BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

TIME SERIES ANALYSIS OF WHEAT PRODUCTION IN ZIMBABWE (1975-2023)

BY

TROUBLE MUTUVA

B201563B

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

SUPERVISOR: MR. B. KUSOTERA

JUNE 2024

APPROVAL FORM

This is to certify, that this research project is the result of my own research work and has not been copied or extracted from past sources without acknowledgement. I hereby declare that no part of it has been presented for another degree in this University or elsewhere.

TROUBLE MUTUVA	....	**10/06/2024**
Student	Signature	Date

Certified by:

MR.B. KUSOTERA		**10/06/2024**
Supervisor	Signature	Date

DR.M. MAGODORA		**10/06/2024**
Chairperson	Signature	Date

DEDICATION

I dedicate this dissertation to my loving parents. Your unwavering support, encouragement, and belief in my abilities have been instrumental in my academic success. Thank you for always being there for me, for instilling in me a love for learning, and for teaching me the value of perseverance. This accomplishment is as much yours as it is mine**.**

ACKNOWLEDGEMENTS

I extend my heartfelt gratitude to the professionals and experts in the field who generously shared their knowledge and expertise throughout the development of my dissertation. I express my deepest appreciation to my supervisor, Mr. B. Kusotera and coordinator Ms. P. Hlupo for consistently providing me with support. Their valuable insights and perspectives have greatly enriched this dissertation. I am also thankful for the administrative support provided by Department of Statistics and Mathematics at Bindura University of Science Education in terms of access to research facilities and administrative assistance. I also thank God for providing me with the care, strength, knowledge, and opportunity to pursue my education to this level. Lastly, I would like to acknowledge the unwavering support of my friends and family, whose encouragement and understanding have been a constant source of motivation throughout this journey.

ABSTRACT

This study focused on time series analysis of wheat production in Zimbabwe from 1975 to 2023. In order to fulfil objectives of this study, FFNN and ARIMA models were developed based on annual wheat production data from 1975 to 2013(training set) and forecasts from 2014 to 2023 were deduced from the two models for comparisons with the actual data from 2014 to 2023(testing set). MAE and RMSE used as performance metrics. Results of this study show FFNN model as a best fit, since it comes out with lowest error values and it mimics the actual values of wheat production compared to ARIMA which shows a linear pattern over time and FFNN detected hidden patterns of wheat production data. Forecasts from the best selected FFNN model show an upward trend to be expected from 2024-2028, indicating that wheat production values are increasing, which is a good indicator to the agricultural sector. Based on the findings, this study recommends the use Artificial Neural Networks (ANN) for time series analysis and the government of Zimbabwe and policy makers to keep on enacting policies and support in agriculture to increase agricultural output and achieving the country's Sustainable Development Goals (SDG) of being a middle-income country by 2030.

Key words: Time series analysis, ARIMA models, FFNN models, forecasting

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

1. ACF - Auto Correlation Function
2. ADF - Augmented Dicky Fuller
3. AIC - Akaike Information Criteria
4. ANN- Artificial Neural Networks
5. ARIMA - Auto Regressive integrated Moving Average
6. BIC - Bayesian Information Criteria
7. FFNN- Feed-Forward Neural Networks
8. GDP-Gross Domestic Product
9. GMB- Grain Marketing Board
10. MAE - Mean Absolute Error
11. PACF - Partial Auto correlation Function
12. RMSE - Root Mean Squared Error
13. SDG-Sustainable Development Goals
14. ZIMSTATS-Zimbabwe National Statistics Agency

CHAPTER 1: INTRODUCTION

1.0 Introduction

National development, food security and domestic income generation are mainly controlled by agriculture in Zimbabwe (Dzvimbo,et al.,2017). The government of Zimbabwe, private institutions, individual farmers, and other stakeholders are interested in knowing the production of crops that could be expected, therefore predictive analytics in agriculture adds economic value. Estimation of crop production outputs leads to assessment of costs versus profits in order to make informed decision, optimize agricultural operations and achieve the country's Sustainable Development Goals (SDG) of being a middle-income country by 2030.

This chapter introduces time series analysis of wheat production in Zimbabwe by outlining the background of the study, problems that driven this study, research objectives, questions, and problem statement, scope, significance, limits, assumptions, delimitation, organisation of the study which gives a brief summary of each chapter since five chapters included in this study and finally a summary.

1.1 Background of the Study

History exposes that approximately 70% of Zimbabweans rely on the country's agriculture sector for their livelihoods, it also contributes to Gross Domestic Product (GDP) by 15% to 20%, export revenue of 40% and 63% of the country's agro-industrial raw materials (ZIMSTATS, 2012). Because of this, the farming industry plays a vital role in developing plans and laws that reduce food insecurity and increase income in the country. After maize, wheat follows as the important crop for food security (Kapuya, et al., 2010). Because of its main product bread that is on demand, the commodity has a high value.

Zimbabwe has a history of wheat production and traditionally been one of the major wheat producing countries in Southern Africa. During winter season wheat is grown, captivating the advantage of the country's climate. However, the production of wheat in Zimbabwe has been facing challenges such as a change of landholding structure through Fast Track Land Reform program which took place in 2000 and 2001, discouraged investments and the government's decision to expropriate land without compensation has led to disputes and legal challenges. Land reform program resulted in the acquisition of large-scale farms, which were then redistributed to small-scale farmers of which many of these small-scale farmers lacked skills and resources, leading to an overall decline in wheat production. Due to trade restrictions imposed by the international community, government adopted policies of self-reliance to improve domestic production. These policies included government grants in development of irrigation farming, loans to support farmers for the importation of farming equipment, commodity markets and prices through Grain Marketing Board (G.M.B) as the major buyer of grain, increased efforts in agriculture research and extension.

Besides the previous investment efforts from the government wheat production has been decreasing, which went from 324 000 metric tonnes in 1999 to 62 000 metric tonnes in 2016. A comparison of the two-year average 2020–21 with 2022–23 shows a 20 percent decrease in production, from 222 000 metric tonnes to 178 000 metric tonnes but with an increase on wheat imports from 200 in 2020 to 230 in 2023 metric tonnes (Index Mundi,2023). In this case Zimbabwe has been importing wheat to meet annual consumption level.

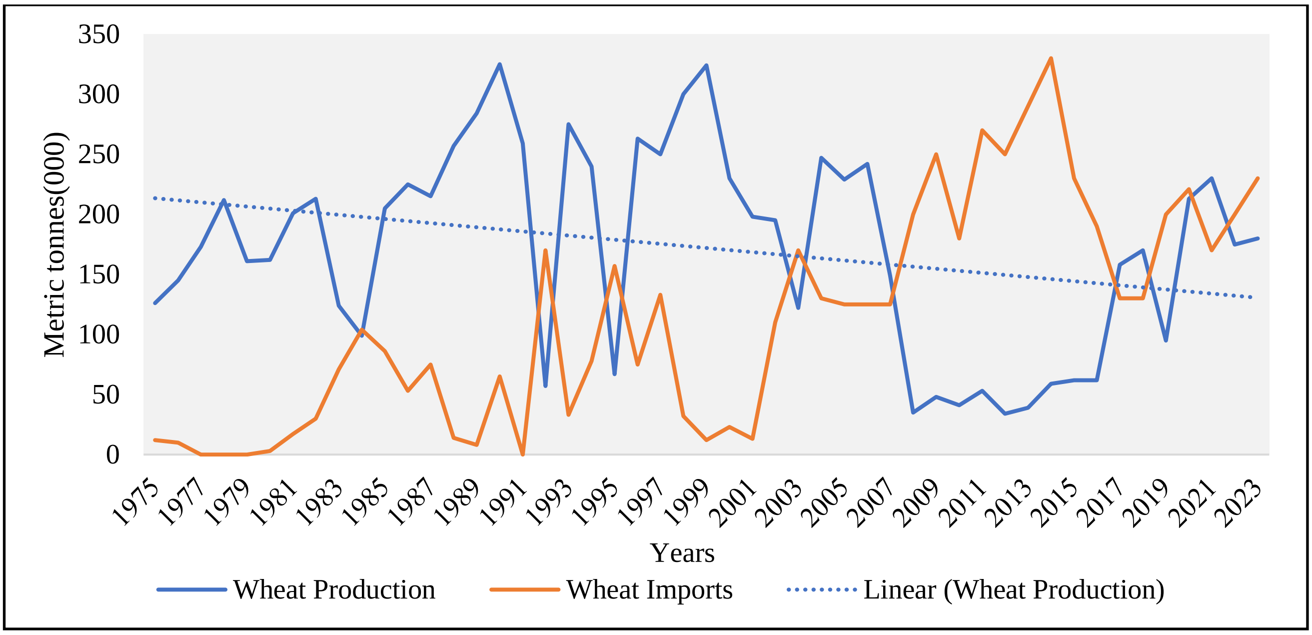

Figure 1.1 Wheat production and imports in Zimbabwe 1975-2023

Source: Index Mundi (2023)

Figure 1.1 shows a declining trend of wheat production and significant fluctuations in both wheat production and imports over the past years from 1975 to 2023. Wheat production appears to have experienced a fluctuation pattern, with the highest production levels observed around the early 2000s and again in the late 2010s. Wheat imports, on the other hand, have also exhibited a volatile pattern, with periods of high imports took place when production decreased alternatively low imports when wheat production increases. That suggests that the country's domestic wheat production has been unable to consistently meet its overall wheat demand, leading to the need for wheat imports to supplement the supply. The fluctuations in both production and imports indicate the complication of wheat market dynamics in the country.

1.2 Problem Statement

Declining trend of wheat production is the major problem in Zimbabwe which led to insufficient supply of domestically grown wheat in Zimbabwe that has led to an increased reliance on expensive imports in order to meet domestic demand. Government cannot accurately forecast wheat production leading to unplanned imports due to shortages that happen because of fluctuating demand and unplanned productivity. Dependence on wheat imports causes negative economic consequences such as trade imbalances leading to potential trade deficit, arising questions such as "what is the future wheat production trends after a certain period of time and which model has the greatest ability in forecasting wheat production for better decision making?".

1.3 Research Objectives

1. To develop time series models based on ARIMA and ANN that explain movements in wheat production in Zimbabwe.
2. To determine which model has the greatest ability in forecasting wheat production, between ARIMA and ANN models.
3. To forecast the pattern for the years from 2024-2028.

1.4 Research Questions

1. Which time series models explain movements in wheat production in Zimbabwe.
2. Which model can be recommended for forecasting wheat production?
3. What is the future wheat production in Zimbabwe?

1.5 Significance of the study

The time series analysis of wheat production data holds significant importance for the student, in academics, the Zimbabwe government, policymakers and other stakeholders in the following ways;

1. **To the student**

The time series analysis of wheat production serves as a foundation for the student research project and thesis work by providing practical learning incorporated into agriculture, to enhance student understanding of real-world challenges and decision-making processes fulfilling the requirements of Bachelors of Science Honours Degree in Statistics and Financial Mathematics.

2. **To other students and researchers**

The insights from the time series analysis can help other students and researchers prioritize research topics and design studies that address the most pressing issues in wheat production and highlighting areas where further research is needed to address challenges or capitalize on opportunities in the wheat sector.

3. **Government of Zimbabwe**

Food security planning, accurate forecasts of wheat production can assist the government in ensuring food security and managing wheat production and imports to meet the country's needs. The insights from the time series analysis can help the government make more informed decisions about budgeting, subsidies, and investments in the wheat sector.

4. **Policymakers**

Alignment with national development plans, the wheat production trends can be incorporated into broader agricultural and economic development strategies for the country and to track the effectiveness of existing policies and programs related to wheat production, enabling policymakers to make evidence-based adjustments.

1.6 Assumptions of the Study

ARIMA model assumptions

1. The time series data is stationary and can be made stationary through differencing.
2. Residuals are normally distributed.
3. The variance of residuals is constant over time.
4. Residuals are uncorrelated with each other.

ANN models assumption

1. The amount of training data is sufficient to learn the underlying pattern.

1.7 Limitations of the Study

The absence of literature review on time series analysis on wheat production in Zimbabwe was a challenge, however this study used international articles and journals for the literature review. This study used secondary data on Index Mundi an international source rather than local sources such as ZIMSTATS and GMB because they failed to provide wheat production data from 1975-2023 for time series analysis. This study failed to analyse and capture all important variables (e.g., labour, capital and other production costs) explained in the theoretical chapter due to lack of time series data. Other factors such as socio-economic indicators, consumer preferences and technological developments were also excluded from this study.

1.8 Scope (Delimitation of the Study)

To ensure focus and clarity, this study focused on time series analysis of wheat production in Zimbabwe and it has a specific time period of historical data from 1975 to 2023 and a short forecast from 2024 to 2028. This delimitation allows for a depth study of wheat production changing aspects within the selected area and period. By setting a defined time frame this study focused on short to medium term forecasting rather than long term predictions which may result in more indecision.

1.9 Definition of terms

Wheat production

The growing of wheat crops for different purposes mainly for food consumption, livestock feed and industrial uses (FAO,2021). It involves the complete process, starting from land preparation and sowing of wheat seed to the managing of crops and the final harvesting wheat grains (Shewry, 2009).

Time series analysis

Refers to a statistical method used to analyse and forecast time series data. It involves examining patterns, trends and cycles of data points collected and recorded over consecutive time intervals, such as months, quarterly basis, or years (Box et al., 2015).

Artificial Neural Networks (ANN)

ANN defined as a computational algorithm consisted of interconnected artificial neurons that train based on the available data to perform certain tasks (Nehra, Sangwan, & Kumar, 2021). ANNs are designed to imitate the operative structure of biological neural networks at recognizing patterns and non-linear relationships in any data set, makes them powerful tools for solving complex problems and making predictions based on patterns and relationships in the data.

Forecasting

Forecasting it's an attempt to forecast a variable's future behaviour in order to make more realistic and well-informed decisions depending on past data and statistical models (Wheelwright and Hyndman, 2016). It includes analysis of past relationship to project future outcomes for organizations and individuals to make decisions, plans and policies.

1.10 Organization of the Study

This study has five chapters. Chapter one introduces this study, by giving a general understanding of the time series analysis of wheat production in Zimbabwe. Followed by chapter two where literature review is carried out presenting researches that were previously done and methods that were used to estimate and predict wheat production and others crops production. Chapter three presents the research methods, data sources, variable and analytical models. ARIMA and ANN models were presented as analytical models. Chapter four then presents the results and outputs comparing two analytical models, then wheat production forecasts done use the best selected model. Chapter five concluded the study. It presents summary of findings, conclusion, recommendations then areas for further research based on the results that were presented from the previous chapter.

1.11 Summary

This chapter introduces time series analysis of wheat production in Zimbabwe by reviewing the study's background, problem statement, objectives, research questions, assumptions, limitations and delimitations and an organization of the study that reviewed what is expected in other chapters. The following chapter discusses previous studies that are related to the one that is being conducted.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

Historically, several studies have been done related to this study. By reviewing those existing literature, it provides valuable understandings into this current study of time series analysis of wheat production and areas that require further research.

2.1 Theoretical Literature Review

2.1.1 Predictive Analytics Theory

Predictive analytics comprises application of several statistical models, data mining techniques and artificial neural networks to get useful insights from data (Kumar and Garg,2018). Rapid transformations of our world led to production of complex data, this data should be analysed to a form which is beneficial to the recipients. Due to development of different algorithms, predictive analytics models took place to analyse and forecast time series data and a shift has been taking place towards computational nonparametric modelling techniques, without prior assumptions about any distributions in the data. These techniques include artificial neural networks (Kumar and Garg,2018). Forecasting of future values helps in solving the projected problems by planning a strategy for the future. Humans has been curious in understanding about the future hence, forecasting considered as crucial component for future planning. All sectors medicine, education, agriculture, finance and government, are focusing towards forecasting (Idrees,2019).

2.1.2 Cobb-Douglas production function theory

Two economists Charles Cobb and Paul Douglas, introduced the production function in the 1920s, the Cobb-Douglas production function describes the relationship between inputs such as labour and capital and outputs in production process (Cobb & Douglas, 1928). In order to come up planned budgets of wheat production, labour and capital should be considered when optimising the level of inputs needed to meet certain outputs based on available resources, enabling informed decisions based on wheat production forecasts. In a hectare of wheat in Zimbabwe, the contributions of labour and capital can vary due to factors such as technology, mechanization, and agricultural practices(Gweru,2019) The typical total cost of wheat production in a hectare is $(USD) 1,788.00 and approximately 31 % of the total costs meant for fertilizers and lime, 20 % for irrigation costs, 13 % tractor power, 9 % seeds, 9 % harvesting, 9 % for fixed costs and 9 % for other costs, which produces approximately to 4-8 metric tonnes of wheat under good management practices (Mutambara,et al ,2013). Consideration of inputs and costs in relation to wheat production forecasts, helps policymakers, government, farmers and other stakeholders to identify the optimal level of inputs to achieve the desired level of wheat production output by minimizing costs such as on fertilizers, lime and irrigation .

2.1.3 Model Comparative Theory

The Akaike information criterion (AIC), the Bayesian information criterion (BIC), the root mean-square error (RMSE), and the mean absolute error (MAE) are some of the evaluation or performance metrics used in time series analysis to assess the performance of ARIMA and ANN models (Wheelwright and Hyndman, 2016). In forecasting using ARIMA, models with higher precision and lower error are selected by AIC and BIC. According to Aras & Kocakoç, (2016), the RMSE assigns a significant penalty to these variances and more suited for detecting outliers, whereas MAE assigns equal weighting to deviations from actual values showing that MAE is a better tool as compared to RMSE for considering the goodness of fit of a model.

2.2 Time series analysis

Aras and Kocakoc (2016), specified that time series analysis consist of three areas which are forecasting, modelling, and description. It is a very important area of data analysis which has got practical interest different sectors and organizations and it allows the prediction of the upcoming values from historical data (Tealab, 2018). Data for time series analysis produced from quantities and measurements of some social, economic, or environmental phenomena of interest. Seasonality, cyclicity, trend, and irregularity are major components of time series data for model development. The ARIMA model presented by Box and Jenkins in 1976 is a statistical model for analysis of time series data. It uses past data to analyse trends and make predictions based on historical data used in the model (Edbrooke, 2017).

2.3 Autoregressive Integrated Moving Average (ARIMA) models

Several studies accomplished to fit ARIMA models for various types of crops. Production for crops like sugarcane (Sankar & Pushpa, 2019) and oilseeds (Mithiya, Datta, & Mandal, 2019) among many others modelled and forecasted using ARIMA. According Aslam, Salman, & Jan, (2019), ARIMA models have been applied and acknowledged in different areas all over the world as well as in economic growth, weather forecasts, stock markets and logistics among many others. However, in recent times of huge and complex data, ARIMA models has been seen incompetent in detection of non-linear data like the recent methods (Mensi et al, 2017). Awe & Dias, (2022) and Kharin, (2018) cited that the progression of information technologies has contributed to generating massive amounts of complex data and problems. This evolution has also introduced the application of Artificial Neural Networks (ANN) that outperforms traditional models, whether it is a question of regression, forecasting, classification or more advanced problems such as the processing of multimedia data.

2.4 Artificial Neural Networks

Although ARIMA models have been used for decades, a change of focus is taking place by the application of ANN models which have outstanding features (Mensi et al, 2017). In the modern world data is associated with nonlinearity behaviours which requires techniques such as ANN which imitates nonlinearity of data. ANN models are able to detect the complexity of non-linear models as they can capture the nonlinearity of data, improving their forecasting accuracy (Rather et al, 2015). The ANN models applied and recommended in time series analysis and forecasting where forecasts of a variable produced relying on historical data. Many studies applied ANN models for forecasting and solving problems.

2.5 Comparison of ARIMA and ANN models.

Over the past years, there has been a large assessment between two models mainly ANN and traditional ARIMA techniques in different fields of applications such as wheat production (Safa,et al, 2015), stock market index (Herliansyah,et al, 2017), electronics (Karami, 2010) and stock management (Mitrea, et al, 2009). The outcomes showed from different areas of applications shows that better performance done using neural networks in comparison with traditional ARIMA methods in forecasting and in more advanced problems such as classification.

2.6 Wheat Production

Wheat is becoming a staple food all over the world because of quick urbanization and income development. African countries produce approximately 30% to 40% of the total demand for domestic consumption, causing dependence on imports in the African region, making it visible to supply disruptions (Negassa, et al., 2013). Since wheat demand in Zimbabwe has grown up to about 450,000 tonnes per year, production of wheat in Zimbabwe has been declining and fluctuating and the country has been in a struggle to fulfil the demand to ensure that the country has suitable food supply (Mutambara,et al., 2013).

2.7 Empirical Literature Review

Awe and Dias, (2022) carried out comparative analysis of ARIMA and ANN techniques for forecasting Nigeria non-stationary agricultural output data. Past trends of agricultural output in Nigeria have been on a downward slope and fluctuating. Using the proposed models, modelling and forecasting was done based on data for a period of 1980 to 2019, and the models were compared using evaluation metrics. The result justifies the superiority ANN model over ARIMA models in non-stationary time series forecasting. Hence, they recommended use of neural network methods in forecasting of non- stationary data because of their computational power of detecting nonlinear patterns without assumptions. It is further proposed that accurate prediction from superstitious model helps farmers and other stakeholders to make best decisions and proper budgets and solving problems agricultural production.

Safa, et al, (2015), conducted a research on prediction of wheat production using ANN and investigating indirect factors affecting it in New Zealand's Canterbury Province. Indirect factors were employed to predict agricultural production by utilising ANN models in order to come up with main factors affecting wheat production fluctuations in Canterbury Province. Results shows that the forecasts produced affected more by indirect variables like social-economic factors and farming practices. They recommended the use of ANN in forecasting and consideration of dissimilar variables to improve the decisions based on the problem of decrease in agriculture production from different perspectives and since ANN models were more effective in prediction so they recommended future studies to be done using artificial neural networks models for accurate results.

A study done by Ghodsi, et al., (2012), used artificial neural networks to estimate Iran's wheat production. ANN model used as a prediction tool and they selected eight important wheat-producing elements to carry out their review since the negative effects of soil and climate on wheat output in many areas of Iran. Annual data span from 1988 to 2006 used in their study. In order to reduce bias and noise the model was run hundred times. Forecast of five years period were done for wheat production and the findings indicated that forecasting helps in informed decision making and supply chain optimisation and further recommended the use of ANN model as a good method for forecasting wheat production. The study also advised other researchers to focus on machine learning in their predictions.

Mapuwei, et al., (2022) carried out a study using time series ARIMA forecasting model in Zimbabwe tobacco production. A time series of Zimbabwe tobacco yield was employed in order to design appropriate methods to tackle this declining pattern. In order to forecast the tobacco yield for 2019–2023, they focused on the yield from 1980 to 2018 to build ARIMA models. Tobacco yield showed a declining trend that changed slightly over the projected years. In order to improve tobacco production researchers recommended the adequate provision of inputs, farmers education and assistance from the government regarding tobacco production. They also recommended other academics to use alternative techniques such recently developed ANN models for comparisons and coming up with the best model.

A comparison of ANN and traditional time series models for timber price forecasting is a study done by Kożuch, et al., (2023), to analyse the average price decrease for timber in Poland. The research used quarterly time series of timber prices from 2005–2021. ANN was found to be far superior to ARIMA and Prophet Model in forecasting price fluctuations. The study recommended that accurate price forecasting is important as prices to understand the market variables concerning suppliers and buyers. The researchers endorsed upcoming studies to continue considering ANN models for development of the best solutions enabling accurate forecasting to solve problems.

2.8 Research Gap

The literature assessment cited above is helpful in recognizing research gaps and challenges, which serve as the main pillars around which the objectives of the current study are based. Two studies have been done to predict wheat output, but focused on other nations New Zealand and Iran and less focus has been on Southern Africa region particularly Zimbabwe. One of the studies used the ARIMA model without ANN to forecast other crops, such as tobacco in Zimbabwe. The following are some ways in which this study adds to the gaps mentioned above, firstly by developing both ARIMA and ANN based on wheat production movements in Zimbabwe since more studies recommended ANN. Secondly, in order to come up with the best model between ARIMA and ANN. Lastly coming up with projected trends of wheat production in Zimbabwe for informed decisions making.

2.9 Proposed Conceptual Model

The causal relationship between wheat production, its restrictions and impacts are proposed by the framework.

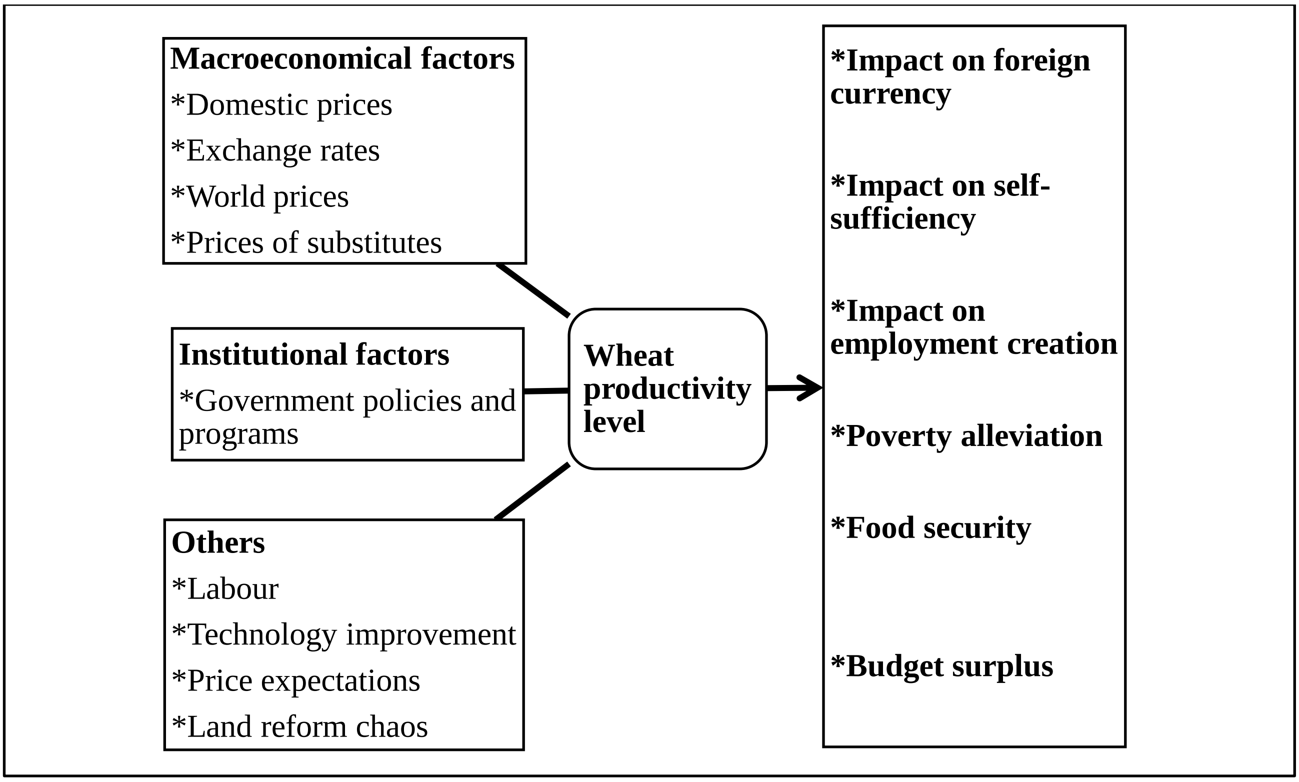

Figure 2.1 Proposed conceptual model

Labour, institutional variables like policy, technology, pricing expectations, and macroeconomic factors like currency rates all affect wheat production. Developments in wheat production can result in several constructive outcomes, including increased foreign exchange revenue, wheat self-sufficiency, reduced poverty, increased employment, and an improved government budget.

2.10 Summary

This chapter provides dynamic reviews of earlier studies related to this study. After the literature was reviewed, various findings were drawn, such as the fact that earlier studies only looked at predicting wheat output in other regions that are outside of southern Africa, particularly Zimbabwe and that some studies even focused on other crops such as tobacco. Because of this, there is a gap where it is preferable for time series analysis of wheat production in Zimbabwe rather than that of other nations or regions. The research procedure, including data collection and analysis methods, is covered in the next chapter.

CHAPTER 3: RESEARCH METHODOLOGY

3.0 Introduction

In this study research methodology involves a systematic approach of collecting, analysing, and interpreting relevant data to produce reliable results.

In this chapter, data sources, target population and sample, research instruments, data collection methods, description of variables and expected relationships, data analysis procedures, model building and selection processes clearly shown. Evaluation metrics of ANN and ARIMA discussed together with their mathematical formulas. Research methodology guides and directs how this study is carried.

3.1 Research Paradigm

One common research paradigm for time series analysis is positivism, which highlights objectivity, quantifiability, and the use of statistical methods to analyse data (Babbie, 2020). Positivist researchers' purpose is to discover patterns within data to make forecasts about future results. In the context of forecasting wheat production, a positivist approach would involve collecting past data, identifying trends and developing models to predict future production values.

3.2 Research design

Research design mentions the plan or framework that guides the whole study process. It outlines the specific steps, procedures, and strategies that a researcher employs to address the research questions or objectives. The research design regulates the value of your conclusion from your findings (Babbie, 2020). Predictive research design, a quantitative research method which uses historical data to predict outcomes used in this study. Statistical analysis tools and steps completed in this study to draw results from the time series data and typically using a complete sample size to ensure that the results suit the whole population. Predictive research design provides valuable insights that enable informed decisions.

3.3 Data collection and sources

The process of gathering, measuring, and recording data for a specific purpose is called data collection. In this study secondary time series data of Zimbabwe's wheat production from 1975-2023 was accessed from an online platform called Index Mundi, using a laptop. The first step was to identify the required data from the portal. The desired data depended on a variable and time period to locate the desired data. Once the wheat production data was identified and accessed, it was downloaded in a compatible file and exported to Microsoft Excel and R Studio. Other wheat production data sources for Zimbabwe's wheat production including GMB and ZIMSTATS were not considered in this study since they failed to provide data from 1975. This would not provide a significant number of time series analysis observations.

3.3.1 Justification

Secondary data sources provide data that is already collected which saves time and costs. Index Mundi is regarded as a reliable and extensive global statistic which provides data that is updated validated on a regular basis and suits the sample size, making it a valuable tool for academic research and analysis.

3.4 Data validity and reliability

When referring to the data reliability and validity, it is all about quality and trustworthiness of data provided by the source.

- **Data Reliability**

Data reliability involves to the extent to which data is provided by the source is consistent, stability and verified by other sources. In the context of Index Mundi, data is updated and validated on regular to ensure that data is free from significant errors, biases, and inconsistencies.

- **Data Validity**

Valid data reflects the intended phenomenon (Weil,2017). In case of Index Mundi data provided aligns with established standards, methodologies and official sources.

3.5 Target population and sample period

Weil (2017), defines population as the entirety of the experimental variable(s) under the study. During collecting data of Zimbabwe's wheat production, the target population refers to the specific population with characteristics of interest required by the research. In this study, the target population included historical wheat production in all provinces of Zimbabwe. A sample refers to portion of the total population that has been chosen to meet the representativeness property. Wheat production from 1975-2023 selected as a sample in this research using purposive sampling method, where data with specific criteria and characteristics such as significant data points, patterns and trends in wheat production, allowed time series analysis to meet objectives of this study.

3.6 Research instruments

Research tools are basic tools used for data collection, presentation and analysis of data related to the overall research. The Internet assisted with data accessibility as it was used to access and download online data from Index Mundi using a laptop. The use of statistical programming languages has taken predictive analysis to a further level where computation is a matter of seconds and the accuracy of results is managed. Microsoft Excel and R Studio were statistical software packages used in data analysis in order to come up with results and conclusions of this study.

3.7 Description of Variables

Annual wheat production of Zimbabwe measured in metric tonnes used as a variable. Brief variable description is given below and a justification.

Table 3.1 Description of variables

Variables	Symbol	Indicator	Source
Wheat production	WhP	National Logarithm of wheat produced in metrics tonnes	Index Mundi

3.7.1 Justification

Wheat production output is the only main variable under study, hence its inclusion. Wheat is important in food security and agricultural economies, making it a variable to monitor and analyse.

3.8 Data analysis procedures

3.8.1 Diagnostic tests

Stationarity tests

In time series analysis, it is vital to test for stationarity before any type of analysis. Wheat production data should be stationary when mean, and variance become constant. Non-stationary data is converted into stationary by differencing. Augmented Dickey-Fuller test (ADF) was used to test for stationarity in this research.

Normality tests

Wheat production data was tested for normality, because incorrect parameter estimate or output prediction may take place. Histograms and quantile-quantile (Q-Q) used in this study to test normality. These are the most useful plots for validating the normality and other probability plot variations (Das and Imon, 2017).

Independence test

Independence and no autocorrelation should be met by residuals of wheat production data. Apart from other tests such as Durbin Watson test, another tool used is to plot the auto-correlation function (ACF) and Partial auto-correlation function (PACF) of the residuals. These two , ACF and PACF were used in this study. Of importance is the fact that when applied, both ACF and PACF should not show any significant term.

Homoscedasticity test

Homoscedasticity test used to confirm that the variance of the residuals of wheat production data is constant. It is often evaluated using statistical tests such as White Test and Breusch-Pagan test or the and visual inspections of the residuals. In this study White test was used to check for homoscedasticity, which ensures that the model is stable and reliable.

3.8.2 Analytical models

This section is devoted to the presentation of the models, which are meant for data analysis to meet the objectives and research questions. Started with the most famous and widely used the ARIMA model then the Feedforward Neural Networks. Figure 3.1 shows models building and selection procedures carried out in this study.

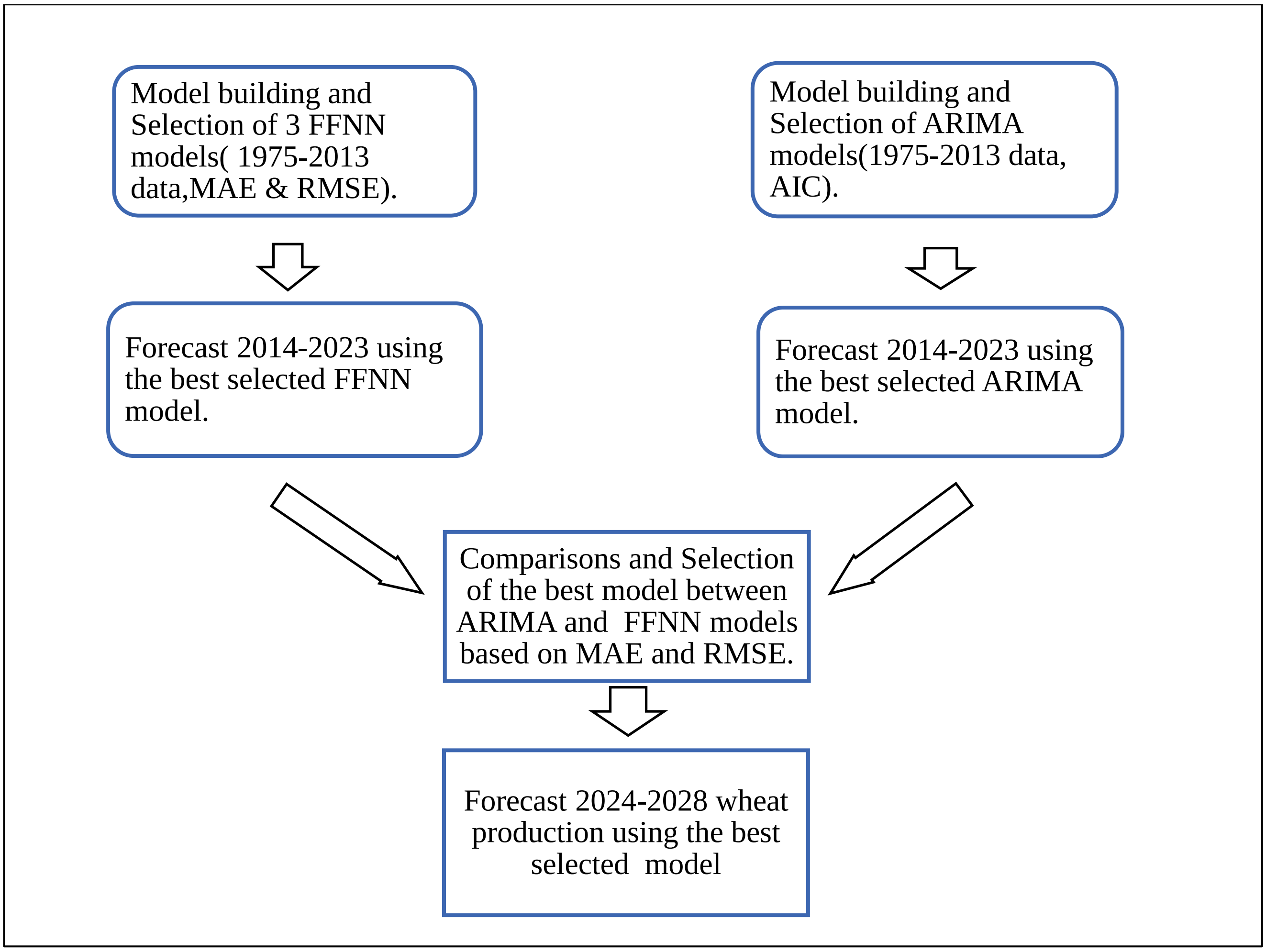

Figure 3.1 Model building and selection procedure

The Box-Jenkins ARIMA models

A well-known family of time series models for analysing and forecasting data with temporal dependencies (Box et al., 2015). Autoregressive (AR), Integrated (I), Moving average (MA) and are three main parts of the ARIMA model. AR detects the linear relationship between observations and lagged observations, component I is used to make the time series stationary by taking the difference between successive data points and component MA represents the dependence between an observation and residual error terms from MA model applied to lagged observations. The ARIMA (p, d, q) model on a Y_t time series is formulated as

$$\Delta^d Y_t = \emptyset_1 \Delta^d Y_{t-1} + \emptyset_2 \Delta^d Y_{t-2} + \cdots + \emptyset_p \Delta^d Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \ldots \quad (3.1)$$

where p, d and q are coefficient of AR, I and MA parts respectively, ϵ_t represents the error term of Y_t which is presumed uncorrelated. Δ represents shift operator, $\emptyset_1$, $\emptyset_2$, ..., θ_q are the AR parameters. θ_1 , θ_2,… , θ_q are represents the MA part (Awe et al., 2020).

Box and Jenkins ARIMA models based on the next iterative three-stage modelling approach.

Model identification

After stationarity test, ACF, PACF and auto.arima function decides which ARIMA models components should be used in the model by checking which model reduces the AIC and BIC values.

Parameter estimation

Involves the derivation of components and parameters of ARIMA models to arrive at coefficients that best fit model.

Model diagnostics

The estimated model should be diagnosed for adequacy to fit to the specifications of a time series analysis. Stationarity should be achieved and the residuals of wheat production data to be independent, normally distributed and homoscedasticity in order to build the best model.

Feedforward Neural Network model (FFNN)

Feedforward neural networks refers to a type of ANN models where information flows in a single direction from input to output without feedback networks (Goodfellow et al., 2016). They consist of interconnected layers of nodes, with each node applying a non-linear activation function to its input to produce an output (Nielsen, 2015). The input layer signifies the features, the output layer produces predictions, and the hidden layers perform in-between computations (Chollet, 2017) Training involves adjusting the network's parameters using techniques like backpropagation, where the error is propagated backward to update the weights and biases. The structure of the FFNN generalised by the following equation,

$$\hat{y} = F\left(v_0 + \sum_{j=1}^{m} H(\lambda_j + \sum_{i=1}^{n} x_i \theta_{ij} \right) v_j \dots (3.2)$$

$\hat{y}$ signifies the network output, H total number of neurons in hidden layer, and F represents the output activation function and x_i is the input vector (Cheng & Titterington, 1994). An example of FFNN with three (3) input neurons, a hidden layer with four (4) nodes lasty one (1) output node presented as 3(4)1 respectively. A FFNN structure of 3(4)1 is shown below,

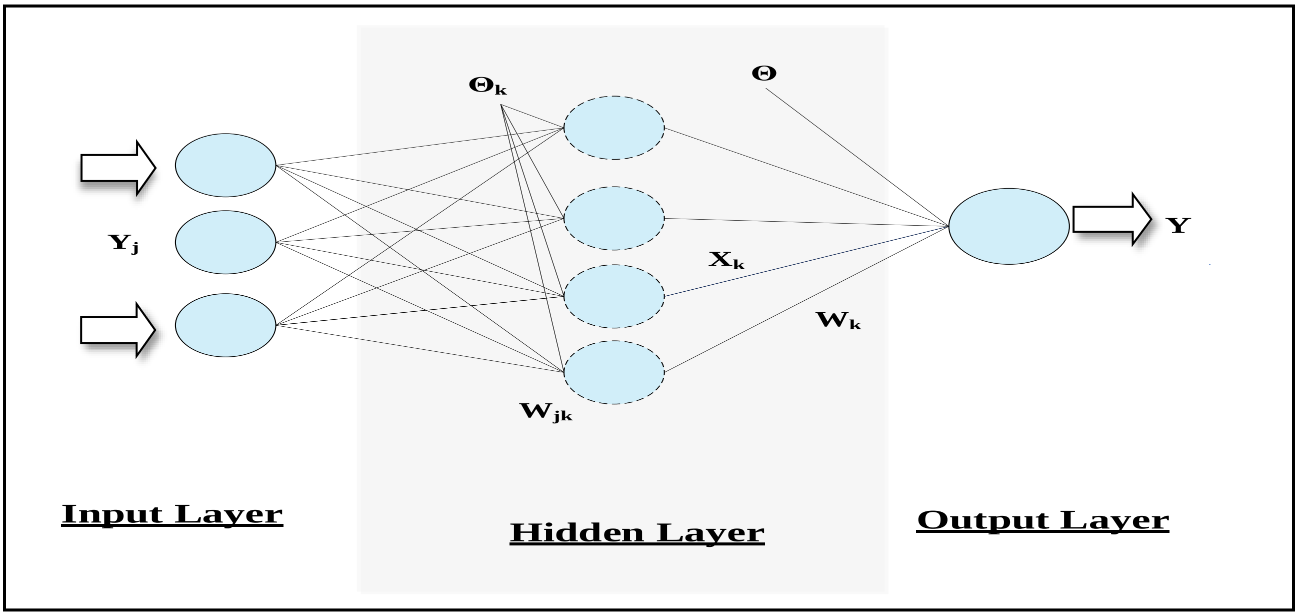

Figure 3.2 FFNN 3(4)1 model structure

The input vector is denoted by $Y_j = \{ Y_1, Y_2, Y_3\}$, $W_{jk}(j = 1 \dots 3; k = 1 \dots 4)$ represents the interconnection weight vector of the *j* neurons from the inputs to the k neurons of a hidden layer, $X_k(k = 1 \dots 4)$ signifies k neurons vector in the hidden layer, $W_k(k = 1 \dots 4)$ signifies the connection weights of the k neurons of a hidden layer to the outputs and Y represents the output vector for the model with an output node. $\Theta_k(k = 1 \dots 3, \dots k)$ represents the bias value in the output layer. The hidden layer is determined by the following formula,

$$X_k = f\left(\sum_{j=1}^{n} W_{jk} Y_J + \Theta_k \right) \dots (3.3)$$

And the output layer is determined by,

$$Y = f\left(\sum_{k=1}^{m} W_k X_k + \Theta \right) \dots (3.4)$$

Where the activation function represented by *f*.

Activation Function

The activation function defined by f represents the neuron output in terms of s the induced local field. This neuron output is used as input of the following neuron and until a desired result to the original problem is solved. It maps the output values into an anticipated range (Schmidhuber, 2015). The "***logistic activation***" function is an example of a sigmoid function used in this study, it is defined by

$$f(s) = \frac{1}{1 + exp\,(-as)} \dots (3.5)$$

Where a is the slope parameter of the sigmoid function.

Data Pre-processing

Before ANN model building and selection wheat production data should be pre-processed. Pre-processing involves data cleaning and normalisation which involves scaling of data. Normalisation is the scaling down of wheat production data to an interval of (0,1) to ensure that all input features have a similar scale. This step helps prevent certain features from dominating the learning process and ensures stable convergence during training. Common normalization techniques include min-max scaling and z-score standardization (Géron, 2019).

Training and Testing Sets

Wheat production data divided into training set and testing set. The training set or model building set of 39 data points(1975-2013) , used to develop the neural network model and the testing set of 10 data points(2014-2023) used to evaluate the forecasting ability of the model. Higher percentages usually 80 percent are basically allocated to the training group and 20 percent to the output set (Tsai & Wu, 2008).

Neural Network Model Training

ANN model training is an iterative process of adjusting the network's parameters to optimize its performance on a specific task by adjusting parameters of a model which are number of hidden layers and hidden neurons to find the best performing model (Goodfellow et al., 2016).

Backpropagation

Backpropagation also known as backward error propagation commonly used in network learning algorithm. It is a law that simplifies gradient iteration method as a form of alternating the weights of in the hidden layer of a model based on wheat production data. It applies flattened averaging to the alteration in weights at the same time avoiding local minima (Goutorbe, , 2006).

3.8.3 Model validation

Model validation techniques are used to assess the performance and effectiveness of predictive or statistical models (Hastie et al., 2009),. These techniques help determine if a model is reliable, accurate and generalizable to wheat production data. Here are some commonly used models validation techniques ,

Akaike Information Criteria and Schwarz –Bayesian Information Criteria

The model can be determined by cumulative its probability of occurring though it results in over fitting or using complications of the structural expressions of the model. The criteria of Akaike and Schwartz, impose a penalty on the number of parameters. They differ in terms of the penalty associated with increasing the order of the model. The following two equations represents AIC and BIC,

$$AIC = 2K - 2In(L) \dots (3.6)$$

$$BIC = In(n) - 2In(L) \dots (3.7)$$

Where L is the maximum likelihood function and k is the number of parameters.

Mean Absolute Error and Root Mean Squared Error

MAE refers to the measure of the average magnitude of the differences between actual and predicted values at the same time. This manner of calculation errors ignores the sign of values, in order to avoid the cancellation of negative values and positive values (Saigal & Mehrotra, 2012).

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - F_T| \dots (3.8)$$

MSE represents the inconsistency in forecast errors between the actual and predicted, it is computed as the squared difference between forecast and actual values and then averaged over the sample (RĂDULESCU & Banica, 2014).

$$MSE = \frac{1}{n}\sum_{t=1}^{n} (Y_t - F_T)^2 \dots (3.9)$$

RMSE measures the magnitude of the error.

$$RMSE = \sqrt{MSE} \dots (3.10)$$

3.9 Ethical Considerations

Access to the relevant sources of wheat production data was considered crucial and ethical considerations were made during the data collection process throughout the research. The permission was sought from the authorities through sign in using email on Index Mundi platform.

3.10 Summary

The research methodology provided a blueprint by clearly illustrating research procedures, techniques and tools used in the study. Research design, research tools, and data sources, analytical models, chosen to produce results and outputs. This chapter opens the door for the following chapter, concentrates on data presentation, analysis and discussion.

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND DISCUSSION

4.0 Introduction

To complete this study successfully in this chapter by data presentations, analysis, interpretation and discussions in a way to answer the research questions and objectives. A reflective time series analysis in wheat production was performed to obtain meaningful results, discussions and conclusions.

4.1 Summary statistics

Table 4.1 Descriptive statistics

Statistical measures	Values
Mean	172 000
Median	180 000
Mode	213 000
Standard Deviation	8 372.07
Kurtosis	-1.0059
Skewness	-0.1860
Range	291 000
Minimum	34 000
Maximum	325 000
Sum	8 428 000
Count	49

Data points of wheat production tend to cluster around the mean value in Table 4.1. Positive mean and negative skewness mean that wheat production is on a declining trend in Zimbabwe. Standard Deviation shows much variability and dispersion of wheat production due to various factors like economic, political and natural factors. Negative kurtosis shown indicates platykurtic distribution, this means that there are fewer values around the mean and more values at the tails of the distribution due to a declining trend in wheat production. The maximum and minimum values give a range value which suggests that Zimbabwe's wheat production has fluctuated significantly over the years, with a variation of nearly 10 times between the minimum and maximum.

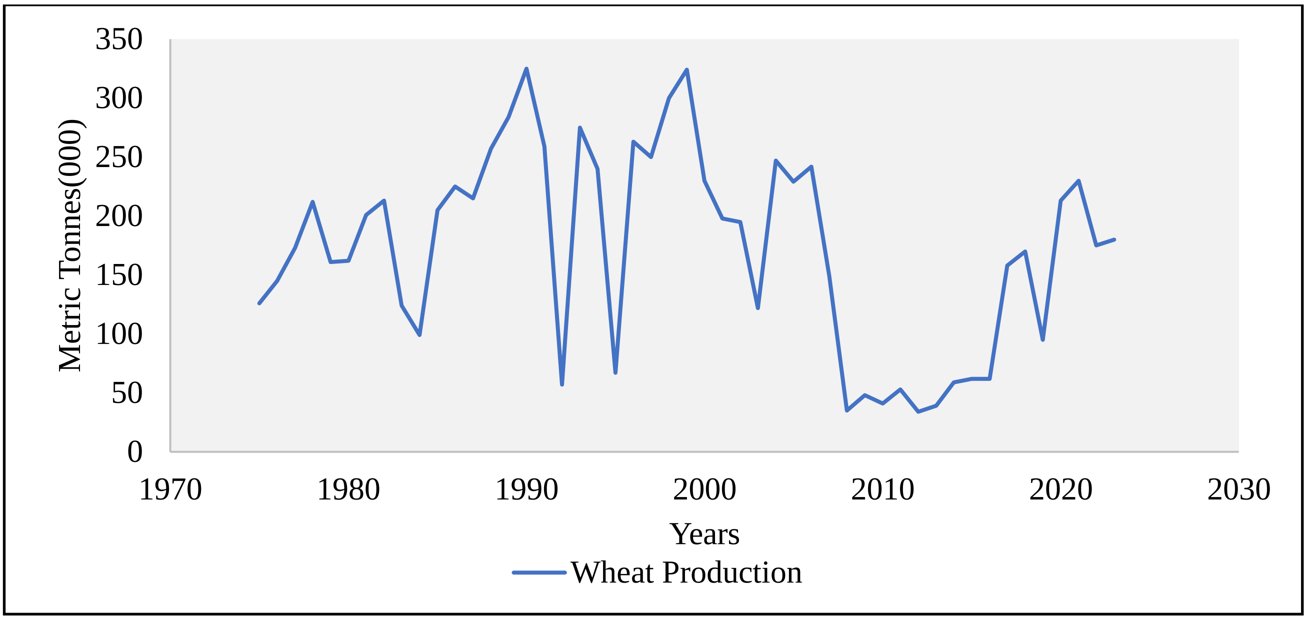

Figure 4.1 Wheat production time series plot

In Figure 4.1, wheat production in Zimbabwe shows significant fluctuations over the time period. There are multiple peaks and valleys in the production levels, indicating high volatility so the data may not be stationary, as the mean and variance of the wheat production change considerably over the years. So, stationarity test was done to check time series data for stationarity or non-stationarity using ADF test.

4.2 Pre-tests

4.2.1 Stationarity

Table 4.2 ADF test results

Augmented Dickey-Fuller Test
data: WhP Dickey-Fuller = -1.8124, Lag order = 3, p-value = 0.6469

To test for stationarity, the ADF test was executed, and it produced the p-value of 0.6469 (> 0.05), and thus rejecting the null hypothesis. This implies that the data is not stationary.

Table 1.3 Differenced data ADF test results

Augmented Dickey-Fuller Test
data: WhP_Diff1 Dickey-Fuller = -4.4837, Lag order = 3, p-value = 0.01

The study detected that after the initial differencing, the data became stationary since the p-value found was 0.01(<0.05). Hence, there was no need for second differencing as stationarity was achieved in this study.

4.3 Model Identification

Since stationarity was achieved, components of ARIMA(p,d,q) had to be identified. The correlograms and the auto-arima function used in this stage to find the best ARIMA model. ACF and PACF plots are shown in the tables below.

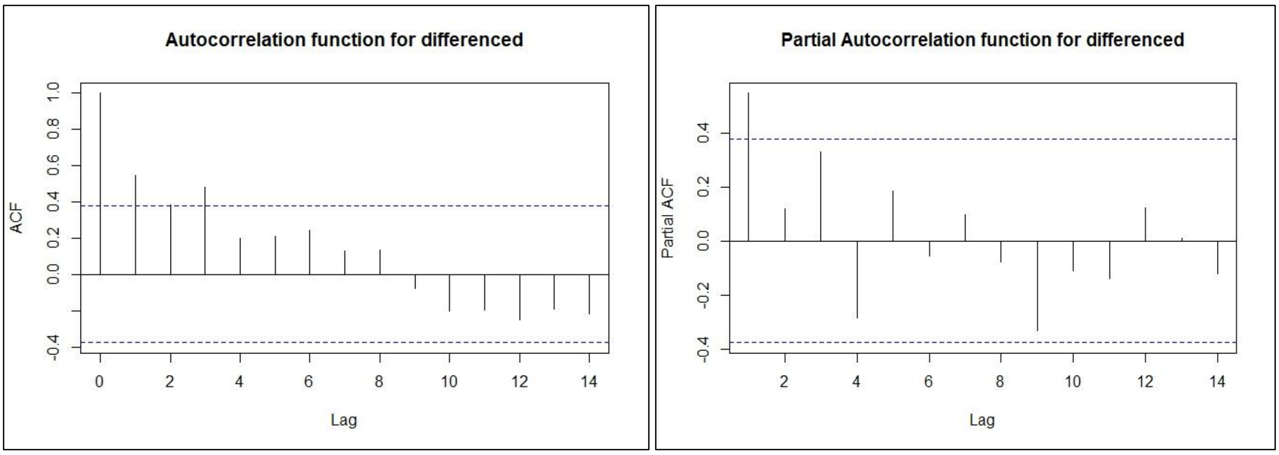

Figure 4.2 ACF and PACF of differenced data

PACF does not show or cut off any significant value excluding the point at lag 0 which should be considered showing that q=0 and p= 1 as the ACF shows off at lag 1 with the highest spike. Finally, the proposed model is ARIMA (1, 0, 0).

Table 4.4 Section of best ARIMA model

	AIC
ARIMA (2,0,2) with non-zero mean	451.8124
ARIMA (0,0,0) with non-zero mean	461.7385
ARIMA (1,0,0) with non-zero mean	450.399*
ARIMA (0,1,1) with non-zero mean	450.73
ARIMA (0,0,0) with zero mean	525.7129
ARIMA (2,0,0) with non-zero mean	452.2253
ARIMA (1,0,1) with non-zero mean	451.9071
ARIMA (2,0,1) with non-zero mean	451.7994
ARIMA (1,0,0) with zero mean	457.0237
Best model: ARIMA (1,0,0) highlighted by * with the lowest AIC value.	

The auto-arima function in R-software evaluated several ARIMA model variants, including models with different orders of autoregressive (p), differencing (d), and moving average (q) terms, as well as models with and without a non-zero mean. The AIC was used to assess the relative quality of the different ARIMA models. The ARIMA (1,0,0) model considered the best fit since it has the lowest AIC value.

4.4 Parameter Estimation.

Table 4.5 Parameter Estimation

Model Information: Series: WhP ARIMA (1,0,0) with non-zero mean		
Coefficients: ar1 mean 0.5529 174.2868 Standard error: 0.1366 25.1156 Sigma-squared = 5433: log likelihood = -222.2		
AIC=450.4	AICc=451.08	BIC=455.39

Parameter estimation meant defining components of AR and MA terms chosen in the best model, model of this study has p=1 for AR without MA since q=0 so the best fit model has Autoregressive with a non-zero mean coefficients only. Also, AIC has the lowest value considered in model identification as compared to BIC which has a high value.

4.5 Model Diagnosis

4.5.1 Test for Normality

Histogram of Residuals

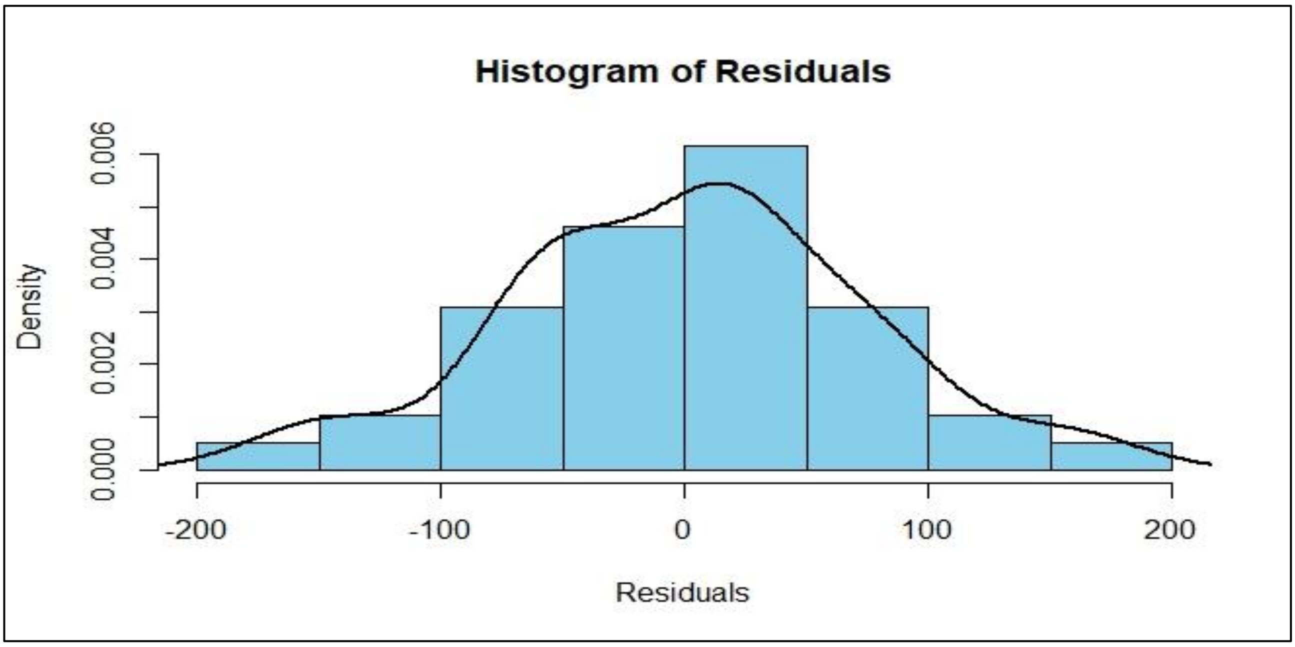

Figure 4.3 Histogram of Residuals

Figure 4.3 shows a histogram of the residuals, which are the differences between the observed values and the predicted or fitted values in a model. The histogram exhibits a bell-shaped, or close to normal distribution. This suggests that the residuals are following a normal probability distribution and is also approximately symmetric around the vertical axis at 0 on the x-axis, indicating that the residuals are evenly distributed around the mean or expected value of 0.

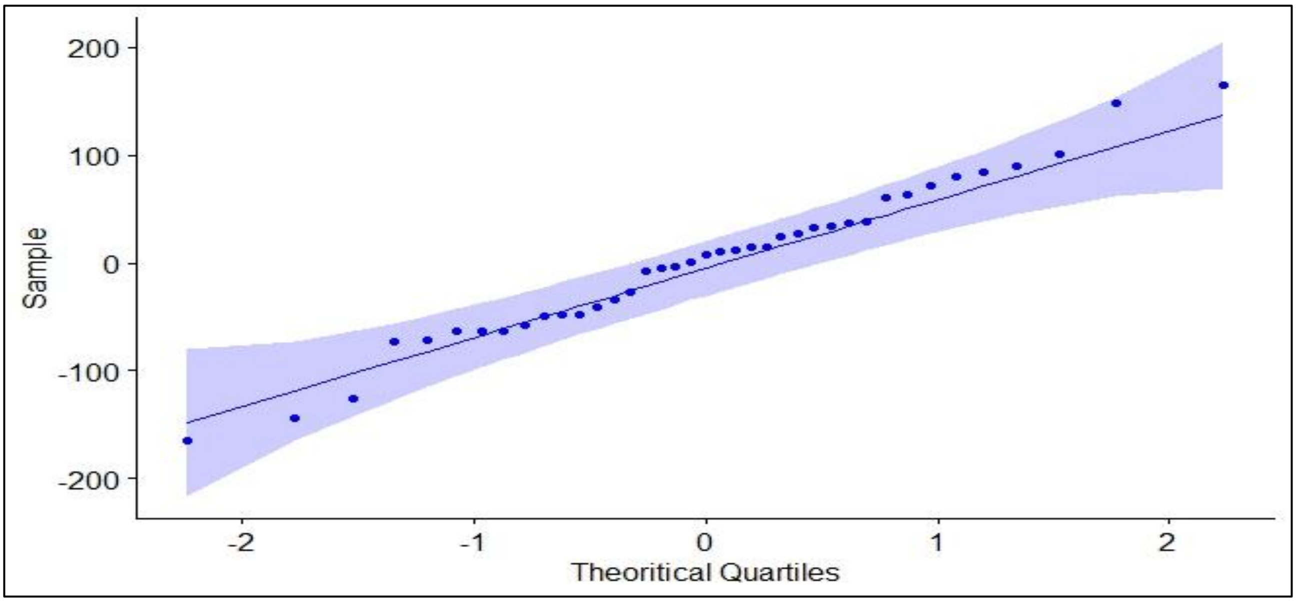

Figure 4.4 A Q-Q plot of residuals

The normal Q-Q shown determines normality of residuals by plotting quantiles (i.e., percentiles). Figure 4.4 shows a distribution which likely follows a normal distribution as shown by data points which are almost linear. Thus, the data points follow a linear trend.

4.5.3 Test for Independence

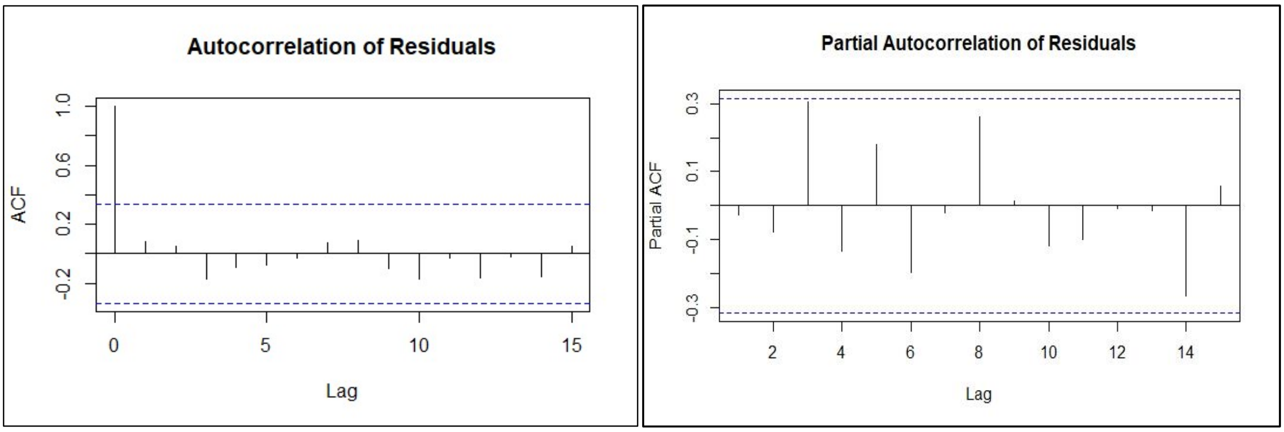

Figure 4.5 ACF and PACF of residuals

Correlograms in Figure 4.5, do not show any structure of arrangement which settles no serial autocorrelations and no significant spikes that cut the blue dotted lines from lag 1 to 15 showing that the fitted model has residuals that are independent and not interrelated in the model.

4.5.4 Homoskedasticity test

Table 4.6 Homoscedasticity test

White Neural Network Test
data: Residuals X-squared = 11.823, df = 2, p-value = 0.0027

The null hypothesis of the White's test is that the residuals have constant variance (homoscedasticity). The alternative hypothesis is that the residuals exhibit heteroscedasticity (non-constant variance). Since the p-value (0.002708) is less than 0.05, this study accepted the null hypothesis. This shows evidence of homoscedasticity of residuals in the model.

4.6 Forecasting from ARIMA (1,0,0)

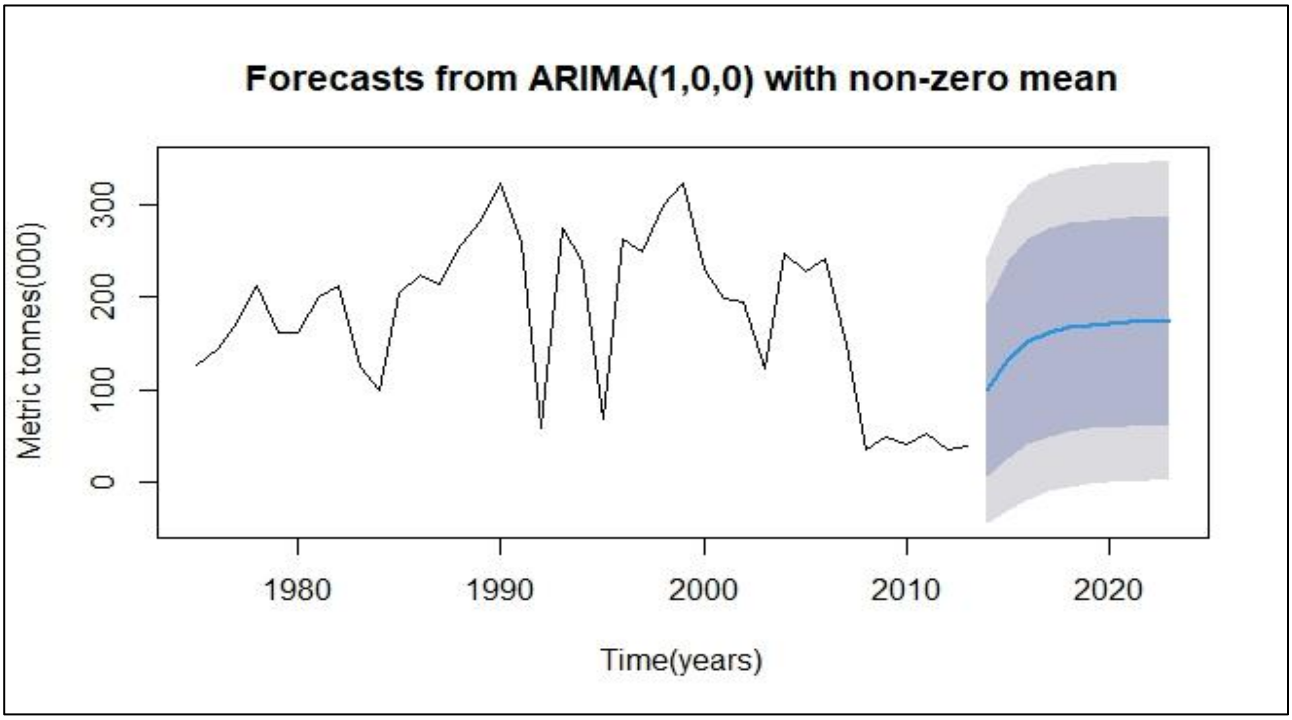

Figure 4.6 Forecasts from ARIMA (1,0,0)

ARIMA (1,0,0) model in-sample forecasts made from 2014-2023 in order to have the predicted values of wheat production, used for the comparisons of actual data. Figure 4.6 shows the forecasted values indicated by a blue line and grey shades indicates confidence intervals for wheat production.

Table 4.7 Predicted values

Year	Point Forecasts (t)	Lo 95(t)	Hi 95(t)
2014	99 490.79	-44 976.87	243 958.53
2015	132 934.33	-32 142.66	298 011.33
2016	151 424.26	-19 457.03	322 305.51
2017	161 646.78	-10 969.74	334 263.35
2018	167 298.51	-5 844.91	340 441.96
2019	170 423.18	-2 880.14	343 727.32
2020	172 150.72	-1 202.64	345 504.01
2021	173 105.82	- 262.54	346 474.13
2022	173 633.87	261.02	347 006.74
2023	173 925.81	551.51	347 300.12

The point forecasts show an increasing trend over the years from 2014 to 2023, suggesting an overall upward movement in the time series. The rate of increase appears to be decreasing slightly over time, as the year-over-year changes in the point forecasts are smaller in the later years from 2020 to 2023.

4.7 FFNN Models Building and Selection

Data Pre-processing

Pre-processing was done for FFNN models through normalisation, using min-max normalisation on R Studio in order to avoid overload or saturation of hidden nodes in model building of the neural network model. Table 4.8 below shows normalised time series data.

Table 4.8 Normalised data

Time Series: Start = 1 End = 49 Frequency = 1
0.316151203 0.381443299 0.477663230 0.611683849 0.436426117 0.439862543 0.573883162 0.615120275 0.309278351 0.223367698 0.587628866 0.656357388 0.621993127 0.766323024 0.859106529 1.000000000 0.773195876 0.079037801 0.828178694 0.707903780 0.113402062 0.786941581 0.742268041 0.914089347 0.996563574 0.673539519 0.563573883 0.553264605 0.302405498 0.731958763 0.670103093 0.714776632 0.395189003 0.003436426 0.048109966 0.024054983 0.065292096 0.000000000 0.017182131 0.085910653 0.096219931 0.096219931 0.426116838 0.467353952 0.209621993 0.615120275 0.673539519 0.484536082 0.501718213

Neural Network Model Training and Testing Sets

When the data has been normalised, the network model building process starts. The processed data in Table 4.9 divided into two sets, namely, training set of 80 percent of total data and testing set of 2 0 percent.

Table 4.9 Training and Testing sets

> training_ data (80%)
0.316151203 0.381443299 0.477663230 0.611683849 0.436426117 0.439862543 0.573883162 0.615120275 0.309278351 0.223367698 0.587628866 0.656357388 0.621993127 0.766323024 0.859106529 1.000000000 0.773195876 0.079037801 0.828178694 0.707903780 0.113402062 0.786941581 0.742268041 0.914089347 0.996563574 0.673539519 0.563573883 0.553264605 0.302405498 0.731958763 0.670103093 0.714776632 0.395189003 0.003436426 0.048109966 0.024054983 0.065292096 0.000000000 0.017182131
> testing_ data (20%)
0.08591065 0.09621993 0.09621993 0.42611684 0.46735395 0.20962199 0.61512027 0.67353952 0.48453608 0.50171821

4.8 FFNN Model Building and Selection

Table 4.10 FNN Models

		Forecasts (t)		
Year	**Actual (t)**	**FFNN model 1 1(4)1**	**FFNN model 2 1(5,4)1**	**FFNN model 3 1(5)1**
2014	59 000	96 634.94	96 820.14	95 986.56
2015	62 000	98 587.25	98 383.06	99 344.14
2016	62 000	98 587.26	98 383.01	99 344.14
2017	158 000	159 961.06	156 922.15	158 588.21
2018	170 000	167 252.73	164 790.73	165 833.93
2019	95 000	120 048.74	116 888.07	119 817.88
2020	213 000	192 189.33	192 099.69	191 147.76
2021	230 000	201 456.52	202 165.63	200 800.45
2022	175 000	170 251.93	168 054.52	168 832.97
2023	180 000	173 226.84	171 303.54	174 819.12
	MAE	**32.5678**	**33.5999**	**31.9721***
	RMSE	**43.6789**	**44.6943**	**42.6754***

Note * represents the minimum value of the performance metrics over all models

Hidden layers and nodes varied in ANN models building by changing hidden layer and nodes parameters to obtain the best models shown in Table 4.10 but the input and output layers remain constant, coming up with three different structures of FFNN models. Out of 3, FFNN model 3 of the structure 1(5)1 considered to be the best because of the lowest values of MAE and RMSE showing that a model with a single hidden layer and 5 nodes performs better in this research as compared to a model of two hidden layers of 5 and 4 nodes and one hidden layer of 4 nodes.

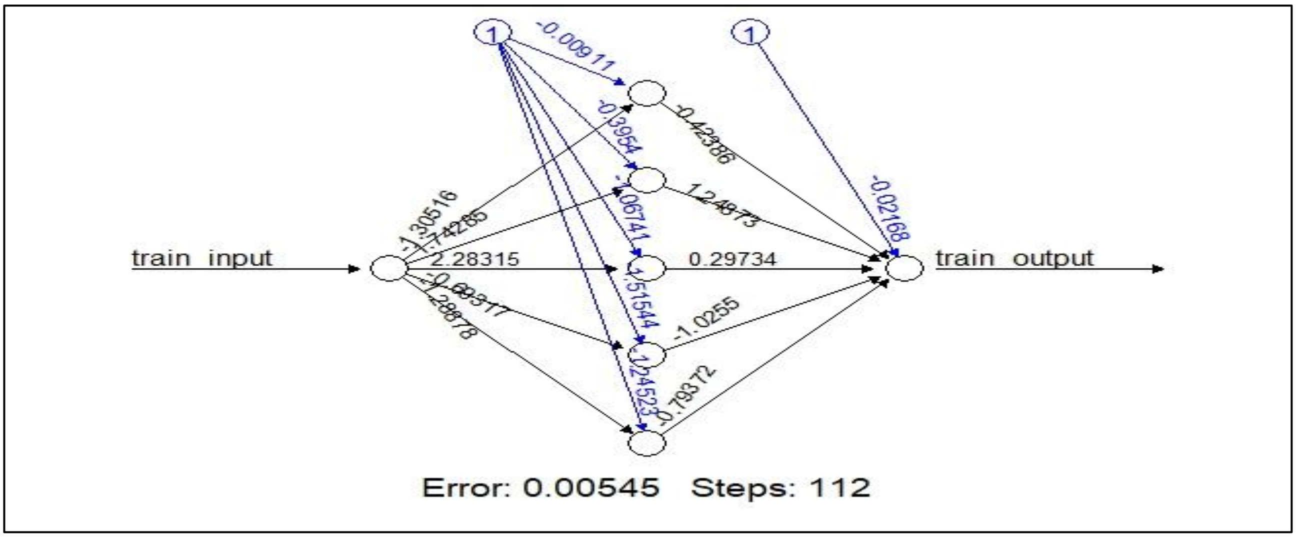

Figure 4.7 FFNN 1(5)1 model structure

The input layer is labelled 'train input', hidden layer with 5 neurons and the output layer is labelled "train output". The connections between the layers are represented by weighted edges, and the numerical values shown on the edges represent the weights of the connections. The network has gone through some training process, as indicated by the error value of 0.00545 and the number of training steps as 112.

4.9 Comparison of the Neural Network and ARIMA Models.

Table 4.11 ARIMA and FFNN models

		Forecasts (t)	
Year	**Actual (t)**	**ARIMA (1,0,0)**	**FFNN 1(5)1**
2014	59 000	99 490.79	95 986.56
2015	62 000	132 934.33	99 344.14
2016	62 000	151 424.26	99 344.14
2017	158 000	161 646.78	158 588.21
2018	170 000	167 298.51	165 833.93
2019	95 000	170 423.18	119 817.88
2020	213 000	172 150.72	191 147.76
2021	230 000	173 105.82	200 800.45
2022	175 000	173 633.87	168 832.97
2023	180 000	173 925.81	174 819.12
	MAE	**56.5831**	**31.9721***
	RMSE	**71.7944**	**42.6754***

Note * represents the minimum value of the performance metrics over all models

After best model selection from FFNN and ARIMA models. The performance of each model assessed using MAE and RMSE as the performance measures. As shown in Table 4.11 FFNN model has the lowest MAE and RMSE values considered to be the best model in this study.

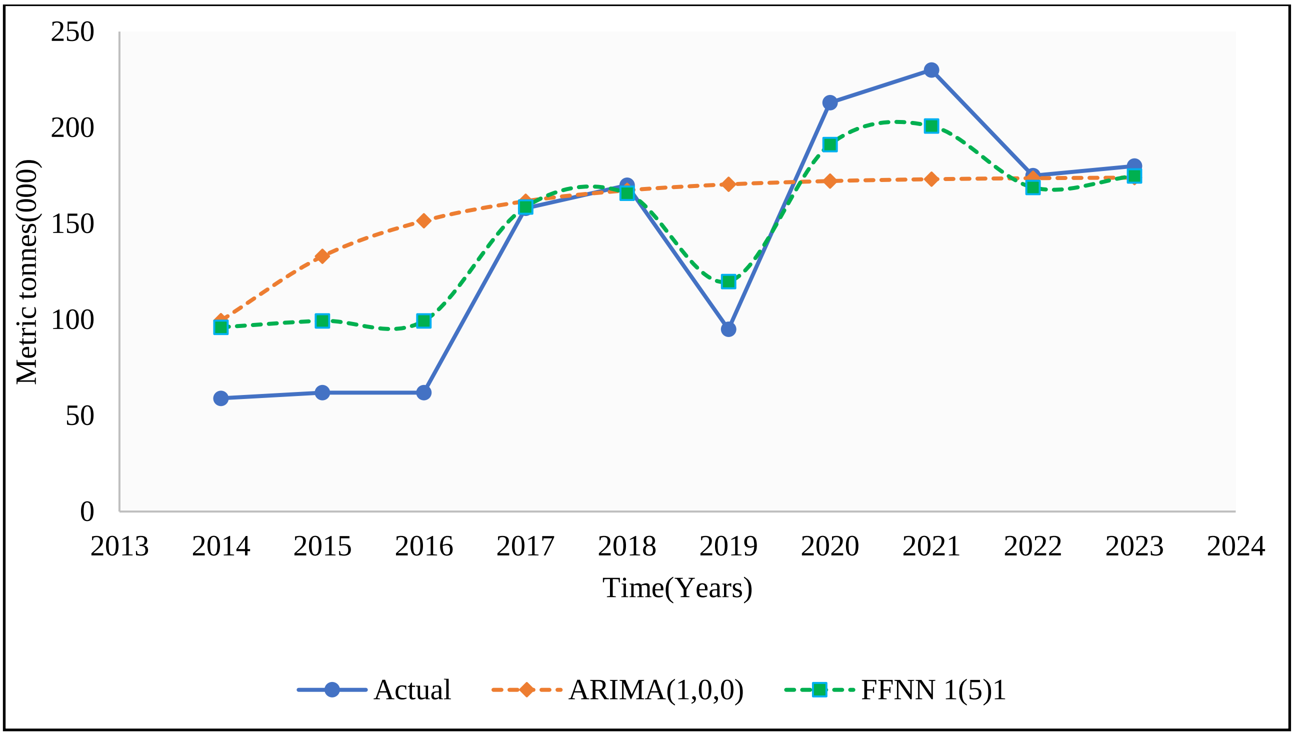

Figure 4.8 Line graph of actual and forecasts values from ARIMA and ANN models

For further model comparison a line graph of three patterns plotted showing the actual and predicted values of ARIMA and FFNN best models. The pattern of the FFNN model mimics the actual values of wheat production as compared to ARIMA which depicted a directional linear pattern. The FFNN detected hidden patterns wheat production data showing that FFNN is considered to be the best model. The ARIMA (1,0,0) model forecast follows a similar trend to the actual values, but it tends to underestimate the peak in 2019 and overestimate the values in the later years. The FFNN 1(5)1 model forecast also follows a similar trend to the actual values, but it more closely matches the peak in 2019 and the subsequent decline in the later years, appears to provide a more accurate forecast compared to the ARIMA (1,0,0) model.

4.10 Wheat Production Forecast for 2024-2028

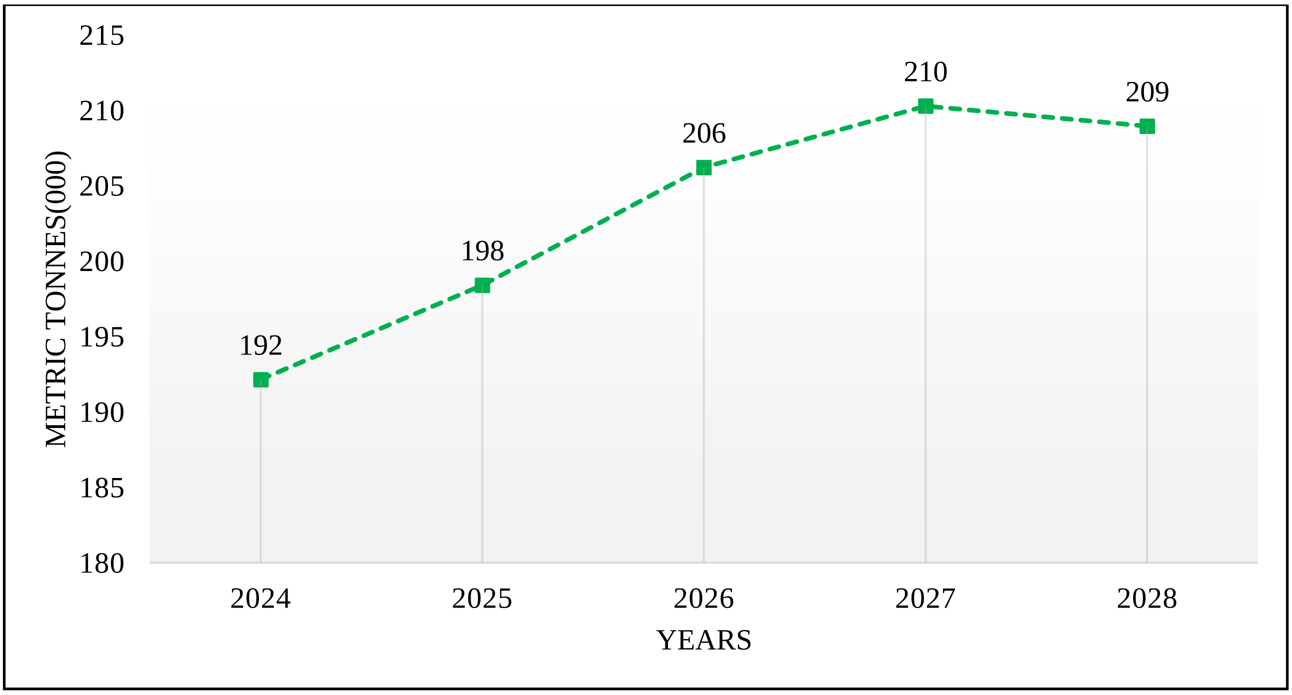

Figure 4.9 Forecasts of wheat production 2024-2028 from FFNN model

The selected FFNN model used to forecast the wheat production for 2024-2028. Figure 4.9 showed that wheat production has the potential to increase as Figure 4.9 displayed an increasing trend to be expected in wheat production therefore correct effective policies and decisions should be implemented by relevant key stakeholders. The metric values show a consistent upward trend, increasing from 192 000 in 2024 to 209 000 in 2028. The year-over-year increase in the metric values appears to be relatively consistent. By 2028, the projected metric value reaches 209 000, representing a significant increase from the starting point in 2024, suggesting a positive outlook for wheat production in Zimbabwe, with a sustained upward trajectory over the next 5-year period.

4.11 Discussion of findings

In this study ARIMA and ANN models were developed based on wheat production and the performance of the models were compared using MAE and RMSE, considering mostly the MAE as cited by Aras & Kocakoç, (2016). ANN models overcome the ARIMA models since the ARIMA models failed to mimic the actual values of wheat production pattern as compared to ANN which detected and mimicked the complexity of nonlinearity pattern of wheat production resulted in accurate forecasts. These findings buttress with findings from Kożuch, et al, (2023) and Awe and Dias, (2022) where ANN comes out as the best model because it has predictive and computational power as compared to traditional ARIMA models.

Forescasts from ANN model shows an upward trend to be expected from 2024 to 2028, therefore the relationship between inputs such as labour and capital and outputs should be considered in wheat production process conjuring with Cobb-Douglas production function of 1928. Additionaly the consideration of inputs and costs in relation to forecasts, helps policymakers, government, farmers and other stakeholders to identify the optimal level of inputs to achieve the desired level of production ouput agreeing with a study done by Mutambara and others in 2013.

4.12 Summary

Various tests needed for the models to be considered fit were carried out and presented in this chapter. FFNN and ARIMA models were developed for time series analysis. The comparison of the two models was done and FFNN proved to be more competitive as compared to ARIMA. Forecasts of wheat production for the next five years done, showing an upward trend over the next 5 years. The next chapter discuss the final conclusions and recommendations.

CHAPTER 5: FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter is a final chapter of this study where summary of findings, conclusion, recommendations, areas for further research and finally the chapter summary based on time series analysis of wheat production in Zimbabwe were discussed. This study provided useful and interesting ideas and facts associated with time series analysis of wheat production by coming up with a best model and projected wheat production in the next five years.

5.1 Summary of findings

Primary objectives of this study hinged on the development of time series models based on ARIMA and ANN, determination of the model that has the greatest ability in forecasting wheat production between ARIMA and ANN models, finally forecasting wheat production pattern from 2024-2028 using the best model. Firstly, time series analysis models based on ARIMA and FFNN were developed and the comparisons between ARIMA and ANN done using evaluation and validation metrics MAE and RMSE. As a result, ANN comes out as the best model with no assumptions required in model building but with greater predictive and computational power as compared to ARIMA model which is also competitive and requires assumptions to be met first.

The last objective was achieved since the study came up with the forecasts of 2024-2028 using the best selected ANN model. An upward trend to be expected from 2024-2028, meaning that general trend of wheat production is increasing which is a good indicator to the agricultural sector. Therefore, the government of Zimbabwe should keep enacting best policies and decisions embarked on many agricultural programmes with the aim of increasing wheat output and decreasing wheat production inputs costs on fertilizers and irrigation by understanding the relationship between inputs such as labour and capital and outputs in a production process as cited by Cobb & Douglas, (1928), resulted in improving the agricultural sector and fulfilling national goals such Vision 2030 in Zimbabwe.

5.2 Conclusions

The study concluded that ANN and ARIMA models were developed based on wheat production data and compared using performance metrics. ANN surpasses ARIMA models in forecasting wheat production. Finally, the best selected ANN model predicted an upward trend from 2024 to 2028 in wheat production which suggested that Zimbabwe has a potential to further increase wheat production.

5.3 Recommendations

Based on the best selected ANN model and projected upward trend in wheat production, potential actions should be done by the government, policymakers, researchers, and other stakeholders.

a. The use of ANN models for accurate forecasts

ANN methods are recommended in this study in forecasting in agricultural outputs because of their predictive power and accuracy for time series data. In the modern world data have nonlinearity features which needs application of nonlinear techniques such as ANN. These nonlinearity models overwhelmed the restriction as they can detect the nonlinear features of data and ANN received recommendations from several studies like Awe and Dias, (2022) and Kumar and Garg, (2018)

b. Increase investment and support

An increase in wheat production over the next five years, suggesting that Zimbabwe has the potential to further expand its wheat output. The government and relevant agricultural agencies could look into providing more financial and technical support to wheat farmers, such as subsidies, improved infrastructure, timely provision on inputs, improved irrigation systems and access to modern farming techniques and technology in order to achieve the country's Sustainable Development Goals (SDG) of being a middle-income country by 2030.

c. Engage with wheat farmers

Regular dialogue with farmers, understanding their challenges, and providing them with the necessary training and extension services can help ensure that the proposed interventions are tailored to their needs and capabilities.

5.4 Areas for further research

The study is not a conclusive one, rather it created a stage for further future studies. The study stretched focus on time series analysis of wheat production in Zimbabwe other studies can focus on other crops or other sectors such as mining, finance and transport but using other ANN models such Recurrent Neural Networks (RNN) to check if they can get the same results.

5.5 Summary

The objectives and questions of the research are answered in this chapter. Summary of findings and conclusions are also part of this chapter discussion. Discussed in the summary is the answers to this research objectives. The recommendations to all the stakeholders were also discussed in this chapter and finally areas for further research.

REFERENCES

Aslam, F., Salman, A., & Jan, I. (2019). Predicting wheat production in Pakistan by using an artificial neural network approach. *Sarhad J. Agric, 35*, 1054–1062.

Awe, O. O., & Dias, R. (2022). Comparative analysis of Arima and artificial neural network techniques for forecasting non-stationary agricultural output time series. *AGRIS on-line Papers in Economics and Informatics, 14*, 3–9.

Awe, O., Okeyinka, A., & Fatokun, J. O. (2020). An alternative algorithm for ARIMA model selection. *2020 international conference in mathematics, computer engineering and computer science (ICMCECS)*, (pp. 1–4).

Babbie, E. R. (2020). *The practice of social research.* Cengage AU.

Cheng, B., & Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical science*, 2–30.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production.

Dzvimbo, M. A., Monga, M., & Mashizha, T. M. (2017). The link between rural institutions and rural development: Reflections on smallholder farmers and donors in Zimbabwe. ***Journal of Humanities and Social Science****, 22*, 46–53.

Edbrooke, J. (2017). *Time Series Modelling Technique Analysis for Enterprise Stress Testing.* Ph.D. dissertation, Thesis, Imperial College London.

FAO. (2021). *Wheat Production* . Rome: Food and Agriculture of the United Nations.

Géron, A. (2019). Hands-on machine learning with scikit-learn, keras, and tensorflow: concepts. *Aurélien Géron-Google Kitaplar, yy https://books. google. com. tr/books*.

Ghodsi, R., Yani, R. M., Jalali, R., & Ruzbahman, M. (2012). Predicting wheat production in Iran using an artificial neural networks approach. ***International Journal of Academic Research in Business and Social Sciences****, 2*, 34.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT press.

Goutorbe, B., Lucazeau, F., & Bonneville, A. (2006). Using neural networks to predict thermal conductivity from geophysical well logs. ***Geophysical Journal International,*** *166*, 115–125.

Herliansyah, R., & others. (2017). Feed Forward Neural Networks for Forecasting Indonesia Exchange Composite Index. ***GSTF Journal of Mathematics, Statistics & Operations Research****, 4*.

Idrees, S. M., Alam, M. A., & Agarwal, P. (2019). A prediction approach for stock market volatility based on time series data. *IEEE Access, 7*, 17287–17298.

IndexMundi. (2023). *Wheat production by year*. Haryana: https://www.indexmundi.com/agriculture/?country=zw&commodity=wheat&graph=production.

Kapuya, T., Saruchera, D., Jongwe, A., Mucheri, T., Mujeyi, K., Ndobongo, L. T., & Meyer, F. H. (2010). The grain industry value chain in Zimbabwe. *Unpublished draft prepared for the food and agricultural organization FAO. Retrieved from www. fao. org/fileadmin/templates/est/AAACP/astafrica/UnvPretoria_GrainChainZimbabwe_2010_1_. pdf*.

Karami, A. (2010). Estimation of the critical clearing time using MLP and RBF neural networks. *European Transactions on Electrical Power, 20*, 206–217.

Kharin, S. (2018). Price Transmission Analysis: The case of milk products in Russia. *AGRIS on-line Papers in Economics and Informatics, 10*, 15–23.

Kożuch, A., Cywicka, D., & Adamowicz, K. (2023). A comparison of artificial neural network and time series models for timber price forecasting. *Forests, 14,* 177.

Kumar, V., & Garg, M. L. (2018). Predictive analytics: a review of trends and techniques. ***International Journal of Computer Applications****, 182*, 31–37.

Mapuwei, T. W., Ndava, J., Kachaka, M., & Kusotera, B. (2022). An Application of Time Series ARIMA Forecasting Model for Predicting Tobacco Production in Zimbabwe. *American* ***Journal of Modeling and Optimization****, 9*, 15–22.

Mensi, W., Tiwari, A., Bouri, E., Roubaud, D., & Al-Yahyaee, K. H. (2017). The dependence structure across oil, wheat, and corn: A wavelet-based copula approach using implied volatility indexes. *Energy Economics, 66*, 122–139.

Mithiya, D., Datta, L., & Mandal, K. (2019). Time series analysis and forecasting of oilseeds production in India: using autoregressive integrated moving average and group method of data handling–neural network. *Asian J. Agric. Ext. Econ. Sociol, 30*, 1–14.

Mitrea, C. A., Lee, C. K., & Wu, Z. (2009). A comparison between neural networks and traditional forecasting methods: A case study. ***International journal of engineering business management,*** *1*, 11.

Mudavanhu, C., & Mandizvidza, C. (2013). Sustaining rural livelihoods through donor funded agricultural inputs scheme in Zimbabwe: The case of Goromonzi district.

Mundi, I. (2023). *Wheat production by year.* Haryana: http//.

Mutambara, J., Zvinavashe, A. P., & Mwakiwa, E. (2013). A critical review of the wheat industry in Zimbabwe. *GJ BAHS, 2*, 23–33.

Negassa, A., Shiferaw, B., Koo, J., Sonder, K., Smale, M., Braun, H. J., . . . others. (2013). The potential for wheat production in Africa: analysis of biophysical suitability and economic profitability.

Nehra, N., Sangwan, P., & Kumar, D. (2021). Artificial neural networks: a comprehensive review. *Handbook of Machine Learning for Computational Optimization*, 203–227.

RĂDULESCU, M., & Banica, L. (2014). Neural networks-based forecasting regarding the convergence process of CEE countries to the Eurozone. *Transylvanian Review of Administrative Sciences, 10*, 225–246.

Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications, 42*, 3234–3241.

Safa, M., Samarasinghe, S., & Nejat, M. (2015). Prediction of wheat production using artificial neural networks and investigating indirect factors affecting it: case study in Canterbury province, New Zealand. ***Journal of Agricultural Science and Technology****, 17*, 791–803.

Saigal, S., & Mehrotra, D. (2012). Performance comparison of time series data using predictive data mining techniques. *Advances in Information Mining, 4*, 57–66.

Sankar, T. J., & Pushpa, P. (2019). Design and development of time series analysis for Saccharum officinarum production in India. ***Journal of Composition Theory****, 12*, 203–211.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85–117.

Sözen, A. (2009). Future projection of the energy dependency of Turkey using artificial neural network. *Energy policy, 37*, 4827–4833.

Tealab, A. (2018). Time series forecasting using artificial neural networks methodologies: A systematic review. ***Future Computing and Informatics Journal**, 3*, 334–340.

Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications, 34*, 2639–2649.

Wheelwright, S. C., & Hyndman, R. J. (2016). *Forecasting methods and applications.* John wiley & sons.

Weil, J. (2017). *Research design in aging and social gerontology: Quantitative, qualitative, and mixed methods.* Routledge.

ZIMSTATS. (2012). Agriculture of Zimbabwe. *Crop production in Zimbabwe,* 6-15.

APPENDICES

APPENDIX A: R Studio codes for FeedForward Neural Networks

```
# Load required packages
library(neuralnet)
library(forecast)
library(nnetar)
# 'ts_data' is the univariate time series data
ts_data=ts(WheatSF$WhP)
# Normalization of  time series data
normalized_data= (ts_data - min(ts_data)) / (max(ts_data) - min(ts_data))
# Splitting data into training and test sets
train_percentage <- 0.8
train_size <- round(train_percentage * length(normalized_data))
train_data <- normalized_data[1:train_size]
test_data <- normalized_data[(train_size + 1):length(normalized_data)]
# Creation of  lagged representations of the time series as input features
create_lagged_data <- function(data, lag) {
  lagged_data <- c()
  for (i in lag:length(data)) {lagged_data <- c(lagged_data, data[(i - lag + 1):i])}
  return(matrix(lagged_data, ncol = lag))}
# Number of lagged time steps to use as input features
lag <- 1
# Lagged input features for training and test data
train_input <- create_lagged_data(train_data, lag)
train_input
test_input <- create_lagged_data(test_data, lag)
test_input
#Target variable for training and test data
train_output <- train_data[(lag + 0):length(train_data)]
test_output <- test_data[(lag +0):length(test_data)]
# Train the feedforward neural network
nnmodel <- neuralnet(train_output ~ train_input, data = data.frame(train_input, train_output), hidden = c(5,5),stepmax=150, linear.output = TRUE)
```

```
# Make predictions on the test data
Predictions=neuralnet:::predict(nnmodel,test_input)
# Denormalize the predicted values
predicted_values <- Predictions * mean(ts_data) + sd(ts_data)
predicted_values
# Checking and evaluating the model accuracy
mae <- mean(abs(Predictions - test_output*mean(ts_data)+sd(ts_data)))
rmse <- sqrt(mean((Predictions-test_output* mean(ts_data) + sd(ts_data) )^2))
#Forecasting
model=nnetar(ts_data)
Forecasts=forecast(model,h=5)
Forecasts
```

APPENDIX B: R Studio codes for ARIMA models

```
library(tseries)
library(forecast)
library(ggpubr)
library(lmtest)
#Converting to time series data
WhP=ts(WheatSF$WhP,start = 1975,end= 2013,frequency = 1)
plot(WhP,xlab='Time(Years)',ylab='Metrics tonnes(000)',main='Wheat production in Zimbabwe',type ='o',pch = 20)
#Stationarity test
adf.test(WhP)
WhP_Diff1=diff(WhP)
adf.test(WhP_Diff1)
#Model building
acf(WhP_Diff1,main = 'Autocorrelation function for differenced')
pacf(WhP_Diff1,main = 'Partial Autocorrelation function for differenced')
WhPmodel=auto.arima(WhP,ic='aic',trace =TRUE)
summary(WhPmodel)
#Diagnostics tests
Residuals=residuals(WhPmodel)
acf(Residuals,main='Autocorrelation of Residuals')
pacf(Residuals,main='Partial Autocorrelation of Residuals')
hist(Residuals, col = "skyblue", main = "Histogram of Residuals", xlab = "Residuals", probability = TRUE)
lines(density(Residuals), col = "black", lwd = 2)
ggqqplot(Residuals,color = ('blue'),xlab='Theoritical Quartiles',type ='o',pch=15)
white.test(Residuals)
#Forecasting
Forecasts=forecast(WhPmodel,h=10)
Summary(Forecasts)
```