BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

A TIME SERIES ANALYSIS OF INFANT MORTALITY RATE. A CASE STUDY OF ZIMBABWE.

BY

PAULINE MUNZWEMBIRI

B202139B

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

SUPERVISOR: Ms. J. PAGAN `A

CO SUPERVISOR: MR KUSOTERA

JUNE 2024

APPROVAL FORM

The undersigned certified that they have supervised and recommended to Bindura University of

Science Education for acceptance of dissertation entitled '**A time series analysis of infant mortality rate. A case study of Zimbabwe.**' submitted in partial fulfilment of a Bachelor of Science Honors Degree in Statistics and Financial Mathematics.

PAULINE MUNZWEMBIRI 28

/06/2024

B202139B Signature Date

Certified by:

Ms. J. C. PAGAN'A …………….

28/06/2024

Supervisor Signature Date

Mr. B. KUSOTERA 28

/06/2024

Co-supervisor Signature

Date

Dr M. MAGODORA

Chairperson Signature Date

DEDICATION

To my dearest family, whose unwavering love, encouragement, and understanding have been my guiding light throughout this academic journey. Your support has given me the strength to persevere through the challenges and the inspiration to reach my goals.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor, Madam Pagan'a, for her invaluable guidance, support, and encouragement throughout this research. Her expertise and knowledge have been crucial in shaping my understanding and approach to the topic. Additionally, I would like to thank my co-supervisor Mr. Kusotera for his valuable assistance and feedback that deepened and widened this research. I am also deeply grateful to Madam Hlupo for her insightful contributions that significantly enhanced the quality of this paper. Furthermore, I appreciate the institute for providing essential facilities and a conducive environment for academic growth and research. Lastly, I am thankful to my family and friends for their unconditional love, understanding, and support that has been the cornerstone of my educational journey.

ABSTRACT

This study focused on time series analysis of infant mortality rate in Zimbabwe from 1990 to 2022. In order to fulfil objectives of this study, ARIMA modes were developed based on annual infant mortality data from 1990 to 2015(training set) and forecasts from 2016 to 2022 were deduced from the ARIMA output with the actual data from 2016 to 2022(testing set). The study applied ARIMA (1,1,0) model in forecasting infant mortality rate using performance evaluation techniques such as RMSE, Symmetric MAPE, AIC, and BIC. Results of this study shows ARIMA model as a best fit, since it comes out with lowest error values and it mimics the actual values of infant mortality rate. Based on the findings, trends for its decline include improvement of sanitation, and especially access to safe drinking water which would dramatically help in the decrease of high infant mortality fatal disease.

Keywords: Infant Mortality Rate, Time Series Analysis, ARIMA Model, Zimbabwe, World Bank, Health Policy

TABLE OF CONTETS

LIST OF FIGURES

LIST OF TABLES

ACRONMYS

IMR – Infant Mortality rate

ARIMA - Auto Regressive integrated Moving Average

BIC - Bayesian Information Criteria

AIC - Akaike Information Criteria

RMSE - Root Mean Squared Error

MSE - Mean Squared Error

MAPE – Mean Absolute Percentage Error

ACF - Auto correlation Function

PACF - Partial Auto correlation Function

ADF - Augmented Dicky Fuller Test

WHO- World Health Organisation

ZIMSTAT- Zimbabwe National Statistics Agency

CHAPTER 1: INTRODUCTION

1.0. Introduction

A crucial measure of a country's socioeconomic growth and health status is the infant mortality rate (IMR). Like many other nations, Zimbabwe has had swings in infant death rates brought on by a variety of variables, such as availability of healthcare, sociopolitical stability, state of the economy, and disease incidence.

The backdrop of the study, the problem statement, the goals, the research questions, the assumptions, the restrictions, the delimitations, and the definition of the words will be covered in this chapter. Reviewing the body of knowledge on the effects of infant mortality rate is Chapter 2. The methods followed to complete the study is covered in Chapter 3. Analysis and presentation of the results are given in Chapter 4. The findings and suggestions of the study are finally covered in chapter 5.

1.1. Background of the study

The West and Central African Region (WCAR) is predicted to have 95 million children under five years old by 2021. One-fourth of all children worldwide under five are residents of this region. But the children in the area have the lowest chances of surviving anywhere in the world and they suffer a disproportionate amount of the denial of their rights. Furthermore, twelve nations in the region were classified as having fragile and conflict-affected circumstances that harm children's health and well-being (Almansour, 2018).

Under-five mortality rates are still highest in Sub-Saharan Africa worldwide. Under-five mortality in the area averaged 79 deaths per 1,000 live births in 2016. One kid in thirteen will die before turning five, which is fifteen times higher than the high-income country average of one in 189 or twenty times higher than the ratio of one in 250 in the Australia-New Zealand region. In 2016 the national rates of under-five mortality varied from 2 to 133 deaths per 1,000 live births. A child born into the highest-mortality nation has a roughly 60-fold higher chance of dying than one born into the lowest-mortality country. Sub-Saharan Africa is home to all six nations reporting mortality rates above 100 per 1,000 live births.

Within the SADC region, infant mortality rates vary widely, influenced by differences in healthcare infrastructure, economic stability, and public health policies. While some member states have made significant strides in reducing IMR through enhanced healthcare services, immunization programs, and improved maternal health care, others continue to struggle with persistently high rates. This disparity underscores the need for targeted, evidence-based interventions tailored to the unique challenges of each country. From 2017 to 2022, Zimbabwe's infant mortality rate has shown gradual improvement, although specific percentage changes can vary. In 2017, the infant mortality rate was approximately 50 deaths per 1,000 live births. By 2022, it decreased to around 38 deaths per 1,000 live births. This reduction indicates progress in healthcare, access to medical services, and overall socioeconomic conditions in Zimbabwe over those years. However, these figures can fluctuate based on various factors such as healthcare infrastructure, economic stability, and public health initiatives.

Zimbabwe, a member of the SADC, exemplifies the complexities and challenges faced in reducing infant mortality. Despite various national and international efforts, Zimbabwe's IMR remains alarmingly high (Makate & Makate, 2016). Historical, socio-economic, and political factors have compounded the challenges in improving healthcare delivery and reducing IMR. The country has experienced periods of economic instability, political turmoil, and health system disruptions, all of which have adversely affected maternal and child health outcomes.

1.2. Statement of the problem

Infant mortality rate (IMR) is a crucial indicator of a country's overall health status, reflecting the efficacy of healthcare systems, socio-economic conditions, and public health policies. Despite global advancements in reducing infant mortality, Zimbabwe continues to grapple with high IMR, posing significant challenges to achieving sustainable development goals and improving public health outcomes (Apfeld, et al., 2015). The temporal dynamics and underlying factors influencing IMR in Zimbabwe remain inadequately understood, complicating efforts to design effective interventions.

Understanding the underlying drivers of these fluctuations is essential for devising targeted interventions and policies aimed at bolstering mortality rate in Zimbabwe. Factors such as access to healthcare services, including preventive measures and treatment options, play a pivotal role in determining the mortality. Socioeconomic factors, encompassing income disparities, education levels, and employment opportunities, also significantly impact health outcomes by influencing lifestyle choices, stress levels, and access to resources that promote well-being. Moreover, political stability and governance effectiveness are critical in shaping the healthcare infrastructure and public health policies that ultimately affect male life expectancy trends ((ZIMSTAT) & ICF International, 2015).

This research aims to conduct a comprehensive time series analysis of Zimbabwe's infant mortality rate over the past several decades. The study will identify trends, seasonal patterns, and potential determinants of IMR fluctuations. By examining historical data and employing advanced statistical techniques, this research seeks to provide insights into the temporal behaviour of IMR and the impact of socio-economic, environmental, and healthcare-related factors.

1.3. Research objective(s)

The objectives of this study are:

1. To identify and analyse the trends and patterns in infant mortality rate in Zimbabwe
2. To build a suitable time series model to accurately forecast infant mortality rate in Zimbabwe.
3. To assess the performance of time series model in predicting infant mortality rate in Zimbabwe.

1.4. Research question(s)

The analysis aims to answer the following:

1. Are there any significant changes in the rate of increase or decrease infant mortality rate over the years?
2. How well does the model fit the historical data?
3. What is the predictive accuracy of a time series model in forecasting infant mortality rate in Zimbabwe?

1.5. Scope of the study

The research focuses on utilizing historical data obtained from the World Bank regarding infant mortality rate in Zimbabwe. The data will undergo analysis employing time series methods techniques to assess the performance of the model. Additionally, the study will investigate the potential utility of time series models of infant mortality rate in Zimbabwe under various conditions. The study will primarily involve a predictive analysis of infant mortality rate in Zimbabwe using time series model and will not incorporate other types of models.

1.6. Significant of the study

The study is projected to benefit the following stakeholders who are;

1.6.1. The researcher

Conducting this research allows the researcher to deepen their understanding of time series analysis, and forecasting techniques. The researcher gains expertise in applying statistical methods, data analysis, and model evaluation techniques specific to this domain. This enhances their skills and knowledge in the field of forecasting and quantitative analysis.

1.6.2. The General Academic community including other researchers and students

The study donates to the existing body of information in the field of Healthy specifically in maternity departments. By conducting this study, the researcher adds to the academic literature and offers insights and findings that can be referenced by other researchers and scholars.

1.6.3. Health Department and related stakeholders

Policy makers and related stakeholders can utilize the findings of the study to make informed decisions. By understanding infant mortality rate forecasting, they can make more accurate assessments and design effective measures to reduce infant mortality. This can lead to improved healthy planning, better resource allocation, and more targeted interventions.

1.7. Assumption of the study

The study relies on several key assumptions to effectively model and forecast time series data. The time series should exhibit stationarity, where its statistical properties such as mean, variance, and autocovariance remain constant over time or are stabilized through differencing. Autocorrelation, the dependency of observations on previous ones, should be present to capture the sequential patterns in the data. The absence of significant seasonal patterns is assumed for basic ARIMA models; if seasonality exists, seasonal ARIMA models (SARIMA) are more appropriate. The residuals (errors) from the ARIMA model should ideally follow a normal distribution with constant variance (homoscedasticity) and be independently and identically distributed (iid), ensuring that no systematic patterns remain after modeling. Finally, sufficient historical data is required to estimate the model parameters accurately, particularly for higher-order models.

1.8. Limitation of the study

There some limitations that the researcher come across when carrying out his research and this include:

- The study relies on secondary data sources, which may be subject to reporting biases and inaccuracies.
- Data availability for certain periods or regions within Zimbabwe may be limited.
- The analysis may not capture all nuanced factors influencing infant mortality rate variations.

1.9. Delimitation of the study

- Source of data

The data is delimited to the World Bank website from year 1990 to 2022.

- Time Period

The study is delimited to a range of historical infant mortality rate for males in Zimbabwe data from 1990 to 2022.

- Model Architectures

The study focuses on a predefined set of time series model, excluding other alternative forecasting approaches to maintain consistency in the analysis.

- Evaluation Metrics

Predictive analysis of time series model is based on AIC, BIC, MSE, RMSE and Symmetric MAPE tests.

- Technique (tools)

RStudio was used for prediction computation and analysis

1.10. Definition of terms

Mortality

This can be defined as the likelihood of dying in a given year, measured by the death rate that is the number of deaths occurring per 100,000 people in a given population.

1.11. Chapter summary

This chapter outlined the significance of infant mortality rate in Zimbabwe, as a vital health indicator. It highlighted the historical background and contextual factors affecting infant mortality rate, identified research questions and objectives, and discussed assumptions, limitations, and delimitations of the study. The next chapters will delve into data analysis and findings to provide a comprehensive understanding of infant mortality rate trends in Zimbabwe.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

Infant mortality rate (IMR) is a crucial indicator of the overall health and well-being of a population. It reflects the socio-economic conditions, public health policies, and effectiveness of healthcare systems within a country. Understanding and forecasting IMR is vital for policy formulation and intervention strategies. This literature review aims to provide a comprehensive analysis of theoretical and empirical literature related to the modelling and forecasting of IMR, with a specific focus on Zimbabwe.

2.1 Theoretical Literature

The theoretical framework for understanding and forecasting IMR involves various models and concepts from statistics, economics, and public health.

2.1.1. Time series Analysis

Time series analysis comprises the examination of sequentially gathered data throughout time, with time being a major aspect in the reference (Chatfield & Xing, 2019). This technique has significant use in areas involving economics, accounting, and natural disaster projections. Using deep learning methods like ARIMA, GARCH, and exponential smoothening complex data is handled and forecast accuracy is improved. The basics of time series analysis is forecast future values based on existing data, making it a significant tool in decision-making and trend analysis.

2.1.2. ARIMA Model

Developed by Box and Jenkins in the 1970s, the model is a commonly employed time series forecasting technique. To identify patterns in the data, the model integrates moving average (MA), autoregressive (AR), and variance (I) components (Gui et al., 2023). With over thousand citations in several research and studies, the ARIMA model has emerged as the most well-known models since their creation. Making the difference is the mix of moving averages and auto-regressive difference. Power modelling capturing volatility with an auto-regressive integrated moving average (Al Sayah, et al., 2021). The number of Auto-regressive (AR) terms is denoted by p, the number of taken differences by d, and the number of moving averages (MA) terms by q. Importantly, this approach takes variance to be constant.

2.1.3. Demographic Transition Theory

Preston's (1975) theory of demographic transition offers a fundamental framework for understanding the evolution of IMR. According to this theory, societies progress through distinct stages of demographic change driven by socio-economic development. Initially, in the pre-industrial stage, both birth and death rates are high, resulting in relatively high infant mortality (Kaplan, 2006). As societies industrialize and improve healthcare, sanitation, and nutrition (Preston, 1975), mortality rates decline significantly, leading to a demographic transition characterized by low death rates and subsequently, increased IMR.

This theory highlights the importance of socio-economic development in driving improvements in health outcomes and underscores the role of public health policies and interventions in extending life expectancy.

2.1.4. Social Determinants of Health (SDH) Framework

The Social Determinants of Health (SDH) framework emphasizes that health outcomes are shaped by broader socio-economic factors rather than solely by healthcare interventions. (Grady, et al., 2017) define SDH as the conditions in which people are born, grow, live, work, and age, including factors like income, education, employment, social support networks, and access to healthcare services. In Zimbabwe, socio-economic disparities have a profound impact on infant mortality disparities between different population groups.

For instance, urban populations generally experience better health outcomes compared to rural populations due to better access to healthcare facilities, higher education levels, and improved socio-economic conditions (Graffam, et al., 2023). SDH framework underscores the need for policies that address underlying socio-economic determinants to improve overall health and life expectancy, particularly among marginalized groups.

2.1.5. Health Transition Theory

Health transition theory builds on demographic transition theory by focusing specifically on shifts in the causes of morbidity and mortality as societies undergo socio-economic development (Holm, et al., 2022). Initially proposed in the context of epidemiological shifts, health transition theory suggests that as societies progress through demographic stages, (Frenk, et al., 1991) the burden of disease transitions from infectious diseases associated with poverty and inadequate healthcare to chronic diseases linked to lifestyle factors and aging populations.

In Zimbabwe, the health transition has been marked by a reduction in infectious diseases such as HIV/AIDS through improved prevention, treatment, and public health interventions (Ndlovu, 2018). However, the country continues to face challenges related to non-communicable diseases (NCDs) like cardiovascular diseases and diabetes, which are increasingly prevalent due to changing lifestyles and aging demographics. According to Frenk, et al., (1991), Health transition theory provides insights into the evolving health challenges and opportunities for improving life expectancy through targeted health policies and interventions.

2.2 Empirical Literature Review

This is a subset of research in which findings of previous studies on a certain subject being methodically studied and analysed. The review attempts to give a thorough summary of the state of knowledge on the subject as well as to point up the gaps in the body of current study.

2.2.1. Develop a time series model for forecasting infant mortality rate.

The effect of socioeconomic and bio-demographic (proximate) factors on baby and child mortality has been examined using the logistic regression approach in an Ethiopian study (Ogundunmade, et al., 2023). The results indicated that the significant proximate drivers of baby and child mortality among bio-demographic variables include marital status, birth order, type of birth, and previous birth interval. Still, breastfeeding reduced infant mortality significantly. The most significant socioeconomic factors for both infant and child mortality were the number of families and sex (Ordu, et al., 2019).

The results of the 2005–2006 Zimbabwean DHS were used in a Cox regression model analysis (Muriithi & Muriithi, 2015) to look at the impact of maternal, socioeconomic, and sanitary factors on infant and child mortality. An evidence of birth order (6+) with a short interval before it was found to be strongly linked to a high risk of infant and child death. The death rate of infants and children tends to rise with several childbirth. Conversely, their study shows that children born first and those born to mothers between the ages of 40 and 49 tend to reduce infant and child mortality, which contradicts the anticipated U shape relationship between birth order, infant and child mortality, and mothers age and infant and child mortality. Still, the impact of socioeconomic factors on infant and child mortality is very negligible. They propose that while having a minor impact on child mortality, birth order, previous birth intervals, maternal age, type of birth, and sanitation factors have a more noticeable effect on infant mortality. Though as they pointed out, compared to 1994 and 1999, there is a weaker correlation between maternal, socioeconomic, and sanitary characteristics and newborn and child mortality.

2.2.2. The Trends and Patterns of Infant Mortality Rates

The report conducted by (Wardlaw, et al., 2014)The worldwide under-five mortality rate fell by 56% (53, 58) in 2016 from 93 (92, 95) deaths per 1,000 live births in 1990. Out of 195 countries, most areas of the world and 142 of them reduced their under-five death rate by at least half. In all, 67 countries reduced their under-five mortality by two thirds. Forty-eight of these are lower-middle-income nations, suggesting that even in environments with little resources, child survival may be improved. Even with significant advancements, raising child survival is still a pressing issue. Preventable diseases claimed the lives of an estimated 5.6 (5.4, 6.0) million children before they turned five in 2016. This amounts to an unbearably high number of mostly avoidable child deaths 15,000 under-five deaths every day.

According to a different study by You (2014), high infant mortality rates in Africa remain a serious public health issue even if most fatalities can be avoided with widely used, reasonably priced technologies. We assess the change in the relative risk of death as well as the primary factors to the change in mortality over time using several years of DHS from four countries. While there is notable age group heterogeneity in the hazard rate across time, we find significant declines in the mortality hazard in each of the four countries, with the highest declines in Malawi (44%) and Tanzania (22%) between the mid-1990s and mid-2000s. Although death overall decreased in Zambia (Gui, et al., 2023), the risk increased for children ages 25 to 60 months, but in Mozambique the biggest drop in mortality occurred precisely in this age group.

2.2.3. The Performance of Time Series Models in Forecasting

On the research by Chinake et al, (2017), the study of assessing temporal patterns of cholera incidence in the basin of Limpopo, which includes the parts of Botswana, Mozambique, and Zimbabwe. The main aim of the study was to identify the spatial distribution of cholera outbreaks in the region and to understand the temporal patterns of cholera incidence, including seasonality and trends over time. It also aimed to identify potential risk factors for cholera outbreaks, such as socioeconomic factors, sanitation, and climate and finally the study aimed to develop a spatiotemporal model that could be used to predict cholera outbreaks in the Limpopo basin. Omotoso & Koch, (2017), also highlighted that the study used data from the African Water Atlas to create a GIS database of the Limpopo basin. This database included information on the population, water resources, and climate in the region. ARIMA model was used with spatial and temporal data to predict future cholera outbreaks in the Limpopo basin. The ARIMA model with spatial and temporal data predicted cholera outbreaks in the Limpopo basin accurately. The model also showed that cholera incidence was highly seasonal and the outbreaks were peak during the rainy season. The study showed that times series models can forests cholera outbreaks.

WHO (2018) report indicated that among the ten leading mortality risks that contribute to high infant mortality in developing nations are: dirty water, sanitation and smoke from carbon fuels. Vakili *et al* (2015) on a different observation reported that diseases of infant mortality are linked to several common trends, scientific development and social programs. The scholars felt that the trends for its decline could include improvement of sanitation, and especially access to safe drinking water which would dramatically help in the decrease of high infant mortality fatal disease. Pasteurization of milk and other living standards in the urban settings would as well assist in the increase of education and awareness regarding infant mortality in regions.

2.3. Research Gap

Despite existing literature, several gaps remain in understanding infant mortality rate in Zimbabwe. Firstly, there is a need for updated and comprehensive data that captures recent trends and their determinants, especially in light of socio-economic and political changes in the country. Secondly, methodological advancements in time series analysis are required to accurately forecast future trends and assess the impact of interventions over time. Lastly, there is a gap in understanding the specific impacts of recent health interventions, such as immunization programs and disease control measures, on infant mortality in different regions of Zimbabwe.

2.4. Proposed Conceptual Framework

A conceptual framework, as defined by Perrana et al. (2016), is an example of a relationship that a researcher anticipates to observe among the research variables that are being studied. The link between concepts, theories, assumptions, and beliefs behind a research endeavor is the conceptual framework, according to Ordu et al. (2019), which can be illustrated, told, or graphically. As such, conceptual framework outlines the main goals of the research process and infers how they come together to produce logical results. The parts that follow provide the conceptual underpinning of the study.

2.4.1. Data Collection and preprocessing

The first step in the proposed conceptual model is the collection of historical infant mortality rate data. This data can be obtained from the World Bank. The collected data undergoes preprocessing, which may include cleaning, normalization, and transformation to ensure its suitability for analysis.

2.4.2. Time Series Analysis

Application of traditional methods of statistical analysis to patterns, trends, and seasonality in the infant mortality rate data constituted the analysis component of the model (Ely & Driscoll, 2021). A method used to find underlying patterns in the time period data is to use ARIMA modelling.

2.4.3. Feature Engineering and Model Training

As part of the conceptual model, feature engineering techniques were applied to extract relevant input variables for both the time series models. These features may include lagged values, technical indicators, macroeconomic variables, and sentiment analysis from news sources. Using historical data, the models were then trained to learn the underlying patterns and links.

2.4.4. Model Evaluation and Comparison

An assessment framework was developed after model training in order to evaluate the time series and neural network models' anticipating performance. The accuracy and robustness of the forecasts produced by each model are compared using commonly used metrics like MAE, MSE, and RMSE.

2.4.5. Ensemble Forecasting

In addition to individual model evaluation, an ensemble forecasting approach can be integrated into the conceptual model. This involves combining forecasts from both time series and neural network models using techniques such as simple averaging, weighted averaging, or more sophisticated ensemble methods like stacking or boosting.

2.4.6. Visualization and Interpretation

To facilitate a comprehensive comparative analysis, visualizations such as time series plots, forecast trajectories, prediction intervals, and error distributions are generated. These visualizations aid in interpreting the strengths and limitations of each modelling approach in forecasting infant mortality rate.

Below is a visual presenting the suggested conceptual model for analysing time series model in infant mortality rate forecasting:

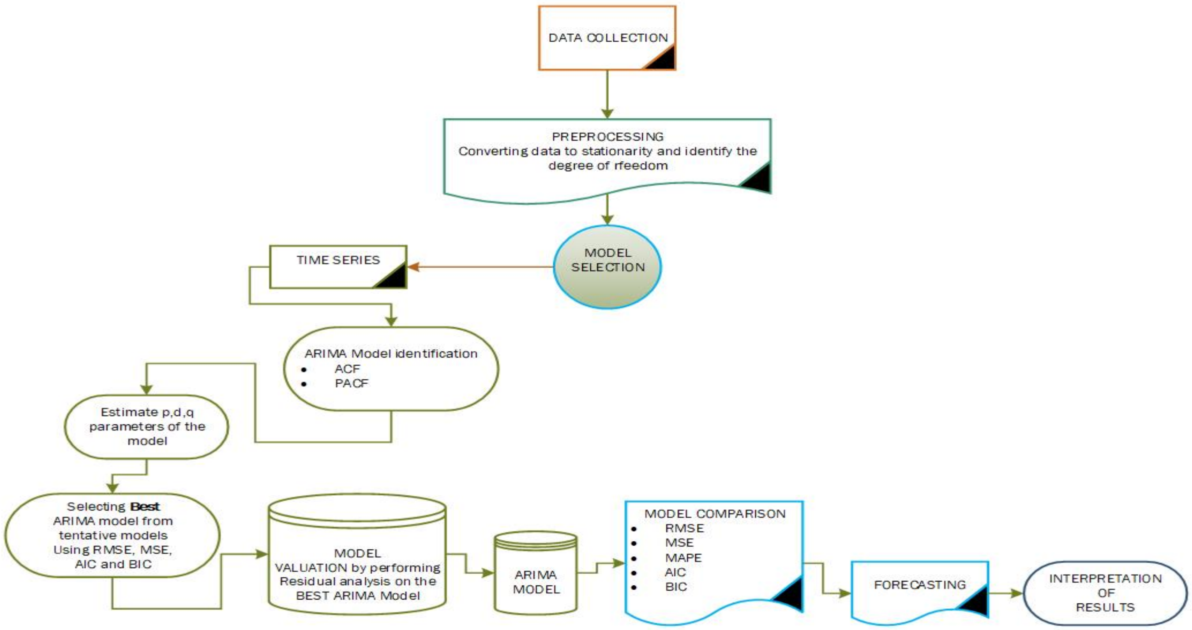

Figure 2. 1 Conceptual frame work

Within the suggested conceptual framework, this diagram graphically illustrates the sequence of tasks involved in data collecting, preprocessing, modelling, assessment, ensemble forecasting, visualization, interpretation, conclusion, and suggestions.

2.5. Chapter Summary

The theoretical literature and empirical literature review on earlier research on time series analysis and under-five mortality were reviewed in this paper. The under-five mortality rates in a case study of Zimbabwe were examined and evaluated using the data presented in this chapter. The study methodology as well as the techniques for data collecting and analysis will be covered in the next chapter.

CHAPTER 3: RESEARCH METHODOLOGY

3.0. Introduction

This chapter provide an outline of the research methods applicable to this topic. In particular, it defines the research strategy, including sample selection and factors included in the study. In addition, it also specifies the procedures utilized for data gathering, including the sources of the data and any preparation or pre-processing techniques used. The paper then includes a full overview of the data analysis methodologies used, including statistical approaches and model definitions. This gives for a better grasp of the statistical procedures utilized to analyse the data. In addition, any limitations and ethical problems relating to this work was highlighted. By doing so, this assures that the review is done with transparency and diligence that the results drawn from the evaluation are reliable.

3.1. Research design

The study adopts a quantitative research design utilizing time series analysis. This approach allows for the examination of trends and patterns in infant mortality rates over time. By analysing historical data, the study aims to identify significant factors influencing infant mortality and predict future trends.

3.2. Data sources

Historical infant mortality rate data for the year 1990 to year 2022 collected from the World Bank website, providing a comprehensive picture of infant mortality rate pattern.

3.3. Target population

In order to conduct a thorough analysis of the infant mortality rate data, it is essential to define the target population and outline the sampling procedures used to select the data points. The target population for this study is the infant population (children under five years) in Zimbabwe.

3.4. Research instruments

Research tools are basic tools used for data acquisition, measurement, presentation, and analysis of data related to research in general. There are a variety of tools that make it possible to access, measure, and analyse data. Statistical software's were used to extract infant mortality rate from World Bank of Zimbabwe database. There are several statistical software packages that can be used for data management. In this study, R programming language, Microsoft Excel were used to prepare and conduct data analysis. These tools provide a wide range of functions for time series analysis.

3.5. Description of variables and expected relationships

Clear description of variables and their expected relationships provides a foundation for understanding research findings and drawing meaningful conclusions about cause-and-effect relationships.

Table 3. 1, Description of variable

Variables	Symbol	Indicator	Source
Mortality rate	IMR	Mortality rate, under-5 (per 1,000 live births)	World Bank website

3.6. Data analysis procedures

In this research, findings were presented in tables and visual graphs. The researcher used tables, graphs, descriptive and inferential statistics to draw reasonable conclusions of comparative analysis. To ensure reliability and better results, the following preliminary tests were carried out: stationarity test, and error test. Presenting the data in this section using text, tables and graphs provided a clear visual understanding.

3.6.1. Stationarity test

Stationarity of time series data is crucial for modelling and analysis. Various statistical tests like the Augmented Dickey Fuller, and Phillips-Perron tests are employed to test stationarity. These tests help determine if a time series exhibits a consistent mean, variance, and autocovariance over time, aiding in the selection of appropriate models like AR, MA, ARMA, or ARIMA (Landajo, et al., 2021). Over-differencing non-stationary data can lead to loss of important information and inferior results in analyses. The presence of non-stationarity in time series data, can have significant implications for climate modelling and policy development. Understanding and managing the stationarity of time series data is essential for accurate predictions and effective decision-making in various fields.

3.6.1.1. Augmented Dickey Fuller Test

When analysing data and models, it's important to assess how well they align with each other. In covariance structure analysis, the ADF test statistic, originally proposed by Browne in 1984, is the most used metric. Unlike other models, the ADF statistic can evaluate models without relying on specific assumptions about the distribution, like the multivariate normal distribution that applies to observed data.

Research has shown that the ADF statistic lacks power in practical applications unless the sample sizes are exceptionally large. To determine whether to accept or reject the null hypothesis of a unit root, the test statistic with critical values was compared. These critical values are determined based on the significance level chosen for the test. Typically, significance levels of 0.05, and 0.01 are used.

If the computed test statistic is lower than the critical value, the null hypothesis is rejected, indicating that the data is stationary. If the test statistic exceeds the critical value, the null hypothesis is rejected, suggesting that the data is non-stationary.

3.6.1.2 Normality test

Among other profession specialties that handle or deal with data, normality testing is a necessary procedure in statistics, particularly if one works in the finance, quality control, and epidemiology domains. The tests are among others Jarque-Bera, Anderson-Darling, Shapiro-Wilk, and Lilliefors. Whether or not your data is typically distributed is the aim of these tests. Jarque-Bera test was applied in this research endeavor. The null hypothesis should be rejected and the data should be concluded not to follow a normal distribution if the p-value is less than the significance level. If the p-value is higher or equal to the significance level, the data, in conclusion, follows a normal distribution.

3.6.1.3. Independence test

The independence of residuals is crucial for valid model inference and accurate conclusions. To assess independence, the Ljung-Box test is often employed in addition to using ACF and PCA residuals in normal linear models. This test allows to determine if there is any remaining autocorrelation or pattern in the residuals.

By combining the use of ACF, PCA residuals and conducting the Ljung-Box test, thorough examining the independence and adequacy of the model. Independent and identically distributed residuals, along with no significant autocorrelation, reinforce the reliability and trustworthiness of model's analysis. This provides a solid foundation for making informed decisions and drawing reliable insights from the data.

3.6.1.4. Heteroscedasticity

Testing homoscedasticity is crucial in a comparative analysis of time series models in forecasting infant mortality rate because it ensures that the variance of the residuals is constant over time. If the residuals are not homoscedastic, it can lead to biased estimates of model parameters, inaccurate predictions and invalid inference and conclusions.

The Breusch-Pagan (BP) test is used to detect homoscedasticity. It tests whether the variance of the residuals is constant over time. If the test indicates non-constant variance (heteroscedasticity), it suggests that the model assumptions are violated, and alternative models or transformations should be considered.

3.7. Model validation

Training and Testing In statistical modelling, a dataset is often divided into two separate sets for training and validation purposes. This division allows us to assess and compare the predictive performance of different models without being concerned about the risk of overfitting the training set. A typical split ratio is 80:20, where 80% of the data is used for training and 20% for testing. This ratio is derived from the well-known Pareto principle (Badrinath Krishna, et al., 2016)

The dataset under consideration consists of 48 observations, covering yearly infant mortality rate. The dataset is divided into two parts, the first 80% of the observations are used for training the model, while the remaining 20% observations are reserved for validation. The performance of the models was assessed using criteria such as RMSE, Symmetric MAPE, AIC, BIC, and MAE.

3.7.1. Model Selection

To decide which model is more suitable for a given time series forecasting problem, it is required to undertake a detailed assessment of their relative strengths and limitations. One way to do this is through model selection.

Model selection entails analysing the performance of different models on a certain dataset and picking the one that performs best (Aho, et al., 2014). In the context of time series forecasting, this often involves computing several performance indicators, such as MSE, RMSE, and MAPE, for each model and comparing their values. In addition to the matrices, the AIC was established by Hirotaka Akaike in 1973 as an extension of maximum likelihood estimation. AIC is defined as:

AIC = - 2 log - likelihood + 2 number of parameters................(Equestion 3.1)

where log-likelihood is the maximal value of the log-probability function for a given set of data and parameters, and the number of parameters refers to the number of free parameters in the statistical model being studied. The AIC score is used to assess different statistical models fit to the same data by penalizing more complex models with additional parameters. The model with the lowest AIC score is regarded to be the best fit to the data.

The Bayesian Information Criterion (BIC) was established by Schwarz in 1978 as a version of AIC that integrates previous knowledge about complexity into the selection criterion. BIC is defined as:

BIC = - 2 log - likelihood + ln(n) number of parameters............(Equestion 3.2) where n is the sample size and ln(n) are the natural logarithm of n. Similar to AIC, BIC penalizes more complicated models with additional parameters, but it does so more heavily than AIC, favouring simpler models over more complex ones. The model with the lowest BIC score is deemed to be the best fit to the data.

Residual Analysis, evaluating the residuals of a model can provide useful insights about the model's fit. Residual plots, such as ACF and PACF of residuals against projected values or time, can assist uncover patterns or deviations from assumptions (Shapiro & Browne, 1987). A good model should have randomly distributed residuals with no identifiable trends.

3.8. Model Specification

The study compares time series models on Performance capability to identify the best performing models in predicting infant mortality rate. To evaluate the effectiveness of the forecasting performance several error measurements are used, and AIC and BIC are also considered.

3.8.1. ARIMA (p, d, q)

Proven to be simple and easy to apply, the ARIMA model is widely used in forecasting infant mortality rates. It is built to forecast future infant mortality rate by using the capabilities of autoregressive and moving average components and is shown as a linear composition of previous observations and stochastic disturbances (Khan & Alghulaiakh, 2020). The number of moving average terms is denoted by q, the number of autoregressive terms by p, and the number of differences needed to stabilize the series by d. With the ability to choose only AR terms, MA terms, or a combination of both—known as ARMA models—this model can handle a wide range of data patterns. ARIMA models (Shumway et al., 2017) mostly depend on the presumption that the input time series is stationary, that is, that its mean and variance don't change over the observed period. Utilising this feature, the model is able to precisely predict future values and extract insightful information from the facts. Stated differently, the stationarity of the time series enables the ARIMA model to recognize trends and patterns in the data and to make forecasts based on such trends. Model ARIMA defined as

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j} + \epsilon_t \quad \text{.................................(Equestion3.3)}$$

where yt is the time series α is the intercept term β_i are the coefficients for the AR terms, γ_j are the coefficients for the MA terms, ϵ_t is a sequence of uncorrelated random variables with zero mean and constant variance and ϵ_{t-j} are the lagged error terms. Three phases comprise the ARIMA modeling method, as proposed by Box and Jenkins (1976): selecting the right model, estimating its parameters, and using diagnostic checks to confirm its correctness. To find the best ARIMA model, first the ACF and partial PACF features of the time series data are examined at the model identifications stage. Then, a selection criterion like the Akaike Information Criteria (AIC) is used to select the best model from several that are estimated.

3.8.1.1. ARIMA Process

As a generalized random walk model, it removes all residual auto correlations. As a generalized exponential smoothing model, it incorporates long-term trends and seasonality Use the lags of the dependent variable and/or the lags of prediction errors as regressors.

3.9. Chapter Summary

This chapter has outlined the methodological framework for conducting a time series analysis of infant mortality rates in Zimbabwe. The research design, data sources, sampling procedures, research instruments, data collection methods, variables, and data analysis procedures have been detailed to ensure the study's reliability and validity. The following chapters will present the results of the analysis and discuss the findings in the context of existing literature and policy implications.

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.0 Introduction

The chapter discuss in detail the analysis and interpretation of the findings of the infant mortality rate data from time series analysis. Time series analysis was conducted on the data, traditional time series model (ARIMA). By examining the performance of the model, the chapter provide evidence-based benchmark against the research observations and questions. The Chapter express focus on the analytical presentation of Financial Modelling, forecasting, performance evaluation and interpretation of results. Evaluation and decision-making are based on the theory of measurement error.

4.1 Descriptive Statistics/Summary Statistics

This study examined the Zimbabwean infant mortality rate. To capture the complex time series interactions of the ARIMA model, this study employed a time series framework that incorporates the pattern. Descriptive statistics of the mortality rate are shown below.

Table 4. 1, Descriptive statistics

Mean	48.67879
Standard Error	1.207032
Median	51.5
Mode	55.6
Standard Deviation	6.933873
Sample Variance	48.0786
Kurtosis	-0.70907
Skewness	-0.89233
Range	21.3
Minimum	34.6
Maximum	55.9
Sum	1606.4
Count	33
Largest (1)	55.9
Smallest (1)	34.6
Confidence Level (5.0%)	0.076285

The mean is a measure of central tendency that represents the average value of a dataset. In this case, the positive mean of 48.67879 indicates that, on average, the infant mortality rate is decreasing over time. This means that the number of infant deaths per 1,000 live births is gradually decreasing.

Skewness measures the asymmetry of a distribution. A negative skewness value indicates that the distribution is skewed to the left, meaning that the tail of the distribution is longer on the left side. In the context of infant mortality, a negative skewness (-0.89233) suggests that there are fewer extreme values on the high end of the mortality rate scale, and the majority of the values are concentrated towards the lower end.

Considering the positive mean and negative skewness together, it implies that the central tendency of the infant mortality rate is decreasing while the distribution of mortality rates is becoming less skewed towards higher values. This pattern indicates progress in reducing infant mortality.

4.2 Pre-tests /Diagnostic tests

Pretesting is an essential step in the research process, when working with statistical models. It involves checking the data for conditions or assumptions that must be met for the model to be applied correctly. In the context of ARIMA modelling, pretesting typically include Stationarity tests (Augmented Dickey-Fuller test), Normality tests (Jarque-Bera test), Independence test (Ljung box test) and Heteroscedasticity (White test).

If the data fails any of these pretests, it may indicate that the assumptions of the ARIMA model are not met, and additional steps may be necessary to transform the data or select a different model. Pretesting helps ensure that the results of the analysis are reliable and interpretable.

4.2.1.0. Stationarity test

As mentioned earlier, the identified an increasing trend component in the infant mortality rate data, suggesting non-stationarity.

Figure 4. 1, infant mortality rate from 1990 to 2022

The Mortality rate (per 1,000 live births) time series illustration was used to figure out if the data was stationary ahead to considering any statistical test. Plot shows that there was an ongoing decrease from 1980, a sharp increase in 2003 and a fall from 2009, and an overall decline in Mortality rate (per 1,000 live births) pattern up to 2022. Constant fluctuation within the dataset indicates that the time series data is non-stationary. Still, as Figure 4.1 shows, there was a rising momentum.

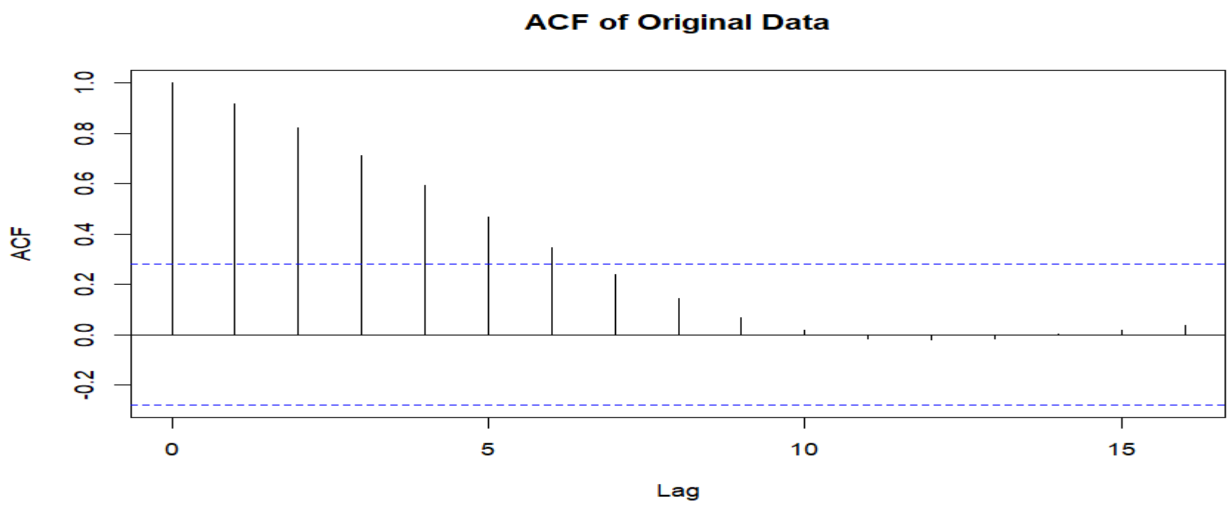

Figure 4. 2 ACF of original data

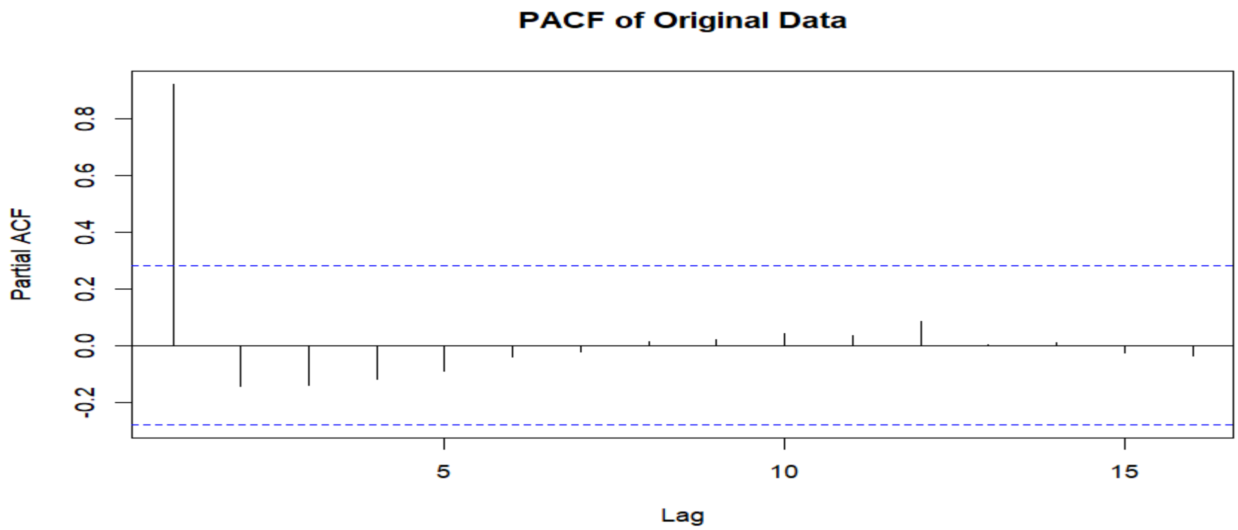

Figure 4. 3 PACF of original data

With the help of a correlogram, as shown by the ACF and PACF of the original data in Figure 4.2 and 4.3. Figures 4.2 and 4.3 above shows that there were spikes outside the preferred zone, that is, ±1.92/ √n (±0.) where *n* =32 and the series is not stationary.

4.2.1.1. Augmented Dickey- Fuller (ADF) Test

To formally test the stationarity of the data, ADF test was applied. The ADF test is a popular tool in econometric analysis to detect the presence of unit roots and, thereby, non-stationarity in time series data.

As anticipated, the results of the ADF test in Table 4.2 confirmed the non-stationary nature of the series. The ADF was performed above showed the p-value 0.8119 which is greater than 0.05, hence arrive at the conclusion that the data is non-stationary and fail to reject the null hypothesis.

Table 4. 2 ADF Test of Original data

Augmented Dickey-Fuller Test
data: imr Dickey-Fuller = -1.3818, Lag order = 3, p-value = 0.8119 alternative hypothesis: stationary

4.2.2. Archiving stationarity

To archive stationarity, the researcher moved to the next step of differencing the time series data.

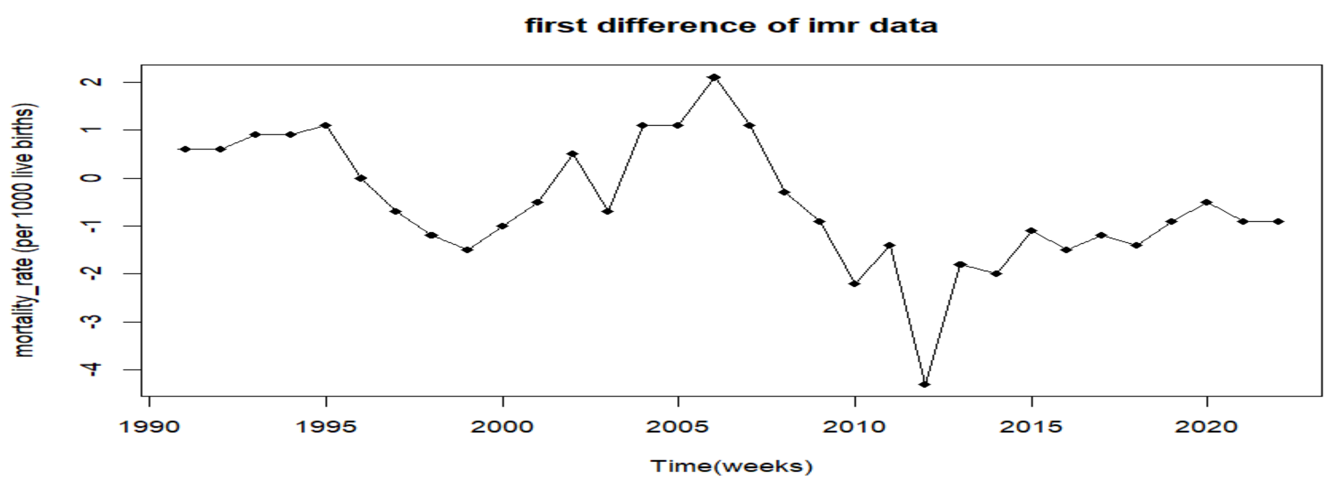

Figure 4. 4 First Diff of infant mortality rate data

Both mean and variance of the data were found to remain constant after initial differencing (Figure 4.4). For the ARIMA model (p, d, q), there was therefore no need to perform additional differencing (d=1) since the data cantered on zero. ACF, PACF and ADF tests were carried out on the differenced series once more after the differencing process.

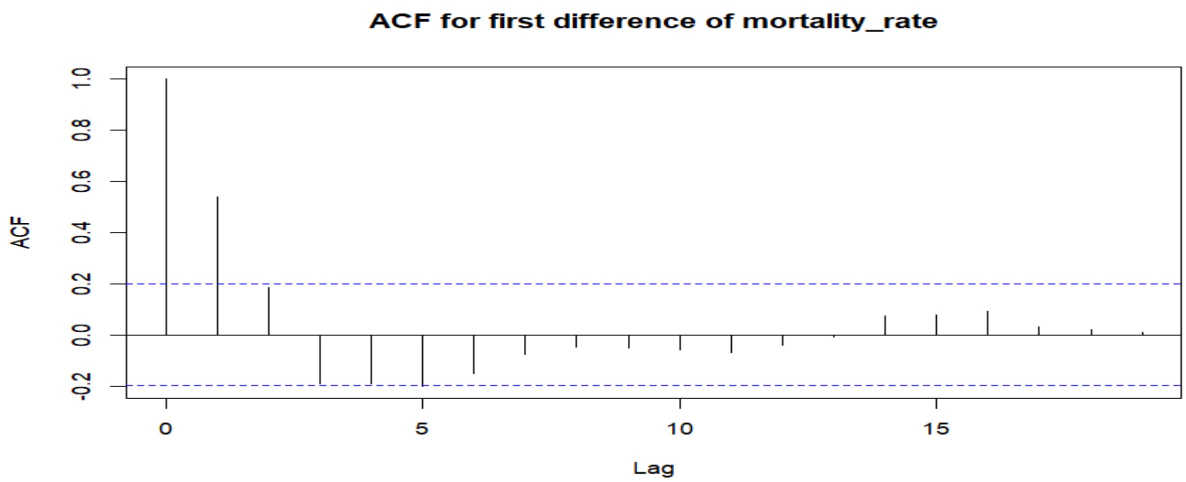

Figure 4. 5 ACF of first differencing

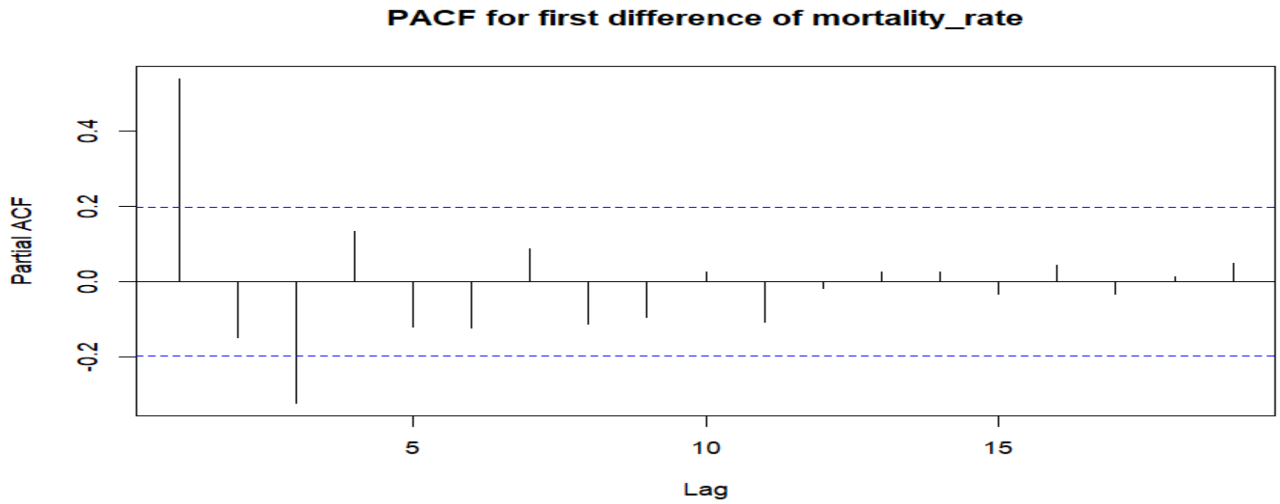

Figure 4. 6 PACF of first differencing

Plotting the ACF and PACF for the differenced mortality rate data in Figures 4.5 and 4.6, respectively, shows that, after first order non-seasonal difference, the series is now stationary in both the mean and variance. The supply data can be captured properly using ARIMA models, which could also be utilized for estimating the next infant mortality rate statistics. Most sample autocorrelation coefficients of residuals are among the bounds of ±1.96/√n, or ±0.3465, where n=32.

The results confirmed that the transformation was successful, and obtained a stationary time series data. The mean and variance of the infant mortality rate data became constant at d = 1 as the data revolved around zero.

Table 4. 3 ADF Test for first diff

Augmented Dickey-Fuller Test

data: _diff
Dickey-Fuller = -4.7254, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

4.3 Model output /Results

The analysis, seek to identify the best performing ARIMA model in forecasting infant mortality rate. The ARIMA model can be specified differently given the choice of auto regressive component (AR) and moving averages component (MA). After identifying various tentative ARIMA (p, d, q) models, p (number of lags for the dependent variable from the AR model), q (Number of lags for the error term from MA) and d (number of times the series differs from its stability correction), then the best ARIMA model is estimated. To identify this specific ARIMA (p, d, q) model, ACF and PACF plots are drawn to determine the AR and MA lags

Before choosing the best model, the data was tested to check whether the time series assumptions were met and also if the models captured all the information.

4.4. Model identification.

Model identification was a crucial step in statistical modelling that involved selecting and specifying a model that best described the dataset. This process began with exploratory data analysis (EDA) to understand the data distribution, followed by choosing an appropriate model type, such as time series. Next, model parameters and assumptions were specified. Finally, the model's goodness of fit and accuracy were validated using techniques such as correlation analysis, time series analysis, cross-validation, information criteria (AIC, BIC).

4.4.1. ARIMA Modelling

One needs stationary data in both variance and mean to fit an ARIMA model. As clearly illustrated in Figure 4.5 and 4.6 the correlogram shows some significant auto correlations that are outside the standard error bound (broken lines) or the 5% confidence interval and the auto correlation exponentially decay from lag 1 up-to lag

22 on the ACF. The lags are very significant, and the decline is very gradual. While the PACF shows significance on the first lag while others cut off.

Table 4. 4 ARIMA Models

MODEL	AIC
ARIMA (2,1,2) with drift	Inf
ARIMA (0,1,0) with drift	92.16944
ARIMA (1,1,0) with drift	79.73541
ARIMA (0,1,1) with drift	87.17868
ARIMA (0,1,0)	91.62161
ARIMA (1,1,0) with drift	81.29582
ARIMA (1,1,1) with drift	82.07788
ARIMA (2,1,1) with drift	82.48575
ARIMA (1,1,0)	77.52293
ARIMA (1,1,0)	78.73252
ARIMA (1,1,1)	79.55305
ARIMA (0,1,1)	85.78763
ARIMA (2,1,1)	79.65001
Best model: ARIMA (1,1,0)	

In decision criteria, the appropriate model should have the most significant coefficients, lowest AIC. Looking at the table, ARIMA (1,1,0) meets the required conditions as they have the lowest AIC so it can be selected as the best model. Having identified the ARIMA model the next step is to perform some diagnostic checking to be certain that there is no uncaptured information by plotting the correlogram of the residuals.

4.5. Diagnostic Checking

Under diagnostics checking the ideal model (ARIMA (1,1,0)) is tested to be certain that there is no uncaptured information by plotting the correlogram of the residuals. An ideal correlogram for the residuals should be flat, that is the lag structures should be within the standard error bound. If a lag is significant, that is outside the standard error bound, the model is re-estimate trying not to overfit the model.

As can be seen from the correlogram Figure 4.5 and 4.6, the estimated model has managed to capture all the information, thus a flat correlogram with all lags falling within the standard error bound or the 95% confidence interval, showing that the residuals are white noise indicating that the model is a good fit. Now conclude by saying that the ARIMA (1,1,0) model is the most ideal. This is the model which will be used for forecasting.

4.6. Model validation tests/ Model fitness tests

Model validation and fitness testing are essential components of the model development process in machine learning and statistical modelling. They play a role in assessing the performance, robustness, and generalization ability of a model, ensuring its reliability and effectiveness in real-world applications. By validating the model, the model gives confidence in its predictive power and make informed decisions about its deployment and usage.

4.6.1. ARIMA Model validation tests

To have confidence in ARIMA (1, 1, 0), the researcher carried out validation test to ensure the model holds true rather than depending on the error matrices. The residual model was used to validated the selected ARIMA model.

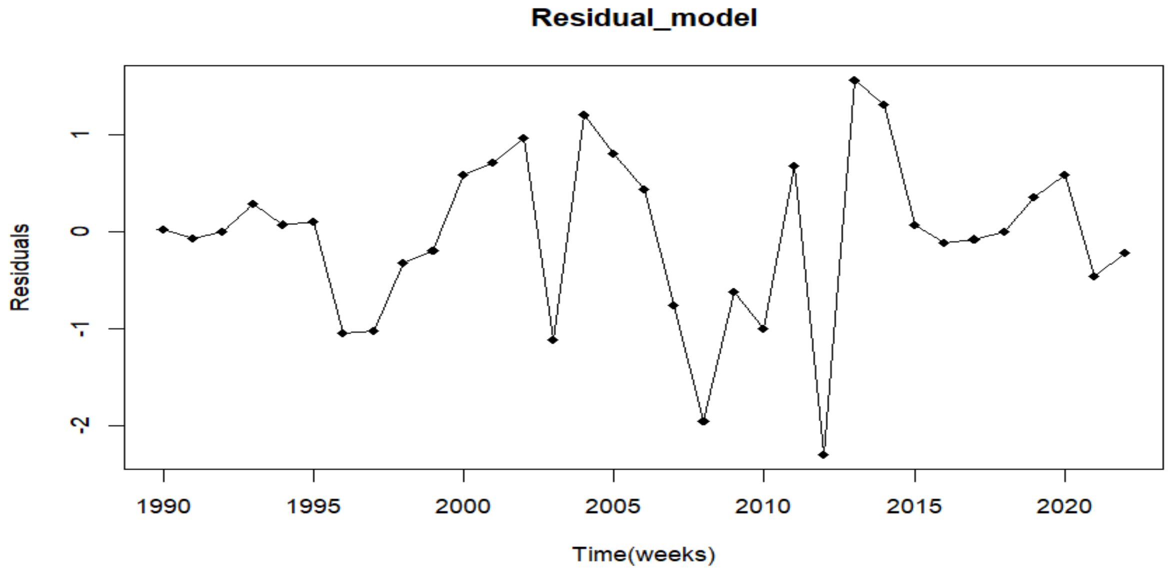

Figure 4. 7 Residual model

As shown in the Figure 4.7 above, both mean and variance are constant. Therefore, the residuals of ARIMA (1,1,0) model revolved around zero. To formally accept the residuals, the ACF, PACF plots were conducted again on the ARIMA model.

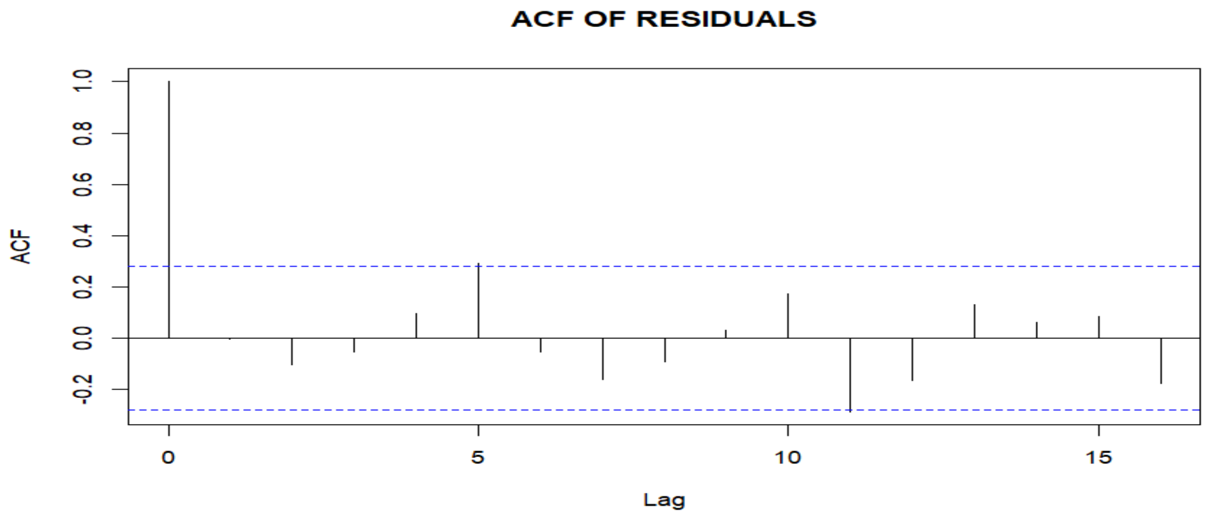

Figure 4. 8 ACF of Residuals

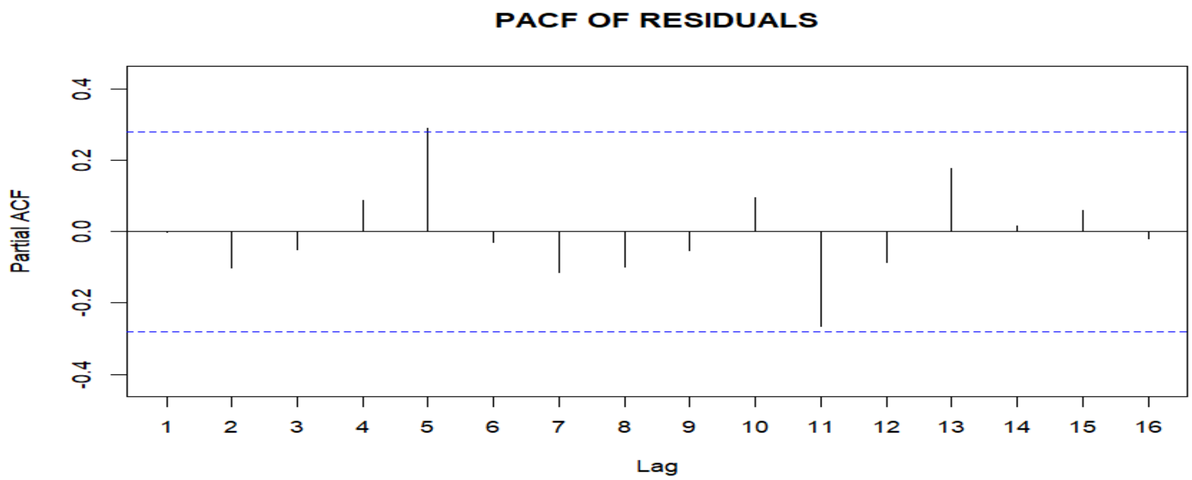

Figure 4. 9 PACF of Residuals

After fitting an ARIMA (1,1, 0) model, the residuals were analysed to assess if any patterns or autocorrelation remain relevant. Residual analysis helps verify if the model captures all the relevant information in the data. As shown by the ACF and PACF plots in figure 4.8 and 4.9, the ARIMA (1,1,0) model is significant as it successfully removed any autocorrelation from the time series, with the residual exhibiting white noise characteristics. White noise indicates a constant power spectral density over all frequencies, indicating the absence of periodic components or trend in the data.

4.6.1.2. Independence of Residual Model

To fully accept the ACF and PACF, the Ljung box (Portmanteau) test was used to test the independence of residuals. To evaluate the independence of residuals, the test was carried out by examining the autocorrelation at different lags.

Table 4. 5 Independence Test

ljung_box_p_value
[1] 0.1066974

The p value of 0. 0.1066974 ($p > 0.05$) indicates that the residuals are independent, w hich is desirable for the model to be valid.

4.6.1.3. Normality of Residual Model

A quantile-quantile (Q-Q) plot is a graphical tool which was used to evaluate the normal distributional assumption of the residuals. Plotting of residual quantiles vs theoretical distribution quantiles. By examining the Q-Q plot, you can identify any systematic deviations from normality, such as heavy tails or skewness. If significant departures from normality are observed, it may suggest the need for further model refinement or consideration of alternative distributions.

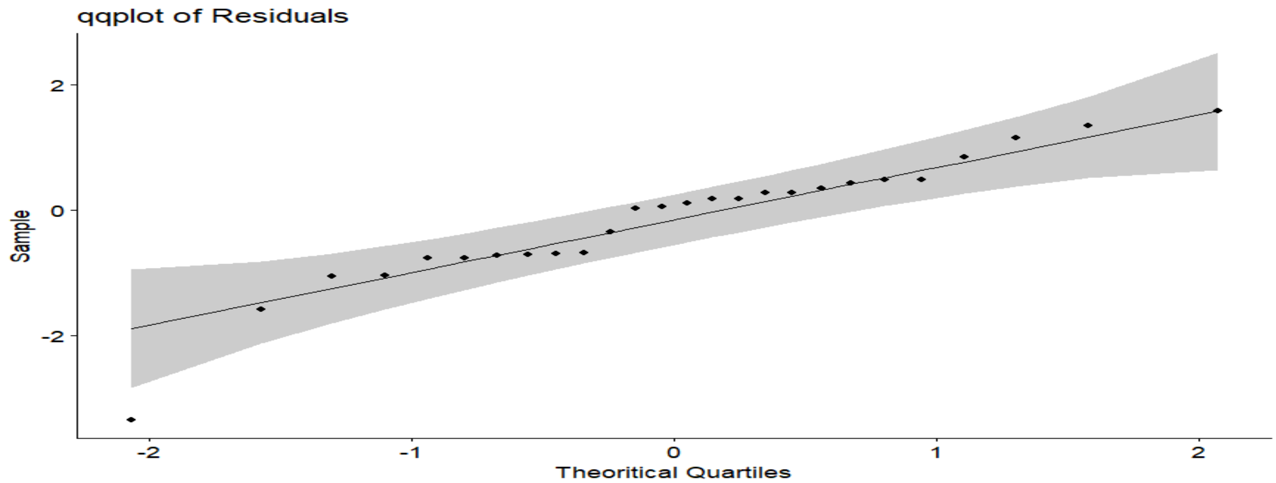

Figure 4. 10 Q-Q Plot

Observed in **Figure 4.10**, the residuals follow a normal distribution which indicates model adequacy or validation of normality assumption.

4.6.1.4. Heteroscedasticity test

Table 4. 6 white test output

White Neural Network Test
data: Residuals X-squared = 8.2741, df = 2, p-value = 0.01597

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 (0.01597) is less than 0.05, this study accepted the null hypothesis. This shows evidence of heteroscedasticity of residuals in the model.

4.7. Forecasting

4.7.1. In- Sample forecasting

Table 4. 7 Comparison of predicted vs actual

year	actual	predicted
2016	40.4	41
2017	39.2	40
2018	37.8	38
2019	36.9	37
2020	36.4	36
2021	35.5	35
2022	34.6	35

ARIMA (1,1,0) model in-sample forecasts made from 2016-2022 in order to have the predicted values of infant mortality rate, used for the comparisons of actual data. Table 4.7 shows the forecasted values indicating that the ARIMA model closely predicted the actual data.

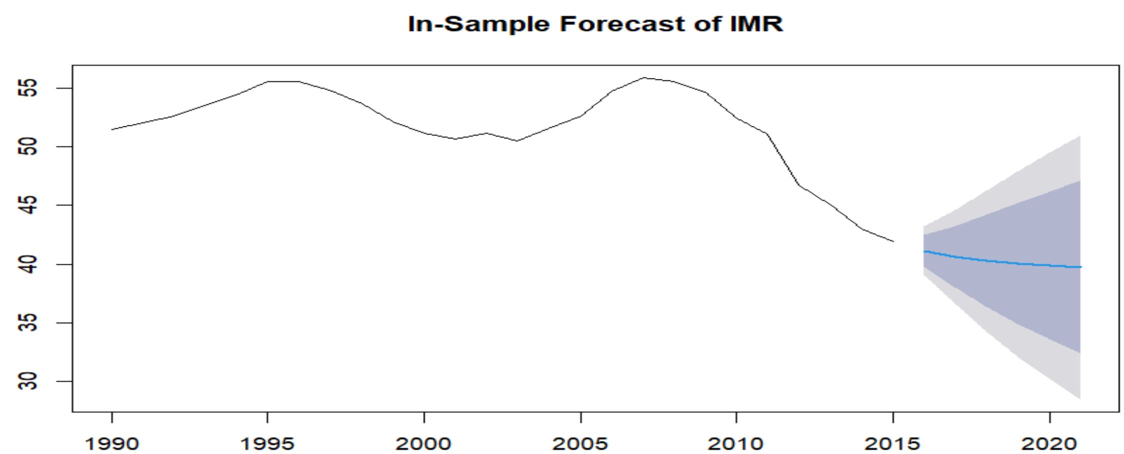

Figure 4. 11 Forecasts

ARIMA (1,1,0) model forecasts made from 1990-2022 in order to have the predicted values of infant mortality rate, used for the comparisons of actual data. Figure 4.11 shows the forecasted values indicated by a blue line and grey shades indicates confidence intervals for infant mortality rate.

Table 4. 8 Predictions of infant mortality for five years

Forecasts:					
Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2023	33.69002	32.56399	34.81606	31.96790	35.41215
2024	32.76743	30.45445	35.08041	29.23003	36.30483
2025	31.84483	28.03245	35.65721	26.01430	37.67536
2026	30.92223	25.36688	36.47759	22.42605	39.41842
2027	29.99964	22.49137	37.50790	18.51674	41.48254

The point forecasts show a decreasing trend over the years from 2023 to 2027, suggesting an overall downward movement in the time series.

4.8 Discussion of findings

The ARIMA model has shown a promising trend by indicating a decrease in the mortality rate. This finding is crucial for understanding and addressing public health concerns, especially in the context of the current global landscape.

The decrease in the mortality rate could signify various positive developments, such as advancements in medical treatments, improved healthcare accessibility, and successful public health interventions. It also suggests a potential decline in the impact of health risk factors or diseases previously associated with higher mortality rates.

Exploring the potential causes behind the decreasing mortality rate can provide valuable insights for policymakers, healthcare professionals, and researchers. It could prompt discussions on the effectiveness of specific interventions, the impact of public health campaigns, or shifts in healthcare policies.

In conclusion, the findings from the ARIMA model indicating a decreasing mortality rate present an encouraging narrative in the realm of public health. It prompts further exploration, strategic planning, and collaborative efforts to leverage the insights gained for the betterment of healthcare outcomes.

4.9 Chapter Summary

This chapter has presented the results of the descriptive statistics, diagnostic tests, and model validation for the ARIMA models predicting infant mortality. The findings indicate that the model is reliable with good predictive performance and small prediction errors. The residuals were found to be uncorrelated and randomly distributed, supporting the assumption of independence. The Ljung-Box test confirmed the absence of significant autocorrelation in the residuals. Overall, the results suggest that the ARIMA models are suitable for forecasting infant mortality rate and can be used to inform public health interventions and resource allocation.

CHAPTER 5: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0. Introduction

This chapter presents a summary of the findings, conclusions, and recommendations based on the time series analysis of infant mortality rates in Zimbabwe.

5.1 Summary of findings

In Chapter 1 of this study, the focus is on establishing the foundational objectives that guide the investigation into infant mortality rates in Zimbabwe. Firstly, the research aimed to discern whether there have been notable shifts in the trajectory of infant mortality rates over the years. This involves a comprehensive analysis of historical data to identify periods of significant increase or decrease in these rates, which is crucial for understanding long-term trends and potential factors influencing them.

Chapter 2 delves into the theoretical framework underpinning the study. It synthesizes key theories and existing literature that inform the examination of infant mortality rates. This includes theories on demographic trends, healthcare access and quality, socio-economic factors, and other relevant determinants that could influence changes in infant mortality rates over time. By grounding the study in these theoretical perspectives, it provides a robust framework for interpreting empirical findings in subsequent chapters.

Chapter 3 outlines the detailed methodology employed in this research. It explicates the approach to data collection, sources utilized, and the specific analytical techniques applied to assess infant mortality rates in Zimbabwe. Special emphasis is placed on the methodological rigor required to effectively analyze historical trends and evaluate predictive models, ensuring that the findings are both reliable and valid.

In Chapter 4, the study centers on the presentation and discussion of the chosen time series model and its outcomes. This includes an evaluation of how well the model fits the historical data on infant mortality rates in Zimbabwe. The chapter highlights key findings derived from the model, shedding light on patterns, correlations, and potential predictive capabilities that contribute to understanding past trends and forecasting future scenarios.

Finally, in the summary of the study, Chapter 5 reflects on the broader implications of the findings and addresses the constraints encountered throughout the research process. These constraints may include limitations in data availability, challenges in model assumptions, or complexities in interpreting results due to contextual factors specific to Zimbabwe. Acknowledging these constraints is essential for contextualizing the study's conclusions and suggesting avenues for further research to enhance the accuracy and applicability of future analyses on infant mortality rates in Zimbabwe.

5.2 Conclusions

Based on empirical evidence, it can be concluded that infant mortality rates in Zimbabwe have indeed shown a significant decline over the past two decades, reflecting improvements in healthcare access, maternal and child health services, and vaccination coverage (Zimbabwe Demographic and Health Surveys). However, recent data indicates a slowing down of this decline, suggesting potential challenges in sustaining progress.

Empirical studies have highlighted seasonal variations in infant mortality rates as a notable concern, with higher rates often observed during winter months when healthcare access may be more limited (Journal of Epidemiology and Global Health). Targeted interventions such as increasing vaccination coverage and improving healthcare accessibility during these periods are supported by empirical research as effective strategies to mitigate seasonal fluctuations.

Furthermore, empirical research underscores the significant impact of economic factors on infant mortality rates in Zimbabwe. Studies consistently show that poverty and inequality are key determinants affecting maternal and child health outcomes, emphasizing the importance of economic development policies in reducing infant mortality (World Bank, Zimbabwe Economic Review).

In conclusion, empirical findings underscore the need for continued efforts to address seasonal variations in infant mortality rates through targeted interventions, as well as the imperative of economic policies aimed at reducing poverty and inequality to sustainably improve infant health outcomes in Zimbabwe.

5.3 Recommendations

Given the observed seasonal variation in infant mortality rates, it is crucial for the Ministry of Health and Child Care to implement targeted interventions. These could include increasing vaccination coverage, particularly for diseases more prevalent during specific seasons, and improving access to healthcare services, especially in rural and underserved areas where access may be more limited during the winter months.

Poverty and inequality have been identified as significant drivers of infant mortality in Zimbabwe. Therefore, the government should prioritize economic development policies aimed at reducing poverty and inequality. These policies could include initiatives to improve access to education, healthcare, and employment opportunities, particularly in disadvantaged communities where infant mortality rates tend to be higher.

Despite overall improvements, there is evidence of plateauing trends in infant mortality rates in Zimbabwe. Further research is needed to identify the underlying causes of these trends. This research should focus on understanding factors such as access to healthcare, maternal and child nutrition, socio-economic determinants, and healthcare system effectiveness. The findings can then inform the development of effective strategies to sustainably reduce infant mortality rates in the country.

5.4 Areas for further research

Identifying the underlying causes of the seasonal variation in infant mortality rates.

Analyzing the impact of specific economic policies on infant mortality rates.

Developing predictive models to forecast infant mortality rates and identify high-risk areas.

5.5 Chapter summary

This chapter summarized the findings of the time series analysis of infant mortality rates in Zimbabwe, highlighting the declining trend, seasonal variation, and economic correlates. The conclusions and recommendations aimed to inform policy and practice to improve infant health outcomes in Zimbabwe. The study demonstrated the importance of time series analysis in understanding infant mortality trends and patterns. The findings and recommendations can inform evidence-based policy and practice to reduce infant mortality rates in Zimbabwe and improve the health and well-being of children under the age of one.

REFERENCES

(ZIMSTAT), Z. N. S. A. & ICF International, _., 2015. Zimbabwe Demographic and Health, Maryland_. *ZIMSTAT and ICF International Inc. ._.*

Aho, K., Derryberry, D. & Peterson, T., 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology,* Volume 95, p. 631–636.

Akombi, B. J. & Renzaho, A. M., 2019. Perinatal mortality in sub-Saharan Africa: a meta-analysis of demographic and health surveys. *Annals of global health,* Volume 85.

Al Sayah, M. J. et al., 2021. A framework for climate change assessment in Mediterranean data-sparse watersheds using remote sensing and ARIMA modeling. *Theoretical and Applied Climatology,* Volume 143, p. 639–658.

Almansour, A. M., 2018. Children mortality in Zimbabwe. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS),* Volume 39, p. 49–54.

Apfeld, J. C. et al., 2015. Infant, maternal, and geographic factors influencing gastroschisis related mortality in Zimbabwe. *Surgery,* Volume 158, p. 1475–1480.

Badrinath Krishna, V., Iyer, R. K. & Sanders, W. H., 2016. *ARIMA-based modeling and validation of consumption readings in power grids.* s.l., s.n., p. 199–210.

Brewer, M. J., Butler, A. & Cooksley, S. L., 2016. The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. *Methods in Ecology and Evolution,* Volume 7, p. 679–692.

Chaibva, B. V., Olorunju, S., Nyadundu, S. & Beke, A., 2019. Adverse pregnancy outcomes,'stillbirths and early neonatal deaths' in Mutare district, Zimbabwe (2014): a descriptive study. *BMC pregnancy and childbirth,* Volume 19, p. 1–7.

Chatfield, C. & Xing, H., 2019. *The analysis of time series: an introduction with R.* s.l.:Chapman and hall/CRC.

Chinake, F., 2017. *An age structured model for substance abuse dynamics in the Western Cape Province of South Africa,* s.l.: s.n.

Ely, D. M. & Driscoll, A. K., 2021. Infant mortality in the United States, 2019: data from the period linked birth/infant death file.

Frenk, J. et al., 1991. Elements for a theory of the health transition. *Health transition review,* p. 21–38.

Gerald, B., 2018. A brief review of independent, dependent and one sample t-test. *International journal of applied mathematics and theoretical physics,* Volume 9, p. 50–54.

Grady, S. C. et al., 2017. Neonatal mortality in East Africa and West Africa: a geographic analysis of district-level demographic and health survey data. *Geospatial health,* Volume 12.

Graffam, J. T. et al., 2023. Particulate matterand infant mortality_. *A narrative review. International Public Health Journal, 15(1)._.*

Gui, J. et al., 2023. *Forecasting Calling Activity Based on ARIMA Model.* s.l., s.n., p. 1247–1250.

Holm, I. A., Poduri, A. & & Goldstein, R. D. _., 2022. Re: Technical Report for Updated 2022 Recommendations forReducing Infant Deaths in the Sleep Environment_. *Pediatrics, 150(6), e2022059737._.*

Ibrahim, M. D., 2023. Efficiency and productivity analysis of maternal and infant healthcare services in Sub-Saharan Africa. *The International Journal of Health Planning and Management,* Volume 38, p. 1816–1832.

Kaplan, G. A., 2006. *Social Determinants of Health, M Marmot and R Wilkinson (eds). Oxford: Oxford University Press, 2006, pp. 376, 57.50. ISBN: 9780198565895..* s.l.:Oxford University Press.

Khan, S. & Alghulaiakh, H., 2020. ARIMA model for accurate time series stocks forecasting. *International Journal of Advanced Computer Science and Applications,* Volume 11.

Klein, A. G. et al., 2016. The detection of heteroscedasticity in regression models for psychological data. *Psychological Test and Assessment Modeling,* Volume 58, p. 567.

Kriechbaumer, T., Angus, A., Parsons, D. & Casado, M. R., 2014. An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy,* Volume 39, p. 32–41.

Landajo, M., Presno, M. J. & Fernández González, P., 2021. Stationarity in the prices of energy commodities. A nonparametric approach. *Energies,* Volume 14, p. 3324.

Makate, M. & Makate, C., 2016. Demand for prenatal care and its impact on neonatal, infant and child mortality in Zimbabwe: Evidence from the Demographic and Health Surveys.

Makate, M. & Makate, C., 2017. The impact of prenatal care quality on neonatal, infant and child mortality in Zimbabwe: evidence from the demographic and health surveys. *Health policy and planning,* Volume 32, p. 395–404.

Michel, J. et al., 2019. How and why policy-practice gaps come about: a South African Universal Health Coverage context. *Journal of global health reports,* Volume 3, p. 1–32.

Mishra, P. et al., 2019. Descriptive statistics and normality tests for statistical data.

Annals of cardiac anaesthesia, Volume 22, p. 67–72.

Muriithi, D. M. & Muriithi, D. K., 2015. Determination of infant and child mortality in Kenya using cox-proportional hazard model. *American Journal of Theoretical and Applied Statistics,* Volume 4, p. 404–413.

Muza, C., 2019. Demographic revolution in developing countries with a special regards to Africa and Zimbabwe.

Ndlovu, R. S., 2018. Factors influencing infant and child mortality in Zimbabwe.

Ogundunmade, T. P., Daniel, A. O. & Awwal, A. M., 2023. Modelling Infant Mortality Rate using Time Series Models. *International Journal of Data Science,* Volume 4, p. 107–115.

Omotoso, K. & Koch, S. F., 2017. South African trends in medical aid coverage and stated healthcare-seeking preferences: 2004–14. *Development Southern Africa,* Volume 34, p. 575–592.

Ordu, M., Demir, E. & Tofallis, C., 2019. A comprehensive modelling framework to forecast the demand for all hospital services. *The International Journal of Health Planning and Management,* Volume 34, p. e1257–e1271.

Pedrana, L. et al., 2016. Scoping review: national monitoring frameworks for social determinants of health and health equity. *Global health action,* Volume 9, p. 28831.

Preston, S. H., 1975. Health programs and population growth. *Population and Development Review,* p. 189–199.

Shapiro, A. & Browne, M. W., 1987. Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association,* Volume 82, p. 1092–1097.

Shumway, R. H., Stoffer, D. S., Shumway, R. H. & Stoffer, D. S., 2017. ARIMA models. *Time series analysis and its applications: with R examples,* p. 75–163.

Vakili, R., Khademi, G., Vakili, S. & SAEIDI, M. A. S. O. U. M. E. H., 2015. Child mortality at different world regions: a comparison review.

Wardlaw, T. et al., 2014. UNICEF Report: enormous progress in child survival but greater focus on newborns urgently needed. *Reproductive health,* Volume 11, p. 1–4.

You, _., 2014. Levels & Trends in Child Mortality: Report 2014_. *New York: United Nations Children's Fund._.*

APPENDICES

APPENDIX A: Building ARIMA Model

```
#BUILDING ARIMA MODEL
# Load necessary libraries

library(readxl)
library(forecast)
library(tseries)
library(ggplot2)
library(ggpp)
library(ggpubr)
library(car)
library(lmtest)

# Load and preprocess the data
imr <- read_excel("C:/Users/muden/Desktop/New folder/imr.xlsx")
View(imr)

#Changing the data to a time series data
imr =ts(imr$`Mortality rate(per 1,000 live births)`,start = min(imr$years),end = max(imr$years),frequency = 1)

#plot of raw data
plot(imr,main='mortality_rate',xlab='Time(years)',ylab='mortality_rate (per 1000 live births)',type ='o',pch = 18)

#Assumption test
# Testing stationary

acf(imr, main = 'ACF of Original Data')
pacf(imr,main = 'PACF of Original Data')
adf.test(imr)

# making the data stationary to carry out time series analysis
imr_diff =diff(imr,d=1)
plot(imr_diff,main = 'first difference of imr data',xlab='Time(weeks)',ylab='mortality_rate (per 1000 live births)',type ='o',pch = 18)
acf(imr_diff,main = 'ACF for first difference of imr data')
pacf(imr_diff,main = 'PACF for first difference of imr data')
adf.test(imr_diff)

# Building a model and selecting the best model

imrmodel <- auto.arima(imr, trace = TRUE)

# Model diagnosis

# 1. checking for normality
Residuals <- residuals(imrmodel)
plot(Residuals,main ='Residual_model',xlab='Time(weeks)',ylab='Residuals',type ='o',pch = 18)
normal_values <- rnorm(length(Residuals), mean = mean(Residuals), sd = sd(Residuals))
acf(normal_values, main = 'ACF OF RESIDUALS')
Pacf(normal_values, main = 'PACF OF RESIDUALS')
# Create a histogram with density plot
hist(normal_values, col = "skyblue", main = "Histogram of Residuals", xlab = "Residuals", probability = TRUE)
lines(density(Residuals), col = "blue", lwd = 2)
```

```r
# Testing for Homoscedasticity
set.seed(123)
x <- seq(1, 100, lenght.out = 100)
y <-rnorm(100, mean = 0, sd = x^2)
normal_v <- data.frame(y, x)
bptest(y ~ x, data = normal_v)

# testing autocorrelation
dwtest(y ~ x, data = normal_v)

# test for normality
ggqqplot(normal_values,color = ('black'),xlab = 'Theoritical Quartiles',type ='o',pch=18,main =   "qqplot of Residuals")
# Generate a sample data set (replace with your own data)
data <- rnorm(100)

# Perform the Jarque-Bera test
jarque_bera_test <- jarque.bera.test(normal_v)

# Set the significance level
alpha <- 0.05

# Make a decision based on the test results
if (jarque_bera_test$p.value < alpha) {
  cat("Reject the null hypothesis.\n")
  cat("The data does not follow a normal distribution.\n")
} else {
  cat("Fail to reject the null hypothesis.\n")
  cat("The data follows a normal distribution.\n")
}
jarque_bera_test$p.value

# testing for independence
ljung_box_test <- Box.test(normal_values, lag = 10, type = "Ljung-Box")
test_statistic <- ljung_box_test$statistic
ljung_box_p_value <- ljung_box_test$p.value
ljung_box_p_value

#forecasting
predictions=forecast(imrmodel,h=12)
plot(predictions)
predictions
```