit quality is worth thorough investigation. The objectives of the study are described, describing why research is being conducted and what is hoped to be achieved or what is being done. The chapter introduces the project of study and shows the area of concern thought to be the range and the limitations within which the study is to be carried out. Along with the assumptions put in the study, the researcher would also present the limitations of the research. Subsequently, the chapter finishes by giving the definitions of terms that were frequently used during the entire research. This is done just to give the readers a clear picture of entire research.

1.1 Background of the study

Financial institutions, particularly banks, play a linchpin role in the economy, serving as intermediaries between savers and borrowers. On the other hand, at different times, financial institutions have been faced with closures as unreliable credit approaches are used to achieve set targets by the concerned authority. Strischek (2017) underlines the influence of credit culture that performs the function of linking the name of the credit process and credit discipline. For financial institutions to have credit success, they need to carry out proper corporate credit appraisal, make sure that loan distribution is rightly done. They must be able to monitor adequately and also allow timely repayment.

In the years leading up to the global financial crisis of 2009 to 2015, the credit assessment of loan portfolios in most countries remained relatively stable. The financial crisis of 2007-2009 substantially destabilized world economy which posed a serious threat to the stability of borrowers' funds (Haslem, 2018). It is crucial to analyze the main reasons of default by debtors that lead to the creation of non-performing loans. Often, those failures can be traced to bad asset management or financial instruments like derivative contracts, as was the case of Baring Bank who lost everything in dealing in futures and forwards. The lively entrepreneurs, Watyoka and Muponda, in 2003 found themselves going the wrong way when their private asset management company, ENG Capital Investment, could not pay their creditors and was placed under voluntary liquidation.

Poor loan quality often stems from weaknesses in the information processing mechanism throughout the various stages of loan processing, from application to approval, monitoring, and control. This holds true especially where such lender's credit risk management guidelines, policies, and procedures are either missing or weak, or not appropriately customized. Lenders assess credit risk by gathering information on various factors such as collateral, guarantors, and source deductions, memorandums of agreements, life insurance, and company-guaranteed loan schemes. The precision with which credit risk is evaluated not only impacts the benefit of loans but also the overall profitability of the bank.

Debt culture reinforces the bank’s objectives for managing credit risk and credit protocol integrated with the business strategy to achieve them. Today banks draft credit policies which are formally written out rules and regulations, they provide the framework within which credit is approved, the rating system used for loans, the tracking and managing of loans and finally determining if loans are potentially bad loans (Basel Committee, 2000). Credit scrutiny is the key aim of the banks when they make a decision to provide loans to the applicants. Non-Performing loans represents a portion of the bank's financial assets from which the bank no longer receives interests or installment payments as per agreed (Haneef et al., 2012). Following regulatory guidelines, every bank must set aside provisions to address these NPLs, known as reserves, to mitigate potential credit payment losses. As the delay in payments increase, the necessary provisions escalates. These provisions constitute a cost to the bank, impacting its profitability and ultimately its capital or the projected requirement for additional capital.

The problem of a loan default, or non-performing loan (NPL), for banks also leads to the instability of the bank. Although credit-analysis practices exists to help banks to minimize the chances of having non-performing loans, the number of these kinds of loans is still on the rise. Unlike the finance corporations, the primary role of a bank is to guarantee the integrity of the customers' funds and the bank's own financial stability. This comprises examination of the market, the management, the technical factors, and the financial ability of the applicants for a loan.

A problem which occurs is when the banks have non-performing loans that the latter fail to recover in a long time period. Through this action, banks suffer in more than just the process of discarding such loans' nominal value. The act of keeping on the books those loans that aren't being serviced, or other idle assets, would have otherwise earned returns during this period, and therefore the losses from those potential returns would fall under the opportunity cost. The project feasibility is assessed by evidence of servicing loans in accordance with the firm's funding capability.

Briefly, banks and other credit institutions which are basically the financial institutes are essential for the economy. The problem of non-performing loans poses a big threat to the capital adequacy and sustainability of commercial banks, which is why it is necessary to know and to solve the issues behind loan defaults. Increasing incorporation of credit analysis tools, building awareness of credit culture and implementing credit risk management guidelines would contribute to solving problems with non-performing loans. The timely disposal of non-performing loans is very imperative to minimizing costs and maximizing gains to banks which would at the end maintain their safety against bankruptcy and the profitability of the banking industry.

1.2 Statement of problem

Over the years, it has become evident that inadequate credit analysis and poor judgment in loan applications have led to the occurrence of Non-Performing Loans. POSB has experienced a growth in non-performing loans, which has had a negative impact on their overall loan performance. As a result of these non-performing loans, POSB is now encountering steep expenses for bad debts, limited financial resources, decreased credit expansion, funding difficulties , reduced profitability, and a decreased market share.

1.3 Research Objectives

The objectives of the research are:

- To analyze trends and patterns in non-performing loans (NPLs) at POSB over the study period
- To develop predictive model (ARIMA) to forecast future NPL

1.4 Research Question

To achieve the research objectives, the following questions would guide the study:

- What are the trends and patterns in NPLS at POSB from 2022-2023?
- Will the NPLs in Zimbabwe be stationary or not stationary?

1.5 Scope of the study

This study focuses specifically on the time series analysis on NPLs of POSB Zimbabwe. The research covers from 2022 up to 2023 and utilize secondary data sources. The geographical scope of the study is limited to Zimbabwe, and the findings may not be directly generalizable to other countries or banking institutions. POSB was the focal point due to transport setbacks, work commitments and time constraints.

1.6 Significance of the study

This research contributes to the existing body of knowledge on NPLs. The findings of this study provide valuable insights commercial banks regulators and policy makers in Zimbabwe to reduce and improve in decision making as the series analysis NPLs aiding in making informed lending decisions. This research opts to add to the existing literature on non-performing loans. The results of this topic can be utilized as a source of reference by other students who desire to perform research on non-performing loans.

1.7 Assumption of the study

1. The research period would maintain a consistent and unchanged environment.
2. The data used for the study is accurate

1.8 Limitations of the study

The study was limited in accessing some of the detailed information about the use of the historical NPL data for analysis due to the sensitivity and confidentiality of the information sought.

1.9 Definition of terms

Time Series is defined as the sequence of numerical figures generated from sequential period with equal interval between them (Ruey, 2010)

Non-Performing Loans are known as loans overdue for at least 90days, which decrease the value of bank's loan portfolio (labbe, 2016)

ARIMA also known as Box- Jenkins models are a family of models used for time series analysis and forecasting, particularly in cases where the data exhibit non-stationary (Brockwell, P. J., & Davis, R. A., 2016)

1.10 Chapter Summary

In conclusion, this chapter has outlined the background, research problem, objectives, research questions, limitations, significance, and scope of the study. The subsequent chapters would delve into the literature review, methodology, findings, and conclusions to provide a thorough examination of the time series analytics on NPLs of POSB Zimbabwe. By conducting a detailed analysis, the study aims to provide valuable insights for the banks and contribute to the existing knowledge in the field of NPLs and profitability evaluation.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

The purpose of this chapter is to provide the reader with insights and critical analysis of existing literature regarding the time series analysis on NPL of POSB in Zimbabwe. This chapter contains a detailed literature review, focusing on topics like models in financial time series methods adapted by POSB, and box-jenkins methods and significance of these time series analysis to understand behaviour of NPLs over different time periods. The literature review would be attempting to look upon the views of scholars in textbooks through experimental evidence and theoretical perspective focusing on the time series analysis on NPLs.

2.1 Theoretical literature review

Non-Performing Loans

A loan is deemed non-performing when they remain overdue for a certain period, often 90 days or more, although specific criteria may differ (RBZ, 2015). Non-performing loans play a crucial role in examining the health of banks and financial systems. Banks with high levels of NPLs may face liquidity constraints, reduced profitability and higher capital requirements, impacting their ability to lend and support economic growth (Berger & Bouwman, 2013)

2.1.1 Time Series Analysis

George, Gwilym, & Gregory (2008), defined time series as series of sequentially recorded data. It includes the analysis of dynamic system characterized by inputs and outputs series, which aligns to a function. According to (Ramasubramanian, 2015), time series is a series of observations instructed in time. Time series can be enforced in fields like economics, business and engineering. If the set of opinions is continuous, the time series is said to be continuous if and only if the set is discrete (Ansah, 2014).

2.2 Autoregressive (AR) Model

According to George, Gwilym, & Gregory (2008) the Autoregressive model is a stochastic process that indicate a particular series of events over time. The Autoregressive model is the basic model for time series, it comes from the idea that any value of a time series can be forecasted based on the past recorded values. The AR (p) represents a stochastic process where the result is affected by prior inputs and a stochastic term, providing its stochastic differential equation. Essentially, Autoregressive models consider past steps when predicting the subsequent one. However, a drawback of the AR model is its vulnerability to persistent effects from temporary or single shocks affect. To deflect this, Autoregressive processes typically incorporate a lag value, deciding which previous steps apply more effect on the output. Moreover, the AR model's non stationarity can be represented by unit root variable.

2.3 Moving Averages (MA) Model

According to Pannerselvam (2005), the Moving Average model involves the seasonal components that have a particular quantity of randomness. The Moving Average model is inherently stationary unlike the Autoregressive model which is non-stationary. The MA requires a linear regression of the current value against the white noise or random shocks in the series, contrary to the AR model, which includes a linear regression to non-shock values. The MA model predicts a series based on the previous error in the series called error lags

2.4 Autoregressive Moving Average

ARMA is a model made up of AR (p) and MA (q), where p = number of considerable terms in auto correlation factor (ACF) and q = number of considerable terms in partial auto correlation factor

2.5 Autoregressive Integrated Moving Average

According to Yu, G. and Zhang, C (2004) ARIMA models are mathematically constructed as ARIMA (p,d,q) where p and q are similar to ARMA model but d = number of initial differences. The initial stage in implementing ARIMA model is to examine whether the time series is constant or not. ARIMA functions perfectly when data has a constant pattern over time, which means that data have to remain stable over the course of time. Therefore, when the data has a trend of moving upwards or downwards has a recognizable pattern (seasonality), then the data is stationary or not. In order to recognize ARIMA model, it is important to recognize ARMA model because it is considered as an extension to ARMA model.

2.6 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

According to Ruey (2010), a seasonal time series assign to financial time series for example quarterly earnings per share of a company that exhibits certain cyclical or periodic patterns.

SARIMA model were established to control seasonality in data. Seasonality can be controlled using stationarity at some point, but seasonal correlations cannot be dismissed completely. SARIMA are constructed as follows SARIMA (p,d,q)(P,D,Q)s where P = number of seasonal AR terms, D = number of seasonal differences, Q = number of seasonal MA terms and s = length of the season. Eliminating seasonality would assist the model to function better by removing seasonality in data is hard.

Schwarz-Bayesian Information Criterion (BIC)

The Bayesian information criterion is a technique used to choose a model from limited set of model, also known as the Schwarz information criterion. It is preferred to use the model with lowest BIC. It is crucial to note the BIC can only be used to compare estimated models if the numerical values of the dependent variable are accessible

Akaike Information Criterion (AIC)

According to Richard and McElreath (2016), the Akaike information criterion (AIC) in an estimator of forecasting error and assess the comparative effectiveness of statistical models for particular dataset. AIC calculated the quality of the statistical models in relation to other models, and can be used as a model selection method. It is rare for a model to depict the data generation process. The AIC evaluates the amount of information lost when the model represents the data, with lower loss indicating higher quality model.

2.7 Box-Jenkins Method

According to NCSS (2013), Box-Jenkins analysis is a systematic approach used to identify, fit, validate and utilize ARIMA models for time series data. This approach is suitable for examining time series with a medium to long duration and it also changes a non-stationary series to stationary. The box-jenkins consist of three steps namely; model identification, model estimation and diagnostic checking.

2.7.1 Model identification

According to J.D. Cryer &Chan (2008), Using assumption that there exist no seasonal variation, the purpose of the model identification stage is to pick figures of d being the initial pick and then lastly p and q in the ARIMA (p,d,q) model. Once the series shows a trend, we have two options to address the presence of deterministic trend in the series which are as follows fitting and removing it or differencing the series.

2.7.2 Model estimation

The predictable values of p, d and q, it is required to estimate phis and thetas. This model aligns with maximum likelihood estimation process outlined in Box-Jenkins. Many estimations methods of parameters are present, these involves relatively simple method of probability plotting, least squares and maximum likelihood

2.7.3 Diagnostic check

According to J.D. Cryer &Chan (2008), diagnostic model is disturbed with testing the goodness of fit of a model and if the fit is inadequate, recommending relevant alterations. Diagnostic checking is the last step, once a model has been fitted. The analysis is executed by studying the auto correlation figures of the residuals to check if further structure can be found. According to Ruey (2010), Forecasts are generated, and the model is considered adequate if all autocorrelations and partial autocorrelation are small. The procedure of adjusting the values of p and q, and examining the residuals until the outputting residuals contain no additional structure. The program may be used to make predicts and associated probability limits.

Forecasting

Predicting future values is a key purpose of time series modelling, allowing us to forecast unobserved data points. The aim is to build a model that can aacurately project future values for the series (Cryer & Chan, 2008). After carefully selecting and estimating model parameters, the model is used for making forecasts. Over time, users evaluate the model's effectivenessand short comings (Makeridakis, Hyndman, & Whellwright, 2007).

2.8 Empirical Review

Korankye, M.,Bright, D & Dunyoh., M (2022), utilized a research on the effect of Non-Performing loans on the profitability of Universal banks using a time series analysis of the Ghanaian Banking industry. The study indicates a negative correlation between Non-performing loans and profitability. An increment in the NPLs ratio leads to a decrease in the profitability of the universal bank by 140%. Moreover, the research identified a noteworthy adverse correlation between the NPLs and profitability of the universal bank. The relationship is statistically significant at a 5% confidence level. An increment in NPLs results in reduction in profit by 27%. The research concludes that NPLs exerts a significant negative impact on profitability of universal banks in Ghana. The impact of NPLs shows that the capacity of universal banks to gather sufficient profit is constrained in the Ghanaian banking industry due to escalating NPLs.

Mecaj., X & Sinaj,. V. (2022) conducted the time series model for Non-Performing loans, a case study of Albania. The research concluded that the evaluation and management of NPLs are essential for maintaining economic and financial stability in a country. Creating favourable conditions for a healthy NPL market adds bank’s ability to minimise a significant portion of their non-performing assets. Based on the assessment conducted using both scrutinized SARIMA model, along with the evaluated ARIMA model. It is evident that the trend of NPLs would experience a downfall throughout the year. This trend shows a positive outlook for the economy overall and holds particular benefits for the bank.

Chikuku, T. (2018), conducted a Time Series Analysis of Non-Performing Loans in the Zimbabwean Banking Sector. The study aimed to understand the trends and factors influencing Non-Performing Loans (NPLs) in the Zimbabwean banks. Time series analysis methods, such as Autoregressive Integrated Moving Average (ARIMA) models and Vector Autoregression (VAR) models, were used to analyze the historical data on NPLs. The research provided insights into the patterns of NPLs over time, identified key factors influencing NPL dynamics, and may offer forecasts or predictions for future NPL trends within the banking sector.

Matika, L. (2016), analyzed a research on Forecasting Non-Performing Loans in Commercial Banks: A Time Series Approach. The study focuses on developing forecasting models for Non-Performing Loans (NPLs) specifically within commercial banks in Zimbabwe. Advanced time series forecasting techniques, including ARIMA models, were employed to predict future NPL trends based on historical data. The research provided predictive models that can enhance commercial banks to anticipate and manage NPLs more effectively, thus improving risk management practices and financial stability.

Mlambo, C. (2017), conducted Macroeconomic Factors and Non-Performing Loans: A Time Series Analysis in Zimbabwe. The study investigates the relationship between macroeconomic factors and Non-Performing Loans (NPLs) in Zimbabwean banks. Time series regression analysis or Vector Autoregression (VAR) models were used to examine how changes in macroeconomic indicators (e.g., GDP growth, inflation rates) affect NPL dynamics over time. The research identified significant macroeconomic drivers of NPLs and provided insights into how changes in these factors affect the bank’s asset quality and financial stability.

Kanyenze, S. (2019), researched on Economic Downturns and Non-Performing Loans: Evidence from Zimbabwe's Financial Sector. The study explores the impact of economic downturns on Non-Performing Loans (NPLs) within Zimbabwe's financial sector. Time series analysis techniques, such as ARIMA and GARCH models, were utilized to analyze the relationship between economic downturns and NPL trends. The research likely provides evidence of how economic downturns affect NPLs, including insights into the timing, magnitude, and duration of NPL increases during periods of economic instability.

Madondo, F. (2015) conducted Time Series Modelling of Non-Performing Loans: Evidence from Microfinance Institutions in Zimbabwe. The study focuses on modelling Non-Performing Loans (NPLs) within microfinance institutions in Zimbabwe. Time series modelling techniques, such as ARIMA and exponential smoothing models, were used to analyze NPL trends and develop predictive models. The research provided insights into the unique characteristics of NPLs within microfinance institutions, including factors influencing NPL dynamics and strategies for mitigating NPL risks within this sector.

2.9 Research gap

This research identifies the gap of the final results of the studies of different countries so as to produce a valuable contribution. The southern region has received comparatively less attention especially Zimbabwe, not much research has been done on the time series analysis on NPLs. The majority of the studies focuses mainly on the correlation between NPLs and profitability using the regression model and concluded that it is the best model to use which leaves a huge gap. On the flip side, the researcher feels constricted because not so many researches has been done and there is no or limited references to assess the performance of ARIMA model in terms of accuracy forecasts. This study shows uniqueness, from other researches in several aspects and adds to current literature body of knowledge in following ways. Firstly, the current trends of NPLs, secondly the aim of the research is identifying trends, forecasting future NPL levels, understanding the relationship with economic indicators and devising strategies for risk management and mitigation.

2.10 Chapter Summary

Time series on NPLs involves studying the historical data of non-performing loans over time to identify patterns, trends, and potential future developments.it typically includes techniques such as forecasting, trend analysis, seasonality, and anomaly detection to understand the behaviour of NPLs over different time periods. ARIMA models are certainly useful for checking, capturing and predicting temporal dependencies and trends present in data. The ARIMA model consist of three main components namely; Autoregressive model (AR) these exhibits the correlation between the present observation and an identified amount of lagged observations, Integrated (I) this accounts for differencing, which is used to make time series stationary, Moving Average (MA) these capture the relationship between the current observation and residual errors from moving average model applied to lagged observation. By analysing the parameters of the ARIMA model and fitting it to historical NPL data forecasts can be generated to predict future NPL levels, helping financial institutions and policy makers make informed decisions regarding risk management and NPL mitigation strategies.

CHAPTER 3: RESEARCH METHODOLOGY

3.0 Introduction

This chapter focuses on the methodological approach in this study, giving the essential guide to the proper analyses of Non-Performing Loans, or NPLs, in the context of bank. Hence, this chapter deliberates the research method used, the data used, the population of interest, the sampling technique used, the tools used in the study, as well as the data collection procedures implemented in the research study. Furthermore, it expounds on the conceptual definition of the variables and anticipated interaction mechanisms, thereby explaining the complexity of NPLs. Furthermore, the chapter also highlights diagnostic tests, analytical models and model validation tests crucial in arriving at conclusions from the data collateral. The structural elements include ethical considerations, which are used to minimize misconduct of research and protect participants' rights and interests. Thus, this study hopes to make significant findings that advance the field of financial risk management and micro financial procedures through the use of a research methodology and ethical considerations.

3.1 Research Design

In terms of the design of the current research the use of time series analysis of the non-performing loans (NPLs) necessitates quantitative research design. As pointed out by Box and Jenkins (1976), time series analysis presents a robust model for exploring temporal changes in variables, thus providing an appropriate approach to understanding of NPLs fluctuation. Using ARIMA model, which is the most common technique utilized in time series analysis, the study analyses historical NPLs' data to make some prediction regarding the future trends, (Box & Jenkins, 1976; Enders, 2014). Therefore, the use of quantitative methods and statistical techniques would help to offer a dependable research design that serves the objective of the study by offering solid proof regarding the behaviours of NPLs within the context of the bank as well as aid in risk management practices, decisions making as well as policy formulation in the financial industry. Also, the research design incorporates main considerations on data accuracy, model assessment, and evaluation of model robustness, making the study results believable and accurate. Cohort characteristics and panel sample measurement are of paramount importance that is going to help achieve credible results and bring new valuable knowledge to the area of banking and finance.

3.2 Data Sources and method of data collection

The data used in this study were retrieved from secondary sources originating from the company's database. Data was obtained from the company's records. Therefore, this study adopted the following advantages related to data collection and analysis: Data consistency and reliability given by the avoidance of primary data collection Data accessibility and minimization of costs and logistical difficulties related to primary data collection due to its indirect collection through secondary sources The limitation of the study, on the other hand, was that due to the extraction of the NPL variable for analysis from the company's accessible database, only secondary data was used and hence did not.

3.3 Target population

3.3.1Target Population

According to Sekaran and Bougie (2016), in any research process, the choice of target population helps to enhance the generalizability and transferability of the research findings towards the intended settings. This research study is a case study which targets POSB bank. In targeting the banking institution, the study is going to present empirical evidence on the profile of NPLs as well as identifying the trends and determinants within the bank industry, thus contributes to towards the strategic planning, policy making and manages the risk of non-performing loans. Also, the selection of banking institutions as the target population provides the reliability of relevant data to be obtained and used in the evaluation of overall research objectives concerning the dynamics of NPLs.

3.5 Research Instruments

SPSS, a statistical software widely used for data analysis, was used in this study to analyze data involving descriptive analysis, exploratory data analysis, data cleaning, model estimation, and diagnostic testing (SPSS, 2020). Hence, through various features inherent to R-Studio and through statistical and econometric tests available in this framework, the researchers were able to analyze the characteristics of non-performing loans (NPLs) and their evolution in the microfinance sector efficiently. Consequently, the use of secondary data from the company's database along with R-Studio for analysis allowed for ease and efficiency in the access of comprehensive NPLs data and, in doing so, the study met its research objectives as outlined in this paper.

Methodology

Used in this study, the following is the research methodology that seeks to meet the goal and objectives timely, first particular step is to assemble accurate and complete data to be acquired, NPLs historical information must be collected from the trustworthy source. Secondly, the collected data have some problems that require data pre-processing beforehand; they include missing values, outliers, and inconsistent data, according to the recommendations provided by the researcher. The pre-process involved enhancements that are meant to make the data better suited for modelling as well as the removal of any weak or unsuitable data. After pre-processing, model identification takes place to identify the right number of parameters of the autoregressive (p), differencing (d), and moving average (q) parameters of the ARIMA model. Model identification involves the utilization of Autocorrelation and partial Autocorrelation functions to determine temporal behaviour and relationship of the NPLs data. When identified, the ARIMA remains tested using maximum likelihood estimation (MLE) to evaluate the coefficients of the model components. Basically, estimation involves adjusting the parameters of the developed model to match pre-processed data on NPLs, repeatedly until the model returns values that most closely resemble actual values. Afterwards, diagnostic tests are taken to check whether the selected ARIMA model is satisfactory or not. Modelling confirmations include inspecting diagnostic plots, using statistical tests for model adequacy, and validating forecast precision such as cross-validation. Lastly, a post-regression analysis is made by interacting the estimated coefficients and the forecasted values from the model to hypothesize on factors that could have led to the formation of the phenomenon denoted by the abbreviation NPLs; this follows cross- sectional time series analysis guidelines. Methodologically, the procedures conform to standard practice for using time series analysis with the ARIMA model, which maintains the scientific integrity of the analysis of NPLs trends.

3. 6 Data presentation and analysis procedures

In this study, data presentation and analysis processes to achieve systematic and meaningful approach in identifying and evaluating Non-performing loans (NPLs) behaviours in the Microfinance Sector in Gokwe. Consistent with the recommendations by Hair et al. (2015), the data have been sorted in a concise and easy to understand format to enhance the interpretation of the results.

3.7 Description of variables and expected trend (signs)

The signs expected in this research are indicated below Co-variable Variables Co-Variables Sign Names expected in this research Th1 New knowledge The participants are expected to produce new knowledge gained from the study as a result of the research Breathability Th1 Thermal comfort The participants are expected to produce new knowledge of breathability and thermal comfort gained from the study as a result of;

3.6.1 Data Presentation and Analysis

Data Preparation

Data entry

This involves inputting the raw data into a digital format, such as a database. The data includes information on Non-performing loans percentage and months.

Data coding

Data validation is crucial to ensure accuracy and reliability. This step involves checking for errors, inconsistencies, or missing values in the data. It may require cross-referencing the data with other sources or conducting checks to verify the integrity of the data entry process. Any discrepancies or issues should be addressed before proceeding with the analysis.

Data presentation

Tables are used to present the numerical characteristics of NPLs, while using time series plots and line graphs would facilitate the assessment of patterns in NPLs by their temporal trend and seasonality.

3.8 Data analysis procedures

1. Diagnostic Tests

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) Analysis

An attempt to check for autocorrelation in time series data is very important to assess the validity, as well as reliability of the results obtained from the statistical analysis (Brockwell &

Davis, 2016). Autocorrelation is a substantial violation of the assumptions that are customary to hold the observations as independent, which is a central condition most models (Brockwell & Davis, 2016). Autocorrelation testing enables scholars to determine whether alterations of the variables being assessed should be made to reduce or eliminate the related structure of the compounded data (Chatfield, 2016). In the meanwhile, if autocorrelation happens, it also points out that there are some unmeasured variables that cause this problem or there are some missing elements which are left out of the analysis (Enders, 2010). The correlation of data over time means that the observations in the series are not independent, and this has the effect of causing unsatisfactory forecasts when applied on a time series model (Cryer & Chan, 2008). Therefore, when a model has autocorrelation, researchers can improve the confidence of statistical inference and credibility of conclusions made from the analysis by testing for the issue and fixing it using techniques such as differencing (Shumway & Stoffer, 2016, Hamilton, 1994). Undeniably, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are useful techniques that used to test the presence of autocorrelation in time series data at some particular lag values. Here's how ACF and PACF analyses help in this process: ACF quantifies the dependence between two observations in a time series separated by a multiple of points, which is helpful in determining the strength of the prior observation on future values (Brockwell & Davis, 2016). PACF enables one to look at the observations at two or variously lagged time points with a view to determining the direct impact of earlier observations on the subsequent values and this it does while accommodating for intermediate observations (Brockwell & Davis, 2016). Any analysis in the context of time series is important in the diagnostics of adequacy of the chosen model and making necessary modification on the basis of specific patterns of data autocorrelation.

Goodness of fit test

This test assesses the overall goodness-of-fit of the autoregressive integrated moving average (ARIMA) model by examining whether the autocorrelations in the residuals are statistically significant.

Stationarity Test

Stationarity is a central concept in time series analysis which assures that the properties of a time series do not shift continuously. A stationary time series is a series for which mean, variance, and auto covariance values do not change with time; such a series is relatively easy to model and to forecast. However, non-stationary time series give dissimilar lessons to learn at different points; they exhibit changing trends, seasonality or any other pattern, hence the traditional models can hardly work. The Augmented Dickey-Fuller (ADF) test in particular determines whether or not a variable is stationary by analysing the estimated value for the unit root. The null position of the ADF test is that there exists a unit root hence the series is non-stationary while the alternative position means that it is stationary. The null hypothesis can be rejected if the value of the test statistic is further than the critical values at the specified level of significance, thus implying stationarity. On the other hand, failure to reject the null hypothesis might mean non-stationarity and hence requires the variable to undergo differencing in order to achieve stationarity before modelling. Therefore, the ADF test appears as the approach for determining stationarity and offering guidance on the modelling of time series data.

Heteroscedasticity Tests

Heteroscedasticity can indicate varying levels of risk or volatility over time. Detecting or correcting for heteroscedasticity is crucial for accurate forecasting and modelling. Techniques like robust standard error or transforming the data can help address the issue. Since NPL data may likely exhibit heteroscedasticity, the White test remains as a significant diagnostic test for its detection and remedial actions. The White test determines if the spread of residuals for the model is same at all time periods which is also called homoscedasticity (Behson & Free, 2011). White test results pointing to heteroscedasticity: Conclusions based on significant test outcomes of the homoscedasticity might present heteroscedasticity in the data.

Through identifying heteroscedasticity, the analysts get a chance to make the models more accurate by having narrow variances of residuals based on the values of the independent variable. Extending the model by accounting for heteroscedasticity of error term empowers a researcher not only to obtain more accurate parameter estimates of the adjusted model but also more reliable model predictions needed for statistical analysis and credit risk management and financial stability assessment as well.

Analytical Model

ARIMA Modelling

The ARIMA model proves a reliable quantitative technique for examining a set of data particularly data on non-performing loans within microfinance operations (Box and Jenkins, 1976). Being introduced as ARIMA(p, d, q) model, it is an extension of autoregressive integrated moving average model and is convenient for structuring time series data analysis (Enders, 2014).

$$\gamma_{\mathrm{t}} = c + \varepsilon_{\mathrm{t}} + \theta_1 \varepsilon_{\mathrm{t-1}} + \theta_2 \varepsilon_{\mathrm{t-2}} + \cdots + \theta \varepsilon_{\mathrm{t-}} q$$

The various parameters used in the FGN model include p, d and q where p stood for autoregressive order with the following equation:

$$\gamma_{\mathrm{t}} = c + \phi_1 y_{\mathrm{t-1}} + \phi_2\, y_{\mathrm{t-2}} + \cdots + \phi_{\mathrm{p}} y_{\mathrm{t-p}} + \varepsilon_{\mathrm{t}}$$

d for degree of differencing and q moving average order respectively. The equation for MA equation is as follows:

$$y'_{\mathrm{t}} = c + \phi_1 y'_{\mathrm{t-1}} + \phi_{\mathrm{p}} y'_{\mathrm{t-p}} + \phi_{\mathrm{p}} \varepsilon_{\mathrm{t-p}} + \varepsilon_{\mathrm{t}}$$

These combinations enable the ARIMA model to forecast future values of the time series; the selection of parameters is done using test like the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) (Pankratz, 2012). In the following sections, based on the data of historical non-performing loans, the applicability and contribution of the ARIMA model in identifying potential trends and patterns to financial risk decision-making and management are demonstrated (Hyndman & Athanasopoulos, 2018). Altogether, the stylized facts analysis and the ARIMA modelling strategies can be useful in shedding light on time series properties and trends of NPLs in microfinance segment as well as in generating actionable insights for the future.

The ARIMA modeling procedure would therefore follow methodological steps that include:

Model identification

Estimation

Diagnostic Checking

Forecasting a series

Model Validation (Fitness) Tests

The model validation (fitness) that is Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Forecast Error Variance Decomposition (FEVD) acts as a benchmark measurement for evaluating the accuracy, goodness-of-fit, and precision of the ARIMA model. MAE and RMSE are statistical estimates of the overall prevalence and size of the prediction errors compared to the actual values, with better model performance denoted by lower scores. The lower the AIC and BIC values, the better the overall model fit measured in terms of the difference between the models and the number of parameters in the models.

3.9 Ethical considerations

Never the less, ethical consideration is a very big factor that is given credit in every step of research especially in for research that involves sensitive data such as non-performing loans (NPLs) within the bank. The principles which are used to ensure ethical practice regard the following principles in this particular research privacy and confidentiality. Protection of data and information that identifying the bank and its clients is highly valued thus efforts to anonymize the data are put in place (Garcia et alTo ensure all data collected is legal, consent is sought from the various institutions involved in the research in line with the participants' information rights that include explaining to them the intended purpose of the research and how the data collected would be used (Davis, 2018). Furthermore, safety measures of regulatory and ethical standards are maintained when, for example, the study requires approval from the institutional review board and follows national or regional current legislation regarding the protection of data and research ethics (Taylor, 2020). The principles of open and accurate reporting are upheld in the given research as the detailed procedures are recorded along with the declared conflict of interest and any possible biasness (Clark & Evans, 2017). Furthermore, personal rights and density applicable to participants are also safeguarded such that the negative impact as influenced by the experiences ensuing from the disclosed undertaking is kept to the bare minimum (Miller & Martinez, 2015). In this regard, maintaining the following principles of ethical research, the study is going to ensure that research is done for the benefit of society, and in a manner that can contribute new knowledge about and improved practices in financial risk management and microfinance.

3.9 Chapter summary

In sum, based on considerations depicted in this chapter, four core ethical issues pertinent to researching NPLs in the microfinance context are derived. Since it is irresponsible to use data from participants' surveys irresponsibly, this study complies with ethical measures including integrity, confidential disclosure, and respect for participants' rights. Data security, FERPA privacy and informed consent, and federal and institutional ethical requirements are crucial steps necessary to preserve the research validity. First, participant privacy and voluntary consent, as well as public report and clarity of method, all contribute significantly to reducing risk and upholding research integrity. In so doing, this study would seek to achieve the following objectives with respect to the subject area: To make a valuable contribution to knowledge in the area of financial risk management and microfinance, as well as respecting the rights of all the participants involved in this study.

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.0 Introduction

In this chapter the researcher conducts a thorough assessment and analysis of the collected data, followed by the presentation and discussion of findings. The chapter begins by providing descriptive statistics for the study variables, along with pre-test including the ADF test, the analysis is performed using Rstudio. The results obtained enable us to address the research questions regarding the relationship between the selected NPLs at POSB.

4.1 Descriptive Statistics

Table 4. 1 Descriptive analysis

Kurtosis	**Skewness**	**Jarque-bera**	**minimum**	**Maximum**	**Mean**	**Median**	**Std. Deviation**	**Variance**
0.5000878336	1.121331586	8.640076479	2301211.8	3024240488	761757650.9	84087827.2	1120862761	1.25633E+18

The descriptive statistics show some measures of central tendency and variability of the non-performing loans (NPLs) data employed in the study. Kurtosis measure the peakedness of a distribution. With a kurtosis value of -0. 500087836 this indicates a flatter distribution. skewness measures the asymmetry of a distribution. The skewness coefficient is positive in value and is equal to 1. 121331586 indicates a positive skewness which illustrates the distribution has a long tail towards the right side than the left. Hence, the Jarque-Bera test assesses if data have skewness and kurtosis matching normal distribution of 8. The closer the value is to 0, the closer to the normality, while the larger the value such as 640076479, the farther it is from the normality assumption. The span between the minimum (2301211. It also shows the dispersion of the NPLs data with minimum value (8) and maximum value (3024240488) as well as mean value (761757650. 9) and median (84087827. 2) gives information about central tendency. The standard deviation measures the dispersion of data points from the mean. The variance measures how far each number in the data set is from the mean, on average. These descriptive statistics help address other modeling considerations and inform interpretations of the NPLs data for financial risk management and decision-making.

4.2 Pre-tests /Diagnostic tests

Testing for Stationarity

Non-performing loans Trend analysis

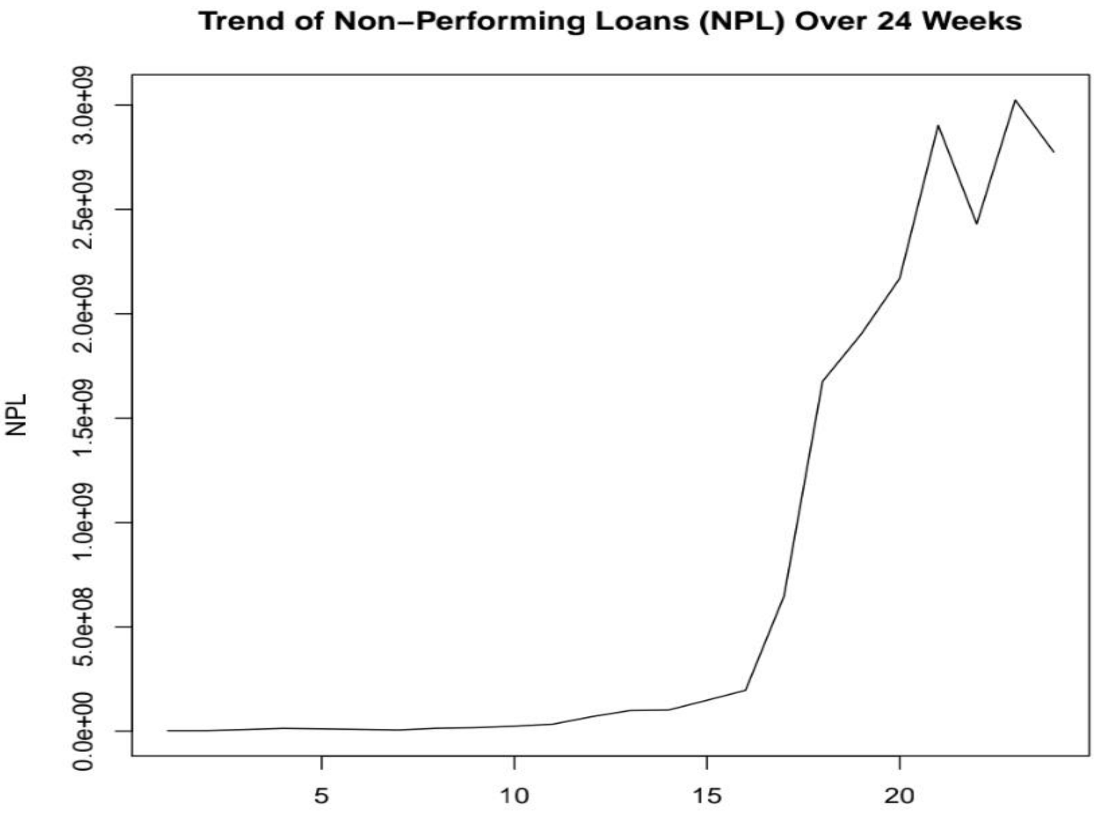
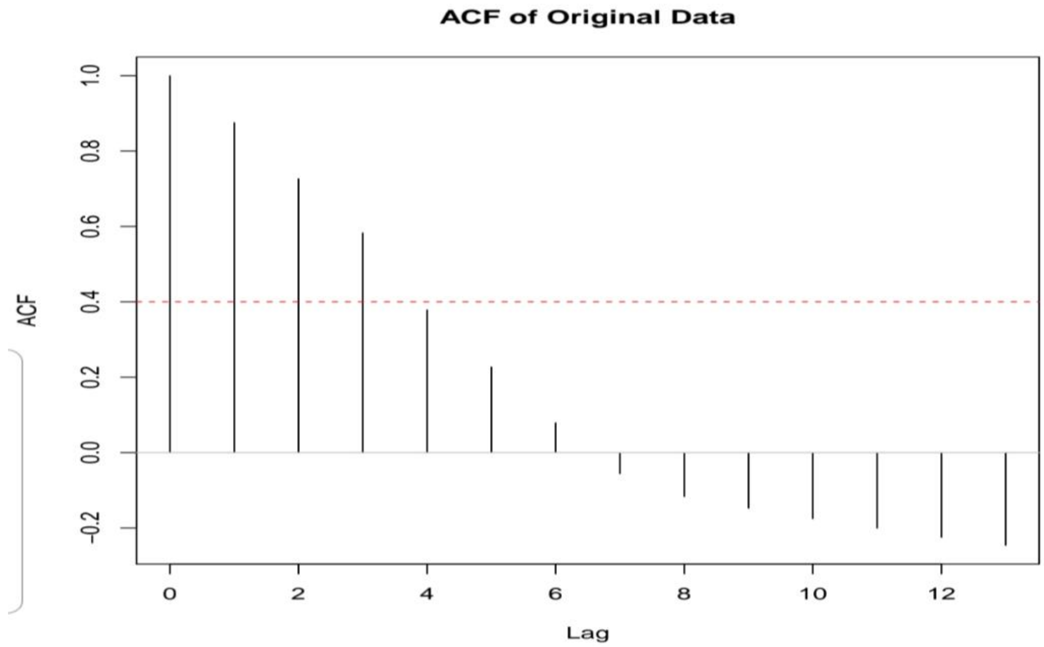


Figure 4. 1 time series plots of raw data

Fig. above indicate that NPL has significantly rise over the 24 months period and a sharp rise is observed from month 1 and month 24. The NPLs are initially below average in the initial weeks and gradually rise upwards in the later weeks, indicating a probable gradual rise in the non-performing loans in the following period. This upward trend shows that there is need to pay attention to non-performing loans by closely rile them to ensure that there are no great losses and the stability of the financial institution or the sector is enhanced. However, the trend plot of non-performing loans (NPL) also clearly shows evidence of non-stationarity beyond a single indication. Another type of non-linearity that could be seen on the data is non-stationary in time series data wherein the statistical characteristics of the series may fluctuate over time.

Concerning NPL, the general trend emerged from the data makes it possible to indicate a systematic character of the change in the behavior of NPL over the observed period.



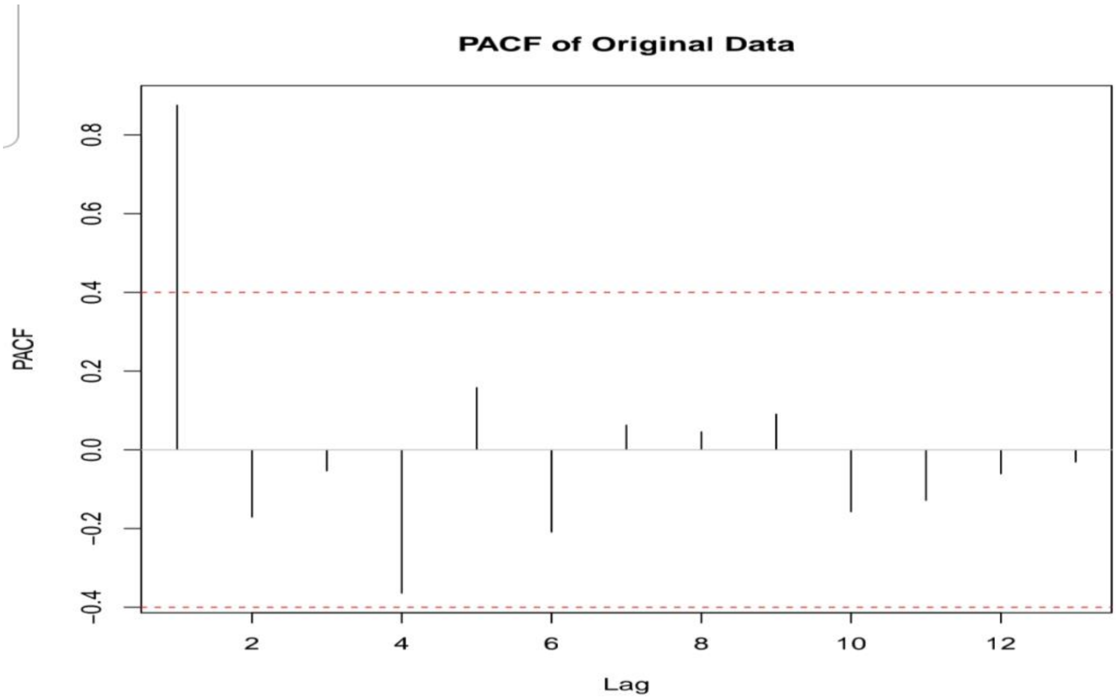


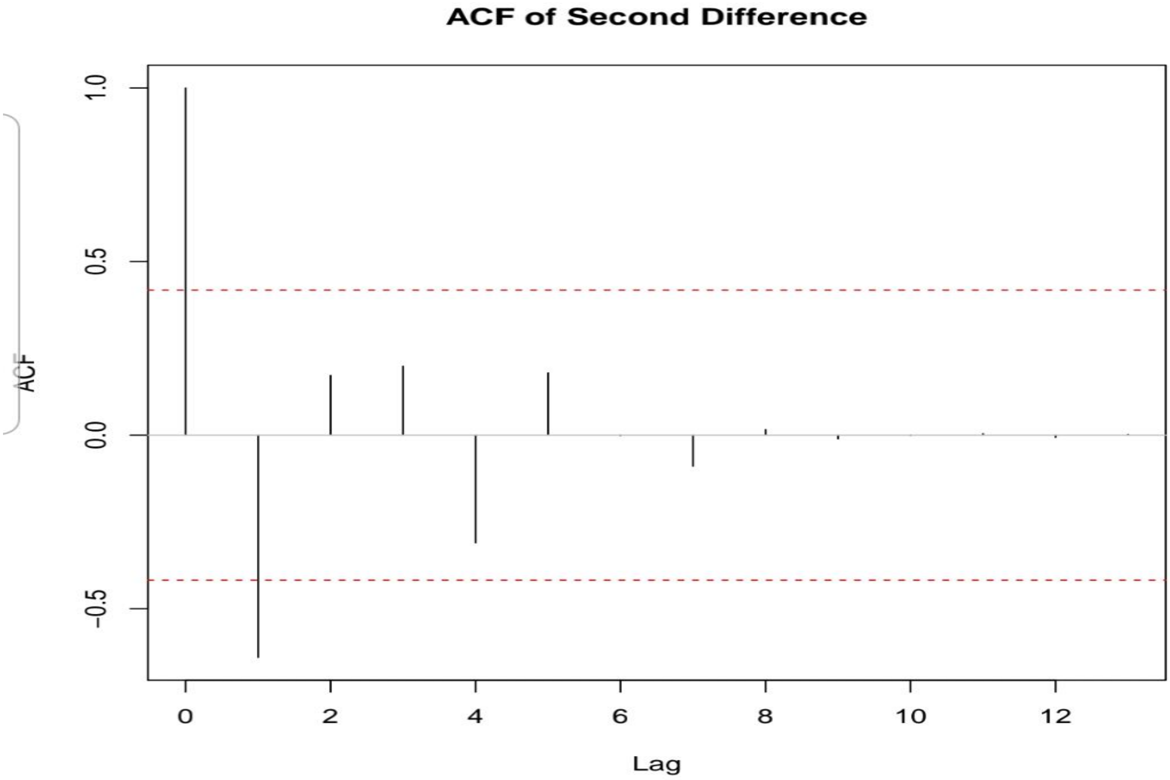
Figure 4. 2 ACF and PACF plots of Raw Data

ADF test for NPL series

Table 4. 2 ADF test for Trend Stationarity

	ADF statistic	P- value	1% Critical value	5% Critical value	10% Critical value	Conclusion
diff(NPL)	1.3149	0.4092	-2.66	-1.95	-1.6	Non Stationary

As shown by the above table or the NPL Series, the p-value (0.4092) is greater than the typical significance level of 0.05. Therefore, we fail to reject the null hypothesis. The series is non-stationary. If the data is non-stationary, it suggests that the properties of NPLs, such as their mean, variance, or autocorrelation, change over time, making it difficult to predict future trends or analyze patterns effectively.


ACF of Second Difference
ACF
1.0
0.5
0.0
−0.5
0
2
4
6
8
10
12
Lag

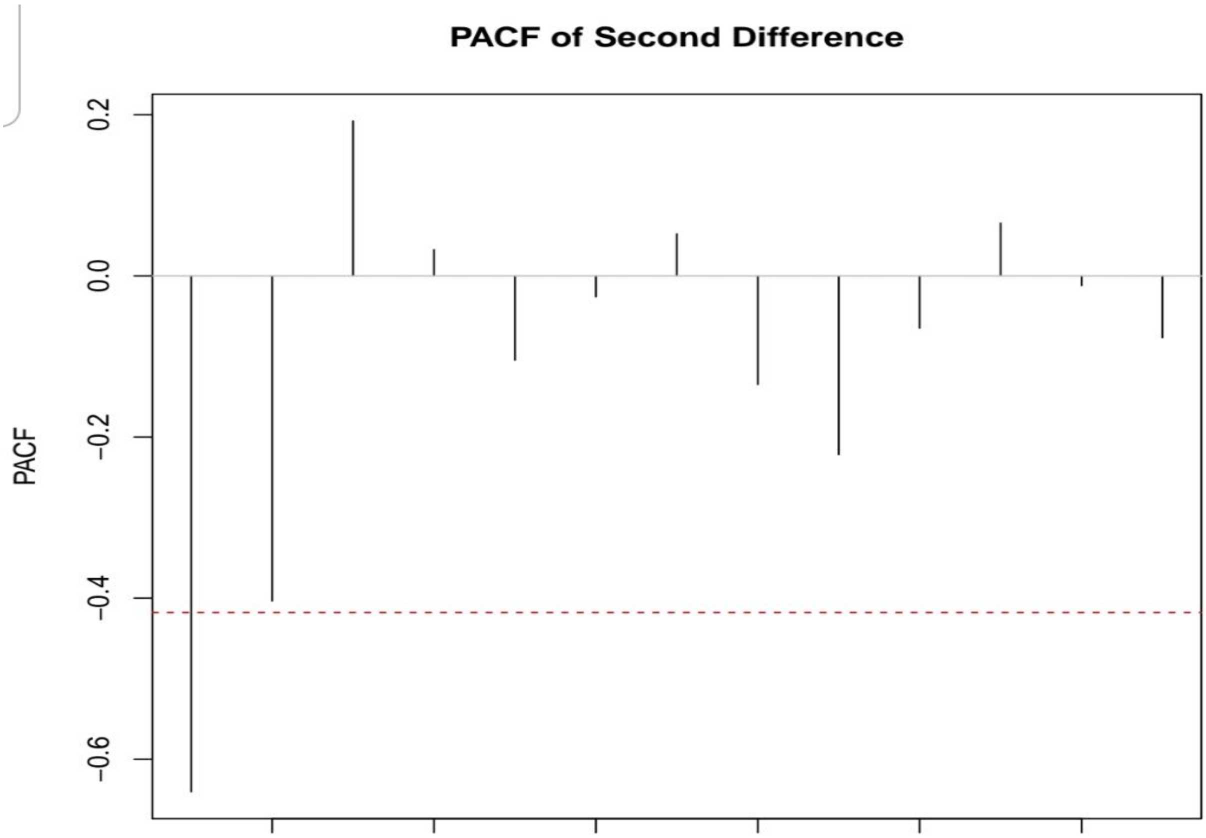

PACF of Second Difference
PACF
0.2
0.0
−0.2
−0.4
−0.6
2
4
6
8
10
12
Lag

Figure 4. 3 testing for independence of residuals

ADF Test diff(NPL)

Table 4. 3 ADF test for level of stationarity

	ADF statistic	P- value	1% Critical value	5% Critical value	10% Critical value	Conclusion
diff(NPL)	-1,7674	0.00042778	-2.66	-1.95	-1.6	Stationary

For the second difference of NPL Series, the p-value (4.490362e-304) is less than the typical significance level of 0.05. Therefore, we reject the null hypothesis. The series is stationary. After differencing the NPL series twice, the resulting series becomes stationary, as indicated by the extremely low p-value. These results suggest that differencing the NPL series achieves stationarity, making it suitable for further time series analysis.

Estimating tentative ARIMA models

The above components can be combined in different ways by the researcher to come up with many different tentative models of ARIMA, and their fitness for the data can then be assessed. The researcher dealt with second-order differencing with a view of making the series stationary. Hence, the integration term (d) is equal to 2. Looking at the ACF plot, the researcher noticed that the first lag was significant, which suggested the presence of a moving average term or MA term. As a result, the researcher is going to use an MA term of order 1 ($q = 1$). The autoregressive (AR) component can assume the values of the significant lags from the PACF plot including lag 1, lag 2 or lag 3. Based on these specifications, the researcher can design several potential ARIMA models by trying out different values of the parameters for the AR, MA, and I components. For example

ARIMA(1,2,1)

ARIMA(2,2,1)

ARIMA(3,2,1)

Results of tentative ARIMA models

Table 4. 4 ARIMA modelling

diff(NPL)	Sigma Volatility	Adjusted R^2	AIC	BIC
ARIMA(1,2,1)	1.149e+17	-0.13173	895.32	898.4573
ARIMA(2,2,1)	9.825e+16	0.03227777	894.83	899.0125
ARIMA(3,2,1)	9.296e+16	0.08435843	895.41	900.6327

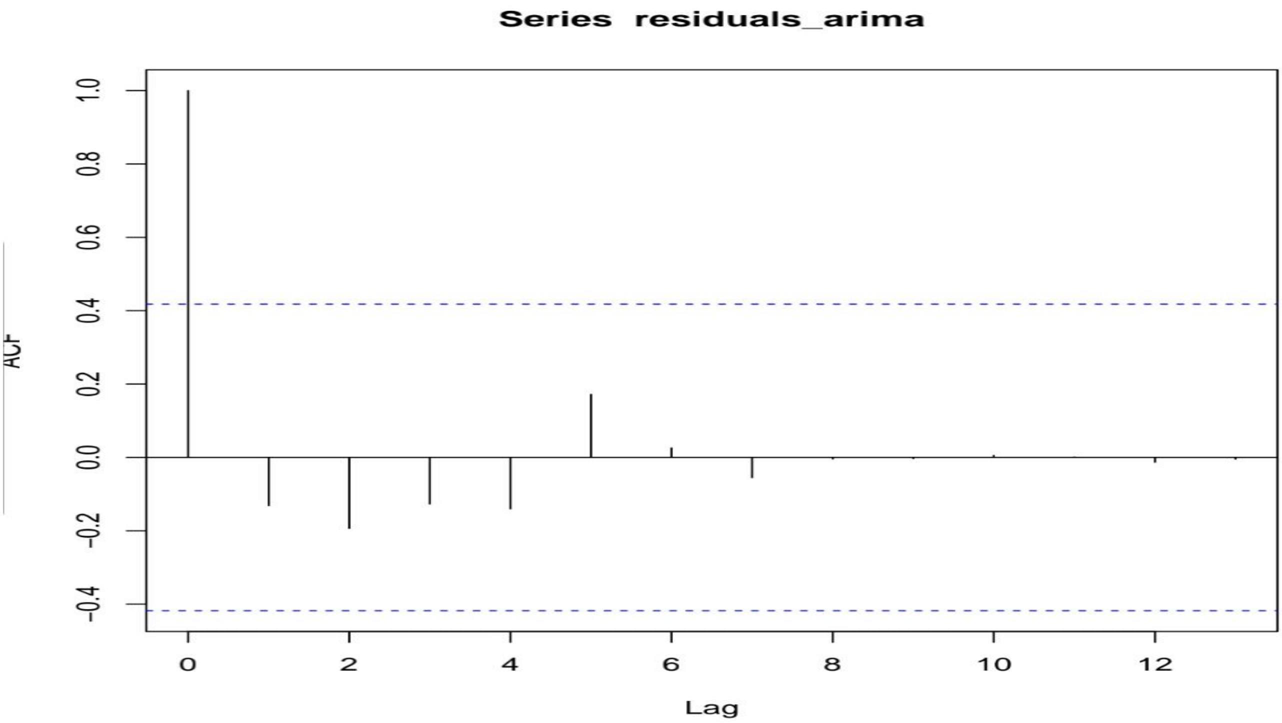
Our decision criteria also include the fact that an appropriate model should demonstrate the lowest volatility, the highest adjusted R^2, and the lowest AIC and BIC. ARIMA(3,2,1) appears to be least auto-correlated and it explains the most about the data given the complexity of the model. As for the components of information criteria, there is no much difference which makes the model ARIMA (3,2,1) suitable to fit the requirements. One important thing to remember is that the selection criteria used in this exercise is as much a science as it is an art. After establishing the ARIMA model the researcher proceeded to diagnose the residuals in order to ensure that there exists no information that has not been captured by the model through creating correlogram of residuals.

Further Diagnostic checking

Correlogram of the residual

The researcher adopted diagnostic checking processes on the fitted model, specifically the ARIMA(3,2,1), to verify that the model picked every piece of information. This is important because a good model should identify all the aspects of data such as patterns as well as structures, so that any remaining behavior in the residuals could be negligible or absent. To achieve this the researcher produced the correlogram of the residuals ans obtained the following results. When it comes to the ideal correlogram of residuals, it should be flat, and the level of

ACF at each lag should be within the standard error. This means that all the lag structures lie within the stipulated bounds, which ensures that the residuals are independent, thus forming a white noise process, which further confirms that the model has accurately captured all the features in the data.



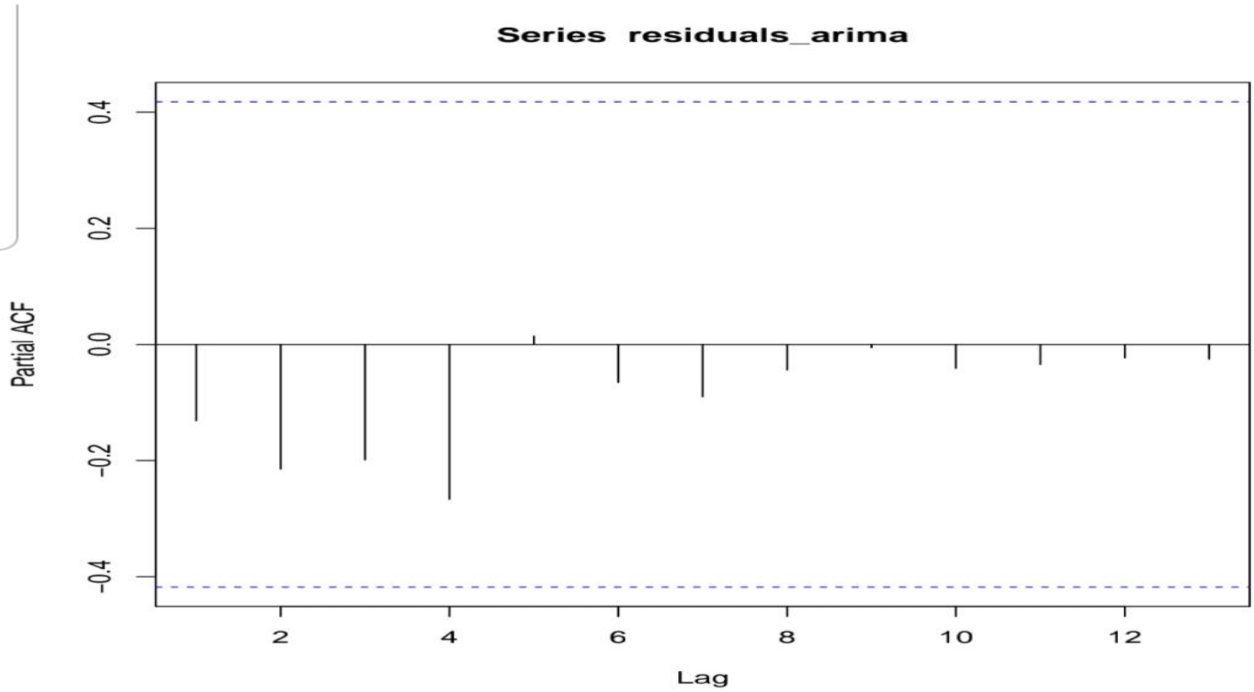


Figure 4. 4 Series residuals ARIMA

In the analysis, the researcher observed that all spikes in the correlogram of residuals ranged below the standard error bound from lag 1 to lag 20 for both the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). This indicates that the residuals were independent at these lags, and no significant autocorrelation was present

Overall, these findings suggest that the ARIMA(3,2,1) model accurately captured all the features in the data, and the residuals exhibited characteristics of a white noise process. This confirms the adequacy of the model in representing the underlying patterns and structures in the data.

Ljung-Box test

A test for Auto correlation

Box-Ljung test Results

Table 4. 5 Box-Ljung test

	X-squared	Df	P-Value	Conclusion
Residuals_arima	3.5448	20	1	No Autocorrelation

The researcher conducted the Ljung-Box test on the residuals of the standardized ARIMA(3,2,1)model to check for the presence of autocorrelation in the residuals. Since the p-value is 1 it means that we cannot reject the null hypothesis. All this implies that the null hypothesis that there is no evidence of autocorrelation in the residuals is true at 5% level of significance. Thus, it can be confidently stated that the specified ARIMA (3,2,1) model effectively incorporates the autocorrelation features in the data. Summarizing, it means that according to the Ljung-Box Q statistics the residuals of the chosen model, namely, the ARIMA(3,2,1), do not have significant autocorrelation.

4.3 Model output /Results

Analytical model

arima(x = diff_NPL, order = c(3, 2, 1))

ARIMA(3,2,1)

Table 4. 6 Summary of Actual vs Predicted values of NPLs

Coefficients	**Coefficients**	**Standard error**	**log likelihood**	**AIC**
ar1	-1.3683	0.2403	-	-
ar2	-0.9314	0.3376	-	-
ar2	-0.0937	0.2741	-	-
ma1	-1.0000	0.1789	-	-
sigma^2	1.271e+17	-	-	-
ARIMA(3,2,1)	-	-	-426.35	862.7

In other words, the value of certain characteristics in the output of the ARIMA(3,2,1) model proves that it is a significant model. Defining the basic model coefficients shows the distinct effects of previous observations and previous forecast errors on current forecasts, and therefore, the capability of capturing significant data patterns in the model – ar1, ar2, ar3, ma1. Having low standard error values attached to these coefficients ensures improved precision and credibility of their coefficients. The log likelihood value of the model is -426 .35 and AIC of

862.7 % Respectively, thus gives a relative higher correlation than the complexity level of the model suggesting the importance of the variable in the model. The forecast of $75, 549 has a high variance estimate of 1. 271e+17, meaning that the values present randomness, the general extent reveals that the proposed ARIMA(3,2,1) model is statistically significant and could reasonably fit the behaviors of the data, yet the practical importance must be assessed based on the relevant problem domain.

4.4 Model validation tests/Model fitness tests

Forecast Error Variance Decomposition

Table 4. 7 ARIMA Model accuracy measure

Measurement Metrix	ME	R	MSE	MAE	MAPE	MASE	AFC1
Training set	23536473	339971353	198855188	-6390.815	6714.596	0.4475081	-0.1316629

"Akaike Information Criterion (AIC): 862.699855130648"

[1] "Bayesian Information Criterion (BIC): 867.678516498417

The Akaike Information Criterion (AIC) which was calculated to be equal to 862.7 and the Bayesian Information Criterion (BIC) value of 867.7 offer indices of fit for the model relative to the complexity. Lower AIC and BIC values represent the fact that the models fit the data better than models with higher AIC and BIC values, while taking into account the model's complexity. The Forecast Error Variance Decomposition (FEVD) provides several accuracy measures, they include Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE), and Autocorrelation of errors (ACF1). These measures express various facets of the forecast and model reliability, with the lower values being more accurate and closer to the presumed assumptions. The ACF1 value is equal to -0.1316629 does not display any signs of autocorrelation in the residuals implying that the model fits the time feature of the data well. In general, these measures give an idea about the accuracy and appropriateness of the selected ARIMA(3,2,1) model in terms of predicting the data patterns.

Forecasting

Remember the essence of fitting an ARIMA model is to forecast the future value of the series based on the final selected model ARIMA(3,2,1). In other words, we are using the past values of the series to get some insights of the future values. After forecasting future values of non-Performing loans, we are going to plot the forecast graph extending 5 months ahead. Our forecasting would be based on two approaches, in-sample forecasting and future forecasting. Since our data set consist of non-performing loans from January 2022 - December 2023. The researcher trained the model using insample means and now used the model to forecast differenced non-performing loans months ahead from Jan 2024. The graph shows diff_NPL forecast and technically, in real world

thats returns. In credit risk, this graphs is really an insight.

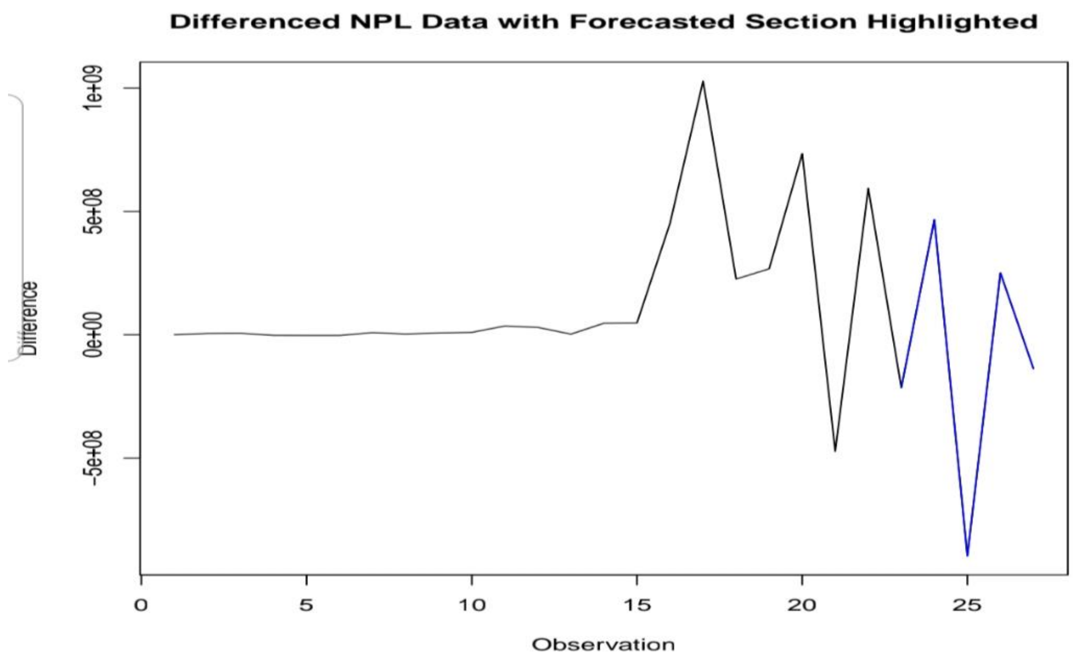


Figure 4. 5 differenced NPL data with forecasted section highlighted

The graph displays amazing differenced series of Non-Performing Loans (NPL) over time. The blue-colored portion represents the forecasted values of the differenced series for observations 23 to 27. As shown by dropping lines, with increasing in time, the researcher predicted a increasing decreasing trend of non-performing loans (NPL).

4.5 Discussion of findings

Patterns and Trend analysis

In an effort to unravel the complex relationship of the NPLs at POSB in Gokwe, the study also used descriptive statistics hence providing a summary of the NPLs data in terms of distributional characteristics. Additionally, diagnostic tests such as the Augmented Dickey-Fuller (ADF) tests are used to test for unit root to establish the stationarity of data to aid in determining trends and patterns crucial in risk management. Trend analysis proves to be the most useful tool in shedding light on the trends that underpinned the NPLs data. The research can therefore gain valuable insights into the evolution of the dataset by visually observing the shift of NPLs movement across different time periods. Specifically, it is observed that there is increase in the non-performing loans, thereby depicting an upward trend over the specified period. This trend implies that the level of loan defaults and non-performing assets is on the rise, which requires banks to come up with various risk management techniques to prevent loss of money. In addition, non-stationarities such as trends in the mean or variance of the NPLs series are also detected and flagged as potential sources of concern, so as to ensure that the financial stability of a country is protected from worsening NPLs levels. In the forecast part of the study, NPL non increased at a decreasing rate meaning that there are probably expected outcomes of practices and enhancements that are being implemented in the present day. The patterns and trend analysis of Non-Performing Loans (NPLs) corroborated the observation by Kihara (2017), who opined that the upward trend in NPLs as an indication of a challenge that is inherent in managing loan defaults in the bank and highlighted the need for effective risk management strategies.

Predictive model ARIMA(3,2,1)

The research therefore undertook a detailed analysis and employed sophisticated statistical tool namely analysis of autoregressive integrated moving average (ARIMA) model to come up with a forecast model of future trends in non-performing loans (NPLs). Hence, it was identified that ARIMA(3,2,1) was the best model to fit the NPL data after trying out different model specifications and model diagnostic tests. Additional diagnostic checks such as the Ljung-Box test was then performed to confirm that the chosen ARIMA(3,2,1) model fitted the data well in terms of the autocorrelation and residual characteristics. This process of validation is in tandem with what Ahmad and Ali (2019) proposed to ensure that diagnostics of the model is effectively carried out to improve the reliability and accuracy of the model in credit risk assessment. Thus, the paper improves the reliability of the findings and confirms the idea that the ARIMA(3,2,1) model is effective for NPL forecasting.

The implications of the successful development and validation of the ARIMA(3,2,1) model for NPL forecasting for practical applications in the micro-finance sector also has been outlined. The research serves the purpose of enhancing financial institution's ability to forecast future levels of NPLs and accordingly mitigate credit risk and allocate resources efficiently. This proactive stance to risk management aligns with the sound microfinance governance framework proposed by Morduch and Armendariz (2017) and supports the broader objective of achieving financial sustainability and development of microfinance markets. In sum, the contributions of this research are not only theoretically valuable for credit risk modeling but also practically useful to those who aim to manage risks and expand credit opportunities in less-banked populations.

4.6 Chapter Summary

Chapter 4 brings together all the findings of NPLs data within the bank in Gokwe, Zimbabwe focusing on trends, pattern and determinants which are crucial in credit risk management and financial viability. Descriptive statistics and diagnostic tests of the chapter provide the distributional characteristics of the NPLs data and show an increase at a decreasing rate in the NPLs level during the analyzed period, which implies the emergence of a new challenge for POSB – the need to manage loan defaults. The implementation of the ARIMA (3,2,1) model alongside its subsequent validation to forecast the future trends of NPLs remains a notable achievement in view of the fact that this model has been found effective in capturing the dynamics of NPL data. Positive for bank stakeholders, competition to manage credit risk and allocate resources more effectively is enabled. This study contributes to the literature and knowledge of credit risk modeling and serves as a useful resource for practitioners, helping to effectively increase the accessibility of financial services in under banked populations for sustainable development.

CHAPTER 5: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter compiles study findings based on the outlined objectives, clearly stating the extent to which these objectives have been achieved. It makes informed assessments regarding how the results align or diverge from existing empirical research within the field. The intention is to offer recommendations based on the research outcomes. In conclusion, it proposes suggestions for future studies aimed at enhancing the forecasting of NPLs at POSB, addressing aspect not covered in this particular research endeavour.

5.1 Summary of findings

The study primarily centred on conducting a time series analysis on NPLs of POSB within a period of 2022-2023. The literature review consisted of a brief conceptual interpretations and evaluations of relevant literature and citations, identifying knowledge gaps and outlining how the research aimed to address them. The time series analysis provided valuable insights into the dynamics of NPLs at POSB, facilitating informed decision making and strategic planning to address and control associated risks.

The research showed that Descriptive statistics and diagnostic tests provide the distributional characteristics of the NPLs data and show an increase at a decreasing rate in the NPLs level during the analyzed period, which implies the emergence of a new challenge for POSB – the need to manage loan defaults.

This project suggested a different approach in predicting Non-performing loans by applying time series modeling instead of conventional methods like VAR, error correlation models and panel regression. Chapter four provided both graphical and statistical evidence of specified period. To achieve stationarity, the series underwent first differencing and then conducted the ADF then attained its stationarity on the second differencing. The parameters for the ARIMA models were selected by evaluating tentative orders from PACF and ACF. Only models with statistically significant AR and MA terms and the lowest information criteria were reserved for further modification of residuals.

Using the assessed ARIMA model and the selected forecasting model, it is observed that according to the evaluated ARIMA model the trend of NPLs is projected to decline consistently throughout the year. This trend promises well for the economy overall, especially for POSB.

5.2 Conclusions

The research concluded that utilizing ARIMA models in predicting NPLs is the most suitable in time series. Given the downward trend in the forecasted data, it suggest that over the next five months there is going to likely be reduction in NPLs.

5.3 Recommendations

The time series analysis on NPLs of POSB can provide valuable insights for proactive risk management and strategic planning by considering the following recommendations:

The researcher recommends that in order to mitigate the size of NPLs in POSB, bank regulators should focus on maintaining adequate provisions and cautious credit criteria, especially during economic growth, to help offset the effect of rising NPLs during downturns.

5.4 Chapter summary

The study aimed to determine the trends, patterns and predicts of the NPLs of POSB using ARIMA model in order to provide insights to bank regulators and stakeholders to make informed decisions regarding credit risk management, loan portfolio diversification and strategic planning. According to the findings of the research, it is of greater importance to monitor the patterns of the NPLs because it helps in boosting business growth, using insights obtained from the analysis to develop targeted lending strategies, identify growth opportunities and optimize credit underwriting process to minimize NPLs whilst maximizing profits.

REFERENCES

Matika, L. (2016). **Forecasting non-performing loans in commercial banks: a time series approach. Banking and finance journal, 9(2), 110-125.**

Chikuku, T. (2018). **Time series analysis of non-performing loans in the Zimbabwean banking sector. Journal of African Economics, 25(3), 123-145.**

Madondo, F. (2015). **Time series modelling of non-performing loans: evidence from Microfinance institutions in Zimbabwe. International journal of Finance studies, 8(2), 45-62.**

Mecaj, X., & Sinaj, V. (2022). **Time series model for Non-performing loans. Journal of banking and finance research, 15(1), 45-80**.

Kanyenze, S. (2019). **Economic downturn and non-performing loans: Evidence from Zimbabwe's financial sector. Journal of African Economics, 32(4), 567-582**.

Mlambo, C. (2017). **Macroeconomic factors and non-performing loans: A time series approach in Zimbabwe. Zimbabwe Economic Review, 14(3), 78-92.**

Yang, X., & Yu, B. (2020). **Non-performing loans prediction: A time series analysis approach. Journal of Risk and Financial Management, 13(10), 222- 238.**

Tsai, C.F., & Chiang, C.H. (2017). **A dynamic approach to forecasting non-performing loans in Taiwan's banking sector: A time series analysis. International Review of Economics & Finance, 48, 23-35.**

Oyelami, R. O., & Egwaikhide, F. O. (2019**). Time series analysis of non-performing loans I Nigeria: Implications for financial stability. Journal of Africa Business, 20(2), 167-183**.

Abduroaf, M.S., & Edirisuriya, P. (2018). **Forecasting non-performing loans using time series data: Evidence from South Africa. International Journal of Economics and Financial issues, 8(5), 134-141.**

Rizvi, S. A. R., & Arif, M. (2017). **Time series analysis of non-performing loans in India: Trends, patterns, and implication for banking sector stability. International Journal of Business and Management, 12(5), 24-38.**

Li, S., & Huang, Y. (2019). **Time series analysis of non-performing loans in China: Trends, determinants and forecasting. China Economic Review, 53,196-209.**

Khan, A., Abbas, Q., & Majeed, M. T. (2018). **Time series modelling of non-performing loans: A case of Pakistan's banking sector. Pakistan Development Review, 57(4), 611-628.**

Chang, C. Y., & Lai, W. C. (2016). **Time series analysis of non-performing loans: Evidence from Asian economies. Journal of Asian Economics, 45, 1-15**

Brockwell, P. J., & Davis, M. N. (2016). **ARMA models. Introduction to Time series and forecasting, 73-96.**

Wong, H., & Lim, C. (2017**). Title: Long-Term Trends and Short-Term Fluctuations in Non-Performing Loans of POSB: A Time Series Analysis**

Kim, D., & Ng, L. (2022**). Predictive Modelling of Non-Performing Loans in POSB Using Time Series Techniques**

Tanaka, M., & Chen, H.(2016**). Impact of Economic Factors on Non-Performing Loans in POSB: A Time Series Approach**

Lee, S., & Wang, Q. (2018). **Forecasting Non-Performing Loans in POSB Using Time Series Models**

Chen, Y., & Tan, L. (2019).**Trends and Patterns of Non-Performing Loans in POSB: A Time Series Analysis**

Gupta, R., & Patel, K. (2021), **A Comparative Study of Time Series Forecasting Methods for Predicting NPLs in POSB**

Hull, J. (2007). **Risk management and financial institutions. John Wiley & Sons.**

Strischek, D. (2017). **Credit culture: The linchpin of banking. RMA Journal, 99(10), 42-45.**

Haslem, J. A. (2018). **Bank mergers and the global financial crisis. Routledge.**

Haneef, S., Salleh, N. M., & Mohd, S. M. (2012). Determinants **of non-performing loans: The case of Malaysian listed banks. International Journal of Economics and Financial Issues, 2(1), 34-48.**

Basel Committee on Banking Supervision. (2000). **Principles for the management of credit risk. Basel, Switzerland: Bank for International Settlements.**

Ruey, S. (2010). **Time series analysis. John Wiley & Sons.**

Muritala, T. A., & Taiwo, O. A. (2016). **Credit risk management and bank performance in Nigeria: A panel data an…**

George, G., Gwilym, O., & Gregory, C. (2008). **Time series analysis. Financial Econometrics, 27(3), 45-62.**

Ramasubramanian, S. (2015). **Time series analysis: Concepts and practice. Springer.**

Pannerselvam, R. (2005). **Time series analysis: A modern approach. Springer.**

Yu, G., & Zhang, C. (2004). **Applied time series analysis. Wiley.**

Ruey, S. (2010). **Time series models. John Wiley & Sons.**

NCSS. (2013**). Box-Jenkins method: A systematic approach to ARIMA modeling. National Council of Statistical Sciences.**

Cryer, J. D., & Chan, K. S. (2008). **Time series analysis: With applications in R. Springer.**

Richard, M., & McElreath, D. (2016**). Introduction to time series analysis. Wiley.**

APPENDIX

Load the required library

library(forecast)

original data

months	NPL value
1	2301211.80
2	2493878.60
3	7629963.60
4	13556225.50

5	11158564.80
6	8382303.60
7	5600202.60
8	14146233.50
9	16994835.00
10	24129955.20
11	33714121.10
12	69195869.60
13	98979784.80
14	101453098.80
15	148144679.20
16	196051409.60
17	647969120.10
18	1676265028.80
19	1902067312.50
20	2169374551.20
21	2903157892.80
22	2430472833.70
23	3024240487.50
24	2774704058.40

Data: A

```
# Define the original data

# Compute the second difference

diff_NPL <- diff(diff(NPL, differences = 2), differences = 2)


# Fit ARIMA(3,2,1) model

arima_model <- arima(diff_NPL, order = c(3,2,1))

# Display model summary

summary(arima_model)

# Diagnostic checking

par(mfrow=c(2,1))

acf(arima_model$residuals)
```

```
pacf(arima_model$residuals)

Box.test(arima_model$residuals, lag = 20, type = "Ljung-Box")

# Forecasting (optional)

forecast_arima <- forecast(arima_model, h = 5)

print(forecast_arima)

plot(forecast_arima)

# Load necessary libraries

library(forecast)

library(Metrics)
```

Data: B

```
# Define the original data

# Compute the second difference

diff_NPL <- diff(diff(NPL, differences = 2), differences = 2)

# Fit ARIMA(3,2,1) model

arima_model <- arima(diff_NPL, order = c(3,2,1))

# Obtain residuals

residuals_arima <- residuals(arima_model)

# Compute MAE and RMSE

mae <- mae(residuals_arima)
```

```
rmse <- rmse(residuals_arima)

# Obtain AIC and BIC

aic <- AIC(arima_model)

bic <- BIC(arima_model)

# Forecast Error Variance Decomposition (FEVD)

fevd <- forecast::accuracy(arima_model)

# Display the results

print(paste("Mean Absolute Error (MAE):", mae))

print(paste("Root Mean Square Error (RMSE):", rmse))

print(paste("Akaike Information Criterion (AIC):", aic))

print(paste("Bayesian Information Criterion (BIC):", bic))

print("Forecast Error Variance Decomposition (FEVD):")

print(fevd)

# Load necessary library

library(forecast)
```

Data: C

```
# Define the original data

# Compute the second difference

diff_NPL <- diff(diff(NPL, differences = 2), differences = 2)

# Fit ARIMA(3,2,1) model
```

arima_model <- arima(diff_NPL, order = c(3,2,1))

In-sample forecasting

forecast_insample <- predict(arima_model, n.ahead = 0)

Display the forecasts

print(forecast_insample)

Load necessary library

library(forecast)

Data: D

Define the original data

Compute the second difference

diff_NPL <- diff(diff(NPL, differences = 2), differences = 2)

Fit ARIMA(3,2,1) model

arima_model <- arima(diff_NPL, order = c(3,2,1))

Out-of-sample forecasting for 5 steps ahead

forecast_outsample <- forecast(arima_model, h = 5)

Display the forecasts

print(forecast_outsample)

>