KWAME NKRUMAH UNIVERSITY OF
SCIENCE AND TECHNOLOGY

TIME SERIES ANALYSIS OF UNDER-FIVE MORTALITY RATE IN GHANA.

By
CHRISTIANA DOKU
TEYE MAXWELL KPAKPAKU

September 4, 2023

Declaration

We hereby declare that this submission is our work for the BSc. ACTUARIAL SCIENCE degree and that, to the best of our knowledge, it does not contain any material that has already been published by another person or that has already been approved for the award of any other degree from the university, unless appropriate citation has been made in the text.

Name :Teye Maxwell Kpakpaku

Index Number: 93767

Sign.....

Date.....

Doku Christiana

Index Number : 9374119

Sign.....

Date....................................

Supervisor

Sampson Twumasi Ankrah (PhD) Sign ....

Acknowledgments

All thanks to God almighty for seeing us through this thesis work.

Great thanks to our supervisor, Dr. Twumasi Ankrah, whose guidance made this thesis work a success.

Finally, we say thank you to other lecturers, Teaching Assistants and colleagues who supported us to make this thesis work successful.

Dedication

We dedicate this undergraduate project work to the God Most High for granting us strength,wisdom and understanding to undertake this academic journey successfully and to our guardians for their prayers, love and support throughout our four years academic journey.

Abstract

Mortality according to the World Health Organization (WHO) is the absence of all traces of life at any time after birth. Under-Five mortality to death recorded under five years. There are economic and social effects on a country. A time series data comprising of yearly recordings of the under-five mortality rate of Ghana which was obtained from the UNICEF DATAWARE HOUSE and used for this study. This study seeks to identify the best fit model to find a suitable Time Series model for predicting the future under- Five Mortality in Ghana. The ARIMA models; ARIMA (3,1,2) with a drift and ARIMA (2,2,2) were chosen based on the autocorrelation and partial autocorrelation function. The best model selected was the ARIMA (3,1,2) with a drift, since it's AIC and BIC values were for lower than the other competing model.

Contents

List of Tables

Chapter 1

Introduction

1.1 Background

Mortality or Death according to the World Health Organization (WHO) is the absence of all traces of life at any time after birth. This can also be referred to as non-functionality of the whole body after birth (Akinwande et al, 2016). Under-five mortality is a good measure of a county's socio-economic status and quality of life (GSS, 2015). Under-five mortality is defined as the probability of a child dying between birth and age five expressed per 1000 live births (UNICEF, 2012; De Onis et al., 2012). Under-five mortality is divided into two: infant mortality and the death of children within the ages of 1 year to 5 years. Infant deaths are divided into two groups: those occurring at less than 28 days after birth, which is called neonatal death; and those between 28 days and one year called post-neonatal deaths.

Infant mortality rates have for a long time been used as country or regional level proxy indicator of socio- economic position. Infant mortality is regarded as an indicator for population health. Neonatal mortality rate is especially sensitive

to event during pregnancy, delivery and the neonatal period, and to the care given to mothers and their babies. Post- neonatal mortality is thought to be influenced by a greater extent of parental circumstances, including the socio-economic position and the care they provide for their infant (Kurinczuk et al 2012).

Globally 2.4 million children died in the first month of life in 2020. There are approximately 6700 newborn deaths every day, amounting to 47% of all child deaths under the age of 5 years, up from 40% in 1990(WHO, 2022). It is a crucial indicator of the health and well-being of children and is used to assess the overall child health status and the effectiveness of healthcare systems in a country.

Under-five mortality continuous to be a very great global health and economic concern. Every minute a significant number of young lives are lost due to causes such as pneumonia, diarrhoea, birth defects, malaria, malnutrition, Poverty and socio-economic factors, lack of access to quality healthcare etc. leaving many families and loved ones in a state of grief and countries with negative economic and social impart. In 2017, there were 56.5 million deaths globally; just over half of these were people who were 70 years or older; 26% were between 50 and 69 years old; 13% were between 15 and 49; only 1% were older than 5 and younger than 14; and almost 9% were children under the age of 5 (Hannah Ritchie et al, 2018). Substantial global progress has been made in reducing childhood mortality since 1990. The total number of under-five deaths worldwide has declined from 12.6 million in 1990 to 5 million in 2020. Since 1990, the global under-five mortality rate has dropped by 60%, from 93 deaths per 1000 live births in 1990 to 37 in 2020. This is equivalent to 1 in 11 children dying before reaching age 5 in 1990, compared to 1 in 27 in 2020 (WHO, 2022).

South Asia and Sub-Saharan Africa which has the highest number of Under-

five mortality has also has seen a significant decrease. South Asia recorded an average of 4.76 million under-five in 1990 and decreased to 1.35 million in 2020. Sub-Saharan Africa recorded an average of 3.84 million in 1990 decreased to 2.79 million deaths in 2020 (ourworldindata.org).

Ghana has experienced a considerable progress in the probability that an infant will survive the first five years of his or her life. In 1988, Ghana's average U5MR was 155 deaths per 1000 live births which decreased to 60 deaths per 1000 live births in 2014 (DHS). According to the UNICEF GLOBAL DATABASE Ghana recorded an average U5MR of 253 deaths per live births which has decreased to an average of 44 deaths per 1000 live births.

1.2 Problem Statement

In recent years, the awareness of under-five mortality has been massively campaigned. A lot of research findings have helped in suggesting proper healthcare methods to secure the lives of children before and after delivery. Immunization policy and interventions has also helped to reduce under-five mortality rate.

Despite the considerable progress in improving child surviving, child mortality remains a matter of urgent concern. In 2021 alone roughly 13,800 under-five deaths occurred every day globally, an intolerably high number largely preventable child deaths(UNICEF, 2022). Sub-Saharan Africa contributed about 58 percent of the Under-five mortality in the year 2021.

Ghana was ranked forty-seventh (47th) in the world in terms of under-five mortality rate in 2002, with a rate of 100 deaths per 1000 live births (down from 126 per 1000 live births in 1990) and an infant mortality rate (IMR) of 57 per 1000 live births (UNICEF, 2004). However, the rates increased slightly to 111 per 1000 live births and 64 per 1000 live births for under-five and infant mortality respectively for the year 2003 (GSS/NMIMR/ORC Macro, 2004) and in

2008, went down to 80 and 50 per 1000 live births respectively (GSS/GHS/ORC Macro, 2008 preliminary results).

1.3 OBJECTIVE OF STUDY

The objectives for this study are as follows:

1. To identify patterns of Under-five mortality rate in Ghana from 1952 to 2021
2. Develop a suitable time series forecasting model for Under-five mortality rate in Ghana from 1952 to 2021
3. To use the suggested model to forecast 10 years of Under-five mortality rate in Ghana

1.4 JUSTIFICATION

The success of this study will help determine the pattern of U5MR in Ghana. Secondly the success of this study will help forecast future U5MR which will help decision making in Ghana.
Lastly will serve as an analytical basis for future study in the fields related to child mortality and child healthcare.

1.5 ORGANISATION OF THESIS

There are five chapters in our research study. The background of study, problem description, goals, methods, justification, and organization of study are all covered in chapter one along with the study's organization. Chapter two reviews the study's relevant literature. The third chapter focuses on the study's

methodology. Analytical framework, data source, sample and sampling technique, time series analysis, and forecasting and data analysis procedures are some of the issues covered. The topics covered in chapter four are data collection, research findings, and outcomes of our findings. Chapter five provides the summary, conclusions drawn from the data, and research recommendations

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews the literature on under-five Mortality rate in general. It also examines a summary of abstracts from various literatures in relation to the model being used and the factors considered.

2.2 State of under-five mortality rate

The under-five mortality rate reflects the likelihood of a child dying before reaching the age of five. It is influenced by various factors, including access to healthcare services, nutrition, clean water and sanitation, immunization coverage, and socioeconomic conditions. High under-five mortality rates are often associated with regions or countries facing significant challenges in terms of healthcare infrastructure and resources.

A global health priority is lowering under-five mortality, and increases in child survival rates are viewed as an indication of development in general public health. Worldwide under-five mortality rates have decreased as a result of inter-

ventions like better maternity and child health care, immunization campaigns, nutrition programs, and activities to prevent and cure prevalent children illness. In the past three decades, there has been a tremendous improvement in child survival, and millions of children today have a higher chance of surviving than they did in 1990. In 2021, 1 in 26 children would die before turning five, compared to 1 in 11 in 1990. Additionally, compared to the 1990s, progress in lowering child mortality rates has increased in the 2000s, with the yearly rate of decline in the world's under-five population.

Globally, the under-five mortality rate decreased by 59% from a rate of 93 deaths per 1000 live births in 1990 to 39 deaths per 1000 live births in 2018. This is equivalent to 1 in 11 children dying before reaching age five (5) in 1990, compared to 1 in 26 in 2018. Increasing child survival remains a pressing priority despite this significant progress. Over 13,800 children under the age of five died per day in 2021 alone, a shockingly high rate of entirely avoidable infant fatalities.

Regional and income gaps in children's chances of survival are still pervasive. African region has made the lowest progress towards reducing under-five mortality rates. The World Health Organization (WHO) Africa region has the highest under-five death rate of 76 deaths per 1000, more than 8 times the rate in WHO Europe region with 9 per 1000 live births. In fact 52% of the global under-five deaths occurred in Africa in the year 2018 (WHO, 2019).

Sub-Saharan Africa continues to have the highest mortality rate for children under five in the world (73) per 1,000 live births, deaths. In sub-Saharan Africa, 1 in 14 children died before turning five in 2021, which is 20 years behind the global average (which reached a 1 in 14 rate by 2001) and 15 times higher than the risk for children born in high-income nations. The risk of dying before the age of five for a child born in the country with the highest mortality rate is

approximately 67 times greater than in the country with the lowest mortality rate, and all five of the countries with mortality rates above 100 deaths per 1,000 live births are in sub-Saharan Africa.

Sub-Saharan Africa bears the brunt of child mortality due to changing demography. Only two geographic areas accounted for 83% of all under-five deaths worldwide in 2021: sub-Saharan Africa (58%) and South Asia (26%). The share of global under-five deaths that occurred in Sub -Saharan Africa increased from 31% in 1990 to 58% in 2021 and is expected to increase even further in the coming decades due to growing child populations and a shift in population distribution towards high-mortality regions.

Ghana made significant progress towards achieving the Millennium Development Goals (Goal 4) aimed at reducing under-five mortality rate by two-thirds between 1990 and 2015, equivalent to an annual average rate of reduction of 4.3 percent. Ghana's under-five mortality rate declined from as high as 155 per 1000 in 1988 to 110 in 1998, 111 in 2003, and 80 in 2008 before reaching the current level of 60 per 1000 (WHO, 2019).

Moreover, within Ghana, disparities exist in under-five mortality rates between regions. In the more resource rich regions of southern Ghana, under-five mortality rate ranges from 75 per 1000 live births whiles in the most impoverish and deprived regions of the north such as the Upper East Region it is as high as 128 per 1000 live births. Therefore, there is the need for concerted efforts especially in resource poor settings if we hope to achieve the desired improvements in under-five mortality.

2.3 Use of ARIMA Model in Under-Five mortality Rate

Techniques used to analyze data and provide practical statistics are referred to as time series analysis. By doing so, it is possible to identify the variables that change throughout time as a result of different factors. In times series models, the natural one-way ordering of time is typically utilized such that values for a given period are expressed as having some connection to previous values rather than future values.

In order to predict the level of infant mortality rate (IMR) in China from 2005 to 2007, Yun-li et al. (2004) analyzed the infant mortality rate (IMR) in China from 1991 to 2004 to determine its current state and trends., did a trend test and calculate the IMR's average growth rate; Between 2005 and 2007, IMR levels were predicted in China's urban and rural areas using the Autoregressive Moving Average (ARIMA) Model. From 1991 to 2004, China's IMR revealed a trend of decline, with average fall rates of 3.77% (µ= 3.9964, p¡ 0.001) and 5.97% (µ = 4.7628, p¡ 0.001), respectively.

Compared to urban areas, rural areas have a significantly higher IMR. Between 2005 and 2007, the rural IMR significantly decreased, whilst the urban IMR mainly stabilized. the infant mortality rate in China is falling in both urban and rural areas, however, there is a distinct difference between the two, compared to the urban IMR, the rural IMR is still high and has more room to decline. Kuhn et al. (1993) used Poisson regression and time series analysis to examine changes in child injury incidence after the establishment of a community-based injury prevention program in Central Harlem, New York City. In Central America, the rate of severe injuries among children under the age of 17 increased between 1983 and 1991.

Opoku et al. (2019) examined monthly infant mortality data from the K.N.U.S.T. hospital for the years 2013 to 2017. The study's objective was to determine which model would best match the data and predict hospital infant mortality. The trend of monthly infant mortality was examined to forecast future monthly newborn mortality. The ARIMA models ARIMA (0,1,1), ARIMA (1,1,0), and ARIMA (1,1,1) were identified based on the autocorrelation and partial autocorrelation functions. The best model was chosen because the AIC and BIC values of the ARIMA (1,1,1) were the lowest in comparison to those of the other competing models. The model was determined to be sufficient and to meet the ARIMA assumptions. It was mentioned that the KNUST hospital's forecasted new-born mortality ranges from zero to one per five months.

2.4 Other Use Of ARIMA Models

The Journal of China Medical University (2011) published a study that examined the viability of using the time series ARIMA model to forecast the maternal mortality ratio (MMR) in the Guangxi Zhuang Autonomous Region in order to lay the theoretical foundation for ongoing MMR reduction efforts. The ARIMA model was created using the monthly MMR of Guangxi from 2002 to 2006. The model's structure was selected to meet the requirements for residual uncorrelation, and the parameters were computed using the least squares method. The goodness of fit of the model was evaluated using the Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (BIC). The autoregressive coefficients (AR1 = -0.708, AR2 = -0.537) and seasonal moving average coefficients (SMA1 = 0.511) of the model ARIMA (2,1,0) (0,1,1) could most effectively fit the time series of MMR with a statistically significant (P0.05) level of significance. AIC and BIC respectively, were 33. 814 and 39.364, and the prediction

error was white noise (P¿0.05). with a relative error of 2.08%, the 2007 MMR prediction and the actual MMR were in harmony. Future MMR may be predicted using the ARIMA model.

The Ethiopian government conducted research through the Ministry of Health to find trends and develop forecasting models for variables related to health (Abraha and Nigatu, 2009). We retrieved and looked at risk factor indicators, health care coverage, health system resources, demographic and socioeconomic indicators, and mortality and morbidity indicators. Approaches to trend analysis were used to find the patterns in these indicators. Using the ARIMA models in STATA software, the root causes of the recognisable patterns were found. Then, using the trendline formulae, future indicator value predictions were made. Only the Maternal Mortality Ratio shown a statistically significant drop during the course of the study among the mortality indicators that were looked at. It has been demonstrated that trends in the total fertility rate, physicians per 100,000 population, skilled birth attendance, and postnatal care coverage are all strongly correlated with trends in maternal mortality ratio. Aikins et al. (2016) used the ARIMA (0,1,1) model with a drift to effectively forecast long-time series data, despite the fact that it cannot guarantee the accuracy of the predicted values. Their goal was to estimate future breast cancer cases and simulate the trend of breast cancer cases recorded at KATH.

Asomaning (2013) examined the maternal mortality ratios at the Okomfo Anokye Teaching Hospital in Kumasi between the years of 2000 and 2010. The study examines the effectiveness of employing time series autoregressive integrated moving average (ARIMA) to model and predict maternal mortality ratios using the Box-Jenkins Approach. The data used in the analyses were provided by the Bio-Statistics Department of the Obstetrics and Gynecology Directorate of the facility. Since there was no clear trend in the quarterly data gathered from 2000

to 2010, maternal mortality ratios (MMRs) remained essentially stable during this time. The results showed that although the hospital's Maternal Mortality Ratio (MMR) was essentially stable, it had a very alarming average quarterly MMR of 967.7 per 100,000 live births, nearly twice as high as the country's average of 451 per 100,000 live births. It was discovered that the ARIMA (1, 0, 2) model with an AIC (581.41) could accurately forecast the hospital's quarterly maternal mortality ratios.

Chapter 3

Methodology

3.1 Introduction

This chapter focuses on the theory underlying the models to be employed and the techniques for assessing the data at hand in order to meet the study's goals. It covers the fundamental definitions and uses of time series analysis as well as the procedures needed to build the Autoregressive Integrated Moving Average (ARIMA) model. This chapter also covers modeling methodologies and software.

3.2 3.2 The Concept of Time Series

A time series analysis is a collection of values for a quantity gathered over time, frequently with equal gaps. A time series, in example, enables one to examine how various variables are affected by various circumstances throughout time. Observing changes over time in a particular asset, security, or economic variable, such as weekly sales, annual rainfall, etc., can be informative. Time series analysis refers to techniques for deriving useful statistics and other aspects of time

series data. In general, time series has four main components namely; trends, seasonality, cyclical, and irregular variations. For example, weekly exchange rates, quarterly rainfall, monthly rate of road accidents, quarterly records of maternal

3.3 Objectives Of Time Series

The Objectives of time series analysis include description, explanation, prediction and control.

3.3.1 Description

Having a time series data, the immediate step in the analysis is mostly to plot the data to get simple descriptive measure of the main property of the series such as seasonal effect and trend. Apart from trend and seasonal variations the outlier to look for in the graph of the time series is the possible presence of turning points, where for example, an upward trend has suddenly changed to a downward trend (fluctuations).

3.3.2 Explanation

When observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in another series. This may lead to a deeper understanding of the mechanism that generated that a given time series. For example, it is of interest to see how sea level is affected by temperature and pressure.

3.3.3 Prediction

Given an observed time series one may want to predict the future values of the series. This is an important task in sales forecasting and in analysis of economic and including time series. Prediction is closely related to control problems in many situations. For example, if one can predict that manufacturing process is going to move off target, then appropriate corrective action can be taken.

3.3.4 Control

When a time series is generated which means the quality of a manufacturing process, the aim of the analysis may be to control the process. Control procedures are of different kinds. In statistical quality control for instance, the observations are plotted on control charts and the controller takes action as a results of studying the charts.

3.4 Categories Of Time Series

There are several categories under which time series can be classified.

1. **Uni-variate Time Series:** This refers to a time series that consists of single (scalar) observations recorded sequentially over equal time increments.

2. **Multivariate Time Series:** A time series is said to be multivariate consists of more than one time-dependent variable and each variable depends not only on its past values but also has some dependency on other variables.

3. **Linear Time Series:** A time series is said to be linear if the current value of the time series is a linear function of the past observations.

4. **Non-Linear Time Series:** A time series is assumed to be non-linear if the current value of the time series is a non-linear function of the past observations.

5. **Discrete Time Series:** A time series is discrete once observations are taken only at specific times, normally equally spaced, even when the measured variable is a continuous variable

6. **Continuous Time Series:** A time series is continuous when observations are made continuously through time, even when the measured variable can only take a discrete set of values.

3.5 Components Of Time Series

Any time series can contain some or all the following components: Trend, cyclical, seasonal and irregularity.

3.5.1 Trend Component

The trend is the long-term increase or decrease in a time series data. The trend can be any function, such as linear or exponential, and can change direction over time.

3.5.2 Cyclic Variation

Any pattern showing an up and down movement around a given trend is identified as a cyclical pattern. The duration of a cycle depends on the type of business or industry being analyzed.

3.5.3 Seasonal Variation

Seasonality occurs when the time series exhibits regular fluctuations during the same month (or months) every year, or during the same quarter every year. For instance, retail sales peak during the month of June.

3.5.4 Irregular Variation

This component is unpredictable. Every time series has some unpredictable component that makes it a random variable. Considering the effects on these four components, two different types of models are generally used for a time series. That is;

- Additive model with an assumption that the four components are independent of each other.
  Y (t) = T(t) + S(t) + C(t) + I(t)
- Multiplicative model with an assumption that the four components of a time series are not necessarily independent, and they can affect one other.
  Y (t) = T(t) * S(t) *C(t) * I(t)

3.6 Examples Of Time Series Model

3.6.1 White Noise

A simple time series could be a collection of uncorrelated random variables w_t, with zero mean and finite variance w2 denoted as
$w_t = w_n(0,\sigma_w^2)$

3.6.2 Gaussian White Noise

It is a particular white noise where the wt are independent normal random variables denoted as

$w_t = iidN(0,\sigma_w^2)$

3.6.3 Random Walk

A random walk is the process by which randomly moving objects wander away from where they started. Generally random walk is not stationary process because it depends on time.

3.7 Ubiquity Of Time Series

Time series data can be seen in several fields of study. Some of these are; • In meteorology, we monitor daily high temperatures, annual rainfall, hourly wind speeds, earthquake frequency amongst others.

- In business, we observe daily stock prices, weekly interest rates, quarterly sales, monthly supply figures, annual earnings.
- In epidemiology, we note the number of flu cases per day, the number of health-care clinic visits per week, annual tuberculosis counts
- In natural sciences, we observe chemical yields, turbulence in ocean waves, earth tectonic plate positions.

3.8 Stationary Time Series

A stationary time series has a constant mean, variance, and autocorrelation through time. The stochastic properties of the stationary process are assumed to be invariant with respect to time.

3.8.1 Weakly Stationary Series

A time is said to be weakly stationary if its mean in constant and its auto covariance function depends only on the lag. Y_t is weakly stationary if:

- $\mu_y t$ is independent of t.
- $y_y(t+h,t)$is independent of t for each h.

It must be noted that however that no assumptions are made about higher order moment. A given series $Y(t_1),(t_2),...,(t_n)$ follows a multivariate normal distribution; since the multivariate normal distribution is completely characterized by the first moments (population mean) and the second moments (population variance). It must again be noted here that the two concepts of strict stationary and weak stationarity are equivalent.

3.9 Non-Stationary Time Series

Most of the time series we encounter are non-stationary. Any series the is not stationary is said to be non-stationary and hence must pass through due process of the Box- Jenkins approach to make it stationary. A simple non-stationary time series model is given by:
$Y_t = \mu_t + e_t$ Where the mean $_t$ is a function of time and et is a weak stationary series. A ramdon noise process y_t defined as
$Y_t = y_{t-1} + \varepsilon_t$

3.10 Box-Jenkins Method Of ARIMA Modelling

The Box-Jenkins methodology is a statistical sophisticated way of analyzing and building a forecasting model which best represents a time series. ARIMA models are basically a class of models that have potential to represent stationary

together with non-stationary time series and to generate precise forecasts based on description of historical data of single variable. Seeing as it does not assume any particular pattern in the observed historical data of the time series that is to be forecast, this model is very dissimilar to other models used for forecasting. Aikins et al. (2016). When fitting an ARIMA model to a set of (non-seasonal) time series data, the BoxJenkins procedure provides a useful general approach.

1. Convert the data to a time series model data. '
2. Plot the data and identify any unusual observations such as trend, seasonality etc.
3. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance
4. If the data is non-stationary, difference the time series data until it is stationary.
5. Examine the ACF and PACF: Is an ARIMA (p, d,0) or ARIMA (0, d, q) model appropriate? OR
6. 6. Try your chosen model(s) and use the AIC, BIC or AICc to find the appropriate model.
7. Try your chosen model(s) and use the AIC, BIC or AICc to find the appropriate model.
8. 7. Check the residuals from your chosen model by plotting the ACF of the residuals and doing a diagnostic test of the residuals. If they do not look like white noise, try a modified model. It can be done using the Ljung-Box statistical test and a correlogram.
9. Once the residuals look like white noise, calculate forecasts. There is an automated algorithm in R software, which only takes care of steps 4 to

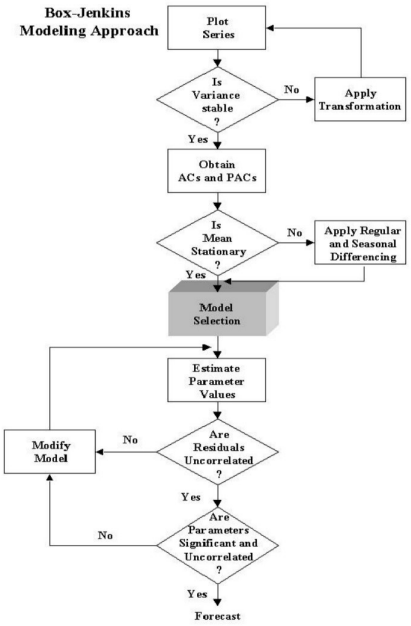

Figure 3.1: General Process for Forecasting Using ARIMA Model

8, the process is summarized in Figure 3.1 Below is the chart of the Box-Jenkins ARIMA modelling approach (Unbehauen and G¨ohring, 1974)

3.10.1 Unit Root Tests

One way to determine whether a time series data is stationary or not is to use the unit root test. The unit root test is performed to determine whether a stochastic or a deterministic trend is present in the series. Several statistical tests are available with different assumptions for testing for the presence of unit root in a series. They include the Augmented Dickey- Fuller (ADF) test, Kwiatkowski-Phillips-Schmidt Shin (KPSS) test and Phillips-Perron (PP) test (Dickey & Fuller, 1979; Kwiatkowski et al., 1992).

- ADF test $x_t = c + \beta t + \alpha x_{t-1} + \emptyset_1 \nabla f X_{t-1} + \ldots + \emptyset p \nabla f X_{t-p} + w_t$

The Unit Root Test as proposed by Dickey and Fuller (ADF), test the Hypotheses below:

H_0 = The time Series data is non-stationary.

H_1 = The time series data is stationary.

If the p-value of the ADF test is less than the significance level (0.05), we reject the null Hypotheses that the series is not stationary.

- Similarly, the unit root test proposed by Kwiatkowski-Phillips-Schmidt-Shin (KPSS), test the Hypotheses below:

  H_0 = The time series data is stationary.

  H_1 = The time Series data is non-stationary.

  If the p-value of the KPSS test is greater than the significance level (0.05), we fail to reject the null Hypotheses that series is stationary.

- The Unit Root Test as proposed below.

  Phillips-Perron (PP), test the Hypotheses below:

  H_0 = The time Series data is non-stationary.

  H_1 = The time series data is stationary.

  If the p-value of the PP test is less than the significance level (0.05), we reject the null Hypotheses that the series is not stationary.

3.11 Autocorrelation Function (Acf)

The Autocorrelation function is extremely useful for describing the general process used to develop a forecasting model. It measures the degree of correlation between neighbouring observations in a time series. The autocorrelation at any lag h (the difference in time between an observation and a previous observation. Thus, X_{t+h} lags X_t by h periods) is defined as Corr (X_t, X_{t+h}).

3.12 Partial Autocorrelation Function

A partial autocorrelation coefficient measures the degree of association between an observation X_t and X_{t+h} when the effects of the other time lags are held constant. We consider partial autocorrelation when we are unaware of the appropriate order of the autoregressive process to fit the time series. PACF is denoted by ϕ_{hh} and defined by

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t$$

3.13 ARIMA(p,d,q) Models

ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. Specifically,

- **AR (Autoregression)**: A model that uses the dependent relationship between an observation and some number of lagged observations.
- **I (Integrated)**: The use of differencing of raw observations to make the time series stationary.
- **MA (Moving Average)**: A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. Each of these components are explicitly specified in the model as a parameter.

Note that AR and MA are two widely used linear models that work on stationary time series, and is a pre-processing procedure to "stationarize" the time series if needed.

Notations

A standard notation is used of ARIMA (p, d, q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

- p - the number of auto regressive lag observations included in the model, also called the lag order.
- d -the number of times that the raw observations are differenced, also called the degree of.
- q - the size of the moving average window, also called the order of moving average. A value of 0 can be used for a parameter, which indicates to not use that element of the model.

In other words, ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

3.13.1 Autoregressive model (AR(p))

Autoregressive models are based on the idea that the current values of the series, X_t can be explained as a linear combination of p past values $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$,together with a random error in the same series. An autoregressive model of order p, abbreviated AR(p), is of the form $X_t = c+\phi_1 X_{t-1}+\phi_2 X_{t-2}+ \ldots + \phi_p X_{t-p} + \varepsilon_t$. Where X_t is stationary on $w_t wn(0, \sigma_w^2)$ and $\phi_1, \phi_2, ..., \phi_p$ are model parameters. The hyper parameter p represents the direct look back in the series. Always, wt is independent of X_{t-1} , $X_{t-2}, ..., X_{t-p}$. An AR(p) process has it's PACF cutting of after lag p and the auto-correlation function (ACF) decays exponentially.

3.13.2 Moving Average MA(q)

A model that uses the dependency between an observation and a residual error from a moving average model applied to lag observation. Moving average of order q is given by $Xt = w_t + \theta_1 wt - 1 + \theta_2 w_{t-2} + + \theta_q w_{t-q}, w_t \sim w_n(0,\sigma_w^2)$. An MA(q) process has it's PACF decays exponentially and the ACF cut of after lag q.

3.13.3 Autoregressive Moving Average Models (ARMA (p,d))

Autoregressive and Moving Average models can be combined to form ARMA models. ARMA model of order p and q is given by

$Xt = X_t = c+\phi_1 X_{t-1}+\phi_2 X_{t-2}+\ldots+\phi_p X_{t-p}+\theta_1 \varepsilon_{t-1}+\theta_2 \varepsilon_{t-2}+\ldots+\theta_q \varepsilon_{t-q}+\varepsilon_t$

For ARMA models both the PACF and the ACF decays exponentially

3.13.4 Autoregressive Integrated Moving Average

The acronym ARIMA stands for "Auto-Regressive Integrated Moving Average." Lags of the differenced series appearing in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. A non-seasonal ARIMA model is as an "ARIMA (p, d, q)" model, where p is the number of autoregressive terms, d is the number of non-seasonal differences, and q is the number of lagged forecast errors (moving average) in the prediction equation. A process, Xt is said to be ARIMA (p, d, q) if $\nabla f_d X_t = (1 - \beta)\mathrm{X}_t$,

where d represents the d_{th} differencing and B denotes the backward-shift operator.

3.13.5 Backward-shift operator (B)

Backward-shift operator B is definned as

$$BX_t = X_{t-1}$$

and can be extended to powers such as $B^2X_t = X_{t2}$ Thus, $B^hX_t = X_{th}$ Now applying backward shift operator on AR (p) , MA (q) and ARMA(p, q)

$$B(AR(p)) : X_t = B(\phi_1 X_{t1} + \phi_2 X_{t2} + \ + \phi_p X_{tp} + w_t)$$

$$= \phi_1 BX_t + \phi_2 B^2 X_t + \ + \phi_p B^p X_t + w_t$$

$$\text{B (MA (q)) :}$$

$$X_t = B(w_t \theta_1 w_{t12} w_{t2q} w_{tq})$$

$$= w_{t1} B w_{t2} B^2 w_{tq} B^q w_t$$

$$\text{B (ARMA (p, q)) : } Xt = B(\phi_1 X_{t1} + \ + \phi_p X_{tp} + w_t \theta_1 w_{t1} \theta_q w_{tq})$$

$$= \phi_1 BX_t + \ + \phi_p B^p X_t + w_t \theta_1 B w_t ... -_q B^q w_t$$

3.13.6 Differencing

One limitation of ARMA models is stationary condition. In many situations, time series can be thought of as being composed of two components, a non-stationary trend series and a zero-mean stationary. Hence to remove the non-stationary trend we apply the following methods;

1. De-trending
2. Moving Average Filter
3. Using Regression Techniques
4. Differencing

Differencing is the easiest way to remove trends and we lose one observation each time we use differencing. One advantage of differencing over de-trending

for trendremoval is that no parameter estimation is required, and differencing operation can be repeated.

- First Difference eliminates a linear trend, and it denotes as:
  $\nabla_X t = (1 - B)X_t$
- Second differencing eliminates a quadratic trend, and it denotes as:
  $\nabla^2 X_t = (1 - B)^2 X_t$
- Differencing for order d is defned as:
  ∇^d =(1-B)^d
  where $(1 - B)^d$ can be expanded algebraically for higher values of d.

3.13.7 Characteristic polynomial

We define MA(q) and AR(p) time series as X_t =ϕ(B)w_t and θ(B) X_t = w_t respectively. The two functions θ(x) and ϕ(x) are known as the characteristic polynomial. If the absolute value of all roots including real and complex roots of the characteristic polynomial are greater than one, then the AR(p) and MA (q) is said to be stationary and invertible respectively. In general, ARIMA model can be written as:
$\phi(B)(1 - B)^d X_t$= $\theta(B)w_t$

3.14 Model Identification and Selection

Building a model consist of the following steps: Model Identification, Model Estimation, Model Checking (Goodness of fit).

3.14.1 Model Identification

The order p and q of an AR and MA time series are unknown and therefore must be specified. We can use the graph of the sample autocorrelation function

and the sample partial autocorrelation function to determine the ARIMA model hyperparameters p and q, which processes can be summarized as follows:

Table 3.1: Model selection

MODELS	ACF	PACF
AR(p)	Tails off	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off
ARMA (p, q)	Tails off	Tails off

3.14.2 Model Estimation

- **Method of Moment estimators**

  The idea behind this is that we equate population moments to sample moments and then solve for the parameters in terms of the sample moments. If the $E(X_t) = \mu$, then method of moment estimator of μ is the sample average X.

- **AIC, BIC, AICc**

  The final model can be selected usingInformation Criterion (AIC) or Bayesian Information Criterion (BIC) or Akaike's information corrected criterion (AICc). When an appropriate differencing is obtained, suggesting models with different orders of both AR and MA should be fitted. The order with the minimum information criteria should be selected as the suitable model. Given a data set, several competing models may be ranked according to their AIC or BIC with the one having the lowest information criterion value being the best.

  - Akaike's information criterion (AIC) (Akaike, 1974) The AIC uses the maximum likelihood method. It can be calculated as AIC = -2loglikehood + 2k Where k = the number of parameters in the model

– Bayesian information criterion (BIC) Schwarz (1978) Like the AIC, the BIC also uses maximum likelihood. And it is given by BIC = -2loglikehood + klog(n) Where k = the number of parameters in the model. n = the sample size.

3.15 MODEL DIAGNOSTICS

After a suitable model is selected, the model is diagnosed to check if it violates any of the assumptions that the residuals from an ARIMA model must have the normal distribution and should be independent.

- Examining the ACF plot of the residuals The ACF plot of the residuals is visually examine and none of the lag spikes should be significant, which means the spikes should all be within the threshold limit. This suggests that the residuals are independent and uncorrelated, which is a good indication that the residuals follow a white noise process (a series with uncorrelated errors)

- **Ljung-Box Test** The diagnostic test provided by the Ljung-Box Q statistic is;

  Where rˆi (e) = the residual autocorrelation at lag i n = the number of residuals k = the number of times lags includes in the test **Hypotheses** H_0: The ARIMA model is appropriate H_1: The ARIMA model is not appropriate Conclusion If the p-value is greater than the significance level of 0.05; we fail to reject the null Hypotheses. Thus, the ARIMA model is appropriate, and the residuals are characterized by white noise process.

- **Normality Test** The normality test determines if the residuals are normally distributed or not. Histogram of residuals data can be used to assess the normality assumption visually. It should be bell-shaped and

look like the normal distribution. We also perform the Shapiro-Wilk test for normality.

Hypothesis H_0: The residuals are normally distributed. H_1: The residuals are not normally distributed.

If the p-value is greater than significant value (0.05), we fail to reject the null hypothesis and conclude that the residuals are normally distributed.

3.15.1 Forecasting

One of the primary objectives of building a model for a time series is to be able to forecast the values for that series at future times. Forecasting is an art of estimating or predicting a trend or future event. Forecasting problems are often classified as short-term, medium-term, and long-term.

1. Short-term forecasting problems involve predicting events only a few time periods (days, weeks, months) up to one year into the future, but it is generally less than 3 months.

2. Medium-term forecasts extend from one to three years into the future. It is useful in sales planning, production planning and budgeting, cash budgeting and analysing various operating plans.

3. Long-term forecasting problems can extend beyond that by many years. They are used in planning for new products, capital expenditures, facility location or expansion and research and development.

The logic behind time series methods is that past data incorporate enduring patterns that will carry forward into the future and that can be uncovered through quantitative analysis. Thus, the forecasting task becomes, in essence, a careful analysis of the past plus an assumption that the same patterns and relationships will hold in the future. Once the appropriate model has been selected

and checked, the residuals are seen to be uncorrelated and the parameters are significant, then we can proceed to forecast.

3.16 Evaluating The Error In Forecasting

- **Bias (Direction)** A forecast is biased if it errs more in one direction than in the other. The method tends to under-forecasts or over-forecasts.
- **Accuracy (Distance)** Forecast accuracy refers to the distance of the forecast from the actual observation ignore the direction of that error.

Forecast Error Measures Error in forecasting is given as; $e_t = Y_t - F_t$ where Y_t=Actual value, F_t=Estimated value

Accuracy measure

- **Mean square Error (MSE)** Average all the error. Measures variance of forecast error.

  MSE $= 1\frac{}{n\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
- **Mean Absolute Error (MAE)** Measures average absolute deviation of forecast from actuals.

  MAE $= 1\frac{}{n\sum_{i=1}^{n}|y_i - \hat{y}_i|}$
- **Mean Absolute Percentage Error (MAPE)** Measures absolute error as a percentage of the forecasts

  MAPE $= 1\frac{}{n\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%}$
- **Root mean square Error (RMSE)** It is the square root of the differences between predicted values and actual values.

RMSE $= \sqrt{\frac{\sum_{i=1}^{n}(actual_i - predicted_i)^2}{n}}$

MAPE, MAE, MSE, RMSE are all accuracy measure. A lower MAPE, MAE,

MSE or RMSE implies a good forecast and we can calculate the error measure for the forecasting methods.

3.17 Statistical Software used

The R software would be used in analysing and fitting the stochastic models. Other statistical and numerical simulation methods of parameter estimation were used as and when necessary.

Chapter 4

RESULTS AND ANALYSIS

4.1 Introduction

This chapter deals with the analysis under-five mortality data obtained from the UNICEF data from the year 1952 to 2021. The Box-Jenkins method is used with the aid of R software to fit and forecast with an ARIMA model, applying change point estimation in this part.

4.2 Description of Data

The total number of datapoints used in this analysis was 72. The median U5MR was 152.99 per 1000 livebirths, while the mean was 144.53. The highest U5MR ever reported during this time period was 252.99 per 1000 livebirths in 1950, while the lowest U5MR ever reported during this time was 43.97 in 2021.
The computed standard deviation of 62.01 indicates a substantial departure

.

Table 4.1: Description of under-five mortality rate data in Ghana .

Statistic	Value
Minimum	43.97
Maximum	252.99
1st Quartile	88.20
Median	152.53
3rd Quartile	202.58
Mean	144.82
Variance	3844.839
Standard Deviation	62.00677
skewness	-0.09327724
Kurtosis	-1.366995

from the mean. 3844.84 was the variance, which is another indicator of dispersion. The skewness rating of -0.09 indicates a minor leftward bias in the data. There aren't many outliers in the data.

4.3 Objective 1: Examine the trend for under-five mortality rate in Ghana.

The Figure 4.1, shows a time series plot that depicts a general downward trend in Ghana's under-five mortality rate from 1950 to 2022.

From figure 4.1, we notice a decrease in the under-five mortality rates over the years. The mean of the under-five mortality rate is not constant as it decreases with time. Under-five mortality rate was high in 1950 and there was a decrease in the rates from 1951 to 1995, where there was a noticeable decrease in 1995 before there was a slightly fall in 2000 and there has been a consistent downward trend from 2000 to 2021. Since the mean of the under-five mortality rate is decreasing with time, the trend of the under-five mortality rate in Ghana is not stationary.

To confirm the trend component present we use the Mann-Kendall, cox and Stuart and the Theil's Sen test.

4.3.1 Graphical Approach

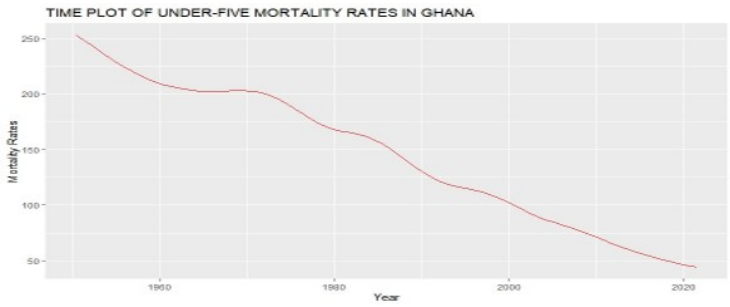

Figure 4.1: Time series plot of the Under-five mortality rate in Ghana from 1950 to 2022

4.3.2 Hypothesis trend tests

The Mann-Kendall, cox and Stuart and Theil's Sen have the hypotheses as follows: H_0: There's no trend in the under-five mortality rate data.

H_1: There's a trend in the under-five mortality rate data.

From table 4.2 below, the p-values for the three tests are less than the significance value (0.05). Hence, we reject H_0 and conclude that there is a trend component in the under-five mortality rate data. Since there is trend in the data it means the under-five mortality rate is nonstationary. Therefore, we perform the unit root tests (ADF, PP and KPSS) to find out if the series is stationary or not.

.

Table 4.2: Trend Tests for mortality rate data in Ghana .

Test	P-value	Conclusion
Mann-Kendall	<0.00001	There's a trend.
Cox and Stuart	<0.00001	There's a trend.
Theil-Sen	<0.00001	There's a trend.

4.3.3 Test For stationarity

The Kwiatkowski Philips-Schmidt-Shin (KPSS) test

H_0: The under-five mortality rate series is stationary.

H_1: The under-five mortality rate series is non-stationary.

Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) have the hypotheses as;

H_0: The under-five mortality rate series is not stationary.

H_1: The under-five mortality rate series is stationary

From the table 4.3, ADF and PP have p-values greater than the significance value (0.05), so we fail to reject the null hypothesis and conclude that Under-five mortality rate series is not stationary. KPSS test also have a p-value less than the significance value (0.05), hence we reject the null hypothesis and conclude that under-five mortality rate series is not stationary.

All the three tests confirmed that the under-five mortality rate series is not stationary, hence we difference the series to achieve stationarity.

.

Table 4.3: Output of the unit root tests for under-five mortality rates in Ghana

.

TEST TYPE	TEST STATISTIC	LAG PARAMETER	P-VALUE	CONCLUSION
KPSS	1.61	3	0.01	Not stationary
ADF	-1.5542	3	0.7551	Not stationary
PP	-5.31	3	0.8036	Not stationary

4.4 Objective 2: Identify a suitable time series model for Under-five mortality rate in Ghana.

4.4.1 First Differencing

Since stationarity was achieved at the first difference for KPSS test and stationarity also achieved at the second difference for ADF and PP tests, two models were generated from the results of their minimum information criteria (AIC, BIC and AICc). The best model was selected based on the error metrics, thus the model with the least error values.

Figure 4.2, is the plot of the under-five mortality series after the first differencing was done. There is a slight trend in the series as seen in figure 4.2. This can be confirmed by the tests for stationarity.

From table 4.4, KPSS test has p-value (0.1) which is greater than the alpha

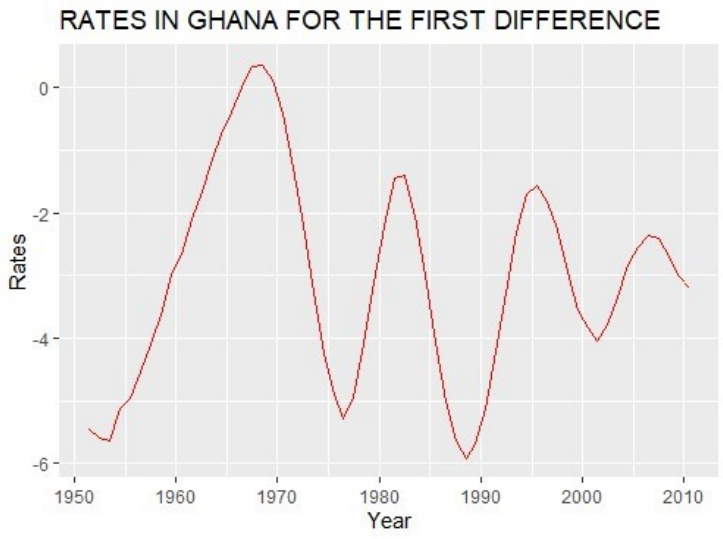

Figure 4.2: Plot of the Under-five mortality rate in Ghana for the first difference.

value (0.05) hence we fail to reject H0 and conclude that, the under-five mortality series is stationary.

According to PP and ADF Test, since their p-values (0.7453) and (0.1748) respectively are greater than the alpha value (0.05), we conclude that the under-five mortality series is non-stationary.

From ADF and PP test, we difference the series the second time to obtain stationarity

.

Table 4.4: Output of the unit root tests for stationarity after the first difference

TEST TYPE	TEST STATISTIC	LAG PARAMETER	P-VALUE	CONCLUSION
KPSS	1.8967	3	0.1	Stationary
ADF	-2.9836	4	0.1748	Notstationary
PP	-6.3091	3	0.7453	Not stationary

4.4.2 Second Differencing

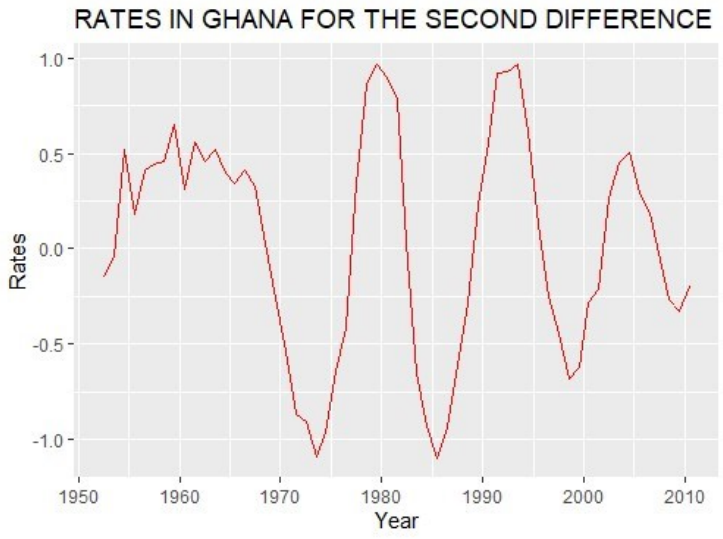

Figure 4.3: Plot of under-five mortality rate times series after second difference.

Plot of under-five mortality rate times series after second difference.

.

Table 4.5: Output of ADF and PP Stationarity Tests after the second difference

TEST TYPE	P-VALUE	CONCLUSION
Augmented Dickey-Fuller	0.01	Stationary
Phillips-Perron	0.035	Stationary

4.4.3 Model Selection

The ACF plot tails off (decays exponentially) which indicates the presence of an Autoregressive (AR) process, looking at the PACF plot tells us the lag of the Autoregressive (AR) process, thus; it cuts off after lag 4. Also, the PACF plot decays exponentially which indicates the presence of a moving Average (MA) process and looking at the ACF plot tells us the its lag. Thus, it cuts off after lag 3. (See figure 4.4)

Therefore, the suggested models from ACF and PACF plot is ARIMA (4,1,3).

The ACF plot tails off (decays exponentially) which indicates the presence of an Autoregressive (AR) process, looking at the PACF plot tells us the lag of the Autoregressive (AR) process, thus; it cuts off after lag 4. Also, the PACF plot decays exponentially which indicates the presence of a moving Average (MA) process and looking at the ACF plot tells us the its lag. Thus, it cuts off after lag 3. (Refer to figure 4.5)

Therefore, the suggested models from ACF and PACF plot is ARIMA (4,2,3).

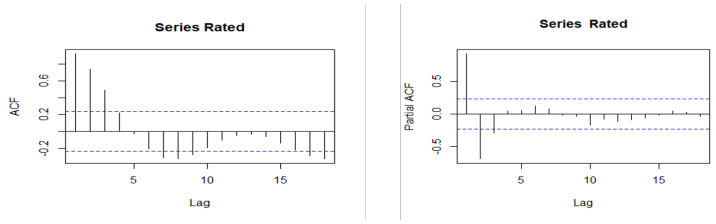

Figure 4.4: ACF and PACF of the first difference Stationary series

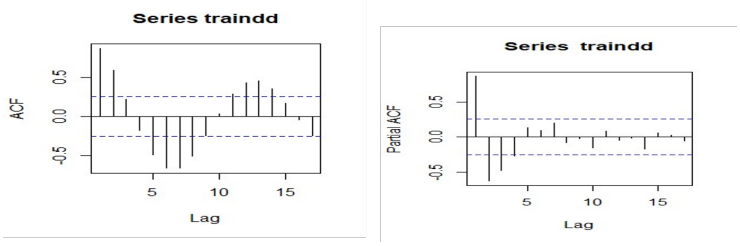

Figure 4.5: ACF and PACF of the second difference Stationary series

4.4.4 USING INFORMATION CRITERION (IC)

The minimum information criteria in statistical model building refers to the standard practice of fitting several candidate models to dataset before selecting the "best" model. When modelling with the Arima model, the ADF and PP tests suggested a differenced order of "2", whiles the KPSS test suggested a differenced order of "1" based on the information about stationarity. As a result, two sets of competing models are constructed using the differencing order and the information criteria for each model are calculated. The best model is selected based on the minimum value of the AIC, AICc and BIC.

Conclusion

According to the law of parsimony, ARIMA (3,1,2) has the least information criterion

Hence the suitable model for the series with a difference order of "1" is ARIMA (3,1,2) with drift. (See table 4.6)

.

Table 4.6: R output for the competing models and their respective AIC, AICC and BIC values at first difference.

MODEL	AIC	AICC	BIC
ARIMA (2,1,0) with drift	-13.38475	-12.77869	-4.334029
ARIMA (2,1,1) with drift	-15.38653	-14.46345	-4.073128
ARIMA (2,1,2) with drift	-26.91691	-25.60441	-13.34083
ARIMA (2,1,3) with drift	-32.37298	-30.5952	- 16.53422
ARIMA (3,1,0) with drift	-19.14346	-18.22038	-7.830061
ARIMA (3,1,2) with drift	-38.6069	-36.82912	-22.76814
ARIMA (1,1,0) with drift	120.9531	120.3721	126.802
ARIMA (1,1,1) with drift	54.37465	54.98071	63.42537
ARIMA (1,1,2) with drift	24.65125	25.57432	35.96465
ARIMA (1,1,3) with drift	0.03128988	1.34379	13.60737

Conclusion

According to the law of parsimony, ARIMA (2,2,2) has the least information criterion (Refer to table 4.7).

Hence the suitable model for the series with a difference order of "2" is ARIMA (2,2,2).

Table 4.7: R output for the competing models and their respective AIC, AICC and BIC values at first difference.

MODEL	AIC	AICC	BIC
ARIMA (0,2,0)	105.2835	107.361	105.3537
ARIMA (0,2,1)	49.82341	53.97849	50.0377
ARIMA (1,2,0)	25.96572	30.1208	26.18001
ARIMA (1,2,1)	14.41302	20.64564	14.84939
ARIMA (1,2,2)	4.280915	12.59107	5.021656
ARIMA (1,2,3)	-3.191138	7.196549	-2.059063
ARIMA (2,2,0)	-1.615002	4.617611	-1.178638
ARIMA (2,2,1)	-14.33063	-6.020479	-13.58989
ARIMA (2,2,2)	-24.43119	-14.04351	-23.29912
ARIMA (2,2,3)	-22.43709	-9.971862	-20.8217
ARIMA (3,2,0)	-19.74028	-11.43013	-18.99954
ARIMA (3,2,1)	19.07341	-8.685719	-17.94133
ARIMA (3,2,2)	-22.43611	-9.970885	-20.82073
ARIMA (4,2,0)	-21.2407	-9.871137	-19.12675

4.4.5 RESIDUAL DIAGNOSTIC TEST FOR ARIMA (3,1,2) WITH DRIFT

We perform various tests on the residuals on the selected model, ARIMA (3,1,2) to determine whether it is an appropriate model. The test is based on finding whether the residuals do not violate any of the assumptions.

- Ljung- Box test

  H_0: The model is appropriate

  H_1: The model is not appropriate

  From table 4.8, since the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis and conclude that ARIMA (3,1,2) with drift is appropriate.

.

Table 4.8: R output for the competing models and their respective AIC, AICC and BIC values at first difference.

Ljung-Box	Test statistics	P-value
ARIMA (3,1,2) with drift	5.0514	0.4096

From figure 4.6, the residuals appear to have constant variance, there is no correlation between the lags from the ACF plot, also none of the lags spike up (they are all in the threshold). Also, from the graph, we cannot actually tell whether the residuals are normally distributed or not. Hence, we used the Shapiro-Wilk test to check if the residuals are normally distributed

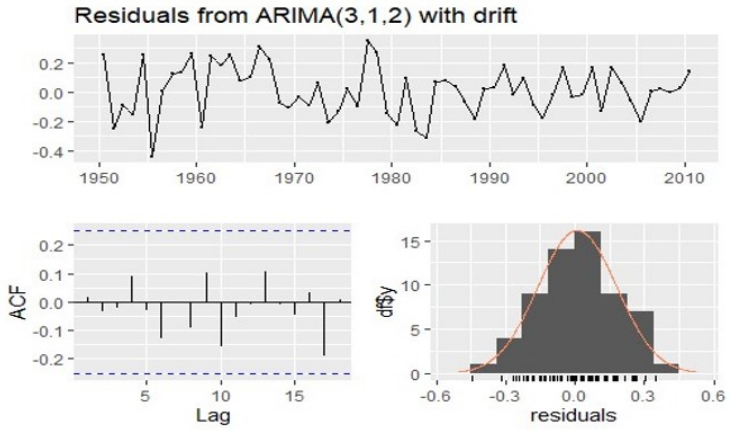

Figure 4.6: Residuals of ARIMA (3,1,2)

Test for normality for residuals

H0= The residuals are normally distributed.

H1= The residuals are normally distributed.

From table 4.9, the p-value is greater than 0.05 and so we fail to reject the null hypothesis. This implies that the residuals are normally distributed.

.

Table 4.9: Results of Shapiro-wilk normality test for ARIMA (3,1,2)

Test	Model	Test statistics	p-value
Shapiro-wilk normality test	ARIMA (3,1,2) with drift	0.98848	0.8369

4.4.6 Residual Diagnostic Test For ARIMA(2,2,2)

From table 4.10, Since the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis and conclude that ARIMA (2,2,2) is appropriate.

From figure 4.7, the residuals appear to have constant variance, there is no correlation between the lags from the ACF plot, also none of the lags spike up (they

are all in the threshold). Also, from the graph, we cannot actually tell whether the residuals are normally distributed or not. Hence, we used the Shapiro-Wilk test to check if the residuals are normally distributed

.

Table 4.10: Results of Shapiro-wilk normality test for ARIMA (3,1,2)

Ljung-Box	Test statistics	P-value
ARIMA (3,1,2)	2.3687	0.8829

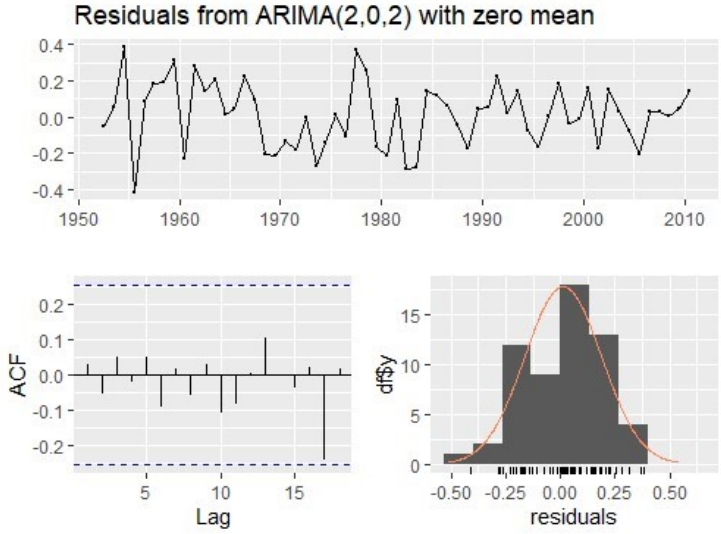

Figure 4.7: Residuals of ARIMA (2,2,2)

Test for normality for residuals From table 4.11, the p-value is greater than 0.05 and so we fail to reject the null hypothesis. This implies that the residuals are normally distributed.

Table 4.11: Shapiro-wilk Test For Normality

Test	Model	Test statistics	p-value
Shapiro-wilk normality test	ARIMA (2,2,2)	0.98605	0.7341

4.4.7 Model Evaluation

Since the two models generated were all appropriate, the two models have to be evaluated to get the overall best fitting model. The data on under-five mortality rate in Ghana was divided into two, that is the training data and the test data. The data from (1950-2011) were used as the training data for fitting the model while the data from 2012 to 2021 were used as test data for the comparison of the forecast performance of the Arima models and therefore their accuracy will be observed and the model with the least error metrics will be selected as the

best fitting model.

4.4.8 Forecast Accuracy Measures For Arima (3,1,2) And Arima (2,2,2)

The forecast values of the two models with different order of differencing were compared in order to choose the best fitting model. The four errors metric were used to calculate their accuracy (ME, MASE, MAE and MAPE).

We observe from Table 4.12 that ARIMA (3,1,2) with drift gives smaller errors as compared to ARIMA (2,2,2). Therefore, we choose ARIMA (3,1,2) as the best model.

Table 4.12: accuracy for ARIMA (3,1,2) with a drift and ARIMA (2,2,2)

Model	ME	RSME	MAPE	MASE
ARIMA (3,1,2) WITH DRIFT	1.833590759	2.2891787	3.72077532	0.59535258
ARIMA (2,2,2)	53.56814666	53.8656056	99.83411	222.6271071

4.5 Objective 3: Forecasting Under-Five Mortality For Ten Years

4.5.1 Forecasting

At the end of the analysis, we perform our main objective by obtaining forecasted values for Under-Five Mortality rate in Ghana for the next ten years. Forecasting is important in decision making.

Forecasting is the process of making predictions about future outcomes based on historical data. It also involves analyzing past and current data to identify patterns, relationships and factors that influence the variable being forecasted. Box and Jenkins (1976) define forecasting as the foundation for economic and business planning, inventory and production control, as well as industrial process control and optimization.

From table 4.13, it can be observed that using the point estimate and the 80%, 95% confidence intervals, the values for the next ten years starting from 2022 to 2031 were predicted.

For the year 2022, the mortality rate will be estimated as 42.54722 with respect to the point estimate and there is 80% and 95% chance of it falling between 42.33201and 42.76244 as well as 42.218079and 42.87636 respectively. The rest of the years follows the same sequence.

Table 4.13: Predicted under-five Mortality rate for the next ten years

Year	Forecast values	80% lower	80% Upper	95% Lower	95% Upper
2022	42.54722	42.33201	42.76244	42.218079	42.87636
2023	41.16818	40.52468	41.81169	40.184027	42.15234
2024	39.73706	38.35485	41.11926	37.623154	41.85096
2025	38.14034	35.67990	40.60079	34.377412	41.90328
2026	36.28321	32.43842	40.12800	30.403113	42.16331
2027	34.11212	28.66215	39.56210	25.777102	42.44715
2028	31.62498	24.46371	38.78624	20.672768	42.57719
2029	28.86833	20.00780	37.72886	15.317318	42.41935
2030	25.92419	15.47484	36.37353	9.943291	41.90508
2031	22.89057	11.02663	34.75451	4.746239	41.03489

From figure 4.8, the graph depicts the trend over a given period of time, thus it shows a downward trend. This decreasing trend reveals that under-five mortality rate will decrease over the cause of time and will cause an increase in Ghana's population.

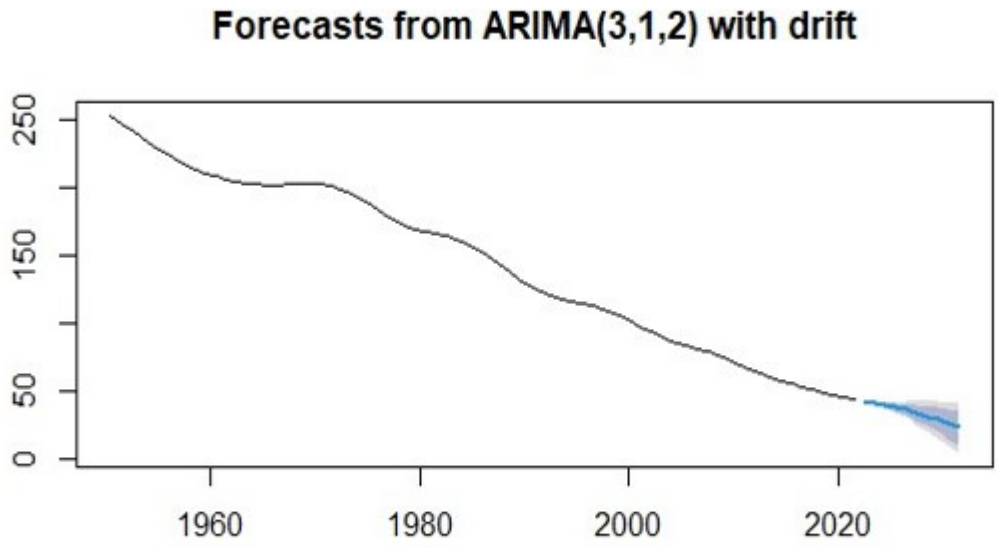

Figure 4.8: Plot of under-five mortality rates from 2022 to 2031

Chapter 5

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Introduction

This chapter summarizes the findings of the study, draws conclusions from the findings and results as well as recommendations for the stakeholders.

5.2 Summary

The aim of the study was to examine the pattern of under-five mortality rate, fit an ARIMA model to the data and also to forecast under-five mortality rate for the next ten (10) years in Ghana. A secondary data obtained from UNICEF was used in the analysis of this thesis. The R software was used in analysing and fitting an ARIMA model to the under-five mortality data. The data was a yearly

mortality rates for the year 1952 to 2021 with seventy- two (72) observations. The time series plot of the under-five mortality data clearly shows unstable variance hence, considered nonstationary. The ARIMA (3,1,2) with drift was selected as the appropriate model for predicting future under-five mortality for Ghana. We concluded statistically that the under-five mortality satisfed all the diagnostic tests for a suitable ARIMA model and was used to forecast under-five mortality rates.

5.3 Recommendation

It is recommended that the stakeholders involved, especially the government put in extra efforts and measures to help reduce under-five mortality to achieve the sustainable development goal. Education on the need to seek prenatal and postnatal care during pregnancy must be intensified, provision of good drinking water should be provided especially in the rural areas, improving healthcare infrastructure and increasing immunization coverage should be implemented. ARIMA (3,1,2) may be adopted for annual under-five mortality projections. Although the ARIMA model adequately fits the data adequately and useful for predicting future mortality rates, it is not recommended for medium and long-term predictions. Finally, this study limited itself to the analysing under-five mortality cases in Ghana regardless of the cause. We recommend further studies into the actual causes of under-five mortality in Ghana.

5.4 References

1. World Health Organization – World Health Statistics. https://www.who.int/news/item/20-05-2022-world-health-statistics-2022

2. MoH Ethiopia, (2000). To analyzing trends and developing a model for

prediction of Health and Health related indicators. The Ethiopian Government through their Ministry of Health, 2000.

3. Yun-li, Ya, L. and Ping, Y. (2004). The trend and forecast of infant mortality rate in China from 1991 to 2004. Department of Epidemiology, West China School of Public Health, Sichuan University, Chengdu 610041, China.

4. Akinwande, M. O., Ibrahim A. and Ibrahim, A. A. (2016). Analysing Infant and Child Mortality (Under-five) in Zaria, Nigeria. A Regression Analysis Approach African Research Review Vol 10(2) Serial No. 42 June 2016,72-95. ISSN 1994-9057(print).

5. World Health Organization. (2019). Maternal mortality. WHO Fact Sheets. (https://www.who.int/news-room/fact-sheets/detail/maternal-mortality) (August 27, 2022, 14:48)

6. Abraha, W. A. and Nigatu T. H. (2009). Modeling trends of health and health related indicators in Ethiopia (1995-2008): a time-series study, Health Research Policy and Systems 7, December 2009, Number 29.

7. Kuhn, L., Davidson, L. L. and Durkin, M. S. (1994). Use of Poisson Regression and Time Series Analysis for Detecting Changes over Time in Rates of Child Injury following a Prevention Program. American Journal of Epidemiology, November 1994, 140(10):943-55.

8. Apanga, P. A. and Awoonor-Williams, J. K. Maternal Death in Rural Ghana: A Case Study in the Upper East Region of Ghana. Front Public Health, 2018 Apr 9, Vol. 6, pp. 101 (Maternal Death in Rural Ghana: A Case Study in the Upper East Region of Ghana - PMC (nih.gov), August 27, 2022, 13:57)

9. Asomaning, S. S. (2012). Analysis of maternal mortality with time; a case study of the Okomfo Anokye Teaching Hospital - Kumasi. Published June 14, 2012.