# BINDURA UNIVERSITY OF SCIENCE EDUCATION

## FACULTY OF SCIENCE AND ENGINEERING

**Department of Mathematics and Statistics**

# Time series analysis for nickel production. A case study for Trojan Nickel Mine

**BY**

**Sean Mdala**

**A PROJECT SUBMITED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF BACHELOR OF SCIENCE HONORS DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS**

**SUPERVISOR: Dr M. Magodora**

**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.

| | | |
|---|---|---|
| Sean Mdala | | 10-06-2024 |
| B200545B | Signature | Date |
| | | |
| Dr. M. Magodora | | 12-06-2024 |
| Supervisor | Signature | Date |
| | | |
| | | 12-06-2024 |
| Dr. M. Magodora | | |
| Chairperson | Signature | Date |

## DEDICATION

In loving memory of my late grandmother Rhodha Allan. Her enduring legacy and the values she instilled in me have been a profound source of inspiration. Though she is no longer with us physically, her spirit continues to guide and motivate me. This work is a tribute to her life and the profound impact she had on mine

**ACKNOWLEDGEMENTS**

First and foremost, I would like to express my deepest gratitude to my Lord and personal Savior, Jesus Christ, for His guidance, strength, and blessings throughout this journey. Without His grace and mercy, this accomplishment would not have been possible. I also wish to give special thanks to two remarkable women, Mrs and Mrs  Mdala, who have been a constant source of inspiration and motivation. Their unwavering faith in my potential and consistent encouragement to stay positive have been invaluable to me. I extend my heartfelt thanks to my supervisor, Dr. M Magodora, for his invaluable support, insightful feedback, and encouragement throughout the course of this study. His expertise and dedication have been instrumental in shaping this research. Additionally, I am profoundly grateful to my classmates for their continuous support, collaboration, and camaraderie. Their shared knowledge and constructive discussions have greatly enriched this work. To all who have contributed to this study in various ways, your support and encouragement have been invaluable. Thank you.

## ABSTRACT

The Trojan mining operation is crucial to the national economy as a major nickel producer, but recent trends show declining yields. To address this, a comprehensive study conducted a time series analysis of nickel production from April 2013 to March 2023, using ARIMA models to predict production from April 2023 to January 2024. The ARIMA model projected that nickel production would reach 496.87 tons by January 2024. Using the Box-Jenkins approach and R software, the study validated the ARIMA (1, 0, 1) model. Predictions showed an initial increase in production from April to July 2023, followed by stabilization with slight fluctuations until January 2024. Boosting nickel production depends on timely resource provision, professional training, effective mine management, and supportive government policies. Future research should consider advanced data analytics methods, such as machine learning and neural networks, for more precise predictions.

Keywords: ARIMA, Box-Jenkins, Forecast, Nickel Production, Zimbabwe

## Contents

## LIST OF TABLES

## LIST OF FIGURES

## ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| A-D Test | Anderson-Darling Test |
| ACF | Autocorrelation Function |
| ADF | Augmented Dickey-Fuller |
| AIC | Akaike Information Criteria |
| AR | Autoregressive |
| ARIMA | Autoregressive Integrated Moving Average |
| ARMA | Autoregressive Moving Average |
| BIC | Bayesian Information Criteria |
| FDI | Foreign Direct Investment |
| GDP | Gross Domestic Product |
| IID | Independently Identically Distributed |
| MA | Moving Average |
| MAPE | Mean Absolute Percentage Error |
| MSE | Mean Square Error |
| MSR | Mean Square Residual |
| PACF | Partial Autocorrelation Function |
| RMSE | Root Mean Square Errors |
| ZIMSTAT | Zimbabwe National Statistic Agency |

# CHAPTER 1

## 1.0 Introduction

Mining has been a cornerstone of Zimbabwe's economy, contributing significantly to the nation's GDP. It also contributes to foreign exchange, government revenues, capital formation, and infrastructure development. According to the 1990 Geological Survey, Zimbabwe is a country rich in a variety of minerals and metals. It has substantial reserves of platinum, gold, diamonds, asbestos, nickel, coal, chrome, iron, and numerous other minerals. These valuable resources hold immense prospects for stimulating economic growth and enhancing overall economic development.

However, despite the importance of these minerals in Zimbabwe, their potential can be overshadowed if mismanaged. In this context, statistical analysis is a valuable tool to understand trends, patterns, and factors influencing mineral production. Time series analysis has gained significant attention in the mining industry to forecast the production trends of minerals (Bauwens & Giot, 2001). Time series techniques, enable users to make accurate forecasts about mining production and anticipate possible future scenarios. This facilitates strategic decision-making and contributes to more efficient management of mineral resources (Aycaya-Paco,Torres, Vilca-Mamani, 2023). The mining industry being complex and dynamic can rely on time series for accurate data analysis and forecasting to optimize operations, plan resource allocation and make strategic decisions.

## 1.1 Background of the study

Nickel holds indispensable importance in contemporary industry as a critical metal, playing a crucial role in the production of stainless steel, alloys, and other materials essential for various sectors, including transportation, construction, and consumer goods (Gunn, 2017). The global demand for nickel is increasing due to its growing applications in renewable energy technologies, such as electric vehicles and wind turbines (IEA, 2020). As a result, nickel mining has become a significant contributor to the global economy, with a market value of over $20 billion in 2020 (Grand View Research, 2020).

Nickel's physical and chemical properties make it essential in many end-use products. This underscores the necessity of predicting mineral production to facilitate future planning. Such planning entails preparing for upcoming requirements, including mining equipment, power supply, fuel, chemicals, and explosives. The determination of yield holds utmost significance as it allows miners to evaluate the costs and benefits of a mine, which informs important decisions about whether to expand production or shut down the mine entirely (Claassen, 2013).

In Africa, nickel mining is a vital sector, with several countries hosting significant nickel deposits. The continent accounts for around 10% of global nickel production, with Zimbabwe being one of the leading producers (USGS, 2022). Nickel mining is essential to Africa's economic development, providing employment opportunities, generating revenue, and contributing to GDP growth (African Development Bank, 2020).

In Zimbabwe, the mining sector is a crucial component of the economy, accounting for over 60% of export earnings and 15% of GDP (ZIMSTAT, 2022). Nickel mining, in particular, is a significant contributor to the country's economy, with Trojan Nickel Mine being one of the largest producers (BNC, 2022). The mine is a major employer and generates substantial revenue for the government, making it a vital component of Zimbabwe's economic development (Ministry of Mines and Mining Development, 2020).

Despite its importance, the nickel mining industry faces numerous challenges, including fluctuating nickel prices, operational inefficiencies, and external factors like regulatory changes and market demands (ICMM, 2020). Accurate forecasting of nickel production is essential to optimize production planning, resource allocation, and strategic decision-making (Sari et al., 2017). Time series analysis has been widely applied in the mining industry to forecast production trends and optimize operations (Jiang et al., 2019). However, existing forecasting methods may not adequately address the complexities and uncertainties of nickel production at Trojan Nickel Mine (Mwanza & Mwale, 2020).

This study aims to identify a suitable time series model to forecast nickel production at Trojan Nickel Mine, addressing the knowledge gap in accurate forecasting and informing strategic decisions to optimize production and resource allocation. The outcomes of this study will make a valuable contribution to the development of effective forecasting methods for nickel production, enhancing the sustainability and profitability of the mining industry in Zimbabwe and beyond.

## 1.2 Statement of the problem

Despite playing a crucial role in bolstering Zimbabwe's economy, the nickel mining industry, particularly Trojan Nickel Mine, faces operational challenges and uncertainties. Fluctuating nickel prices, inefficiencies, and external factors such as regulatory changes and market demands affect production dynamics. To enhance decision-making and drive sustainable growth, there is a pressing need to develop an accurate forecasting model using time series analysis. However, existing forecasting methods may not adequately address the complexities and uncertainties of nickel production at Trojan Nickel Mine. Therefore, this study aims to identify a suitable time series model to forecast nickel production at Trojan Nickel Mine, addressing the knowledge gap in accurate forecasting and informing strategic decisions to optimize production and resource allocation.

## 1.3 Objectives of the study

The paramount objectives of this project are:

1. To identify a model for forecasting Trojan Mine nickel production
2. To forecast nickel production patterns using the model for 10 months
3. To show and explain the Trojan nickel production mine production trend for the past

## 1.4 Research questions

1. Which model will be incorporated into Trojan’s nickel production forecasting?
2. What will be the future Nickel production at Trojan?
3. What is the trend of historical nickel production at Trojan?

## 1.5 Significance of the study

This study holds immense significance for the nickel mining industry in Zimbabwe, particularly focusing on the operations at Trojan Nickel Mine. By applying time series analysis models to the production data, valuable insights into trends and patterns of nickel production are offered which can aid in strategic decision-making. It is also essential for the reader because it can improve students' knowledge of financial time series and serve as a starting point for individuals interested in learning more about financial time series in forecasting production or similar topics. Furthermore, the research assists other students in evaluating the suitability of college-taught theories and putting them into practice.

## 1.6 Assumptions of the study

For this study to be relevant, we assume that the sample selected will represent the Trojan Nickel Mine company, data collected from the company is accurate, and tools and methods used to collect data are valid and reliable.

## 1.7 Limitations of the study

This study's findings are bound by certain limitations. The first one is that ARIMA models used to forecast are known to have some potential biases and errors such as overfitting and underfitting which could affect the accuracy of the predictions. More so, the results of the model may not be perfectly generalizable to other nickel mines as the model is specific to the data and conditions of the Trojan Nickel Mine.

### 1.8 Definition of terms

**Time Series**

A time series consists of observations taken at regular intervals that are equally spaced over a period (IBM, 2013). In mathematical terms, time series is a sequence of data points usually measured at successive intervals, defined as a set of vectors $x(t)$, where $t$ represents elapsed time, such as $t = 0,1,2\ldots$(Adhikari, 2014).

**Forecasting**

Forecasting involves predicting or estimating future events by analyzing past and present data. It involves using statistical models, algorithms, and techniques to analyze historical trends, patterns, and relationships in data to make predictions about future outcomes (Hermadi et al., 2020).

**Production**.

Nickel production refers to the extraction and processing of nickel ore to obtain nickel metal or its various forms, such as nickel powder, nickel cathodes, or nickel alloys (Contreras et al., 2007)

### 1.9 Chapter Summary

The chapter has provided an overview of the research topic on time series analysis for nickel production outlining the background, statement of the problem, research objectives, research questions, significance of the study, delimitation of the study, limitations to the study, assumptions of the study and definition of key terms with a specific focus on Trojan Nickel Mine as a case study. The importance of forecasting nickel mine production through time series was highlighted emphasizing accurate predictions for planning and decision-making in the mining industry.

# CHAPTER 2

# LITERATURE REVIEW

## 2.0. Introduction

In this chapter, we examine empirical studies and theoretical literature about time series analysis and its practical application in analyzing prior research on production. Additionally, this chapter explores significant findings in this field and provides an overview of time series analysis.

## 2.1 Theoretical Literature Review

### 2.1.1 Time Series Components

Time series analysis involves examining four main components: trend, seasonality, irregularities or cycles, and random variation (Agbo, 2021). Each element has a distinct influence on the observed data, and forecasting can project these patterns into the future. Time series plots provide a means to detect diverse patterns, encompassing random fluctuations, trends, level shifts, periodic or cyclical behavior, and exceptional observations of these patterns. The trend component of a time series offers valuable insights into the overarching long-term trajectory of the data. It indicates whether the data is consistently moving upward or downward in a predictable manner during each period. The trend may exhibit either a linear or non-linear nature, contingent upon the variables under examination. Identifying the trend is crucial as it helps in making forecasts and understanding the overall behavior of the data.

Seasonality refers to the occurrence of systematic fluctuations in data that can be attributed to specific factors. These variations exhibit regular patterns at fixed intervals, such as monthly, quarterly, or yearly. They manifest as recurring patterns within the data. By analyzing the seasonality component, we can discern these consistent patterns and make appropriate adjustments to our analysis. Seasonal fluctuations are often observed in data related to sales, weather, or stock prices.

The cyclical component, sometimes referred to as irregular cycles, signifies the presence of oscillatory movements in the time series that span beyond a one-year timeframe. These fluctuations are not periodic and repeat over longer time spans. They involve rises and falls that are not easily predictable or explainable. These cycles may not have a fixed duration and are often influenced by economic and business cycles. Understanding the cyclical component is important for predicting long-term trends and making informed decisions.

Data fluctuations that cannot be ascribed to the trend, seasonality, or cyclical components are deemed as random or irregular. These disruptions deviate from the prevailing trend observed in the time series data and are characterized by their unpredictable nature. Also known as the random or error component, it captures the unexplained variability in the data and is essential for assessing the accuracy of our time series models.  By comprehending and accurately measuring these components, we can acquire invaluable perspectives on the fundamental patterns inherent in the time series data. This knowledge enables us to make informed forecasts and better analyze the behavior of the data over time.

## 2.2 ASSUMPTIONS REGARDING TIME SERIES DATA

### 2.2.1 Stationary Assumption

Prior to analyzing time series data, it is crucial to perform stationarity tests. Time series modeling, assumes that the data remains stationary meaning its mean, variance, and autocorrelation structure are consistent over a period. This assumption is vital to prevent misleading outcomes when forecasting time series. To ascertain the absence of trend or seasonality in the data, we can visually examine the time series plot. When data is not stationary, methods such as differencing and log transformation are used to make it stationary. ADF method is commonly employed to assess the presence of stationarity. These tests provide statistical evidence to support the stationarity assumption and guide us in making accurate and reliable analyses and forecasts in time series modeling.

### 2.2.2 Normality

One more assumption that the data must satisfy is that they adhere to a normal distribution. Deviation from this assumption can lead to incorrect estimation of parameters or inaccurate forecasts. To assess the normality assumption, common tools include histograms, stem-and-leaf plots, box plots, P-P plots, Q-Q plots, and plots of the empirical cumulative distribution function. Alongside these visual methods, we can also utilize analytical tests that rely on the empirical distribution function (EDF) to examine the normality of the data. One commonly used tests is the kolmogorov-Smirnov test and it relies on the empirical distribution function to assess the normality assumption.

### 2.2.3 Independence

To guarantee the precision and dependability of our time series analysis, the residuals or disturbances must demonstrate independence and the absence of autocorrelation. The Durbin-Watson test is extensively employed to assess the presence of positive autocorrelation in the residuals. Alternatively, we can plot the residuals against the fitted values as another approach. It is argued that provided the model is accurate, this graph should appear lacking structure. Additionally, the autocorrelation function (ACF) of the residuals can be plotted to assess the presence of significant terms. Generally, only around 1/20 of the terms is expected to be above $\pm 2/n$, where n denotes the number of data points used in the time series. By examining these diagnostic tools, we can ensure that the residuals meet the assumptions of independence and lack of autocorrelation, thus enhancing the reliability and validity of our time series analysis.

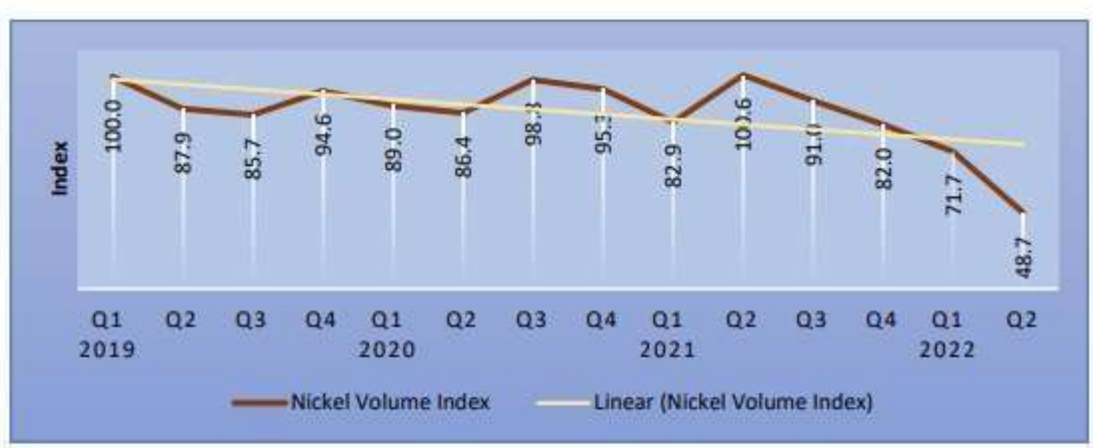
### 2.2.4 Homoscedasticity

To ensure the constancy of variance in residuals, we can examine a scatter plot of the residuals. This plot should reveal a rectangular shape around the zero horizontal levels, indicating that the variance remains constant. The scatter plot needs to be free from any discernible trends. By assessing the scatter plot, we can verify that the assumption of constant variance is upheld, further enhancing the reliability of our analysis.

## 2.3 Examples of Time series data

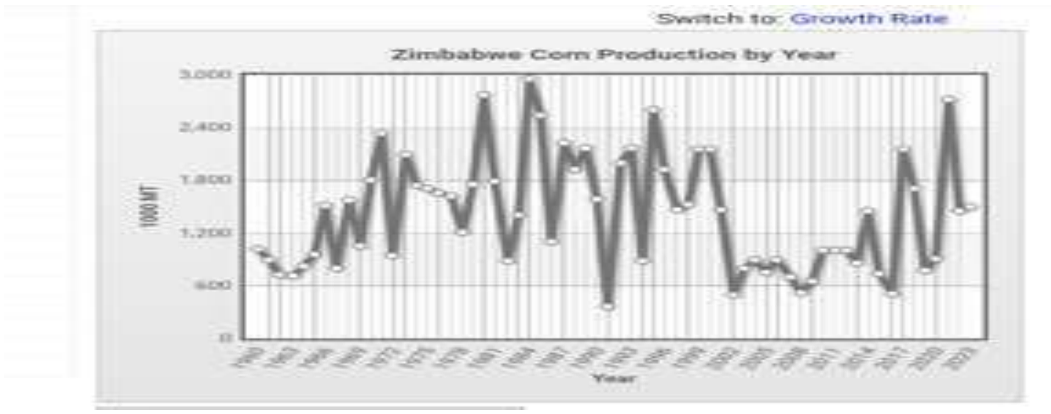
Time series data, comprising regularly recorded observations over time, forms an essential element in diverse domains including healthcare, mining, and finance. This type of data allows analysts to uncover trends, patterns, and relationships that can inform decision-making and forecasting. o illustrate the fundamental trend of the data, typically, a time series is depicted through a graph, where the observations are plotted against the respective time (Adhikari, 2016). Shown below are the two plots of time series data:

**Figure 2.1: Nickel volume produced by Zimbabwe in 2019**




*Source: ZIMSTAT 2019*

**Figure 2.2 Corn produced from 1960 to 2023 in Zimbabwe**




*Source ZIMSTAT 2019*

The first time series is taken from the ZIMSTAT index mineral production and it represents the Zimbabwean nickel output index from 2019 to 2023. Second one represents a seasonal time series, deemed in Index Mundi and it shows the numbers of corn production in Zimbabwe from 1960 up to 2023.

## 2.4 TIME SERIES MODELS

### 2.4.1 Autoregressive Models

An autoregression is a time series model that predicts future values based on past observations. An example of this is the AR(p) model, where p indicates the order.

$$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \cdots \cdots + \emptyset_p \emptyset_{t-p} + a_t$$

Here, $\emptyset_0$ represents the constant term, $\emptyset_p$ are the model parameters, and $a_t$ is assumed to be a white noise series. The benefits of the AR model include its ability to determine the extent to which previous values in the time series can explain current values. Autoregressive models compute the polynomial likelihood of the following symbol. While this is appealing, it means they won't be able to create a distribution model with difficulty in calculating the next symbol probability.

### 2.4.2 Moving Average

An MA is calculated by averaging a particular quantity of time series data points surrounding each point t, excluding the first and last few terms. This technique is utilized for smoothing time series data and for forecasting, but it is only applicable to time series without a trend. An example of a MA series of order q denoted as MA(q)

$$Y_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots \cdots - \theta_q a_{t-q}$$

Where $c_0$ is a constant, $a_t$ is a white noise series, and $\theta_1$, $\theta_2$,⋯ $\theta_q$ are model parameters. One advantage of using MA is that it is resistant to outliers, thus it can still give accurate predictions even if there are anomalies in the data. However it can not capture sudden changes in the data, thus it may not be able to predict sudden spikes or drops in production.

### 2.4.3 ARIMA Models

The ARIMA model is a time series forecasting technique that integrates the benefits of both autoregressive (AR) and moving average (MA) models. The Time series should be differentiated until it is stationary, eliminating patterns and seasonal effects. The "explanatory variables" consist of lagged values of t and errors, forming an ARIMA (p, d, q) model.

$$y'_t = c + \emptyset_1 y'_{t-1} + \cdots \cdots + \emptyset_p y'_{t-p} + \emptyset_p \varepsilon_{t-p} + \varepsilon_t \cdots \cdots$$

Integration (I) is applied to make a time series stationary. If the series is not stationary, first-order differencing (d = 1) or higher-order differencing can be used to achieve stationarity. The general formula for differencing is:

$$\Delta(t) = y(t) - y(t-1)$$

"$\Delta(t)$" represents the variance between the present and preceding values in the time series. The ARIMA model integrates the historical dependence on past values (AR), reliance on past errors (MA), and differentiation to analyze and predict time series. The ARIMA model is recognized for its unit-root nonstationary characteristic (Ruey, 2010). ARIMA forecasting offers a significant advantage by solely utilizing data from the specific time series under consideration. This characteristic proves beneficial when forecasting numerous time series. Additionally, it mitigates potential issues that can arise with multivariate models. Nevertheless, ARIMA may face constraints in predicting extreme values. Despite its proficiency in capturing seasonal patterns and trends, forecasting outliers can be challenging due to their deviation from the typical trend represented by the model.

## 2.5 Empirical literature

Patrícia Ramos, José Manuel Oliveira, and Paula Silva (ND) conducted a study on predictive analysis for manufacturing equipment to forecast malfunctions and enable proactive maintenance. They compared ARIMA forecasting methods and neural network models using data from a continuous monitoring system. The study used the Box Jenkins ARIMA methodology, including a Box-Cox transformation for variance stabilization, and cross-validation to find the optimal ARIMA model and neural network architecture. The best ARIMA model had the lowest RMSE and passed the Ljung-Box test at 5% significance. Time series analysis was done with R and the forecast package. RMSE, MAE, and MAPE were used to evaluate model performance. Both models could predict disc replacement events, but ARIMA provided more accurate forecasts, particularly in predicting the increased distance between discs before and after replacement. The study concluded that ARIMA models were more effective than neural networks.

Tichaona W. Mapuwei, Jenias Ndava, Mellissa Kachaka, and Brain Kusotera (2022) conducted a study using the Box-Jenkins ARIMA approach to predict tobacco production in Zimbabwe, using annual data from 1980 to 2018 from ZIMSTAT. They found that the data became stationary after first differencing, stabilizing both mean and variance. Diagnostic tests using ACF, PACF, and the auto.arima function in R determined that the ARIMA (1, 1, 0) model was the best fit. The study's four-year forecast (2019-2023) showed a slightly declining trend in tobacco production, suggesting opportunities to enhance yield with appropriate measures. The prediction aimed to support the idea that the tobacco industry can grow, despite the downward trend, through strategic interventions by government, private institutions, and farmers.

Nyoni (2019) analyzed Zimbabwe's electricity demand using annual data from 1971 to 2014. Using the Box-Jenkins ARIMA methodology, the study forecasted electricity demand for the next decade. Data from the World Bank showed non-stationarity, resolved by first differencing. Evaluation metrics like AIC, Theil's U, and ME identified ARIMA (1, 1, 6) as the best model, with a stable and suitable fit confirmed by diagnostic tests. The study found that electricity demand peaked at 1038 kWh in 1976 but has since declined. The ARIMA model predicts this decline will continue from 2015 to 2025, primarily due to reduced economic activity. The study recommends that electricity producers and distributors in Zimbabwe adjust their capacities to prevent power outages.

Paraskevi Klazoglou and Nikolaos Dristakis (2018) used the Box-Jenkins ARIMA method to forecast US health expenses from 1970 to 2015, using annual data on health expenditures as a percentage of GDP from the OECD. Autocorrelation plots indicated nonstationarity, which was corrected by differencing. Unit root tests confirmed stationarity and ARMA parameters were defined using autocorrelation coefficients. The model with the lowest AIC, SC, and HQ values was ARIMA (0,1,1), identified as the best fit. The model's forecasting performance was assessed using RME, MAE, and TIC. The study aimed to create a predictive model for US health expenditure from 1970 to 2015, concluding that ARIMA (0,1,1) was optimal

Aycaya-Paco Yhack Bryan, Vilca-Mamani Lindell Dennis, and Torres-Cruz Fred (2023) used the Box-Jenkins ARIMA method to forecast Peru's mineral extraction from 1980 to 2027, using monthly and yearly data from MINEM. This allowed them to examine recent and long-term trends in the mining industry. They used residual analysis, including the Ljung-Box and Shapiro-Wilk tests, to assess the ARIMA model's fit. Using the "forecast" R package, they identified the best model and predicted future mineral extraction. The forecast indicated an increase in annual production over the next five years. The study also recommended addressing mining safety and health issues, particularly fatal accidents, which can reduce overall production.

### 2.6 Research issues and research gap

Despite the significance of time series analysis in the field of production forecasting, there is a dearth of studies focusing specifically on nickel production. While there are numerous research papers exploring time series analysis in other industries, such as manufacturing, agriculture, the health sector, and finance, the application to the nickel mining industry remains relatively unexplored. This research gap presents an opportunity to investigate the potential benefits of employing time series analysis techniques in optimizing nickel production at Trojan Nickel Mine. The existing literature on time series analysis primarily focuses on general theory and applications in various sectors.

However, there is a lack of empirical studies that directly apply these techniques to the unique challenges faced by the nickel mining industry. By conducting a case study at Trojan Nickel Mine, this research aims to bridge the existing gap and provide valuable insights into the effectiveness of time series analysis in improving production forecasting accuracy and operational efficiency

### 2.7 Summary

In this chapter, an examination of the existing literature and empirical studies related to production data and time series analysis was conducted. The information presented in this chapter served as the foundation for assessing the influence of time series on the prospects and performance of Trojan Nickel Mine. The subsequent chapter will delve into the methodology employed in this study, including the research design, data collection methods, and the planned approach for data analysis

# CHAPTER 3

# RESEARCH METHODOLOGY

## 3.0 Introduction

This chapter focuses on the techniques employed to achieve the specified study objectives. The methodology extends beyond research methods, serving as a foundation for strategies utilized in this study's discourse and elucidating the rationale behind the selection of certain techniques over others. Consequently, the researcher assumes the responsibility of evaluating the study's results (Kothari, 2017). The research methodology provides a structure and instruments for data collection and outlines the plan for data analysis (Dawson, 2013). The incorporation of a time series model aims to reveal and forecast future nickel production values.

## 3.1 Research Design

The research design acts as a roadmap for navigating the intricacies of the study, providing a systematic approach to reduce bias and uphold the credibility of findings (Plot et al., 1999). A quantitative research design is adopted to unravel the underlying trends and patterns in nickel production at Trojan Nickel Mine. This approach facilitates a meticulous examination of time series data, enabling the precise measurement of production trends and forecasting accuracy. By employing statistical methodologies, the study aims to rigorously evaluate the effectiveness of time series forecasting techniques in optimizing production outcomes. The quantitative research design inherently brings objectivity to the study, relying on numerical data to drive analysis and minimize subjective biases.

This objectivity not only enhances the integrity of the research findings but also fosters replicability, allowing other researchers to replicate the study's methodology and validate its outcomes. Given the data-intensive nature of mining operations, particularly in a dynamic environment like Trojan Nickel Mine, the quantitative approach is well-suited to analyze large datasets and extract meaningful insights. Ultimately, the research design is tailored to explore theoretical constructs in time series analysis and assess the practical implications of forecasting techniques on nickel production at Trojan Nickel Mine. Through a systematic and data-driven approach, the study endeavors to contribute valuable insights to the field of mining industry optimization and strategic decision-making.

### 3.2 Data Sources and Collection

Data source indicates the origin of information, while data collection encompasses the process of acquiring that information (Baker, 2018). For the study on time series analysis at Trojan Nickel Mine, the researcher is using data directly from Trojan Mine itself, covering a period of 10 years from April 2013 to March 2023. Taking a secondary data collection approach means that new data is not being collected but existing data is being used. The main source of data is directly from Trojan Mine's records, including production numbers and dates. By using data directly from Trojan Mine, the researcher can be confident in the quality of the information being analyzed. This approach also allows for the coverage of a significant period, providing a comprehensive view of production trends over the past decade. Having a substantial amount of data is important for making accurate predictions and understanding long-term patterns.

### 3.3 Research instruments

Research instruments are implemented and utilized to gather and assess data about the research subject (Nieswiadomy, 2018). Laptops played a vital role in handling complex analytical tasks throughout the study. Their robust computing power enabled the execution of statistical analyses and intricate modeling processes. Given the extensive datasets involved in mining operations, laptops provided the necessary capability to process data efficiently, enhancing the precision and thoroughness of the analysis conducted by the researcher.

## 3.4 DESCRIPTION OF VARIABLES AND EXPECTED RELATIONSHIPS

### 3.4.1 Nickel Production

This variable represents the quantity of nickel extracted from Trojan Nickel Mine over time. It serves as the primary focus of the study and is measured in tons.

### 3.4.2 Time

Time serves as the independent variable in the time series analysis, influencing the fluctuations in nickel production. The temporal dimension of the data is represented in months.

### 3.4.3 Trend Analysis

The researcher anticipates observing a trend in nickel production over the study period. This trend may exhibit either an upward, downward, or stable trajectory, indicating long-term patterns in production output.

### 3.4.4 Seasonality

Seasonal variations in nickel production may be evident, influenced by factors such as weather conditions, market demand, or operational cycles. The researcher expects to identify recurring patterns or fluctuations within specific time intervals.

### 3.4.5 Correlation with External Factors

The researcher hypothesizes that nickel production at Trojan Nickel Mine may be correlated with external factors such as economic conditions, commodity prices, or regulatory changes. By analyzing these relationships, the researcher aims to understand the external drivers impacting production dynamics.

### 3.4.6 Cyclical Patterns

Cyclical patterns, characterized by periodic fluctuations in production output, may be present in the data. These patterns may be impacted by elements like investment cycles, technological progressions, or geopolitical occurrences.

### 3.4.7 Forecasting Accuracy

The researcher anticipates assessing the precision of time series forecasting models in predicting forthcoming nickel production levels at Trojan Nickel Mine. By comparing forecasted values with actual production data, the researcher aims to assess the reliability and effectiveness of forecasting techniques.

## 3.5 Diagnostic tests

A thorough evaluation of model adequacy will be carried out to guarantee that the model fitted effectively captures the inherent dynamics of the time series data. This assessment will involve a meticulous analysis of residuals from autoregressive (AR) and moving average (MA) models. A comprehensive analysis of these residuals can ascertain how well the fitted model captures the fundamental patterns and dynamics within the time series data. An important component of this analysis involves investigating whether the residuals, also known as disturbances, demonstrate attributes indicative of a white noise process. This is a process where the residuals are random and exhibit no discernible patterns or trends (Montgomery et al., 2015). To evaluate this, scatter plots of the residuals will be reviewed. If the model is suitable, the scatter plot of residuals displays a rectangular form, indicating the absence of any systematic patterns.

Furthermore, the sample autocorrelation function of the residuals will be carefully examined. In a reliable model, the autocorrelation function of the residuals should show no identifiable structure, suggesting no residual correlation remains (Montgomery et al., 2015). To enhance the evaluation of the fitted model's suitability, statistical assessments like the ADF,A-D test, $X^2$ test of model adequacy, and the Ljung-Box will be utilized. These tests provide valuable insights into the validity and reliability of the fitted model, ensuring that it accurately captures the underlying dynamics of the data. These tests provide quantitative measures of how well the model fits the data (Montgomery et al., 2015), helping to ensure the reliability of the forecasting results. Once an adequate model is identified and validated through diagnostic checks, it can be confidently used for forecasting future values of the time series data.

## 3.6 ANALYTICAL MODEL

The analytical framework utilized in this investigation of time series analysis for Trojan Nickel Mine predominantly relies on the ARIMA model. ARIMA is a commonly utilized approach for examining and predicting time series data, particularly in disciplines like finance, economics, and engineering. The ARIMA model consists of three primary elements: autoregression (AR), differencing (I), and moving average (MA). It is defined by three parameters: p, d, and q, representing the autoregressive order, differencing order, and moving average order, respectively. These parameters are determined based on the characteristics of the time series data and are crucial for accurately capturing its underlying patterns and dynamics.

### 3.6.1 Moving Averages (MA)

Moving averages (MA) are a fundamental component of time series analysis, offering a method to smooth out fluctuations and identify underlying trends. This technique involves computing the average of a specified number of consecutive observations in a time series, centered around each value of *t*. However, the first few and last few terms are often excluded from the calculation to mitigate boundary effects. MA models are particularly useful for analyzing time series data that do not exhibit a clear trend. By averaging out random fluctuations and noise in the data, MA models help to highlight underlying patterns and relationships (Tsay, 2010). Mathematically, an MA model of order *q*, denoted as MA(*q*), can be represented as:

$$Y_t = C_0 + \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \cdots - \theta_q \alpha_{t-q}$$

Where: $Y_t$ represents the value of the time series at time *t*

$C_0$ is a constant term

$\alpha_t$ represents the observed value at time t

$\theta_1, \theta_2, \ldots . \theta_3$ are model parameters, and q is the order of the moving average

The parameters represent the weights given to lagged observations when calculating the moving average. These weights dictate the impact of past observations on the current value and are usually estimated from the data using optimization methods

### 3.6.2 Autoregressive (AR)

Autoregressive (AR) models are essential tools in time series analysis, particularly for understanding and predicting sequential data points. These models leverage past observations within the time series to forecast future values (Tsay, 2010). An autoregressive model of order *p*, denoted as AR(*p*), is expressed as

$$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p}$$

Here: $Y_t$ represents the value of the time series at time t

$\emptyset_0$ represents the constant term

$\emptyset_1, \emptyset_2 \ldots . \emptyset_p$ are coefficients representing the influence of past observations

$\varepsilon_t$ is a random error term

The autoregressive model represents the linear relationship between the current value of the time series $Y_t$ and its lagged values $Y_{t-1}, Y_{t-2} \ldots, Y_{t-p}$. The coefficients $\emptyset_1, \emptyset_2 \ldots, \emptyset_3$ determine the strength and direction of this relationship. These coefficients are estimated from the data using methods such as least squares estimation.

### 3.6.3 Differencing (I)

Differencing, often denoted as *I* in the context of time series analysis is a fundamental technique used to covert data which is not stationary to being stationary. Stationarity is a key concept in time series analysis, where the statistical properties of the data remain constant over time (Box et al., 2010). A stationary time series maintains a consistent mean, variance, and autocovariance, simplifying the process of modeling and analysis. The *I* operator is applied to the time series data to achieve stationarity by taking the difference between consecutive observations. Mathematically, differencing can be represented as:

$$\Delta Y_t = Y_t - Y_{t-1}$$

Where: $Y_t$ is the value of the time series at time *t*

$\Delta Y_t$ represents the differenced series at time *t*.

Differencing effectively eliminates trends and seasonal patterns from the data, rendering it stationary. This technique is especially beneficial for time series data with trends or seasonal variations.

### 3.7 Box Jenkins Methodology

The Box-Jenkins methodology is a systematic approach used for constructing and implementing ARIMA time series models. This method is particularly suitable when working with datasets that have a minimum of 30 observations. The methodology consists of three iterative steps, each essential for conducting effective time series analysis using ARIMA models:

1. Identification of Model - This initial stage entails determining a suitable ARIMA model through an analysis of past data. This involves scrutinizing the time series plot to detect trends, seasonal patterns, and any other underlying structures. Furthermore, ACF and partial and PACF plots are utilized to ascertain the order of autoregressive (AR) and moving average (MA) terms within the model (Montgomery et al., 2015).

2. Parameter Estimation: After determining the model structure, the subsequent step is to estimate the undisclosed parameters of the ARIMA model. This commonly entails employing statistical methods like maximum likelihood estimation to adjust the model to the available data. The objective is to ascertain the values of the AR, MA, and differencing parameters that most accurately depict the inherent patterns in the time series data (Montgomery et al., 2015).

3. Diagnostic Checking: Following parameter estimation, diagnostic evaluations are conducted to evaluate the adequacy of the fitted model. This entails scrutinizing the residuals, which represent the disparities between the observed values and those predicted by the model. Diagnostic assessments aid in verifying that the model effectively captures the inherent patterns in the data and that the residuals display attributes of white noise, suggesting the absence of systematic patterns within the data (Montgomery et al., 2015).

## 3.8 Conclusion

In conclusion, the research methodology outlined in this chapter lays the groundwork for a comprehensive and rigorous analysis of time series data for Trojan Nickel Mine. By adopting a quantitative research design, the study aims to uncover trends and patterns in nickel production over 10 years from April 2013 to March 2023. Utilizing laptops for complex analytical tasks. The diagnostic tests, including residual analysis and model validation, will help assess the adequacy of the selected time series models, such as AR and MA models. Moreover, the Box-Jenkins methodology furnishes a methodical framework for recognizing, fitting, and validating ARIMA models, crucial for precise production trend forecasting. By engaging in an iterative cycle or model identification, estimating parameters, and diagnostics evaluation, the study seeks to create resilient models capable of comprehensively capturing the fundamental dynamics of nickel production at Trojan Nickel Mine.

# CHAPTER 4

# DATA ANALYSIS

## 4.0 Introduction

To ensure the comprehensive completion of this study, an examination of data was conducted to assess the hypothesis and address previously mentioned research inquiries. The data is depicted descriptively. This section encompasses the examination, portrayal, and understanding of the study's outcomes. A retrospective analysis of the time series was executed to acquire pertinent data for the research aims. The primary objective of the study is to select the optimal model and utilize it to predict nickel production at Trojan Nickel Mine. Visual aids including time series plots, seasonal decomposition charts, and histograms are utilized to improve the lucidity and comprehension of the analysis, employing Excel, R, and R-studio.

## 4.1 Descriptive Statistics

**Table 4 1 Descriptive statistics on nickel production for 10 years**

| Minimum | 1st Quartile | Median | Mean | 3rd Quartile | Maximum | Standard deviation | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| 0 | 405 | 500 | 499.7 | 619.5 | 993 | 189.07 | -0.3 | 0.7 |

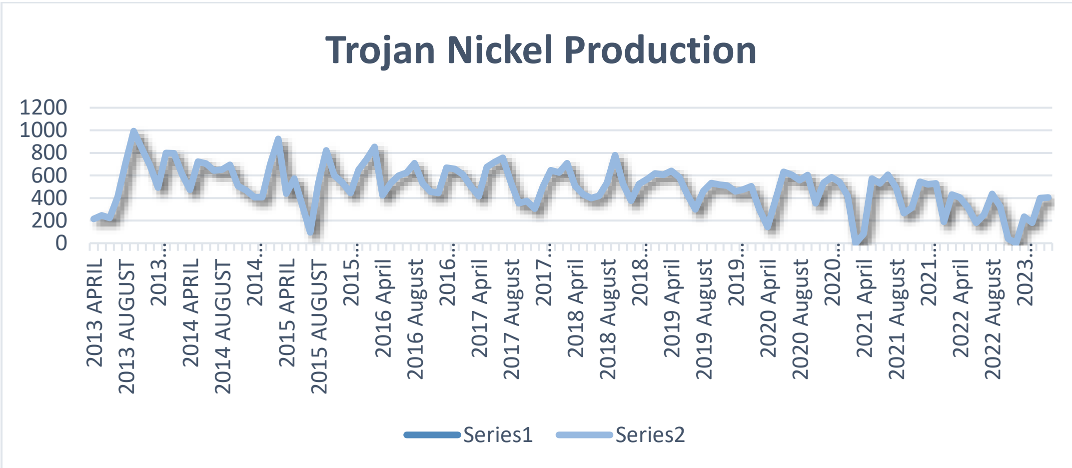
The summary statistics provide valuable insights into the production trends of nickel at Trojan Mine. The minimum value of 0.0 suggests that there were periods during which no nickel was produced, indicating potential downtime or operational issues. At the first quartile (25th percentile), which stands at 405.0 tons of nickel, it is observed that during the lowest production periods, approximately a quarter of the time, Trojan Mine produced less than or equal to this amount. The median value, representing the middle point of the dataset, is 520.0 tons of nickel, indicating that production fell below this value half of the time. The mean value of 499.7 tons provides an average measure of production, giving a sense of the central tendency of the data, with a standard deviation of 189.07 tons reflecting the degree of dispersion or variability around the mean.

As for the third quartile (75th percentile), standing at 619.5 tons, it indicates that during the majority of production periods, approximately three-quarters of the time, Trojan Mine produced less than or equal to this amount. The maximum value observed is 993.0 tons, representing the peak production level recorded during the observed period. The negative skewness (-0.3) suggests that the data is skewed to the left, indicating that the production values are more concentrated on the lower side. Additionally, a kurtosis value close to zero like 0.4 indicates that the data has a similar distribution to a normal distribution (mesokurtic). Overall, these summary statistics offer a comprehensive overview of nickel production by Trojan Mine, aiding in understanding its production dynamics and variability over the observed period.

### 4.1.2 Trojan Production History

A time series plot of the total production spanning from 2013 to 2023 was generated to assess the stationarity of the data before considering any statistical tests. The line graph depicting nickel production at Trojan Nickel Mine from April 2013 to January 2023 provides a clear view of the production trends over the ten years. Initially, there is a marked fluctuation in production levels, with several peaks and troughs, reflecting significant variability. The production started high, nearing 1000 tonnes in 2013, but quickly experienced sharp declines. The time series data has constant variation, indicating stationarity. However, the graph suggests a general decline in nickel production at Trojan Nickel Mine over the past decade, presented in Figure 4.1

**Figure 4.1: Trojan Nickel Mine Production Trend**




The Augmented Dickey-Fuller (ADF) test was conducted (Table 4.2) to evaluate stationarity, yielding a p-value of 0.01, which is below 0.5. Consequently, the null hypothesis is rejected, indicating that the data is stationary. The researcher then proceeded to the model identification phase.

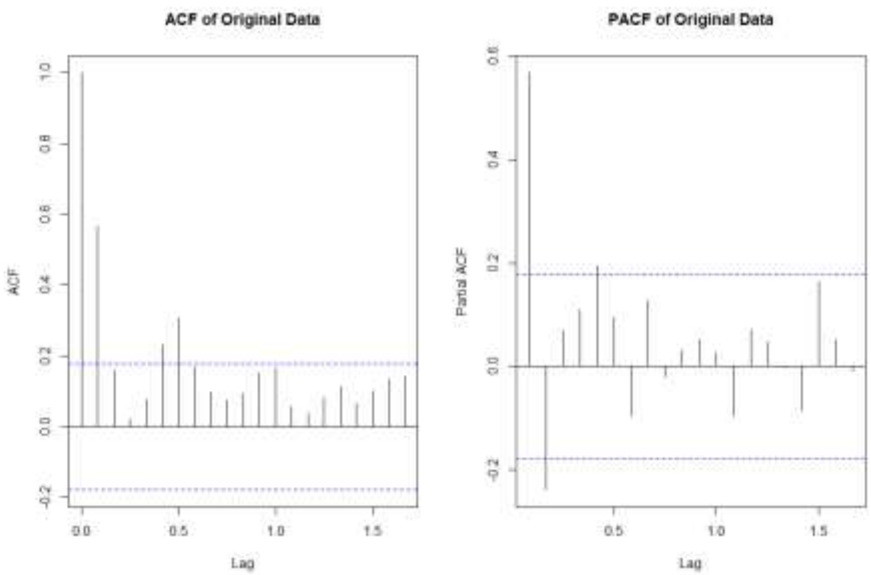
**Table 4.2 for data showing the ADF Test**

| Dickey-Fuller | Lag Order | P-Value |
|---|---|---|
| -5.5021 | 4 | 0.01 |

### 4.2 Model Identification

The primary objective of this stage is to ascertain the AR and MA components to define the ARIMA (p, d, q) model. The ACF and PACF was reviewed and depicted in Figure 4.2.

**Figure 4.2 showing ACF and PACF of original data**




It is evident that p=1 because the partial autocorrelation function (PACF) in Figure 4.2 (ACF) has a notable spike at lag 1 before cutting off. Likewise, q=1 since the autocorrelation function (ACF) in Figure 4.2 (PACF) shows a prominent spike at lag 1 followed by a rapid decline. Given that the data is already stationary and does not require differencing (d=0), the suggested model is ARIMA(1,0,1).

### 4.3 Parameter Estimation

The next step involves determining the parameters of the AR and MA terms included in the fitted model.

**Table 4.3: Model parameters**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Model:ARIMA(1,0,1) | | | | | | | |
| Standard error = 57011.49 | | | | | | | |
| Estimated variance = 22941 | | | log likelihood=-771.46 | | | | |
| AIC = 1550.91 | | | AICc =1551.26 | | | BIC = 1562.06 | |
| Training and Test set error measures: | | | | | | | |
| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
| Training set | 1.21398 | 149.5589 | 118.1564 | -Inf | Inf | 0.6766114 | 0.002769046 |

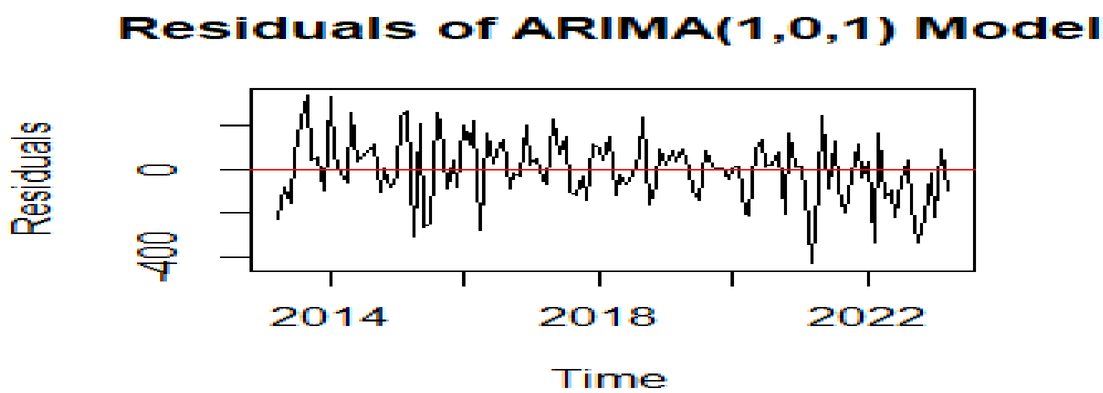
The model has d = 0, indicating no differencing was performed. The optimal model features an autoregressive order of 1 (p = 1), a moving average order of 1 (q = 1), and a SE of 0.1285.

## 4.4 Model Diagnosis / Pretests

### 4.4.1 Test for stationarity

Figure 4.3 displays residuals that exhibit characteristics akin to white noise, fluctuating around a mean of zero with consistent dispersion. The model in question is considered to be stationary. based on the pattern of these errors.

**Figure 4.3 Residuals of ARIMA (1,0,1)**




Ho: The time series data for Nickel production at Trojan Mine is non-stationary.

H1: The time series data for Nickel production at Trojan Mine is stationary.

**Table 4.4 Showing test for stationary outputs obtained from Augmented Dickey-Fuller Test**

| Dickey-Fuller | Lag order | p-value |
|---|---|---|
| -7.2377 | 4 | 0.01 |

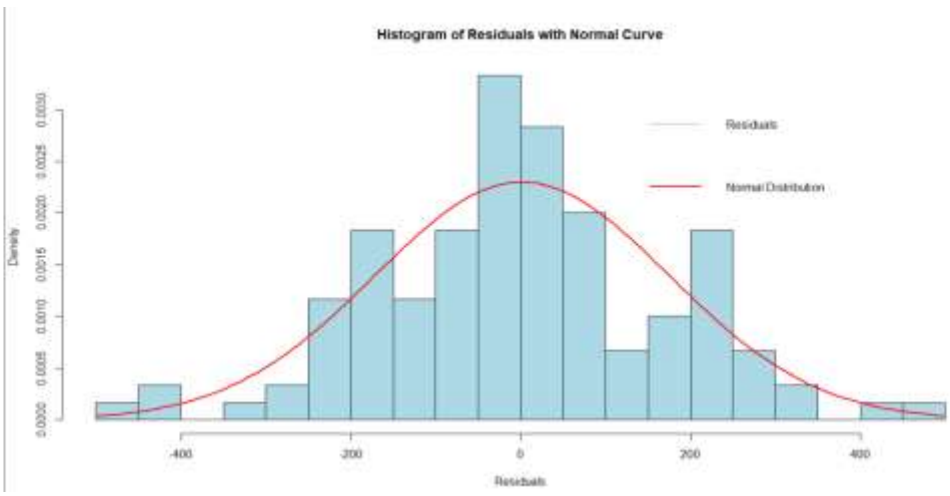
The Dickey-Fuller test statistic is -7.2377, with a corresponding p-value of 0.01, demonstrating statistical significance. Since the p-value is below the typical significance level of 0.05, we reject the null hypothesis of non-stationarity, supporting the alternative hypothesis that the data is stationary.. This implies that the time series data for Nickel production at Trojan Mine is stationary, meaning that its statistical properties such as mean and variance are the same over a period. Lag order of 4 indicates that the test considered up to 4 lags to account for any autocorrelation in the data. Overall, the results suggest that there is evidence to support the stationarity of the Nickel production time series at Trojan Mine.

### 4.4.2 Normality tests

H0: The nickel-produced data follows a normal distribution.

H1: The nickel-produced data does not follow a normal distribution.

**Figure 4.4 Histogram of residual**



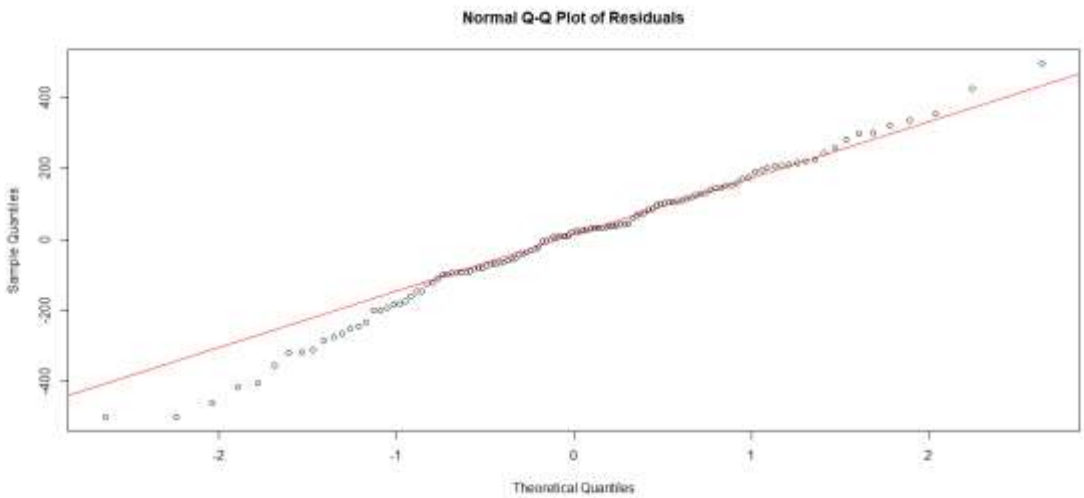

The histogram depicting the residuals (Figure 4.4), displays a symmetrically distributed pattern resembling a normal distribution. Anderson-Darling test for normality was conducted, generating A and P-values to assess normality

**Table 4.5 Showing A and P-values obtained from the Anderson-Darling normality test**

| Test statistics ($A^2$) | P-value |
|---|---|
| 0.52472 | 0.1779 |

The test statistic represents the discrepancy between the observed data and the expected normal distribution. In this case, the value of 0.52472 indicates a relatively small discrepancy. The p-value is a measure of the evidence against the null hypothesis of normality. A small p-value (typically less than the chosen significance level, often 0.05) suggests evidence to reject the null hypothesis, indicating a departure from normality. Conversely, a large p-value indicates insufficient evidence to disregard the $H_0$, suggesting that the dataset is consistent with a normal distribution. Since the p-value (0.1779) is greater than the typical significance level of 0.05 is p-value > 0.05, we do not reject the null hypothesis of normality. Consequently, the dataset follows a normal distribution.

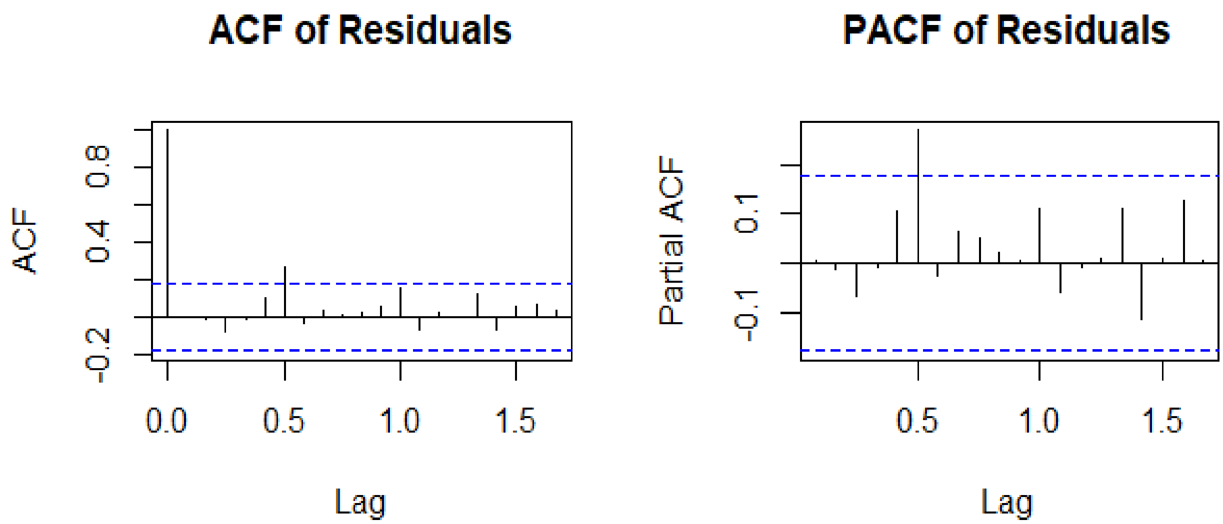
**Figure 4.5 A normal QQ-plot of residuals**




The normal Q-Q plot serves as a tool to assess the normal distribution of the dependent variable. It accomplishes this by plotting quantiles from our dispersion against a theoretical distribution. As depicted in Figure 4.5, the plot indicates that the allocation adheres to a normal distribution since the graphed data predominantly form a straight line. This is concluded by the Anderson-Darling normality test done above.

### 4.4.3 Test for Independence

**Figure 4.6 Showing ACF and PACF of residuals**




The correlograms in Figure 4.6 does not exhibit any discernible structure, confirming the absence of serial autocorrelations. The absence of significant peaks and the lack of any discernible pattern in both the ACF and PACF plots of the residuals indicate that the residuals are consistent with white noise. This suggests that the model has successfully captured the underlying structure of the data, leaving only random noise the plots confirm that the residuals do not show any structural pattern, implying that the model's assumptions about the residuals being uncorrelated are valid.

### 4.4.4 Test for Serial Autocorrelation

The Box-Ljung test was conducted to examine serial correlation, with the hypotheses stated as follows:

Null Hypothesis (Ho)**:** The time series has no serial autocorrelation.
Alternative Hypothesis (H1)**:** The time series exhibits serial autocorrelation.

**Table 4.6 Box- Lung test**

| Chi-square value(Q) | Df | p-value |
|---|---|---|
| 20.504 | 20 | 0.4269 |

This test is used to assess whether there are significant serial correlations in the residuals of the time series model . Given that the p-value is 0.4269, which is greater than 0.05, we do not reject the null hypothesis at a standard significance level. Therefore there is no serial correlation in the residuals.

## 4.5 Forecasting

**Figure 4.7 Nickel production forecast**

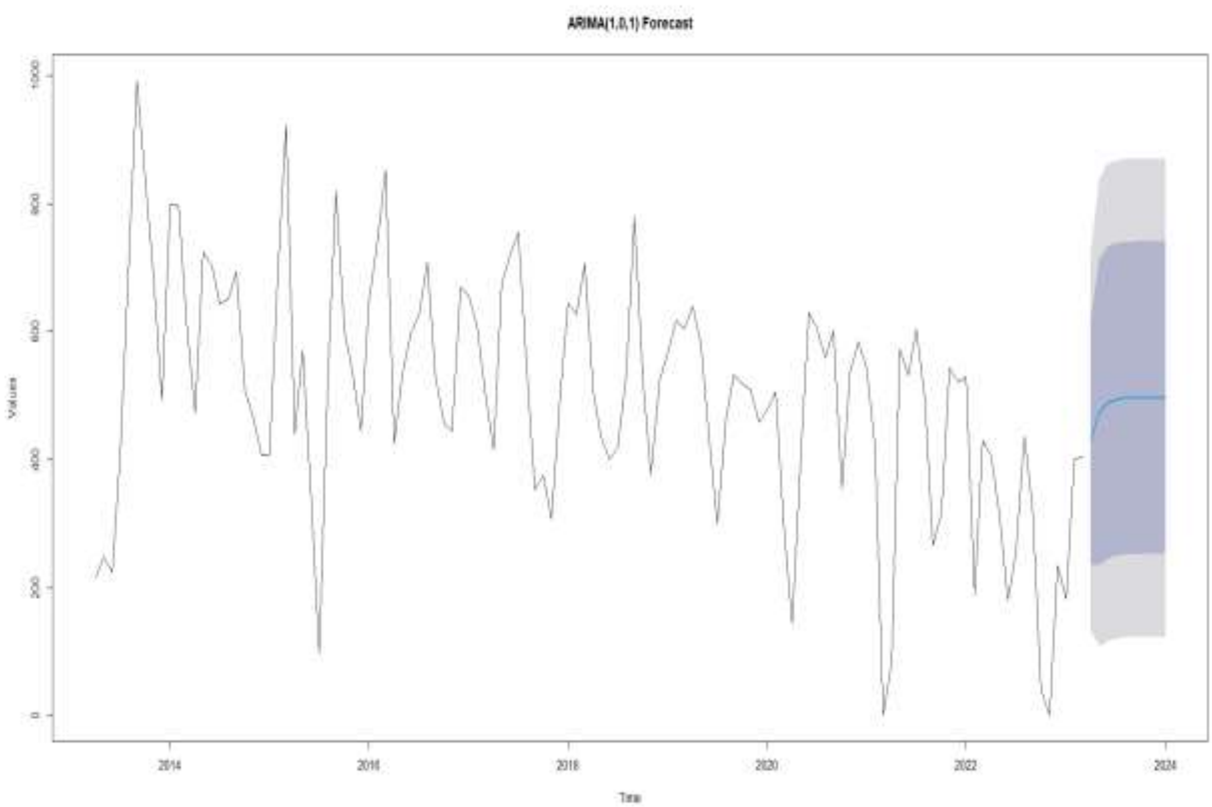


**Table 4.7 Forecasted 10 month values**

| Month/year | Point Forecast | Lo 95 | Hi 95 |
|---|---|---|---|
| 04/2023 | 429.5896 | 132.7251 | 726.4540 |
| 05/2023 | 474.0080 | 108.1328 | 839.8831 |
| 06/2023 | 489.1052 | 116.0794 | 862.1311 |
| 07/2023 | 494.2366 | 120.3936 | 868.0797 |
| 08/2023 | 495.9807 | 122.0434 | 869.9181 |
| 09/2023 | 496.5735 | 122.6253 | 870.5218 |
| 10/2023 | 496.7750 | 122.8255 | 870.7245 |
| 11/2023 | 496.8435 | 122.8938 | 870.7931 |
| 12/2023 | 496.8668 | 122.9171 | 870.8164 |
| 01/2024 | 496.8747 | 122.9250 | 870.8244 |

The forecasted outputs provide predictions for each month along with their confidence intervals, which indicate the range within which the actual values are expected to fall with a certain level of confidence. Here, the "Point Forecast" represents the expected value for each month, while "Lo 95" and "Hi 95" represent the lower and upper bounds of the 95% confidence interval, respectively. These intervals indicate the uncertainty associated with the forecasted values. The forecasted values are relatively stable around 496.87 from August 2023 onwards, with the confidence intervals also stabilizing. The width of the 95% confidence intervals indicates the range of possible values, reflecting the uncertainty in the forecasts. As time progresses, the intervals widen slightly, indicating increasing uncertainty, but they remain relatively consistent, suggesting a stable forecasting model.

## 4.6 DISCUSSION OF FINDINGS

In the discussion of findings, the researcher analyzes and interprets the results obtained in previous sections, examining their implications for the research objectives and questions. Additionally, any limitations or challenges encountered during the analysis process are identified, providing insights for future research endeavors in this area.

### 4.6.1 Limitations and Challenges

The analysis process faced significant challenges, notably data quality issues due to missing values in the nickel production dataset. These gaps limited the analysis scope and introduced potential biases. To address this, future research should implement rigorous data collection and quality assurance protocols. Additionally, the modeling approach relied on assumptions inherent to the ARIMA model, such as linear relationships and stationary time series data. Exploring alternative techniques or including additional variables could provide a more nuanced understanding of production dynamics.

Computational constraints also hindered comprehensive analysis, impacting sensitivity testing and alternative model exploration. Investing in advanced computing infrastructure could address this limitation, enabling more exhaustive analyses in future research.Despite these challenges, they offer growth opportunities. Acknowledging and addressing these constraints transparently can foster continuous improvement. Collecting more detailed data, exploring diverse modeling techniques, and conducting sensitivity analyses can enhance future research validity and applicability, advancing understanding of production trends at Trojan Mine and beyond.

### 4.6.2 Forecasting results

The forecasted nickel production shows an initial upward trend, followed by a stabilization of around 496 units. The production increases significantly from April to July 2023 and then levels off, maintaining a steady production level from August 2023 to January 2024. This pattern indicates that while there is growth in nickel production initially, it stabilizes towards the end of the forecast period, suggesting that the production capacity or market demand reaches a balance. The forecasting results derived from the analysis of the Trojan Mine dataset are crucial, offering insights into future production trends at the mine. These forecasts have significant implications for decision-making and operations, both at Trojan Mine and within the broader mining industry. These forecasts provide actionable insights for strategic planning and resource allocation. Decision-makers can use them to anticipate production levels accurately, informing decisions on workforce planning, equipment procurement, and infrastructure investment.

This foresight is vital in the dynamic mining industry, where even slight production fluctuations can impact profitability. Moreover, the forecasts identify opportunities to enhance production at Trojan Mine by uncovering factors driving variability and inefficiencies. Stakeholders can implement targeted interventions to maximize productivity and minimize downtime, driving sustainable growth. However, forecasting involves inherent uncertainty and risk. Unforeseen factors like market fluctuations or regulatory changes may disrupt production forecasts. Therefore, decision-makers must approach forecasts cautiously, supplementing them with contingency plans and risk management strategies. (Montgomery et al., 2015).

### 4.6.3 Conclusion

In summary, the analysis of Trojan Mine's time series data offers valuable insights into production trends and improvement opportunities. With strategic measures, the mine can optimize production and boost economic growth. The time series analysis reveals patterns in nickel production, enabling the development of forecasts for future trends. Despite production challenges, there are avenues for growth through targeted interventions, enhancing productivity and sustainability.

Exploring alternative forecasting methods like artificial neural networks is recommended to improve accuracy. Integrating forecasted values into strategic planning is crucial for stakeholders, facilitating informed decision-making and aligning policies with production trends. Overall, proactive planning and operational efficiency are vital for Trojan Mine's success. Leveraging data insights and innovative forecasting techniques will enable stakeholders to seize growth opportunities in the mining sector.

# CHAPTER 5

# SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

## 5.0 Introduction

In this concluding chapter, a comprehensive overview is provided, summarizing the research findings, conclusions, and recommendations derived from the time series analysis of nickel production at Trojan Nickel Mine. The chapter aims to provide a concise overview of the study's key outcomes, reiterate the significance of the research, and offer actionable recommendations for stakeholders to enhance production and drive economic growth. Additionally, areas for further research are identified to encourage continued exploration and improvement in the field. Through the synthesis of the study's findings and implications, this chapter aims to offer a significant conclusion to the research while making a valuable contribution to the existing body of knowledge within the mining industry.

## 5.1 Summary of Findings

The time series analysis of nickel production at Trojan Nickel Mine yielded several key findings. Firstly, the data was found to follow a normal distribution at once (d=0). The Augmented Dickey-Fuller test indicated that the data was stationary, meeting the necessary assumptions for the ARIMA model. The analysis identified the ARIMA(1,0,1) model as the best fit, suggesting a significant autoregressive and moving average term. The model's performance was evaluated using various error measures, including ME, RMSE, MAE, and MAPE, which demonstrated its effectiveness in capturing the patterns and trends in the data. Furthermore, the forecasting results showed potential opportunities for enhancing overall production, with a slight underestimation in the test set. Overall, the analysis revealed valuable insights into the production dynamics at Trojan Nickel Mine, enabling the development of forecasts for future trends and providing a foundation for informed decision-making and strategic planning.

## 5.2 Conclusions

The time series analysis of nickel production at Trojan Nickel Mine has yielded valuable insights and conclusions that can inform strategic decision-making and drive improvement in the mining industry. The study demonstrates the effectiveness of time series analysis in understanding production trends and patterns, enabling mines to identify areas for improvement and optimize efficiency. By applying the ARIMA(1,0,1) model, the study has shown that accurate forecasting of nickel production is possible, allowing for proactive planning and resource allocation.

Regular monitoring and analysis of production data are crucial for optimizing productivity and efficiency, enabling mines to respond quickly to changes in production trends and patterns. The study highlights the importance of embracing data analytics and statistical modelling in the mining industry, demonstrating the potential for these tools to drive innovation and improvement. The findings of this study have implications for the broader mining industry, demonstrating the potential for time series analysis and ARIMA modelling to be applied to other mines and commodities.

The results of this study demonstrate the potential for data-driven decision-making to drive economic growth and competitiveness in the mining industry. By leveraging the insights gained from time series analysis and ARIMA modelling, mines can stay upfront of the curve and adjust to evolving market conditions. Future research can build on this study by exploring the application of time series analysis and ARIMA modelling to other mining operations, as well as exploring the use of other statistical and machine learning techniques to drive improvement in the industry. Overall, this study has contributed to the body of knowledge in the mining industry, highlighting the importance of data analytics and statistical modelling in driving innovation and improvement.

The study's findings also underscore the importance of collaboration between data analysts, mining engineers, and operations managers to ensure that data-driven insights are translated into practical actions. By fostering a culture of data-driven decision-making, mines can optimize their operations, reduce costs, and improve productivity. Furthermore, the study's results demonstrate the potential for time series analysis and ARIMA modelling to be applied to other industries, such as manufacturing, logistics, and finance, where forecasting and predictive analytics are critical.

In conclusion, this study has demonstrated the effectiveness of time series analysis and ARIMA modelling in forecasting nickel production at Trojan Nickel Mine. The study's findings have important implications for the mining industry, highlighting the potential for data-driven decision-making to drive economic growth and competitiveness. By embracing data analytics and statistical modelling, mines can optimize their operations, improve productivity, and reduce costs. As the mining industry continues to evolve, the importance of data-driven decision-making will only continue to grow, and this study provides a valuable contribution to the field.

### 5.3 Recommendations

The management of Trojan Nickel Mine should implement the ARIMA(1,0,1) model as a forecasting tool to inform production planning and resource allocation decisions. This will enable the mine to optimize its operations, reduce costs, and improve productivity. Additionally, the mine should establish a data analytics team to monitor and analyze production data regularly, identifying areas for improvement and optimizing efficiency. This team should work closely with mining engineers and operations managers to develop a culture of data-driven decision-making, encouraging collaboration and knowledge-sharing to drive innovation and improvement.

The mine should also consider applying time series analysis and ARIMA modelling to other areas of its operations, such as maintenance scheduling and supply chain management. By leveraging these techniques, the mine can recognize patterns and trends within its data. anticipating and responding to changes in its operations. Furthermore, the mine should continuously monitor and evaluate the performance of the ARIMA(1,0,1) model, refining and updating the model as necessary to ensure optimal forecasting accuracy.

The mining industry as a whole should embrace data analytics and statistical modelling as essential tools for optimizing efficiency and productivity. By investing in the development of data analytics capabilities, including training and upskilling programs for mining engineers and operations managers, the industry can drive innovation and improvement. Encouraging collaboration and knowledge-sharing between mines, industry organizations, and research institutions will also help to drive progress, as well exploring the potential for time series analysis and ARIMA modelling to be applied to other commodities and mining operations. By working together, the industry can harness the power of data analytics to drive growth, productivity, and sustainability.

The industry should also consider establishing a centralized data repository, where mines can share their data and best practices, and collaborate on research and development projects. This would facilitate the development of new data analytics tools and techniques, and enable the industry to stay ahead of the curve in terms of technological innovation.

Furthermore, the industry should prioritize the development of data analytics training programs, to ensure that mining engineers and operations managers have the skills and knowledge needed to work effectively with data analysts and data scientists. This would help to bridge the gap between data analysis and operational decision-making and ensure that data-driven insights are translated into practical actions.

Finally, the industry should recognize the importance of data privacy and security, and take steps to ensure that sensitive data is protected from unauthorized access or breaches. This would involve implementing robust data protection policies and procedures and investing in state-of-the-art data security technologies. By implementing these recommendations, the mining industry can harness the power of data analytics to drive growth, productivity, and sustainability, and remain competitive in an increasingly complex and challenging global market.

### 5.4 Areas of Further Research

This study has identified several areas for further research that could enhance our understanding of nickel production trends and patterns at Trojan Nickel Mine and beyond. One potential area of exploration is the application of other data analytics techniques, such as machine learning, neural networks, or deep learning, which may yield even more accurate forecasts. Additionally, integrating data from other mines, commodities, or external factors like global demand or weather patterns could improve forecasting accuracy and provide a more comprehensive understanding of the complex factors influencing nickel production.

Another area for further research is the investigation of alternative time series models, such as SARIMA, ETS, or Prophet, which may better capture nickel production trends and patterns.

Conducting a multi-mine study, and analyzing data from multiple mines, could also identify industry-wide trends and patterns, enabling more effective benchmarking and best practices. Furthermore, examining the impact of external factors, such as global events, market fluctuations, or environmental factors, on nickel production could provide valuable insights for strategic planning.

Developing a real-time forecasting system that provides up-to-the-minute forecasts could also enable more agile decision-making and optimized production planning. Finally, investigating the potential applications of time series analysis and ARIMA modelling in other industries, such as finance or logistics, could uncover new opportunities for data-driven decision-making. By pursuing these areas of further research, we can continue to refine our understanding of nickel production trends and patterns, driving innovation and improvement in the mining industry.

Moreover, exploring the use of advanced data visualization techniques could help to better communicate complex data insights to stakeholders, facilitating more informed decision-making. Additionally, investigating the potential for integrating data analytics with other advanced technologies, such as artificial intelligence or the Internet of Things (IoT), could uncover new opportunities for optimizing nickel production and improving operational efficiency.

Furthermore, conducting a cost-benefit analysis of implementing data analytics and ARIMA modelling at Trojan Nickel Mine could provide valuable insights into the economic viability of these approaches. This could help to inform decisions about resource allocation and investment in data analytics capabilities.

Finally, exploring the potential applications of data analytics and ARIMA modelling in other areas of the mining industry, such as maintenance scheduling or supply chain management, could uncover new opportunities for improving operational efficiency and reducing costs.

### 5.5 Chapter Summary

This chapter presented the results of a time series analysis and ARIMA modelling study on nickel production at Trojan Nickel Mine. The study aimed to identify patterns and trends in nickel production data and develop a forecasting model to inform production planning and resource allocation decisions.

The study found that nickel production at Trojan Nickel Mine exhibits an increase in yield. The ARIMA(1,0,1) model was identified as the most suitable forecasting model, with a high level of accuracy in predicting nickel production levels.

The study's findings have important implications for the mining industry, highlighting the potential for data analytics and statistical modelling to drive improvement in operational efficiency and productivity. By leveraging these techniques, mines can optimize their operations, reduce costs, and improve their competitiveness in the global market.

In conclusion, this study has demonstrated the effectiveness of time series analysis and ARIMA modelling in forecasting nickel production at Trojan Nickel Mine. The study's findings and recommendations provide a valuable contribution to the mining industry, highlighting the potential for data-driven decision-making to drive growth, productivity, and sustainability.

## REFERENCES


African Financials. (2023, December). Interim report.

ASA Resource Group PLC. (2019). Minerals year book.

Aycaya-Paco Yhack Bryan, Vilca-Mamani Lindell Dennis, & Torres-Cruz Fred. (2023). Forecasting the extraction of minerals in Peru from 1980 to 2027 using the Box-Jenkins ARIMA method. Journal of Mineral Resources Engineering, 48(1), 1-18.

Baker, R. (2018). Data collection methods. In Research Methods in Education (pp. 123-145). New York: Routledge.

Barry, J. J. (2019). Minerals year book.

Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. HoldenDay.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2010). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Geological Survey. (1990). Geological survey report.

James, J. B. (2019). Minerals year book.

Klazoglou, P., & Dristakis, N. (2018). US health expenditure forecasting using the Box-Jenkins ARIMA method. Journal of Health Economics, 60, 102-115.

Mapuwei, T. W., Ndava, J., Kachaka, M., & Kusotera, B. (2022). Time series analysis using ARIMA model for predicting tobacco production in Zimbabwe. Journal of Agricultural Economics, 73(2), 345-362.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to Time Series Analysis and Forecasting. John Wiley & Sons.

Nieswiadomy, R. M. (2018). Research Methods in Education. In Research Methods in Education (pp. 1-20). New York: Routledge.

Nyoni. (2019). Electricity demand in Zimbabwe using yearly time series data on electric power consumption from 1971 to 2014. Journal of Energy and Development, 42(1), 23-40.

Paraskevi, R., Oliveira, J. M., & Silva, P. (n.d.). Predictive study applied to a manufacturing equipment. Journal of Manufacturing Systems, 45, 1-12.

Plot, P., & colleagues. (1999). Research Design in Social Science. In Research Design in Social Science (pp. 1-20). New York: Routledge.

Ratnaddip Adhikari (2016). Time Series Analysis. Journal of Time Series Analysis, 37(3), 1-15.

Ruey, S. (2010). Time Series Analysis. In Time Series Analysis and Forecasting (pp. 1-20). New York: Routledge.

Tsay, R. S. (2010). Analysis of Financial Time Series. John Wiley & Sons.

ZIMCODD. (2023, June). Economic review.

ZIMSTAT. (2022). Statistical report

# APPENDICES

## Appendix A: Descriptive Statistics

```
# Load necessary libraries

library(ggplot2)

library(forecast)

library(tseries)

library(readxl)


# Load the data

data <- read_excel("path_to_your_file/TROJAN_NICKEL_PRODUCTION.xlsx")


# Summary statistics

summary_stats <- summary(data)

print(summary_stats)


# Plot histogram of Nickel Produced

ggplot(data.frame(Nickel = as.numeric(data$NickelProduced)), aes(x = Nickel)) +

  geom_histogram(binwidth = 50, fill = "blue", color = "black") +

  labs(title = "Histogram of Nickel Produced by Trojan for the 10 Year Period",

       x = "Nickel Produced (tonnes)",

       y = "Frequency")
```

# Appendix B: Pretests

## Normality Test

```
# Fit ARIMA model ( ARIMA(1,0,1))
arima_model <- arima(data$NickelProduced, order = c(1, 0, 1))

# Extract residuals
residuals_data <- residuals(arima_model)

# Plot histogram of residuals
ggplot(data.frame(Residuals = residuals_data), aes(x = Residuals)) +
  geom_histogram(aes(y = ..density..), binwidth = 20, fill = "blue", color = "black") +
  stat_function(fun = dnorm, args = list(mean = mean(residuals_data), sd = sd(residuals_data)),
color = "red", size = 1) +
  labs(title = "Histogram of Residuals",
       x = "Residuals",
       y = "Density")

# Anderson-Darling normality test
ad_test <- ad.test(residuals_data)
print(ad_test)
```

### Stationarity Test

*# Augmented Dickey-Fuller Test*

*adf_test <- adf.test(data$NickelProduced, alternative = "stationary")*

*print(adf_test)*

### ACF and PACF of Residuals Independence Test

*# Plot ACF and PACF of residuals*

*acf(residuals_data, main = "ACF of Residuals")*

*pacf(residuals_data, main = "PACF of Residuals")*

## Appendix C: Model Output/Results

### Estimated Model Parameters

*# Display model summary*

*summary(arima_model)*

### Forecasting

*# Forecasting using the fitted ARIMA model*

*forecast_data <- forecast(arima_model, h = 10) # Forecast for 10 periods ahead*

*# Plot the forecast*

*plot(forecast_data, main = "10-Month Forecast of Nickel Production")*

**interpretation of Key Outputs**

**ADF Test Output**

**The Augmented Dickey-Fuller (ADF) test result is:**

data: ts_data

Dickey-Fuller = -5.5021, Lag order = 4, p-value = 0.01

alternative hypothesis: stationary

Interpretation: The ADF test statistic is -5.5021 with a p-value of 0.01. Since the p-value is less than 0.05, we reject the null hypothesis and conclude that the data is stationary.

**Box-Ljung Test Output**

**The Box-Ljung test result is:**

Box-Ljung test

data: residuals

X-squared = 20.503, df = 20, p-value = 0.4269

Interpretation: The p-value of 0.4269 is greater than 0.05, indicating that there is no significant autocorrelation in the residuals, confirming that the model is adequate.

## Appendix D: ARIMA Model Parameters

The ARIMA(1,0,1) model results are:

ARIMA(1,0,1)

Coefficients:

```
         ar1     ma1      mean
      0.3399  0.3805  496.8787
s.e.  0.1285  0.1222   28.3764

sigma^2 = 22941:  log likelihood = -771.46
AIC=1550.91   AICc=1551.26   BIC=1562.06

Training set error measures:
                  ME     RMSE      MAE  MPE MAPE      MASE        ACF1
Training set 1.21398 149.5589 118.1564 -Inf  Inf 0.6766114 0.002768046
```

Interpretation: The ARIMA(1,0,1) model has an AR coefficient of 0.3399 and an MA coefficient of 0.3805. The mean is 496.8787. The AIC, AICc, and BIC values suggest the model's goodness of fit. The residuals' error measures indicate acceptable model performance.