TIME SERIES MODELS FOR THE DECREASE IN UNDER-FIVE MORTALITY RATE IN GHANA

CASE STUDY 1961 - 2012

BY

PRINCE EDWARD OPARE (B. Ed MATHEMATICS)

PG6323211

A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE: Industrial Mathematics

College of Science/Institute of Distance Learning

NOVEMBER, 2014

DECLARATION

I hereby declare that this submission is my own work towards the MSc, and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any degree of the University, except where due acknowledgement has been made in the text

PRINCE E. OPARE (20250164)

…….. ……. …….

Student Name and ID Signature Date

Certified by:

NANA KENA FREMPONG

..... ……. …….

Supervisor(s) Name Signature Date

Certified by:

PROF. S.K. AMPONSAH

…… …… ……

Head of Dept. Signature Date

Certified by:

PROF. I.K. DONTWI

……. …… ……

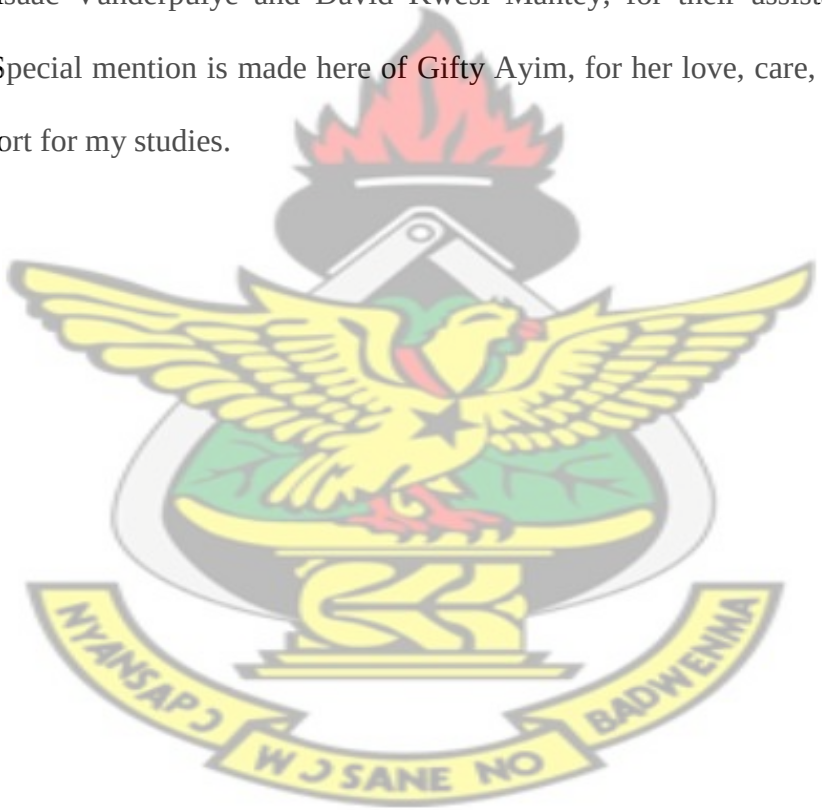
Dean, IDL Signature Date

DEDICATION

Firstly, to **God Almighty** for the strength and guidance that enabled me to finish this work and secondly, to the memory of **Madam Leticia Adukai Addo,** my beloved grandmother for the excellent care she gave me and the sponsorship for my University Education.

ACKNOWLEDGEMENTS

I would want to thank God Almighty for the grace to undertake this course and for a successful completion. I am also thankful to my supervisor, Nana Kena Frempong, of the Mathematics Department of the Kwame Nkrumah University of Science and Technology, whose guidance, encouragements and directions made it possible for me to complete this work. I say God richly bless you sir. I am also grateful to my course mates; Samuel Asante, Isaac Vanderpuiye and David Kwesi Mantey, for their assistance during my studies. Special mention is made here of Gifty Ayim, for her love, care, encouragements and support for my studies.



ABSTRACT

A time series data, comprising of annual estimates of Under-five Mortality rates for Ghana from the year 1961 to 2012, obtained from the Worldbank website is used for the analysis. Three time series models; the Box-Jenkins (ARIMA), the Bayesian Dynamic Linear Model, and the Random walk with drift models are built for the decline of Ghana's under-five Mortality. Each model is built with data values from 1961 to year 2000, and an in-sample forecasting is made with each model from year 2001 to 2012. The Mean Squared Error (MSE) and the Mean Absolute Percentage Error (MAPE) as a measure of accuracy are used to determine the best fit model. The Random Walk with drift model produced the least values for both the MSE and the MAPE and is selected the best fit Model, and used for an out-of-sample forecasting for the years (2013 – 2016), producing respectively; 69.3, 66.6, 64.0 and 61.3 deaths per 1,000 live births. The forecast value of 64.0 deaths per 1000 live births for year 2015 shows that Ghana may not be able to realize her Millennium Development Goal four (MDG 4) target of reducing her Under-five Mortality rate to about 42.7 deaths per 1,000 live births by the year 2015.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- ARIMA – Autoregressive Integrated Moving Average
- DLM – Dynamic Linear Model
- GHS – Ghana Health Service
- GNA – Ghana News Agency
- HDR – Human Development Report
- HIV/AIDS – Human Immuno Virus/Acquired Immune Deficiency Syndrome
- MDG – Millennium Development Goals
- UN – United Nations
- UNDP – United Nation Development Programme
- UNICEF – United Nations Children's Fund
- WHO – World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 Background

Under-five (Infant and Child) Mortality Rate is the probability of a child dying within the period of birth and his/her fifth birthday, expressed per 1,000 live births. (UNICEF, WHO, Worldbank). Infant and Child Mortality Rate is one of the most important measures of child health and an indication of the overall development level of a nation. High Rate of Under-five mortality is therefore undesirable as it does indicate falling living standards of a country.

Available reports on under-five mortality rate show that in the 1990s, about twelve million children died annually in the world, out of which ten million were in Developing countries. Infant and child deaths in developing countries constitute the largest age category of mortality and this is because children under the age of five years are the group vulnerable to diseases caused by health risks and poor environmental conditions (UNICEF, 1998). High Infant and Child Mortality Rates are also the result of high levels of poverty and deprivation, malnutrition, poor access to basic education, the spread of HIV/AIDS, malaria, tuberculosis as well as unhealthy conditions during the time of birth (Asante and Asenso-Okyere, 2003).

Children under five years make up to fourteen percent of the population in Africa and that accounts to about fifty percent of all deaths annually (Kessel, 2000). Estimates of Infant and Child Mortality Rates are high in developing countries and especially in the Sub-Saharan African Nations because basic necessities for infant survival are lacking or unevenly distributed. As a result, infectious and communicable diseases are very common in these countries even though

sound sanitary practices and proper nutrition are given. The WHO attributes seven out of ten childhood deaths in developing countries to five main causes: Pneumonia, Diarrhea, Measles, Malaria and Malnutrition.

In response to the above concerns, the WHO and UNICEF, in the early 1990's led the development and promotion of the Integrated Management of Childhood Illness(IMCI) strategy (UNICEF,1999), aimed at reducing mortality and morbidity associated with the major causes of diseases in children under age five and to contribute to their healthy growth and development.

One of the Eight Millennium Development Goals (MDG's), adopted after the Millennium Summit in 2000 is to reduce child mortality (MDG4). Hence, donor and Development Agencies and Governments around the world committed themselves to the goal of reducing under-five mortality rate by two-thirds between 1990 and 2015(UN Millennium Declaration). Reports on Infant and Child Mortality rates have it that twelve years after the world leaders committed themselves to the Millennium Goal 4 (MDG4), which sets out to reduce under-five mortality rate by two-thirds between 1990 and 2015, the world made substantial progress by reducing the number of under-five deaths by forty-seven percent from1990 to 2012 (i.e. 90 deaths per 1,000 live births in 1990 to 48 in 2012. In 2012, an estimated 6.6 million children – 18,000 a day-died from mostly preventable diseases. The progress made however, has not been enough and the target risks being missed at the global level with only a year remaining for the 2015 deadline. Also, Infant and Child Mortality Rates are still very high in Sub-Saharan Africa where 1 in 9 children dies before age five, more than 16 times the average for the developed regions (UNICEF, 2011).

Data on under-five mortality rate for Ghana can be had from different sources. These include the Ghana Demographic and Health Survey data set, produced jointly by the Statistical Service of Ghana and the Ghana Health Service as well as the Worldbank and sister agencies such as the UNICEF, WHO and the UNDP. The Worldbank and the other agencies give yearly estimates of the rates whereas the Ghana Demographic and Health Survey estimates are done in a five years interval.

1.2 Background of Study Area

Infant and Child deaths in developing countries constitute the largest age category of mortality. This is because children under the age of five years are the group vulnerable to diseases caused by health risks and poor environmental conditions (UNICEF, 1998).

Ghana, the country whose data on under-five mortality rate is used for the study, is a lower middle-income country located in the West African Sub-region. She is bounded on the north by Burkina Faso, on the west by Cote d'Ivoire and Togo on the east. The first Black African Nation to obtain her independence from British colonial rule on 6th March 1957, Ghana covers a total area of about 238,305km squared and divided into ten regions. The Country has a tropical climate with temperature ranging from 21-32 degrees Celsius (70-90F). Ghana has two rainy seasons from March to July and from September to October, separated by a short cool dry season in August. With a current population of about 25 million people and Accra as its capital, Ghana is inhabited mostly by people from different ethnic or tribal backgrounds among which are the Akans, the Ga-Adangbes, Mole-Dagbane, Guans, Ewes etc. with different dialects. However, her official language is English. The Country is also rich in mineral resources such as Gold,

Diamond, Manganese, Bauxite, Iron Ore, granite deposit as well as Oil and Gas. Ghana seeks to reduce by two-thirds her infant and child mortality rate between 1990 and 2015 in line with the MDG4. However, child mortality rate in the country remains very high although some improvements have been made over the years. (HDR, 2007).

1.2.1 Trends in Mortality Rates in Ghana

Referring to the Worldbank data on Ghana's under-five mortality rate, it is observed that Ghana had a gradual decline in her rates since the year 1961 to 2012 .From 1990, the rate fell from 128.1 deaths per 1,000 live births to 72 deaths per 1,000 live births in 2012, indicating a 43.8 percent reduction in the rates. However, many professionals have been very skeptical of Ghana's realization of her 2015 target rate of about 42.7deaths per 1,000 live births.

The regions with significant reduction in under-five mortality rate between 1998 and 2008 were Upper-East Region (reduction of up to 77.6 per1,000 live births),Western, Brong Ahafo and Volta Regions (up to 52.7 per 1,000 live births reduction),while those that recorded the least improvements over the same period were Ashanti (increased by 1.8 per 1,000 live births), Eastern, Greater Accra and Upper West (reduced by up to 13.3 per 1,000 live births only). An observation of the trend showed that Upper East, Western, Brong Ahafo and Volta Regions were on track to achieving the MDG on under-five mortality, while the rest were off track (Ghana MDG report, 2010).

1.3 Problem statement

Studies conducted as well as available information and data from the Worldbank show that there is a decline in the Under-five Mortality Rates for many nations, attributable to realization of the MDG4 target by these nations.

Ghana as a lower- middle income nation is also experiencing this decline (i.e.128.1 per 1,000 live births in 1990 to 72.0 per 1,000 in 2012). However, the extent of the decline is not known and that's the problem this study seeks to look into.

1.4 Objective of the study

The objectives of this study are:

- To model the Under-five Mortality Rates for Ghana using the Box - Jenkins (ARIMA), the Bayesian Dynamic Linear Model (DLM), and the Random walk with drift modelling methods.

- To choose the best model based on measures of forecast accuracy, and to use the selected model to predict the Under-five Mortality Rates for Ghana over a period of four years.

1.5 Methodology

As a result of unhealthy environmental conditions as well as the prevalence of some curable and preventable diseases, a great number of Ghanaian children do not live to see their fifth birthday, hence the very high rates of Under- five Mortality in Ghana. (Ghana's Integrated Child Health Campaign, 2006)

The Box-Jenkins (ARIMA), the Bayesian Dynamic Linear Model (DLM) and the Random Walk with drift methods are used for the analysis. The ARIMA procedure analyze and forecasts equally spaced univariate time series data, transfer function data, and intervention data using the Autoregressive Integrated Moving-Average (ARIMA) or autoregressive moving-average (ARMA) model. An ARIMA model predicts a value in a response time series as a linear combination of its own past values and past errors (also called shocks or innovations).

Dynamic linear models (DLMs) also, are a broad class of models with time varying parameters, useful for modelling time series data. They are parametric models where the parameter variation and the available data information are described probabilistically. They are characterized by a pair of equations, named observational equation and parameter evolution or system equation. The DLM can be seen as a generalization of the regression models allowing changes in parameters values throughout time. The DLM follows the usual steps in Bayesian inference, combining two main equations; evolution to build a prior and updating, to incorporate a new observation arrived at time t.

A random walk is a process where the current value of a variable is composed of the past value

plus an error term defined as a white noise (a normal variable with zero mean and variance one).

Algebraically, a random walk is represented as $Y_t = Y_{t-1} + \epsilon_t$. The implication of a process of this type is that the best prediction of y for the next period is the current value. The mean of a random walk process is constant but its variance is not. Therefore a random walk process is nonstationary and its variance increases with time.

A secondary data comprising fifty-two data points of annual estimates of Under-five mortality rate per 1,000 live births for Ghana from 1961 to 2012, obtained from the Worldbank website in the year 2013, is used for the analysis by the R statistical software. Three different procedures for modelling and forecasting a univariate time series data; the classical or Box - Jenkins (ARIMA), the Bayesian Dynamic Linear (DLM) and the Random walk with a drift methods are used to model the Under-five mortality rates for Ghana, from 1961 to 2000. The models are each used to make an in-sample forecasting for the years 2001 to 2012. The Mean squared error (MSE) and the Mean absolute percentage error (MAPE), as a measure of accuracy are used to determine the best fit amongst the three fitted models, and the best model (i.e. model with the least deviations) is then used to make an out of sample forecast for the years 2013 to 2016, which provide (statistical) evidence as to whether or not Ghana is able to realize her MDG4 goal target.

1.6 Justification

Infant and Child Mortality Rate reflects a country's level of socioeconomic development and quality of life and is used for monitoring and evaluating population and health programmes and policies. High rates of infant and child mortality are undesirable as they tend to reduce the progress a nation makes in her developmental project.

Most stakeholders in the Health sector, without any proof have made statements that Ghana may not be able to achieve her MDG 4 target. This study mainly, will provide a statistical model for predicting the Under-five mortality rate for Ghana at any instant and also bring to the fore, whether or not Ghana is able to achieve her target of reducing under-five mortality rate to about 42.7 deaths per 1,000 live births per the Worldbank data set. The study will provide policy makers with the evidence and the idea of possible future values of the rates, and thus help them to revise their childhood death intervention strategies so as to maintain and sustain the rates, or to reduce further, if the 2015 rate is found to be far from the target value. All these will ensure that Ghanaian children are healthier and grow to realize their talents and potentials and thus help maintain a strong labour force for the future that will continue with the developmental program of the nation Ghana.

1.7 Thesis organization

This study is in five Chapters. Chapter one considers the Introduction of the study, its background, trends in mortality rates in Ghana, the problem Statement and the objectives of the study. It also considers the justification for the study, the methodology and the thesis organization. Chapter two covers the review of available literature that is relevant to the study. Chapter three is devoted for the research methodology. Chapter four focuses on the data analysis, which involves the estimation of the model parameters, the in-sample forecasting by each of the three models, the estimation of the best fit model by a comparison of the Mean square error (MSE) and the Mean Absolute Percentage Error (MAPE) for the in-sample forecast values by

each model. The chapter ends with a four years ahead forecast values by the best fitted model. Chapter five looks at the conclusion and recommendations of the study.

CHAPTER TWO

2.0 Literature Review

This chapter provides a review of some previous studies on infant and child Mortality, conducted at different places across the globe and some of which involves the modelling and prediction of infant and child mortality rates.

Sankrithi et al. (1991), developed a product form multivariate regression models (multiplicative exponential) with infant and child mortality as outcome, and national economic, health, nutrition, education, and demographic statistics as predictor variables. The models were applied to data from 129 countries, resulting in R-square values for the product form models of infant and child mortality of 0.77 and 0.80. For comparison purposes, more conventional sum form models (additive linear) were also estimated, and yielded R-square values (0.22, 0.29) markedly lower than the product form models. The product form models also a much more uniform distribution of residuals and provided improved model fit across the different categories of nations. An inherent advantage to the product form models was that they did not predict negative mortality rates, in contrast to the sum form models which did predict negative mortalities for some of the more developed nations.

Having the objectives of developing and evaluating a model that predicts mortality risk based on admission data for infants weighing 501 to 1500 grams at birth, and to use the model to identify neonatal ICUs where the observed mortality rate differs significantly from the predicted rate, Horbar et al.(1993), undertook a validation cohort study involving a sample of 3,603 infants with birth weight 501 to 1500 grams who were born at seven National Institute of Child Health and Human Development (NICHHD) Neonatal Research Network Centers, over a 2-yr period of time.

Based on logistic regression analysis, admission factors associated with mortality risk for inborn infants were: decreasing birth weight, appropriate size for gestational age, male gender, non-black race, and 1- min Apgar score < or = 3. The mortality prediction model based on those models had a sensitivity of 0.50, a specificity of 0.92, a correct classification rate of 0.82, and an area under the receiver operating characteristic curve of 0.82 when applied to a validation sample. Goodness-of-fit test showed that there was a marginal of fit between the observations and model predictions (chi 2 =15.4, p=.06). There were no statistically significant differences between observed and predicted mortality rates at any of the centers. In their conclusion, the researchers noted that mortality risk for infants weighing 501 to 1500 grams could be predicted base on admission factors but until more accurate predictive models are developed and validated and the relationships between care practices and outcomes are better understood, such models should not be relied on for evaluating the quality of care provided in different neonatal ICUs

Hussein (1993), used two models, one with the natural logarithmic transformation of infant mortality time series and the other with successive differences to provide infant mortality projections for the period 1983 – 2000 in Egypt. Data were obtained from CAPMAS and UNICEF on the Egyptian infant mortality rate for the period 1947 – 82.The best model was determined by successive steps of model specification, estimation, and comparison. Plots of the data were provided for the original data for 1947 – 82, the degree of non-seasonal differencing, and a natural log transformation of the data. Plots were also provided of the sample autocorrelation function and the sample partial autocorrelation function for the original data, the degree of differences, and the natural logarithmic transformations. The preferred model was an autoregressive integrated moving average one for a first difference model (model 1) and a natural logarithmic model (model 2). Parameter estimates in model 2 were more significant and therefore preferred. Goodness of fit

comparisons and comparisons of plots of sample autocorrelation functions for the errors with their probability limits showed both models to be adequate. The two models were used to forecast infant mortality between 1983 and 2000. Model 1 showed a faster decline in mortality than model 2: a decline of 44.4% compared to 25.9%. Model 2 results were preferred because of the known inaccuracies in infant mortality data and the initially sharp decline between 1984 and 1985, which was due to implementation of government health programs.

For the prediction of subsequent mortality among very low birth weight infants (< 1500grams) on days of life 3 and 14 using the Score for Neonatal Acute Physiology (SNAP) and traditional risk factors, Ellington Jr. et al. (1997), prospectively abstracted clinical and demographic data on a cohort of 1670 infants (< 1500grams) at seven regional NICUs from October 1994 to July 1996 and identified all NICU deaths. The researchers measured severity of illness using the Score for Neonatal Acute Physiology (SNAP) at day 3 and 14. The risk of subsequent mortality at day 3 and 14 was determined using sequential logistic models of the traditional risk factors – male sex, white race, SGA status, birth weight and low 5 – minute Apgar – as well as SNAP at days 1,3 and 14 (SNAP1, SNAP3,SNAP14). Receiver Operator curves (ROC) for the competing models were constructed and the differences in area under the ROC curves were compared. The results were that; of the 198 deaths in the cohort, ninety-three occurred after day of life 3 and 43 after day of life 14. On both days only birth weight and SNAP improved mortality prediction. Sequential addition of increasingly proximate SNAP improved the predictive power of the models at both days 3 and 14. ROC- areas of day 3 were significantly less for the traditional model (0.82±.02) than for the traditional model plus SNAP1 and SNAP3 (.88±.02) with p-value <.02. ROC- areas at day 14 were .80±.02 for the traditional model and .84±.03 for the traditional model plus SNAP1, SNAP3 and SNAP 14(p-value=NS). The conclusions were that SNAP does improve the prediction

of subsequent mortality at day 3 over traditional risk factors and that serial measurement severity of illness yields additional information about evolving mortality risk among infant < 1500grams.

With the view to investigate the feasibility of developing an objective tool for predicting death and severe disability using routinely available data, including an objective measure of illness severity, in very low birthweight babies, Fowlie et al. (1997), used a cohort study of 297 premature babies surviving the first three days of life. Predictive variables considered included birthweight, gestation, 3 day cranial ultrasound appearances and 3 day CRIB (clinical risk index for babies) score. Models were developed using regression techniques and positive predictive values (PPV) and likelihood ratios (LR) were calculated. Among the results were that; on univariate analysis, birthweight, gestation, 3 day CRIB score and 3 day cranial ultrasound appearances were each associated with death. On multivariate analysis, 3 day CRIB score and 3 day cranial ultrasound appearances remained independently associated. A 3 day CRIB score > 4 along with intraventricular haemorrhage (IVH) grade 3 or 4 was associated with a PPV of 64% and an LR of 9.8 (95% confidence limits 3.5,27.9). In conclusion, the researchers observed that, incorporating objective measures of illness severity may improve current prediction of death and disability in premature infants.

In a study that describes the time trends for infant mortality in Hong Kong and aims to develop statistical models that could be used to predict changes in infant mortality in places already having low levels of infant mortality, Wong et al. (1997), annually analyzed data on births and deaths of infants in Hong Kong during the years 1956 – 90 as well as aggregating the data into seven consecutive quinquennia. To assess the contribution of preventable infant deaths, causes for infant deaths were classified into two broad categories: (i) congenital anomalies; and (ii) preventable diseases. A simple linear regression model was used to analyze the time trend of the

of the mortality rate of the preventable diseases (PIMR) over the seven quinquennia. Their findings were that; during the period 1956 – 90, the infant mortality rate fell from 60.9 in 1956 – 5.9 per 1000 in 1990 and the neonatal mortality rate fell from 24.2 – 3.8 per 1000. There was no clear time trend observed for infant mortality of congenital anomalies. However, the time trend for PIMR (log scale) was very close to a straight line and simple linear regression modeling showed a R^2 of 0.9970. The conclusions were that; as the infant mortality rate (IMR) falls to below 30 per 1000, the further rate of decrease becomes less predictable from the regression model of the IMR and by removing the portion of deaths attributable to congenital anomalies, the further decrease in infant mortality became more predictable down to very low levels of IMR

In order to predict the individual neonatal mortality risk of preterm infants using an artificial neural network" trained" on admission data, Zernikow et al. (1998), enrolled a total of 890 preterm neonates (< 32 weeks gestational age and/ or 1500g birthweight) in their retrospective study. The neural network trained on infants born between 1990 and 1993. The predictive value was tested on infants born in the successive three years. The results were that; the artificial neural network performed better than the logistic regression model (area under the receiver operator curve 0.95 vs. 0.92). Survival was associated with high morbidity if the predicted mortality risk was greater than 0.50. There were no preterm infants with a predicted mortality risk of greater than 0.80. The mortality risk of two non-survivors with birthweight > 2000g and severe congenital disease had largely been underestimated. The researchers concluded that an artificial neural network trained on admission data can accurately predict the mortality risk for most preterm infants. However, the significant number of prediction failures renders it unsuitable for individual treatment decisions.

In order to determine the interrelationships between potential predictors of infant mortality, Terra de Souza et al. (1999), undertook an ecological study across 140 municipalities in the state of Ceara, Brazil. The researchers classified 11 variables into proximate determinants (adequate weight gain and exclusively breastfeeding), health services variables (prenatal care up-to-date, participation in growth monitoring, immunization up-to-date, and decentralization of health services), and socioeconomic factors (female literacy rate, house income, adequate water supply, adequate sanitation, and per capita gross municipality product), and included the variables in each group simultaneously in linear regression models. Included in their findings were that; only one of the proximate determinants (exclusively breastfeeding (inversely), R^2 =9.3) and one of the health services variables (prenatal care up-to-date (inversely), R^2=22.8) remained significantly associated with infant mortality. In their conclusion, the researchers stated that their results suggested that promotion of exclusive breastfeeding and increased prenatal care utilization, as well as investments in female education would have substantial positive effect in further reducing infant mortality rates in the state of Ceara.

With a goal to generate a preoperative risk-of-death prediction model in selected neonates with congenital heart disease undergoing surgery with deep hypothermic circulatory arrest, Clancy et al.(2000), completed a single-centre, prospective, randomized, double-blind, placebo-controlled neuroprotection trial in selected neonates with congenital heart disease requiring operations for which deep hypothermic circulatory arrest was used. An extensive database was generated that included preoperative, intraoperative, and postoperative variables (delivery, maternal, and infant related) were evaluated to produce a preoperative risk-of-death prediction model by means of logistic regression. An operative risk-of-death prediction model including duration of deep hypothermic circulatory arrest was also generated. Among the results were that; between July

1992 and September 1997, 350(74%) of 473 eligible infants were enrolled with 318 undergoing deep hypothermic circulatory arrest. The mortality was 52 of 318(16.4%), unaffected by investigational drug. The resulting preoperative risk model contained 4 variables: (1) cardiac anatomy (two-ventricle vs. single ventricle surgery, with/ without arch obstruction), (2)1-minute Apgar score (< / =5 vs. > 5), (3) presence of genetic syndrome, and (4) age at hospital admission for surgery (< / = 5 or 5 days). Mortality for two-ventricle repair was 3.2% (4/130). Mortality for single ventricle palliation was 25.5% (48/188) and was significantly influenced by Apgar score, genetic diagnosis and admission. The preoperative risk model had a prediction accuracy of 80%.The operative risk model included duration of deep hypothermic circulatory arrest, which significantly (p=.03) increased risk of death, with a prediction accuracy of 82%. Among the conclusions made was that; postoperative mortality risk was significantly affected by preoperative conditions.

To test and compare published neonatal mortality prediction models, including Clinical Risk Index for Babies (CRIB), Score for Neonatal Acute Physiology (SNAP), SNAP-Perinatal Extension (SNAP-PE), the National Institute of Child Health and Human Development (NICHD) network model, and individual admission factors such as birth weight, low Apgar score (< 7 at 5 minutes), Pollack et al. (2000), collected data on 476 VLBW infants admitted to 8 neonatal intensive care units between October 1994 and February 1997. The calibration (closeness of total observed deaths to the predicted total) of models with published coefficients (SNAP-PE, CRIB, and NICHD) was assessed using the standardized mortality ratio. Discrimination was quantified as the area under the curves. Calibrated models were derived for the current database using logistic regression techniques. Goodness-of-fit of predicted to observed probabilities of death was assessed with the Hosmer-Lemeshow goodness-of-fit test. Among the observations made by

the researchers was that; the calibration of published algorithms applied to the data was poor. The standardized mortality ratios for the NICHD, CRIB, and SNAP-PE models were .65, .56, .82 respectively. Discrimination of all the models was excellent (range: .863 - .930). The conclusions made were that; published models for the severity illness over predicted hospital mortality in the set of VLBW infants, indicating a need for frequent recalibration. Discrimination for the severity of illness score remained excellent. Included in the conclusion was that birth variables should be reevaluated as a method to control for severity of illness in predicting mortality.

In order to compare the prediction of mortality in individual extremely low birth weight (ELBM) neonates by regression analysis and by artificial neural networks, Ambalavanan et al.(2001), used a database of 23 variables on 810 ELBW neonates admitted to a tertiary care center and which was divided into training, validation, and test sets. Logistic regression and neural network models were developed on the training set, validated, and outcome (mortality) predicted on the test set. Stepwise regression identified significant variables in the full set. Regression models and neural networks were then tested using data sets with only the identified significant variables, and then with variables excluded one at a time. The results were that; the area under the curve (AUC) of receiver operating characteristics (ROC) curves for neural networks and regression were similar (AUC 0.87+/- 0.03; p= 0.31). Birthweight or gestational age and the 5-min Apgar score contributed most to AUC. Their conclusions were that both neural networks and regression

analysis predicted mortality with reasonable accuracy and that for both models, analyzing selected variables was superior to full data set analysis. They speculated that neural networks may not be superior to regression when no clear non-linear relationships exist.

To test a paediatric intensive care mortality prediction model for UK use, Pearson et al. (2001), analyzed a total of 7253 admissions using tests of the discrimination and calibration of the logistic regression equation from a prospective collection of data from consecutive admissions to five UK paediatric intensive care units (PICUs), representing a broad cross section of paediatric intensive care activity. It was observed that the model discriminated and calibrated well, and the area under the ROC plot was 0.84 (95% CI 0.819 to 0.853). The standardized mortality ratio was 0.87 (95% CI 0.81 to 0.94). There was remarkable concordance in the performance of the paediatric index mortality (PIM) within each PICU, and in the performance of the PICUs as assessed by PIM. In conclusion, the researchers recommended that UK PICUs use PIM for their routine audit needs and that PIM was not affected by the standard of therapy after admission to PICU.

With the aim of developing a mortality prediction score for retrieved neonates based on the information given at the first telephone contact with a retrieval services, Broughton et al. (2004), examined data from the New South Wales Newborn and Pediatric Emergency Transport Service database. Analysis was performed with the results for 2504 infants and whose outcome (neonatal death or survival) was known. The study population was divided randomly into 2 halves, the derivation and validation cohorts. Univariate analysis was performed to identify variables in the derivation cohort related to neonatal death. The variables were entered into a multivariate logistic regression analysis with neonatal death as the outcome. Receiver operator characteristics (ROC) curves were constructed with the regression model and data from the derivation cohort and then the validation cohort. The results were used to generate an inter-based score, the Mortality Index for Neonatal Transport (MINT) score. ROC curves constructed to assess the ability of the MINT score to predict perinatal and neonatal death. A 7 – variable (Apgar score at 1 minute, birth

weight, presence of a congenital anomaly, and infants age, pH, arterial partial pressure of oxygen, and heart rate at the times of the call) model was constructed that generated areas under ROC curves of 0.82 and 0.83 for the derivation and validation cohorts respectively. The seven variables were then used to generate the MINT score, which gave areas under ROC curves of 0.80 for both neonatal and perinatal death. Their conclusion was that; data collected at the first telephone contact by the referring hospital with a regionalized transport service could identify neonates at the greatest risk of dying.

Slater et al. (2004), conducted a two-phase prospective observational study to compare the performance of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III in Australia and New Zealand. The study involved two phases where phase 1 assessed the performance of PIM, PRISM, and PRISM III between 1997 and 1999 and phase 2 assessed PIM 2 in 2000 and 2001. Discrimination between death and survival was assessed by calculating the area under the receiver operating characteristic plot for each model. The areas (95% confidence interval) for PIM, PIM2, PRISM, and PRISM III were 0.89(0.88-0.90), 0.90(0.88-0.91), 0.90(0.89-0.91) and 0.93(0.92-0.94). The calibration of the models was assessed by comparing number of observed to predicted deaths in different diagnostic and risk groups. Prediction was best using PIM2 with no difference between observed and expected mortality (standardized mortality ratio [95% confidence interval] 0.97 [0.86 – 1.05]). PIM, PRISM III and PRISM all over predicted death, predicting 116%, 130 % and 189% of observed deaths, respectively. The performance of individual units was compared during phase 1, using PIM, PRISM, and PRISM III. There was agreement between the models in the identification of outlying units; two units performed better than expected and one unit worse than expected for each model. The conclusions were that; of the models tested, PIM 2, was the most accurate and

had the best fit in different diagnostic and risk groups; therefore, it is the most suitable mortality prediction model to use for monitoring the quality of pediatric intensive care in Australia and New Zealand. However, more information about the performance of the models in other regions is required before those results could be generalized.

In their bid to compare multiple logistic regression and neural network models in predicting death for extremely low birth weight neonates at 5 time points with cumulative data sets: scenario A, limited parental data, scenario B, scenario A plus additional parental data, scenario C, scenario B plus data from the first 5 minutes after birth, scenario D, scenario C plus data from the first 24 hours after birth; scenario E, scenario D plus data from the first 1 week after birth, Ambalavanan et al.(2005) , used data for all infants with birth weights of 401 to 1000g who were born between January 1998 and April 2003 in 19 National Institute of Child Health and Human Development Neonatal Research Network centers. Twenty-eight variables were selected for analysis, and logistic regression and neural network models for predicting subsequent death were created with training data sets and evaluated with test data sets. The predictive abilities of the models were evaluated with the area under the curve of the receiver operating characteristic curves. The data sets for scenarios A, B and C were similar, and prediction was best with scenario C (area under the curve: 0.85 for regression; 0.84 for neural networks), compared with scenarios A and B. The logistic regression and neural network models performed similarly well for scenarios A, B, D and E, but the regression model was superior for scenario C. Included in the conclusions was that; prediction of death is limited even with sophisticated statistical methods such as logistic regression and nonlinear modeling techniques such as neural networks.

To compare two models (The Pediatric Risk of Mortality III score and Pediatric Index of Mortality) for prediction of mortality in a pediatric intensive care in Hong Kong, Choi et al. (2005), used a prospective case series design for their study in a five-bed pediatric intensive care unit in a general hospital in Hong Kong. All patients were consecutively admitted to the unit between April 2001 and March 2003 and the scores for both models compared with observed mortality. The results showed that; a total of 303 patients were admitted to the pediatric intensive care unit during the study period. The median age was 2 years, with a interquartile range of 7 months to 7 years. The male to female ratio was 169:134 (55.8%: 44.2%). The median length of hospital stay was 3 days. The overall predicted number of deaths using The Pediatric Risk of Mortality III score was 10.2 patients whereas that by Pediatric Index of Mortality was 13.2 patients. The observed mortality was eight patients. The area under the receiver operating characteristics curve for the two models was 0.910 and 0.912 respectively. The researchers concluded that the predicted mortality using both prediction models correlated well with the observed mortality.

To develop and validate a model for very low birth weight (VLBW) neonatal mortality prediction, based on commonly available data at birth, in 16 neonatal intensive care units (NICU) from five South American countries, Marshall et al. (2005), prospectively collected bio - demographic data from the Neonatal del Cono Sur (NEOCOSUR) network between October 2000 and May 2003 on infants with birth weight 500 to 1500g. A testing sample and cross validation techniques were used to validate a statistical model for risk of in-hospital mortality. The new risk score was compared with two existing scores by using area under the receiver operating characteristic curve (AUC). The findings were that, the new NEOCOSUR score was highly predictive for in hospital mortality (AUC= 0.85) and performed better than the Clinical

Risk Index for babies (CRIB) and the NICHD risk models when used in the NEOCOSUR network. The new score was also well calibrated; it had a good predictive capability for in-hospital mortality at all levels of risk (HL test= 11.9, p=0.85). The new score also performed well when used to predict in hospital neurological and respiratory complications. Among their conclusions was that the new and relatively simple VLBW mortality risk score had a good prediction performance in South American network population and the score may prove to be a better model for application in developing countries.

To use the Canadian Neonatal Network (CNN) database to validate the Score for Neonatal Acute Physiology, Version II (SNAP-II) for prediction of mortality among CDH infants admitted to a neonatal intensive care unit (NICU), and to compare that to the predictive equation developed by the Congenital Diaphragmatic Hernia Study Group (CDHSG), Skarsgard et al. (2005) identified infants with CDH in the CNN database. Bivariate and multivariable logistic regression models were used to identify risk factors predictive of mortality. Model predictive performance and calibration were assessed using the area under the receiver operator characteristic curve and the technique of the Hosmer-Lemeshow, respectively, and compared with the CDHSG predictive equation. Among the 19,507 admissions to CNN hospitals, there were 88 patients with CDH. The mortality rate among the CDH patients surviving to NICU admission was 17 %, and 12.5% received extracorporeal membrane oxygenated therapy. Gestational age and admission SNAP-II Score predicted mortality. Model predictive performance and calibration were optimized with those variables combined. The CDHSG equation was equally predictive of mortality, but was only marginally calibrated. The conclusion was that SNAP-II was highly predictive of mortality among patients with CDH, and could be used to risk-adjust those patients.

Bitwe et al. (2006), in their bid to find a simple mortality prediction model based on nutritional and infection indicators for the assessment of the care of children admitted to hospital in central Africa, conducted a cohort study of 414 children admitted at Goma Hospital between 1.4.2003 and 31.3.2004. The researchers did a univariate analysis and logistic regression, computed adjusted odds ratios and constructed a prognostic score from the coefficients of logistic regression. The performance of logistic model and score were evaluated by the calculation of areas under the ROC curves. The intrahospital mortality rate reached 15.9%. In the univariate analysis, age, WAZ, arm circumference, neurological status (Blatyre coma score), stiff neck, subcostal indrawning, and infection were significantly associated with mortality. Logistic regression model analysis and adjusted odds ratios (AOR) confirmed higher risks of death for young (AOR 3.4(1.4-8.8) and underweight children (WAZ - 2 - > - 3 and WAZ < or = - 3, AOR 3.2 (1.4-7.6) and AOR 4.4(1.7-11.2)), for children with arm circumference under 115mm (AOR 3.4(1.5-7.3)); impaired consciousness (AOR 9.6(3.1-29.9)) and bloodstream infections(AOR 6.6(2.1-21.1)). The area under the ROC curve of the prognostic model was 0.83(0.78-0.88), and that of the prognostic score, 0.80(0.75-0.86). In conclusion, the researchers noted that the study provided a simple mortality prediction model for hospitalized children in central Africa and that, the model and scoring system could be used to evaluate programs set up to reduce intrahospital mortality in that region.

In a document that presents a Bayesian approach to forecasting mortality rates, an approach that formalizes the Lee- Carter method as a statistical model for all sources of variability, Pedroza (2006), used Markov chain Monte Carlo methods to fit the model and to sample from the posterior predictive distribution. The document shows how multiple imputations could be readily incorporated into the model to handle missing data and presented some possible extensions to the

model. The methodology was applied to U.S. male mortality data. Mortality rate forecast were formed for the period 1990 – 1999 based on data from 1959 – 1989. Those forecasts were compared to the actual observed values. Results from the forecasts showed the Bayesian prediction intervals to be appropriate wider than those obtained from the Lee-Carter method, correctly incorporating all known sources of variability. An extension to the model was also presented and the resulting forecast variability appeared better suited to the observed data.

In a study that aimed at providing estimates of diarrhea mortality at country, regional and global level by employing the Child Health Epidemiology Reference Group (CHERG) standard, Boschi-Pinto et al. (2008), undertook a systematic and comprehensive literature review of all studies published since 1980 reporting under-5 diarrhea mortality. Information was collected on characteristic of each study and its population. A regression model was used to relate these characteristics to proportional mortality from diarrhea and to predict its distribution in national populations. Among the findings were that; global deaths from diarrhea of children aged less than 5 years were estimated at 1.87 million (95% confidence Interval, CI:1.56-2.19). In their conclusion, the researchers noted that planning and evaluation of interventions to control diarrhea deaths and to reduce under-5 mortality was obstructed by the lack of a system that regularly generates cause-of death information.

Blot et al. (2009), with the objective to develop a user-friendly model to predict the probability of death from acute burns soon after injury based on burned surface area, age and presence of inhalation injury, conducted a population- based cohort study which included all burned patients admitted to one of the six Belgian burn centres. Data from 1999 to 2003 (5246 patients) were used to develop a mortality prediction model, and data from 2004 (981 patients) were used for validation. The results were that mortality in the derivation cohort was 4.6 per cent. A mortality

score (0 – 10 points) was devised: 0 – 4 points according to the percentage of burned surface area (less than 20, 20 – 39, 40 – 59, 60 – 79 or at least 80 per cent), 0 – 3points according to age (under 50, 50 – 64, 65 – 79 or at least 80 years) and 3 points for the presence of an inhalation injury. Mortality in the validation cohort was 4.3 per cent. The model predicted 40 deaths, and 42 deaths were observed (p=0.950). Receiver – operator characteristic curve analysis of the model for prediction of mortality demonstrated an area under the curve of 0.94 (95 per cent confidence interval 0.90 to 0.97). The conclusion by the researchers was that an accurate model was developed to predict the probability of death from acute burn injury based on simple and objective clinical criteria.

For the validation of Clinical Risk Index for Babies (CRIB II) score in predicting the neonatal mortality in preterm neonates ≤ 32 weeks gestational age, Rastogi et al. (2009), used a prospective cohort study. The five variables related to CRIB II were recorded within the first hour of admission for data analysis. The receiver operating characteristics (ROC) curve was used to check the accuracy of the mortality prediction. H-L Goodness of fit test was used to see the discrepancy between observed and expected outcomes. Among the 69 neonates completing the study, 24(34.8%) had adverse outcome during hospital stay and 45(65.2%) had favorable outcome. CRIB II correctly predicted adverse in 90.3% (Hosmer-Lemeshow) goodness – of – fit test p=0.6). Area under curve (AUC) for CRIB II was 0.9032. The conclusion made by the researchers was that; CRIB II score was found to be a good predictive instrument for mortality in preterm infants ≤ 32 weeks gestation.

Sergio et al. (2009), analyzed the annual mortality rates from infectious diarrheic diseases in children under 5 years of age in Brazilian municipalities. The rates from 1990 to 2000 were analyzed using multilevel model, with years as first level units nested in municipalities as second

level units. The dependent variable was the yearly mortality rate by municipality, on the log scale. Polynomial time trends and indicator variables to account for differences in geographic regions were used in the modeling. Time trends were centered on 1995, so they could be modeled differently before and after 1995. From 1990 to 1995 there was a sharp decrease in mortality rates by diarrheic diseases in most Brazilian municipalities, while from 1995 to 2000 the decrease was more heterogeneous. In 1995 the North and Northeast of Brazil of Brazil had higher mortality rates than the Southeast, and the differences were statistically significant. Most importantly, the study concludes that there was an important difference in the pattern of mortality rate decrease over time, comparing the country's five geographic regions.

As mortality improvement has become an increasingly significant source of financial risk, it has become important to measure the uncertainty in the forecasts. Probabilistic confidence intervals provided by the widely accepted Lee- Carter model are known to be excessively narrow, due primarily to the rigid structure of the model. In their study, Siu-Hang Li et al. (2009), relaxed the model structure by considering individual differences (heterogeneity) in each age-period cell. The proposed extension not only provided a better goodness-of-fit based on standard model selection criteria, but also ensured more conservative interval forecasts of central death rates and hence could better reflect the uncertainty entailed. The researchers illustrated the results using the US and Canadian mortality data.

Amouzou et al. (2010), used the Lives Saved Tool (LiST) to model neonatal and under-5 mortality levels among the highest and lowest wealth quintiles in Bangladesh based on national and wealth- quintile-specific coverage of child survival interventions. The cause-of-death structure among children under-5 was modeled using coverage levels. Modeled rates were compared to the rates measured directly from the 2004 Bangladesh Demographic and Health

Survey and associated verbal autopsies. Modeled estimates of mortality within wealth quintiles fell within the 95% confidence intervals of measured mortality for both neonatal and post-neonatal mortality. LiST also performed well in predicting the cause-of-death structure for those two age groups for the poorest quintile of the population, but less well for the richest quintile. The conclusion was that, LiST holds promise as a useful tool for assessing socio-economic inequities in child survival in low-income countries.

Fry-Johnson et al. (2010), used zero-corrected, negative binomial multivariable modeling to predict Black infant mortality (1999-2003) in all US counties with reliable rates. Independent variables included county population size, racial composition, educational attainment, poverty, income and geographic origin. Resilient counties were defined as those whose Black infant mortality rate residual score was < 2.0. Mortality data was accessed from the Compressed Mortality File compiled by the National Center for Health Statistics and found on the CDC WONDER website. Demographic information was obtained from the US Census. Among the results were that; the final model included the percentage of Blacks, age 18 to 64 years, speaking little or no English ($p<.008$), a socioeconomic index comprising educational attainment, poverty, and per capita income ($p<.001$) and household income in 1990 ($p<.001$). In their conclusion, the researchers stated that models for reduction/elimination of racial disparities in US infant mortality, independent from county-level contextual measures of socioeconomic status, may already exist.

In a work that aimed at comparing the performances of ARIMA, Neural Network and Linear Regression models for the prediction of Infant Mortality Rate, Purwanto et al. (2010), compared the models using performance measures such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), using the Infant Mortality Rate

data collected in Indonesia during the years 1995-2008. The results showed that the Neural Network Model with 6 input neurons, 10 hidden layer neurons and using hyperbolic tangent activation functions for the hidden and output layers was the best among the different models considered.

In a study that presents a predictive cause of death model for under-five mortality based on historical vital statistics and discusses the utility of the model in generating information that could accelerate progress towards MDG4, Rao et al. (2010), analyzed over 1400 country years of vital statistics from 34 countries collected over a period of nearly a century , to develop relationships between levels of under-five mortality, related mortality ratios, and proportionate mortality from four cause groups: perinatal conditions, diarrhea and lower respiratory infections; congenital anomalies; and all other causes of death. A system of multiple equations with cross-equation parameter restrictions and correlated error terms was developed to predict proportionate mortality by cause based on given measures of under-five mortality. The strength of the predictive model was tested through internal and external cross-validation techniques. Modeled cause-specific mortality estimates for major regions in Africa, Asia, Central America, and South America were presented to illustrate its application across a range of under-five mortality rates. Consistent and plausible trends and relationships were observed from historical data. High mortality rates were associated with increased proportions of deaths from diarrhea and lower respiratory infections. Internal and external validation confirmed strength and consistency of the predictive model. Among the conclusions made were that; the predictive model could help set broad priorities for interventions at the local level based on periodic under-five mortality measurement.

Although there has been substantial reduction in infant and child mortality rates in most developing countries in the recent past, infant mortality remains a major public health issue in developing countries where it is estimated that over 10 million preventable child deaths occur yearly. With special reference to Nigeria, available statistics suggested that infant mortality levels continue to be high and exhibit wide geographic disparities. In a study that attempted to estimate infant mortality rate in Nigeria using linear regression model, Mojweku et al. (2011), selected crude death rate (CDR) as the minimum relevant parameter (independent variable) needed for estimating Infant Mortality Rate (IMR) which was the dependent variable, because it represented the 'end result' of development. The IMR derived model was checked for adequacy by comparing the estimates of the present study with the estimates from other sources. The diagnostic test showed that the regression derived, was quite adequate and reflected the true picture of Nigerian Infant Mortality Rate pattern.

Nakwan et al. (2011), used a prospective cohort study of 41 infants with persistent pulmonary hypertension of the newborn (PPHN) admitted to a neonatal intensive care unit between June 2008 and March 2010, and underwent a SNAP-II test within 12h of admission, with a view to evaluate the ability of the Score for Neonatal Acute Physiology- Version II (SNAP-II) to predict mortality in infants with PPHN. Of the 41 infants, 14 died (34.1%) and 27 survived (65.9%). The SNAP-II Scores were significantly higher in infants who died (50.1±18.5 vs. 35.7±16.8, P=0.02). Each point increase in the SNAP Score increased the odds of mortality by 1.04 [95% confidence Interval (CI) 1.01 – 1.07, P < 0.01]. Infants who had a SNAP-II Score of ≥ 43 had the greatest mortality risk with an odds ratio (OR) of 10.00 (95% CI 1.03 – 97.50). The SNAP-II model showed moderate discrimination in predicting mortality with a result of 0.72(95% CI 0.56 –

0.88) under the receiver operating characteristic curve. Among the conclusions were that; the SNAP-II scoring system significantly predicted mortality.

With an objective to develop a predictive model and identify maternal, child, family and other risk factors associated with U5M in Nigeria, Kayode et al. (2012), used a population- based cross-sectional study which explored the 2008 demographic and health survey of Nigeria (NDHS) with multivariable logistic regression, Likelihood Ratio test, Hosmer-Lemeshow Goodness-of-Fit and variance Inflation Factor were used to check the fit of the model and its predictive power was assessed with the Receiver Operating Curve (ROC curve). The study yielded an excellent predictive model which revealed that the likelihood of U5M among the children of mothers that had their first marriage at age 20-24 years and greater or equal to 25 years declined by 20% and 30% respectively compared to children of those that married before the age 15 years, Also, the following factors reduced odds of U5M: health seeking behavior, breastfeeding children for > 18 months, use of contraception, small family size, having one wife, low birth order, normal birth weight, child spacing, living in urban areas, and good sanitation. In their conclusion, the researchers indicated that, the study revealed that maternal, child, family and other factors were important risk factors of U5M in Nigeria.

In a latest estimates of the causes of child mortality in 2010 with time trends since 2000, Liu et al. (2012), used vital registration data for countries with an adequate vital registration system. A multinomial logistic regression model was applied to vital registration data for low-mortality countries without adequate vital registration. A similar multinomial logistic regression with verbal autopsy data for high mortality countries was used. For India and China, national models were developed and country results were aggregated to generate regional and global estimates. Among the findings were that; out of 7.6 million deaths in children younger than 5 years in 2010,

64.0% (4.879 million) were attributable to infectious causes and 40.3% (3.072 million) occurred in neonates. Preterm birth complications (14.1%; 1.078 million, uncertainty range [UR] 0.916 – 1.325), intrapartum-related complications (9.4%; 0.717 million, 0.610 – 0.876), and sepsis or meningitis (5.2%; 0.393 million, 0.252 – 0.552) were the leading causes of neonatal deaths. In their conclusion, the researchers advocated that child survival strategies should direct resources toward the leading causes of child mortality, with attention focusing on infectious and neonatal causes.

The WHO has released prescriptive child growth standards for, among others, BMI- for-age (BMI-FA), mid-upper arm circumference-for-age, and weight velocity. In a study that aimed, firstly, to assess in children under 2, the independent and combined ability of those indices and of stunting to predict all cause mortality within 3 mo, or secondly the comparative abilities of weight-for-length (WFL) and BMI-FA to predict short term (< 3 mo) mortality, O'neil et al. (2012), used anthropometry and survival data from 2402 children aged between 0 and 24 mo in rural area of the Democratic Republic of Congo with high malnutrition and mortality rates and limited nutritional rehabilitation. Analysis used Cox proportional hazard models and receiver operating characteristic curves. Univariate analysis and age-adjusted analysis showed predictive ability of all indices. Multivariate analysis without age adjustment showed that only very low weight velocity [HR= 3.82(95% CI=1.91, 7.63); P <0.001] was independently predictive. With age adjustment, very low weight velocity [HR=3.61(95% CI=1.80, 7.25); P <0.001] was again solely retained as an independent predictor. There was no evidence for a difference in predictive ability between WFL and BMI-FA.

To predict neonatal mortality and length of stay (LOS) from readily available perinatal data for neonatal intensive care unit (NICU) admission in Southern African private hospitals, Pepler et al. (2012), did a retrospective observational study using perinatal data from a large multicentre sample. The researchers used 2376 infants born between 1 January – 31 December 2008 to build regression models, and a further 1578 infants born between 1 January – 31December 2007 to test the models. Outcome measures were mortality and length of hospital stay for NICU admissions. Included in the results were that; of the infants included in the 2008 dataset,(3.8%) died after being admitted to NICU centres. An analysis of the structural peculiarities of the data showed high correlations between groups of the perinatal variables pertaining to the size and apgar scores of the newborn infants, respectively. The logistic regression model to predict neonatal mortality had a good fit (AUC: 0.8507, misclassification rate = 13.6%), but the low positive predictive value of the model reduces its usefulness.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This section examines some basic definitions and concepts of time series analysis and the processes involved in the building and application of autoregressive integrated moving average (ARIMA) models, the Bayesian Dynamic Linear Models (DLM) as well as the random walk with drift models for forecasting future values of an investigated variable.

3.1 SOME BASIC CONCEPTS AND DEFINITIONS OF TIME SERIES

3.1.1 Time-Series

A time series is a set of observations measured sequentially through time. (Chatfield, 2001),

3.1.2 A Time Series Plot

A time series plot is a graph with a dependent variable plotted as ordinates against an independent variable (time) as abscissa. There is usually a single value of the dependent variable for each value of the independent variable and those values are typically equally spaced. The time series plot could be done in various ways, and it's used to evaluate patterns and behaviors in the data over time.

3.1.3 Time-Series Model

A time series model establishes a relationship between the present value of a time series and its past values so that forecasts can be made on the basis of the past values alone. A time series model uses a model for explanation that is based on theoretical foundations and mathematical representations. Time series data could be modelled by several different approaches or methods, among which are:

- Autoregressive (AR) models
- Moving Average (MA) models
- Autoregressive Moving Average (ARMA) models
- Autoregressive Integrated Moving Average (ARIMA) models
- The Bayesian Dynamic Linear (DLM) Models
- The Random Walk with or without drift

3.1.4 Stationary Time Series

A time series is said to be stationary if the mean, variance and autocovariance (correlation) structure do not change over time. This basically means, the series has a constant mean and has no trend overtime. A common theoretical example of a weak stationary process is the white-noise process, $(\varepsilon_t)_{t\epsilon T,}$ which has a an expected value of zero, $E(\varepsilon_t) = 0$, a constant variance $Var(\varepsilon_t) = \sigma^2$ and a covariance of zero, $Cov(\varepsilon_t , \varepsilon_s) = 0$ for all $t\epsilon T$. If the time series to be modelled is not stationary, it is often possible to transform it to stationarity with one of the following techniques:

- Differencing the data in the following way, $Y_t = X_t - X_t - _1$ to give a new series Y_t. Consequently, the new data set contains one less point than the original. Although one can difference the data more than once, the first difference is in most cases sufficient.

- Fitting some type of curve (line) to the data, if there is a trend in the data and then by modelling the residuals obtained from that fit. Since the purpose of making a time series stationary is to remove its long-term trend, a simple fit such as a straight line is typically used.

- Taking the logarithm or square root, if the series has no constant variance- this may stabilize the variance

3.1.5 Autocorrelation Function (ACF)

A time-series is assumed a realization of a stochastic process. In the context of time series analysis, the relationships between observations in different time periods play a very important role. These relationships across time can be captured by the time series correlation respectively (resp.) covariance, known as autocorrelations resp. -covariances.

The autocovariance function (γ_k) of a time-series is defined as:

$$\gamma_k = E\{[X_t - E(X_t)][X_{t-k} - E(X_{t-k})]\},$$

Where X_t stands for the time-series. The autocorrelation function (ρ_k) is defined as:

$$\rho_{(k)} = \frac{\gamma_k}{\gamma_o} \quad \text{.....(1)}$$

The graph of this function is called correlogram. The correlogram has an essential importance for the analysis, because it comprised time dependence of the observed series. Since γ_k and ρ_k only differ in the constant factor γ_o i.e. the autovariance of the time-series, it is sufficient to plot just one of these two functions. One application of autocorrelation plots is for checking the randomness in the data set. The idea is, that if these autocorrelations are near zero for any and all time lags then the data set is random. Another application of this correlogram is for identifying the order of an AR and an MA process. Technically, these described plots are formed by displaying on the vertical axis the autocorrelation coefficients (γ_k) and on the horizontal axis, the time lag.

3.1.5.1 Partial Autocorrelation function (PAFC)

The partial autocorrelation function (π_K), where $k \geq 2$, is defined as the partial correlation between X_t and X_{t-k} under holding the random variables in between X_u , where $t - k < u > t$, constant. It seems to be obvious, that the PACF is only defined for lags equal to two or greater, because considering the following example: if one calculates π_2 of X_t and X_{t-2} under holding X_{t-1} constant then the correlation of X_{t-1} disappears. But if one wants to calculate the π_1 of X_t and X_{t-1} it is the same as computing the ACF at lag one, i.e. ρ_1. The partial autocorrelation plot or partial correlogram is also commonly used for model identification in Box and Jenkins models. On the y-axis they display the partial autocorrelations coefficients at lag k and on the x-axis the time lag k .

3.1.5.2 Autoregressive (AR) Models

A common approach for modelling a univariate time series is the AR model. The intuition behind this model is that, the observed time series X_t depends on weighted linear sum of the past values, p, of X_t and a random shock ε_t. Thus, the name "autoregressive" derives from this idea. Technically, one can therefore formulate the AR(p) model as follows:

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

Where X_t denotes the time series and ε_t indicates a white-noise process. The value of p is called the order of the AR model. If $p = \infty$, then the process is called an infinite AR process. An autoregressive model corresponds simply to a linear regression of the current value of the series against one or more prior values of the series. Often, a formulation of the AR(p) model is made by using the lag operator L, which is defined as $LX_t = X_{t-1}$.

Consequently, $L(LX_t) = L^2X_t = X_{t-2}$ and in general, $L^sX_t = X_{t-s}$ and $L^0X_t = X_t$. This means operating L on a constant leaves the constant unaffected. Using the lag operator, one can rewrite an AR(1) model, $X_t = \phi X_{t-1} + \varepsilon_t$ in the following way:

$$X_t = \phi L X_t + \varepsilon_t \Leftrightarrow X_t (1 - \Phi L) = \varepsilon_t$$

Similarly, using the lag operator we can write the general AR(p) model in equation (2) above as:

$$X_t = \phi_1 L X_t + \phi_2 L^2 X_t + \ldots + \phi_p L^p X_t + \varepsilon_t$$

$$X_t = X_t(\phi_1 L + \phi_2 L^2 + \ldots + \phi_p L^p) + \varepsilon_t$$

$$X_t(1 - \phi_1 L - \phi_2 L^2 - \ldots \phi_p L^p) = \varepsilon_t$$

In the brief form; $X_t\phi(L) = \varepsilon_t$, where $\phi(L)$ is a polynomial of order p in the lag operator

$\Phi(L) = 1 - \phi_1 L - \phi_2 L - \ldots - \phi_p L^p$.

3.1.5.3 Moving Average (MA) Models

Another common approach for modelling a univariate time-series is the MA model. The intuition behind this model is that, the observed time series X_t depends on a weighted linear sum of past, q, random shocks. This means that at period t, a random shock ε_t is activated and this random shock is independent of random shocks of other periods. The observed time–series X_t is then generated by a weighted average of current and past shocks – this explains the name "moving average".

Technically, one can therefore formulate the MA(q) model as follows:

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots\ldots\ldots + \theta_q\varepsilon_{t-q} \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where, X_t denotes the time-series and ε_{t-q} indicates a white – noise process. The value of q is called the order of the MA model. If $q = \infty$, then the process is called an infinite MA process. An MA model corresponds simply to a linear regression of the current value of the series against the random shocks of one or more prior values of the series. But fitting a moving average is more complicated than fitting an autoregressive model, because it depends on the error terms that are not observable. Therefore in opposite to an AR model, one has to use an iterative non-linear fitting procedure and the resulting estimation of the parameter has less obvious interpretation than in the case of AR models. As before, one can rewrite an MA model in brief by using the described lag operator in the following way:

$X_t = \theta(L)\varepsilon_t$

Where $\theta(L)$ is a polynomial of order q in the lag operator

$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots\ldots\ldots\ldots + \theta_q L^q$.

3.1.5.4 Autoregressive Moving (ARMA) Model

An ARMA model consists according to its name of two components: the weighted sum of past values (autoregressive component) and the weighted sum of past errors (moving average component). Formally, an ARMA model of order (p, q) can be formulated as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \ldots\ldots\ldots\ldots (4)$$

An important assumption of the ARMA (p, q) model is that the time-series is stationary and so if the series is not stationary, Box and Jenkins recommend differencing the time series to achieve stationarity. Doing so produces a so-called ARIMA model, where " I " stands for integrated.

3.2 Akaike's Information Criterion (AIC)

Akaike's Information Criterion (AIC) provides a means of selecting a best fit model from a set of candidate models. The (AIC) offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. The chosen model is the one that minimizes the Kullback - Leibler distance between the model and the truth. It is based on

information theory and it's a criterion that seeks a model that has a good fit to the truth but few parameters. It is defined as:

AIC = -2 (ln (Maximum likelihood)) + 2r ……(5)

$\approx$ n ln(σ_a^2) + 2r

where n is the number of observations, r the number of parameters estimated in the model including a possible constant term and σ_a^2 is the maximum likelihood estimate of residual variance. The best model has the least AIC value.

3.2.1 Akaike's Bias Corrected Information Criterion (AICC)

This is AIC with a correction for finite samples sizes.

AICC = -2ln (Maximised Likelihood) + $\frac{2rn}{n-r-1}$ ……(6)

$\approx$ n ln(σ_a^2) + $\frac{2rn}{n-r-1}$

Where r is the number of parameters and n, the sample size. Thus, AICC is AIC with a greater penalty for extra parameters.

3.2.2 Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) proposed by Schwarz (1978) is yet another criterion, which attempts to correct for AIC's tendency to overfit. This criterion is given as

BIC = -2 ln(maximized likelihood) + r ln(n) ……....(7)

$\approx n \ln(\sigma_a^2) + r \ln(n)$

One could rewrite this as BIC $\approx n \ln(\sigma_a^2)$ + PBIC where PBIC = r ln(n) is a penalty function. As for all the criteria, the preferred model is the one with a minimum BIC.

3.2.3 Ljung- Box Test

The Ljung – Box test, is applied to the residuals after an ARIMA model has been fitted to test for randomness in the residuals. The Ljung – Box test is based on the autocorrelation plot. However, instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags. For this reason, it is often referred to as a "portmanteau" test

The Ljung – Box test can be defined as follows:

H_0 : The data are Random

H_a : The data are not Random

The test statistic is:

$Q_{LB} = (n (n + 2) \sum_{j=1}^{h} \frac{\rho 2(j)}{n-j}$

Where n is the sample size, ρ(j) is the autocorrelation at lag j, and h is the number of lags being tested

Significance level: α

Critical Region: The hypothesis of randomness is rejected if $Q_{LB} > \chi^2_{\alpha,(h-p-q)}$

Where α is taken to be 5% (0.05), h is the maximum lag being considered and p and q are respectively the order of the AR and MA processes.

3.2.4 The Augmented Dickey-Fuller Test (ADF)

Stationarity test of a differenced time series utilizes the Augmented Dickey-Fuller (ADF) technique (Dickey and Fuller (1981), which is a generalized auto-regression model formulated in the following regression equation

$$\Delta X_{i,t} = KX_{i,t-1} + \sum_{k=1}^{5} \omega i\,, k^{\Delta xi}{}_{,t-k} + \varepsilon_{k,t}$$

The Model hypothesis of interest are: The series is:

H_0 : Non-stationary

H_A : Stationary

ADF Statistics is compared to critical values to draw conclusions about stationarity.

3.2.5 Mean Squared Error (MSE)

The Mean squared Error (MSE) measures the quality of an estimator of a parameter. It thus, measures of how close a fitted line is to data points. For an observed time series data ($Y_1, Y_2, \ldots\ldots Y_N$) and a vector of N predictions ($\hat{Y}_1, \hat{Y}_2, \ldots\ldots \hat{Y}_N$), The Mean Squared Error is given by:

MSE = $\frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i - Y_i)^2$……(8)

The smaller the Mean squared Error, the closer the fit is to the data. The MSE has the units squared of whatever is plotted on the vertical axis.

3.2.6 Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE), is a measure of accuracy of a method for constructing fitted time series values. MAPE measures the size of the error in percentage terms. It is calculated by the formula:

MAPE = $\left(\frac{1}{n}\sum\frac{|Actual-Forecast|}{|Actual|}\right) * 100$ ……(9)

The value of MAPE is zero when there is a perfect fit.

3.3 THE BOX-JENKINS (ARIMA) PROCEDURE

The first requirement for a univariate Box-Jenkins modelling is that the time series data are either stationary or can be transformed into one. The Box-Jenkins modelling of a stationary time series involves the following four steps:

I. Model Identification

II. Model Estimation

III. Diagnostic checking

IV. Forecasting

3.3.1 Model Identification

The first step in the Box-Jenkins model identification is the time plot of the series. This is a visual inspection of the time series to ascertain the stationarity (if there is a seasonality or trends) in the series. If the visual inspection indicates non stationarity in the time series, a confirmation is sought by examining the autocorrelation Function (ACF) and partial autocorrelation Function (PACF) plots, known as the correlogram. A more formal approach of detecting stationarity such as the Augmented Dickey-Fuller test which has a null hypothesis that the data under consideration has a unit root and hence non stationary against an alternative that the series is stationary, could be used for an enhanced confirmation.

Box and Jenkins recommends the differencing approach to achieve stationarity. However, a curve could be fitted and the residuals from the fit modelled. The next step involves the identification of the time lags (p, q), for the AR and the MA processes and that is done by comparing the ACF and the PACF plots to a theoretical behaviour of these plots when the order is known.

3.3.2 Order of an autoregressive process (p)

In the modelling process, if the (ACF) shows a sinusoidal pattern or an exponential decay to zero, and the (PACF) has one large spike, we will choose an AR(1) model for the data. The "1" in parenthesis indicates that the AR model needs only one autoregressive term, and the model is an AR of order 1. For an AR(p) process, the (ACF) shows exponential or oscillating decay and the (PACF) become zero at lag p+1 and higher lags.

3.3.3 Order of a Moving average process (q)

If the PACF depicts an oscillating decay with the ACF displaying one large spike at lag one, we will choose an MA(1) model for the series. For an MA(q) process, the PACF shows an exponential or oscillating decay and the ACF cuts off after time lag q. An infinite damped exponentials and or damped sine waves that tails off, in both the ACF and the PACF portray an ARMA process.

However, for ARMA models picking the right orders for MA and AR components are not that straightforward and ACF and PACF offer little help except for potentially revealing that the model we should entertain is not a pure MA or AR model. There are other tools besides ACF and PACF such as extended sample autocorrelation function (ESACF), generalized sample partial autocorrelation function (GPACF), and inverse autocorrelation function (IACF) that can be of help in determining the order of the ARMA model.

3.4 Model Estimation

Once a tentative model has been identified, the next is the estimation of the model parameters. Some simple AR models are linear and can be estimated with Ordinary Least Squares (OLS) procedure. However, for models with MA components, an iterative method is used. This involves starting with a preliminary estimate, and refining the estimates iteratively until the sum of squared errors is minimized. Another method of estimating the parameters is the Maximum Likelihood procedure which is usually favored because it has some desired statistical properties.

However, the availability of statistical software packages currently, has made the estimation of model parameters and fitting of time series models require only few seconds, making it very easy to consider various models at once.

There may be more than one plausible model identified and the need to determine which of them is preferred. Here, the Akaike's Information Criterion (AIC) test is used, and the model with least AIC value is selected as the best fit model. However, occasionally it might be necessary to adopt a model with not quite the smallest AIC value but with better behaved residuals.

3.5 Diagnostic checking

Before using the model for forecasting, it must checked for adequacy, and the Ljung –Box test could be applied to the residuals. A model is adequate if the residuals left over after fitting the model is simply white noise. In other words, the residuals should be uncorrelated with constant variance. The pattern of ACF and the PACF are helpful in detecting any misspecification which will result in identifying a different and a better model.

The R^2 could be used to measure the degree of correlation between the dependent variables and the independent variables; the t-statistics to test the significance of the coefficients and the standard error to measure how closely the model fits the data.

3.6 Forecasting

The ARIMA model obtained from the differenced series W_t is given by:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \mu + \varepsilon_t$$

Since the model will be used to forecast the observed series Y_t, there is the need for a transformation from W_t to the Y_t form using the substitutions:

$$W_t = (1 - B)^d Y_t$$

Where $W_t = (1 - B)Y_t$, $W_{t-1} = (1- B)Y_{t-1}$ and $W_{t-p} = (1 - B)Y_{t-p}$

3.7 THE BAYESIAN ANALYSIS

Bayesian analysis quantifies information about an unknown parameter vector of interest, θ, for a given data set, y, through the joint posterior probability density function (pdf), $p(\theta|\mathbf{y})$, which is defined such that $p(\theta|y) \propto p(y|\theta) \times p(\theta)$, where $p(y|\theta)$ denotes the pdf of y given θ and $p(\theta)$ is the prior pdf for θ.

3.7.1 The Bayesian Dynamic Linear Models

The Dynamic Linear Models (DLM) or the State Space modelling has been used mainly in time series data analysis. It has found its application in many areas, such as economics, engineering, biology etc. The phrase 'state space' derives from a class of models developed by control engineers for systems that vary through time. When a scientist or engineer tries to measure a signal, it will typically be contaminated by noise so that;

Observation = signal + noise ……(10)

In state-space models, the signal at time t is taken to be a linear combination of a set of variables, called state variables, which constitute what is called the state vector at time t. Denoting the number of state variables by m, and the (m × 1) state vector by θ_t, (10) may be written as:

$$Y_t = F_t\theta_t + v_t \quad \ldots\ldots\ldots\ldots\ldots\ldots (11)$$

Where, F_t is assumed to be a known (m × 1) vector, and v_t denotes the observation error, assumed to have zero mean. The set of state variables may be defined as the minimum set of information from present and past data such that the future behaviour of the system is completely

determined by the present values of the state variables (and of any future inputs in the multivariate case). Thus the future is independent of past values. This means that the state vector has a property called the Markov property, in that the latest value is all that is needed to make predictions. It may not be possible to observe all (or even any of) the elements of the state vector, θ_t, directly, but it may be reasonable to make assumptions about how the state vector changes through time. A key assumption of linear state-space models is that the state vector evolves according to the equation:

$$\theta_t = G_t\theta_{t-1} + w_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

where the (m×m) matrix G_t is assumed known and w_t denotes an m-vector of disturbances having zero means. The two equations (11) and (12) constitute the general form of a univariate state-space model. The equation modelling the observed variable in (11) is called the observation (or measurement) equation, while (12) is called the transition (system /evolution) equation. The 'error' terms in the observation and transition equations are generally assumed to be uncorrelated with each other at all time periods and also to be serially uncorrelated through time. It may also be assumed that v_t is $N(0, \sigma^2)$ while w_t is multivariate normal with zero mean vector and known variance-covariance matrix W_t. If the latter is the zero matrix, then the model reduces to time varying regression.

A Dynamic Linear Model (State Space Model), is characterized by an initial Normal prior distribution for the parameter vector, $\theta_0 \sim N(m_0 ; C_0)$,where m_0 and C_0 are the mean and variance, respectively, and the dynamic set of four matrices $\{F_t; G_t; V_t; W_t\}$, that for each time

$t \geq 1$ are known matrices of appropriate dimensions. The set $\{F_t; G_t; V_t; W_t\}$ defines the model relating the observation vector Y_t to the state vector θ_t at time t, and the θ_t sequence through time

by satisfying the equations (11) and (12) above. Furthermore, it is assumed that θ_t is independent of both (V_t) and (W_t), the independent noise sequences. From equation (12) it is easy to see that, given the known matrices G_t and W_t , θ_t depends only on the previous state θ_{t-1} and not on earlier information. From equation (11) it is clear that, conditionally on (θ_t), the Y_t's are independent and Y_t depends on θ_t only. The DLM is completely specified by the conditional densities $Y_t|\theta_t \sim N(F_t\theta_t; V_t)$ and $\theta_t|\theta_{t-1} \sim N(G_t\theta_{t-1}; W_t)$ combined with an initial prior distribution. The state space model provides rich covariance structures for the observations Y_t. It indeed covers the covariance structure of the ARMA (Autoregressive Moving Average) model as the latter can be written in the state space form. Another interesting aspect of the state space model is its flexibility in modelling the underline mechanism (state equation) and the observations (observation equation) separately.

Having expressed a model in state-space form, an updating procedure can readily be invoked every time a new observation becomes available, to compute estimates of the current state vector and produce forecasts. This procedure, called the Kalman filter, only requires knowledge of the

most recent state vector and the value of the latest observation. If the matrices F_t and G_t are constant for all t, the model is referred to as time series DLM (TSDLM). A TSDLM with constant variance matrices V_t and W_t for all t is called a constant DLM. A constant DLM is characterized by a single set of matrices {F; G; V; W} for all times t, and this special case of DLMs includes essentially all classical linear time series models.

One interesting example of constant DLMs, such as the random walk plus noise, and also known as local level or steady model, arises when θ_t is a scalar, μ_t, denoting the current level of the process, while F_t and G_t are constant scalars taking the value one. Then the local level, μ_t, follows

a random walk model and depends on two parameters which are the two error variances, namely V_t and $Var(w_t) = W_t$. This model is used effectively in numerous applications, particularly in short-term forecasting for production planning and stock control.

3.7.2 A Parameter Prior Distribution

A prior, as in the Bayesian Inference is one's initial probability statement about the parameter under study. Thus, prior subject-matter knowledge about a parameter is an important aspect of the inference process since the Bayesian models are typically concerned with inferences on the parameter set $\theta = (\theta_1, \ldots\ldots\ldots\theta_d)$ of dimension d , that includes uncertain quantities, whether fixed and random effects, hierarchical parameters, unobserved indicator variables and missing data. Thus, it represents all the available relevant starting information that is used to form initial views about the future, including history and all defining model quantities. In the Bayesian inference, a prior amounts to a form of modelling assumption or hypothesis about the nature of the parameters and it's often summarised by the density $p(\theta)$.

In many situations, existing knowledge may be difficult to summarize or elicit in the form of an 'informative prior', and to reflect such essentially prior ignorance, resort is made to non-informative priors. How to choose the prior density or information is an important issue in Bayesian inference, together with the sensitivity or robustness of the inferences to the choice of prior, and the possibility of conflict between a prior and data. In some situations it may be possible to base the prior density for θ on cumulative evidence using a formal or informal meta-analysis of existing studies. A range of other methods exist to determine or elicit subjective priors. A simple technique known as the histogram method, divides the range of θ into a set of

intervals (or 'bins') and elicits prior probabilities that θ is located in each interval; from this set of probabilities, p(θ) may be represented as a discrete prior or converted to a smooth density. Another technique uses prior estimates of moments along with symmetry assumptions to derive a normal N(m, V) prior density including estimates m and V of the mean and variance.

3.7.3 Role of the Bayesian Priors

The role of the prior is to capture knowledge about a parameter θ, denoted D_0, which existed prior, in time order, to consideration of a new empirical data D_1. This can be acknowledged by denoting the prior as $p(\theta | D_0)$ instead of $p(\theta)$ and thus, the posterior of a study depends on both D_0 and D_1. Suppose a researcher conducts a further study and obtains a new empirical data D_2, then the information about θ prior to that study is shown in the previous posterior $p(\theta | D_0, D_1)$. Hence this defines a sequence of posterior analyses, each informing the next:

$$p(\theta|D_1,D_0) \propto Lik(D_1|\theta)p(\theta|D_0)$$

and $p(\theta|D_1,D_2,D_0) \propto Lik(D_2|\theta)p(\theta|D_1,D_0)$. This embodies the Bayesian cycle of learning, where the researcher's understanding of parameters $p(\theta|\cdot)$ is continually updated, and we explicitly state that inference is predicated on specific sets of data D_0, D_1, etc. Hence the formulation of priors within this Bayesian cycle of learning provides a flexible basis for accumulating learning across several sources of information.

3.7.4 A Posterior Distribution

To estimate the quantity θ, from a set of measurements of the quantity, y, Bayesian estimation starts by defining the conditional density of the variable to be estimated given the measurements, p(θ | y), which is called the posterior. The posterior is a density function that describes the behavior of the quantity, θ, after observing the measurements. Using Bayes rule, the posterior can be written as follows:

$$P(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad \text{.....(13)}$$

The first term in the numerator of Equation (13) denotes the likelihood function, which is the conditional density of the observations given the true value of θ. According to the likelihood principle (LP), the likelihood function contains all information brought by the observations, y_t about the quantity, θ. The second term in the numerator is the prior, which is the density function of the quantity θ. It is called a prior since it quantifies our belief or knowledge about θ, before observing the measurements. Through the prior, external knowledge about the quantity θ can be incorporated into the estimation problem. Finally, the denominator term is the density function of the observation, which can be assumed constant after observing the data.

The posterior density can be written as,

$p(\theta|y) \propto p(y|\theta)p(\theta)$, or

Posterior ∝ Likelihood × Prior,

which is sometimes referred to as the unnormalized posterior. Thus, the posterior combines the data information and any external information. Having constructed the posterior, a sample from

it is selected as the final Bayesian estimate of the quantity θ. In contrast to non-Bayesian or frequentist approaches, which rely only on the data for inference, Bayesian approaches combine the information brought by the data and any external knowledge represented by the prior to provide improved estimates.

3.7.5 Inference in Dynamic Linear Modelling

The inference in DLMs follows the usual steps in Bayesian inference. It explores the sequential aspects of Bayesian inference, combining two main operations: evolution, to build up a prior and updating, to incorporate a new observation arrived at time t. Let $D_t = \{D_{t-1}, y_t\}$ denote the information until time t , including the values of θ_t and G_t , for every t , which are supposed to be known, with D_0 representing the prior information. Then for each time t, the prior, predictive and posterior distribution are respectively given by:

$$P(\theta_t | D_{t-1}) = \int P(\theta_t | \theta_{t-1}) P(\theta_{t-1} | D_{t-1})\, d\theta_{t-1},$$

$$P(y_t | D_{t-1}) = \int P(y_t | \theta_t) P(\theta_t | D_{t-1})\, d\theta_t, \quad \text{and}$$

$$p(\theta_t | D_t) \propto p(\theta_t | D_{t-1}) p(y_t | D_{t-1}),$$

where the last one is obtained via Bayes theorem. The constant of integration in the above specification is sometimes easily obtained. This is just the case when (F, G, V, W)t are all known and normality is assumed.

Suppose interest lies in a scalar series Y_t (which could be multivariate) and that at time t−1 the current information set is D_{t-1}. The first step in the Bayesian approach is to examine the

forecasting context and to select a meaningful parametrisation, θ_{t-1}, such that all the historical information relevant to predicting future observations is contained in the information about θ_{t-1}. In particular the modeller represents this relevant information in terms of the probability distribution $(\theta_{t-1} \mid D_{t-1})$. The parameter together with this probability distribution defines how the modeller views the context at time $t - 1$. The next modelling step is that of relating the current information to the future so that predictive distributions such as $(Y_{t+k} \mid D_{t.-1})$ can be derived. This is accomplished by specifying a sequential parametric relation $(\theta_t \mid \theta_{t-1}, D_{t-1})$ together with an observation relation $(Y_t \mid \theta_t, D_{t-1})$. In combination with $(\theta_{t-1} \mid D_{t-1})$ these distributions enable the derivation of full joint forecast distribution.

If the posterior for θ_{t-1}, given data observed to time $t - 1$, is

$$\theta_{t-1}|D_{t-1} \sim N(m_{t-1}, C_{t-1}).$$

Then the prior for the next state θ_t given D_{t-1} operates via $\theta_t = G_t\theta_{t-1} + w_t$ and includes extra uncertainty from the state errors w_t , namely,

$$\theta_t |D_{t-1} \sim N(G_t m_{t-1}, G_t C_{t-1} G'_t + W_t).$$

A prediction for the next value of y_t given D_{t-1} can then be made, operating via $y_t = F_t\, \theta_t + v_t$, namely

$$y_{new,t} |D_{t-1} \sim N(F_t G_t m_{t-1}, F_t\, R_t\, F'_t + V_t),$$

where $R_t = G_t C_{t-1} G'_t + W_t$. The posterior for θ_t, given an extra observation to form $D_t = (y_t, D_{t-1})$, includes forecast error $e_t = y_t - F_t G_t m_{t-1}$. Writing $Q_t = F_t\, R_t\, F'_t + V_t$, one obtains

$$\theta_t |D_t \sim N(m_t , C_t)$$

where

$m_t = m_{t-1} + A_t\, e_t$

$C_t = R_t V_t Q_t^{-1}$,

$A_t = F_t\, R_t\, Q_t^{-1}$

This posterior is then used to provide the next prior $(\theta_{t+1}|Dt)$, and so the cycle repeats; at any stage one can subjectively interact with the prior to produce alterations in the one step-ahead forecast for Y_t, in the light of any relevant information that may have arisen. The only prerequisite of the system is that, it is started by defining initial priors m_0 and C_0 such that

$(\theta_0|D_0) \sim N(m_0, C_0)$. These are chosen purely on the basis of the initial available information D_0 which may or may not include some data already, and will - almost by definition - usually include the subjective opinions of the practitioner on the nature of the data evolution.

3.8 The constant DLM Model

The observation equation is expressed as:

$Y_t = \mu_t + \nu_t \quad , \quad \nu_t \sim N(0, V)$ ……..…… (14)

And the system equation is expressed as:

$\mu_t = \mu_{t-1} + \omega_t \quad , \quad \omega_t \sim N(0, W)$ ……..….....(15)

Initial information:

$(\mu_0 | D_0) \sim N [m_0, C_0]$,

3.8.1 The State Space Model Initial Information

The Initial information $(\mu_0|D_0)$ is the probabilistic representation of the forecaster's beliefs about the level μ_0 at time t = 0. The mean m_0 is a point estimate of this level, and the variance C_0 measures the associated uncertainty. Each information set Dv comprises all the information available at time v, including D_0, the values of the variances $\{V_t, W_t : t > 0\}$, and the values of the observations $Y_v, Y_{v-1}, \ldots, Y_1$. Thus, the only new information becoming available at any time t is the observed value Y_t, so that $D_t = \{Y_t, D_{t-1}\}$.

3.8.2 UPDATING OF A PRIOR TO A POSTERIOR AND THE ONE STEP AHEAD FORECAST DISTRIBUTION IN THE CONSTANT DLM MODEL

The updating of a prior to a posterior distribution and the one-step-ahead forecast distribution is illustrated in the following.

Observation Equation:

$$Y_t = \mu_t + \nu_t \quad , \qquad \nu_t \sim N(0 , V) \quad \ldots\ldots (14)$$

and the system equation:

$$\mu_t = \mu_{t-1} + \omega_t \quad , \qquad \omega_t \sim N(0 , W) \quad \ldots\ldots (15)$$

Let the posterior distribution for μ_{t-1} at time step (t - 1) be $P(\mu_{t-1} / D_{t-1}) \sim N[m_{t-1},C_{t-1}]$ with some mean m_{t-1} and variance C_{t-1}. From Equation (15), it is seen that μ_t is the summation of two normally distributed random variables: $N[m_{t-1},C_{t-1}]$ and $N[0,W]$. Assuming that the error term is independent of the level of the process, the summation will be another normally distributed random variable with mean m_{t-1} and variance $(C_{t-1} + W)$. Thus, the prior distribution of μ_t for the time step t is $P(\mu_t/D_{t-1}) \sim N[m_{t-1}, R_t]$, where, $R_t = C_{t-1} + W$. In a similar way, from equation (14), Y_t is the summation of $N[m_{t-1}, Rt]$ and $N[0,V]$, which is also normally distributed with mean m_{t-1} and variance $(R_t + V)$. Thus, the one-step- ahead forecast distribution is $P(Y_t/D_{t-1}) \sim N[f_{t-1},Q_t]$, where $f_t = m_{t-1}$ and $Q_t = R_t + V$. It can be noted that, until now, information up to time step

(t -1) is available, which is denoted D_{t-1}. At the end of the time step t, the observed value of Y_t (denoted as y_t), for this time step, is available. Thus, the available information is improved, and becomes D_t (consisting of D_{t-1} and y_t). Thus finally, $P(\mu_t/D_t) \sim N[m_t, C_t]$, where $m_t = m_{t-1} + A_t e_t$ and $C_t = A_t V$

3.9 THE GENERAL RANDOM WALK MODELS

A random walk is a special case of an AR (1) model with $\phi=1$, and a classic example of a non-stationary stochastic process. The random walk implies that the value of Y at time t is equal to its value at time (t – 1) plus a random shock μ_t. The sequence $\{\mu_t\}$ is white noise error terms with mean zero and variance σ^2. Y_t is the value of the series at time t. Asset prices such as stock prices or exchange rates follow a random walk and thus are non- stationary (Gujarati, 2004). There are mainly two types of random walk models namely; random walk with a drift (i.e. a constant term is present) and random walk without a drift (i.e. no constant or intercept term)

3.9.1 RANDOM WALK MODEL WITH DRIFT (CONSTANT)

Random walk with a drift is a special form of an AR (1) model:

$Y_t = \alpha + \phi Y_{t-1} + \varepsilon_t$, where $\alpha \neq 0$ and $\phi = 1$

This can be written as:

$Y_t = \alpha + Y_{t-1} + \varepsilon_t$ ..... (16)

where α is the drift parameter.

Equation (16) can be written as:

$Y_t - Y_{t-1} = \Delta Y_t = \alpha + \varepsilon_t$ ..... (17)

It shows that Y_t drifts upwards or downwards, depending on α being positive or negative. The expected value of the series at time t is:

$E(Y_t) = t\alpha + Y_0$ ..... (18)

It's variance at time t is:

$Var(Y_t) = t\sigma^2$ ..... (19)

Thus the mean and variance of the random walk with a drift increases with time t, again violating the conditions of weak stationarity. In short a random walk model with or without drift parameter is a non- stationary process.

1.9.2 RANDOM WALK MODEL WITH NO DRIFT PARAMETER

A time series (Y_t) is a random walk with no drift if it satisfies the following equation

$Y_t = Y_{t-1} + \varepsilon_t$ ..... (20)

A random walk model with no drift is also an AR (1) model with $\alpha = 0$ and $\phi = 1$. It is easy to see that $Y_t = Y_0 + \Sigma\varepsilon_t$ and $E(Y_t) = E(Y_0 + \Sigma\varepsilon_t) = Y_0$, $Var(Y_t) = t\sigma^2$.

The mean of Y_t is constant for a random walk without drift but the variance increases with time t. The increasing variance violates a condition of weak stationarity. Thus, the random walk model without drift is a non- stationary stochastic process. A random walk model remembers the shocks δ (random errors) forever and is said to have an infinite memory (Gujarati, 2004).

Equation (20) can be written as:

$(Y_t - Y_{t-1}) = \Delta Y_t = \varepsilon_t$ ..... (21)

Thus, while Y_t is non-stationary, its first difference is stationary since ε_t is a stationary white noise process i.e. $\varepsilon_t \sim N(0, \sigma^2)$. In other words, the first difference of a random walk time series is stationary.

CHAPTER FOUR

DATA COLLECTION, ANALYSIS AND RESULTS

4.0 INTRODUCTION

This Chapter deals with the analysis of the time series data of annual under-five mortality rates for Ghana from 1961 to 2012, which was obtained from the website of the World Bank. It also contains the procedures the researcher used in the analysis of the data.

4.1 Time Series Plot of the Data

This is a plot of the annual estimates of under-five mortality rates for Ghana, otherwise known as observations and displayed as ordinates against equally spaced time intervals as abscissa, which is used to evaluate the pattern and behavior in the data over time. A visual inspection of the plot shows a downward trend, which may be indicating non-stationarity in the series data.

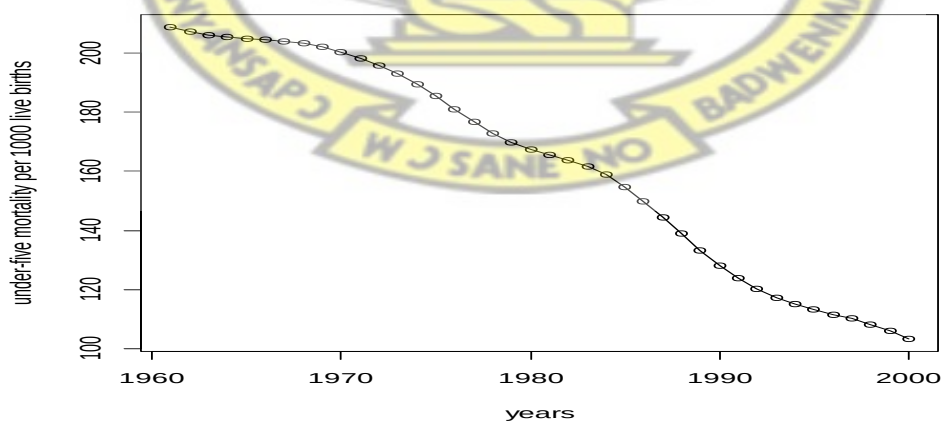


Figure 4.1: A time plot of the series (Under-five Mortality Rates for Ghana) 1961 to 2000

A confirmation is sought by plotting the correlogram for the time series and shown in the figure below.

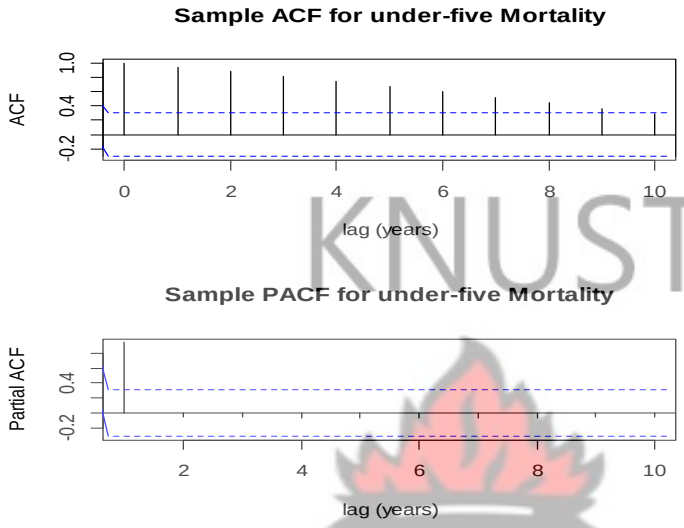


Figure4.2 Correlogram for the Under-Five Mortality Rates for Ghana

The figure 4.2 above (correlogram) displays graphically and numerically the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for the time series. The figure shows large significant ACF for the time lags which gradually decreases in size, but do not decay to zero (slow decay). The ACF thus shows a pattern typical of a non-stationary time series.

In the PACF plot, the partial autocorrelation at time lag 1 is close to one and the partial autocorrelations for the time lag 2 through 10 are close to zero which is also typical of non-stationary series. Researcher transforms the series by $W_t = \Delta y_t = y_t - y_t$, in order to stabilize the variance before proceeding with the model building. The ACF plot for the first differenced series, fig 4.3 (see appendix) dies out relatively quickly which may be indicating a stationary process.

In addition to the observed flat level of the first differenced series fig 4.4 (appendix), its stationarity is confirmed by the KPSS and ADF tests. The KPSS test statistic with a p-value of 0.351 which is greater than the 5% level of significance does not reject null hypothesis that the first differenced series is level stationary. The ADF test also produced a p-value of 0.02625 which is less than the 5% significance level and thus rejects the null hypothesis that the first differenced series is non stationary, hence confirming the stationarity of the first differenced series of the Under-five Mortality Rates for Ghana.

4.2 Model Identification

The autocorrelation function (ACF) for the differenced data (see appendix) diminishes quickly indicating an autoregressive (AR) model. Tentatively, researcher fits different possible models and uses equations (5), (6) and (7), the Akaike Information Criterion (AIC), the Akaike bias corrected Information Criterion with a correction (AICC) and the Bayesian Information Criterion (BIC) to select the best fit model.

Table 4.1: Possible fitted Models

MODELS	AIC	AICC	BIC
ARIMA(1,1,1)	58.47	59.15	63.46
ARIMA(2,1,1)	46.89	48.06	53.54
ARIMA(2,1,2)	39.04	40.86	47.36
ARIMA(3,1,0)	46.21	47.38	52.86
ARIMA(3,1,1)	44.28	46.1	52.6
ARIMA(3,1,2)	38.91	41.53	48.89

A comparison of the AIC, AICC and the BIC values for the possible models shows that ARIMA (3, 1, 2) has the least AIC value. However, the ARIMA (2, 1, 2) produced least values for both the AICC and the BIC and is therefore selected as the best fit model for the data.

Table 4.2: Estimation Summary for the ARIMA (2,1,2) Model

Model Term/Coefficient	Estimate	Standard Error
AR 1: $\hat{\phi}_1$	1.5614	0.2422
AR 2: $\hat{\phi}_2$	-0.5950	0.2394
MA1: $\hat{\theta}_1$	0.0550	0.3355
MA2: $\hat{\theta}_2$	0.4829	0.1483

4.2.1 The ARIMA Model

The ARIMA (2, 1, 2) Model for the Series is:

$$Y_t = 2.5614Y_{t-1} - 2.1564Y_{t-2} + 0.5950Y_{t-3} - 0.0550\varepsilon_{t-1} - 0.4829\varepsilon_{t-2}$$

The model adequacy is further checked to draw empirical conclusions regarding the model as a good fit and for its use for forecasting the time series. These tests are performed using the Ljung-Box test in addition to the ACF plots of the residuals

4.2.2 Residual Diagnostics of the ARIMA (2, 1, 2) Model

A diagnostics of the residuals by the ACF shows that the ACF values are all within the 5% zero-bound - indicating that there is no correlation amongst the residuals. This plot is used as an indicator of the independence of the residual terms.

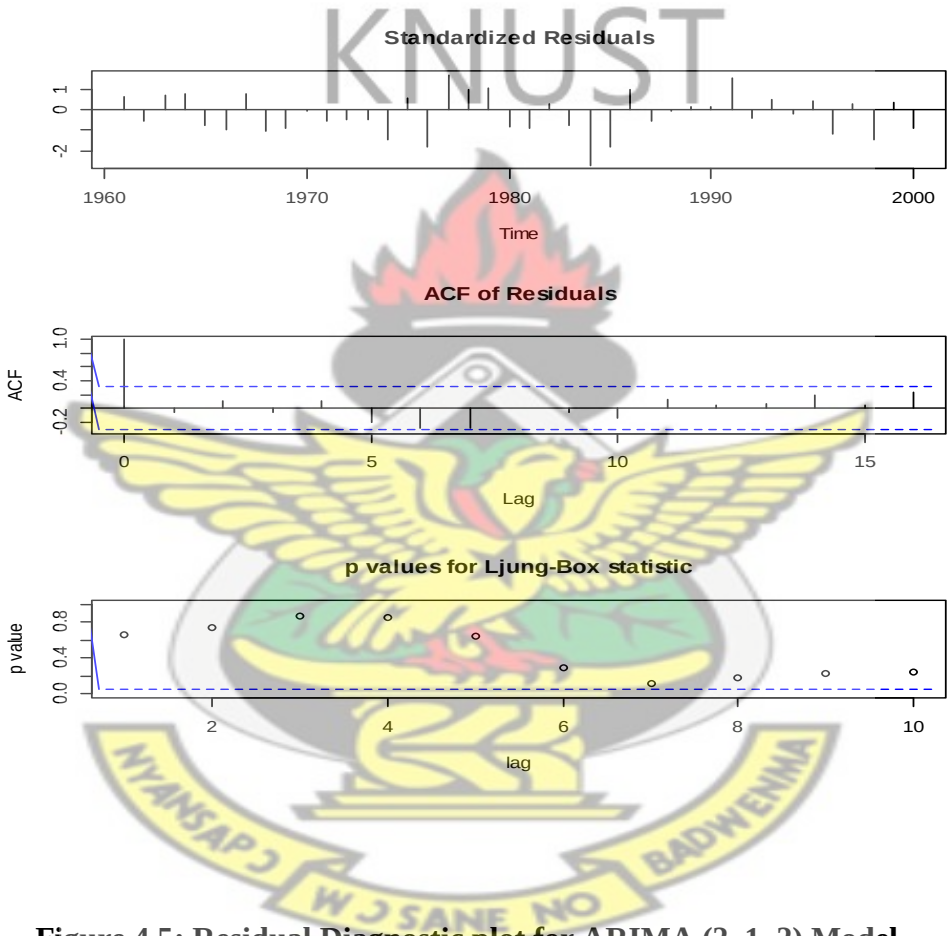


Figure 4.5: Residual Diagnostic plot for ARIMA (2, 1, 2) Model

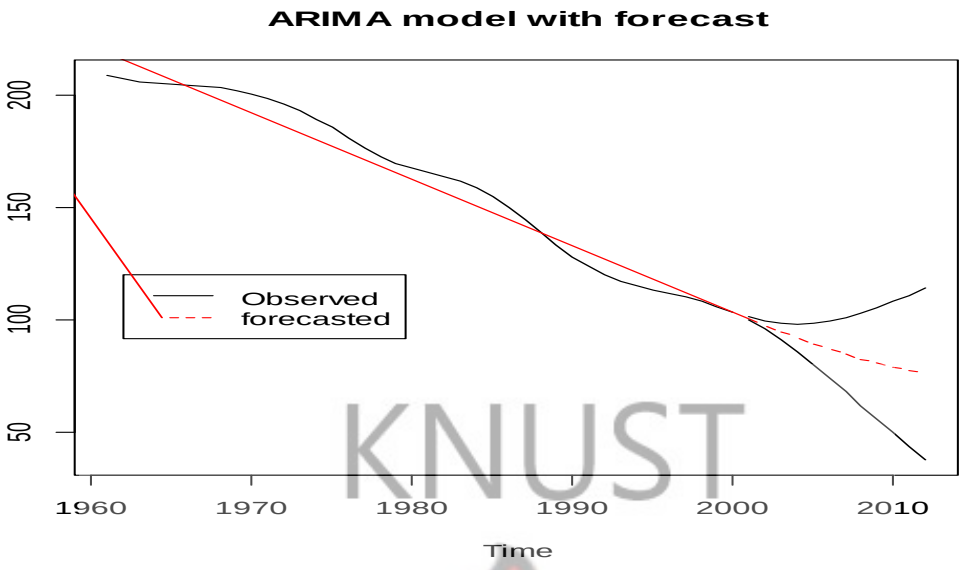


Figure 4.6: Plot of the observed series (1961 – 2000) and the in-sample forecast values

Figure 4.6 above shows the plot of the (observed) time series data from 1961 to 2000, which was used for the modelling. From the figure, it can be observed that Under-five mortality rates in Ghana were high in 1961 (i.e. 208.7 deaths per 1,000 live births), which decreased gradually over time to about 103.4 per 1,000 live births in 2000. The portion of the figure from the year 2001 to 2012 shows the in - sample forecast values by the ARIMA (2,1,2) Model produced from the data with the 95% Confidence Interval for each of the forecast value, depicted by the two opening lines.

Table 4.3: A 95% Confidence Interval for the In- Sample forecast Values by the ARIMA (2,1,2) Model

Year	Observed Value	Forecast Value	Confidence Interval
2001	100.6	100.6	100 – 101.3
2002	97.5	97.7	95.9 – 99.5
2003	94.3	94.8	91.1 – 98.5
2004	91.3	92	85.8 – 98.2
2005	88.4	89.4	80.1 – 98.6
2006	85.7	86.9	74.1 – 99.7
2007	83.2	84.6	68.1 – 101.2
2008	80.7	82.5	61.9 – 103.2
2009	78.8	80.6	55.8 – 105.5
2010	76.4	78.9	49.8 – 108.1
2011	74.2	77.4	43.8 – 110.9
2012	72	75.9	37.9 – 114

4.3 Analysis of the Constant DLM Model

In the constant (DLM) model, the model parameters are updated by utilizing the observed values, at each time step. However, the assumption of normality is the basic assumption of this procedure. With a first data (observed) value of 208.7, a mean value $m_0 = 205$ which is closer to the first data value is chosen. However, as the precision of this value is not known, a high value of initial variance $C_0 = 10000$ is chosen. Considering these points, at time step 0 (1961) m_0 is

assumed to be 205 and variance C_0 as 10000. Thus, the initial information on model parameter (μ_0/D_0) also known as the forecasters initial beliefs is $(\mu_0/D_0) \sim N(205, 10000)$.

To specify the observational and evolution variances, the performance of the model was observed by some trial values (Kumar, 2008), where 2000 for the observation and 5000 for the evolution were found suitable, i.e. V= 2000 and W=5000.

4.4 THE SEQUENTIAL UPDATING OF THE TIME SERIES BY THE CONSTANT DLM MODEL WITH V = 2000 AND W = 5000

At time t = 1 (i.e. 1961),

(a) Posterior for μ_0 :

$(\mu_0 / D_0) \sim N(205 , 10000)$

(b) Prior for μ_1:

$(\mu_1 / D_0) \sim N(205 + 0 , 10000 + 5000)$

$\Rightarrow \quad (\mu_1 / D_0) \sim N(205 , 15000)$

(c) 1-step forecast:

$(Y_1 / D_0) \sim N(205 + 0 , 15000 + 2000)$

$\Rightarrow \quad (Y_1 / D_0) \sim N(205 , 17000)$

(d) Posterior for μ_1:

$(Y_1 / D_1) \sim N(208.3 , 1760)$

At time t = 2 (i.e. 1962)

Prior for μ_2:

$(\mu_2 / D_1) \sim N(208.3 + 0 , 1760 + 5000)$

$\Rightarrow (\mu_2 / D_1) \sim N(208.3 , 6760)$

1-step forecast:

$(Y_2 / D_1) \sim N(208.3 + 0 , 6760 + 2000)$

$\Rightarrow (Y_2 / D_1) \sim N(208.3 , 8760)$

Posterior for μ_2:

$(Y_2 / D_2) \sim N(207.4 , 1540)$

At time t= 3 (i.e. 1963)

Prior for μ_3:

$(\mu_3 / D_2) \sim N(207.4 + 0 , 1540 + 5000)$

$\Rightarrow (\mu_3 / D_2) \sim N(207.4 , 6540)$

1- step forecast:

$(Y_3 / D_2) \sim N(207.4 + 0 , 6540 + 2000)$

$\Rightarrow \quad (Y_3 / D_2) \sim N(207.4 , 8540)$

Posterior for μ_3:

$(Y_3 / D_3) \sim N(206.2 , 1540)$

At time t=4 (i.e. 1964)

Prior for μ_4:

$(\mu_4 / D_3) \sim N(206.2 + 0 , 1540 + 5000)$

$\Rightarrow \quad (\mu_4 / D_3) \sim N(206.2 , 6540)$

1- step forecast:

$(Y_4 / D_3) \sim N(206.2 + 0 , 6540 + 2000)$

$\Rightarrow \quad (Y_4 / D_3) \sim N(206.2 , 8540)$

Posterior for μ_4:

$P(Y_4 / D_4) \sim N(205.5 , 1540)$

Figure 4.5 (appendix) shows the combined plot of the data and the performance of the DLM model. A suitable DLM model for the series (Under-five Mortality Rates for Ghana) is:

{1, 1, 2000, 5000}

The in-sample forecast values by the DLM Model was a constant value of 104.1 deaths per 1,000 live births for the years 2001 to 2012. The constant in-sample forecast value (104.1 deaths per 1,000 live births), which is different from the observed values of the data for that specified period, disqualifies the DLM as a suitable model for the four years ahead forecast of the Under-five Mortality rates for Ghana.

4.5 The Random Walk with Drift Model Analysis

Due to the non-stationarity in the time series data (Under-five Mortality Rates for Ghana), it is just as good to predict the change that occurs from one period to the next i.e. the quantity $Y_t - Y_{t-1}$, as to directly predict the level Y_t of the series at each period. This is because the predicted change can always be added to the current level to yield a predicted level.

$\Rightarrow$ 1^{st} difference series: $Y_t - Y_{t-1} = \alpha$,

Where α is the mean of the first differences known as the drift parameter.

$\Rightarrow$ $Y_t = \alpha + Y_{t-1}$

The R statistical software was used for the analysis of this simple model, and below is the output from the analysis.

The drift parameter $\alpha = -2.7$ with Standard Error (se) = 1.514839

Thus, the Random Walk with drift model is:

$$Y_t = -2.7 + Y_{t-1}$$

Table 4.5: A 95% Confidence Interval for the In- Sample forecast Values by the Random Walk with Drift Model

Year	Observed value	Forecast Value	Confidence Interval
2001	100.6	100.7	100.2 – 101.2
2002	97.5	98.0	97.5 – 98.5
2003	94.3	95.3	94.8 – 95.8
2004	91.3	92.6	92.1 – 93.1
2005	88.4	89.9	89.4 – 90.4
2006	85.7	87.2	86.7 – 87.7
2007	83.2	84.5	84.0 – 85.0
2008	80.7	81.8	81.3 -82.3
2009	78.8	79.1	78.6 -79.6
2010	76.4	76.4	75.9 – 76.9
2011	74.2	73.7	73.2 – 74.2
2012	72	71.0	70.5 – 71.5

4.6 Forecast Assessment of the Random Walk and the ARIMA (2, 1, 2) Model

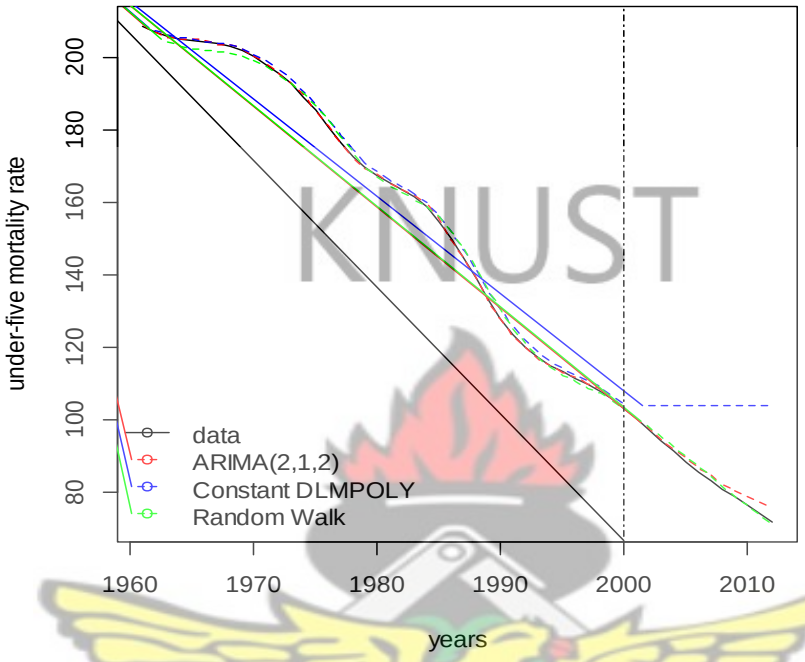


Figure 4.8: A combined plot of the time series and performance of each of the model

Figure 4.8 shows a combined plot of the observed time series and the three models; the ARIMA (2,1,2), the Dynamic Linear (DLM) and the Random Walk with drift Models. The performance of each model with respect to the decline in the Under-five Mortality rates from 1961 to year 2000 is shown in the figure. The portion of the plots from year 2001 to 2012, shows the in-sample forecast values by each of the models displayed on tables 4.3 and 4.5, which was used to select the best among the three models. The horizontal line shown from 2001 to 2012 in the fig 4.8 above, depicts the constant in-sample forecast value of 104.1 deaths per 1,000 live births by the DLM Model, also shown in table 4.4 (appendix).

4.7 DISCUSSION OF RESULTS

After the in-sample forecast by each of the models, a forecast assessment was made to determine the best fit model among the three models. From fig 4.8, it is observed that the constant DLM model which mainly produces a short term forecast produced constant forecast values of 104.1 deaths per 1,000 live births for the in-sample forecast period of 2001 to 2012. This constant value of (104.1 deaths per 1,000 live births) is very different from the (observed) data values for that period. Hence, that disqualifies the constant DLM as a candidate model for the out-of sample forecast. Also, the visual inspection of each of the models shows that the in-sample forecast values of the Random Walk with drift model lie very close to those of the observed data values. A determination was therefore made between the ARIMA (2,1,2) and the Random Walk with drift model, using the Mean Squared Error (MSE) and the Mean Absolute Percentage Error (MAPE) statistics, the results shown on table 4.6 below.

Table 4.6: Results of the In - Sample Forecast Analysis.

Statistic	**Arima (2,1,2)**	**Random Walk**
MSE	3.6478	0.9742
MAPE	0.0915	0.0099

The results of the forecast assessment for the two models on table 4.6 show that, the Random walk with drift model produced comparatively lower values for both the MSE and the MAPE statistic. These indicate that the in-sample forecast values by the Random Walk with drift model has lower deviations from the data values for that specified period. As a result, the Random Walk

with drift model is selected the best fit model for the under – five mortality rates for Ghana, and it's used to make a four years out-of sample forecasting for the 2013 to 2016, producing respectively 69.3, 66.6, 64.0 and 61.3 deaths per 1,000 live births, shown on table 4.7 below.

Table 4.7: 95% Confidence interval for the four year out-of sample forecast values

Year	Forecast	Confidence Interval
2013	69.3	69.0 – 69.7
2014	66.6	66.3 – 67.0
2015	64.0	63.6 – 64.3
2016	61.3	60.9 – 61.6

A forecast value of 64 deaths per 1,000 live births for 2015 with a 95% confidence interval (63.6– 64.3) is an evidence that Ghana may not be able to realize her MDG4 target of reducing her under-five mortality rates to about 42.7 deaths per 1,000 live births by 2015.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.0 INTRODUCTION

This final chapter of the study is based on findings from the analysis of the data by the various modelling procedures. It thus, deals with the conclusions drawn from the study and the recommendations made.

5.1 CONCLUSIONS

Based on the objectives and the analysis of the times series data, one can draw the following conclusions:

In sections 4.2.1, 4.4 and 4.5, the ARIMA (2,1,2), the Bayesian Dynamic Linear (DLM) and the Random Walk with drift models have been specified after the analysis of the time series by each procedure. In view of this, researcher concludes that the first objective of modeling the under-five mortality rates for Ghana by the stated methods has been achieved.

Secondly, among the three models, the Random Walk with drift model is selected the best fit model for the Under-five mortality Rates for Ghana and the four years out-of-sample forecast values for the Under-five mortality rates for Ghana; 69.3, 66.6, 64.0 and 61.3 deaths per 1,000 live births, respectively for the years 2013 – 2016 also shows a decline.

Finally, researcher concludes that the forecast value of 64.0 deaths per 1,000 live births with a 95% confidence interval of (63.3 – 64.3) for year 2015, shows that Ghana may not realize her MDG4 target of attaining an Under-five mortality rate of about 42.7 deaths per 1,000 live births per the Worldbank data.

5.2 RECOMMENDATIONS

Based on the analysis and conclusions drawn from the study, the following recommendations are worth considering:

The decline in the forecasted values of the Under-five Mortality Rates for Ghana by the Random Walk Model, shows that the government's policies and strategies towards the realization of the MDG4 goal are actually working positively towards the goal and would recommend that the government continues with those policies and strategies. However, the extent of the decline based on the model, may not be able to achieve the MDG4 target value at the specified time and would recommend also, that the government of Ghana puts in more efforts and resources to facilitate the realization of the MDG4 target value.

Appendix

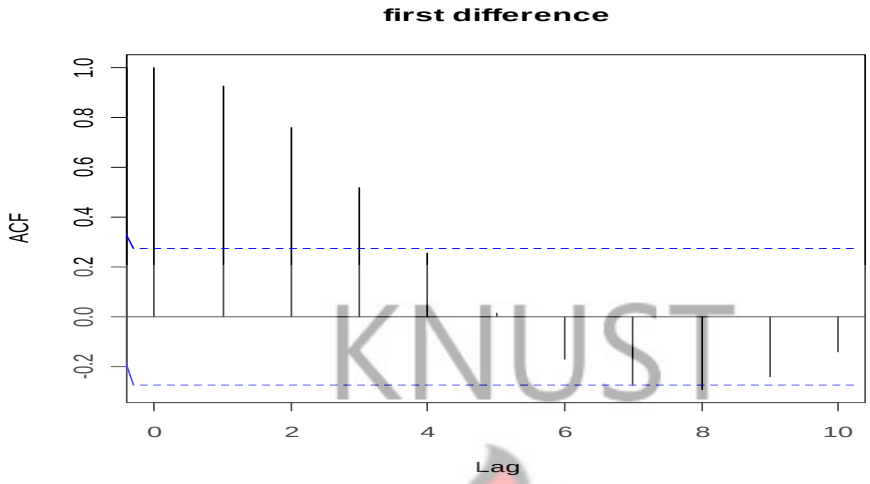


Figure 4.3: ACF of the first differenced series of Under-five Mortality Rates for Ghana

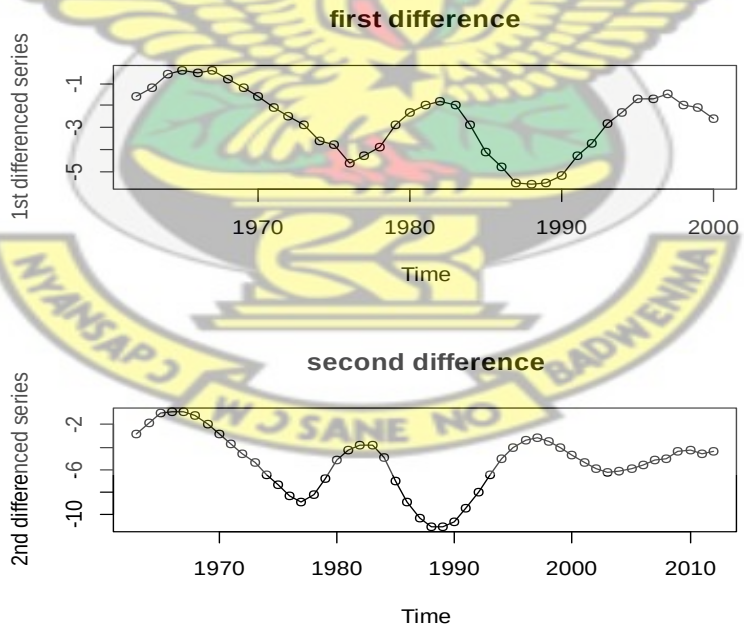


Figure 4.4: Plot of the first & second differenced series of the Under-five Mortality Rate

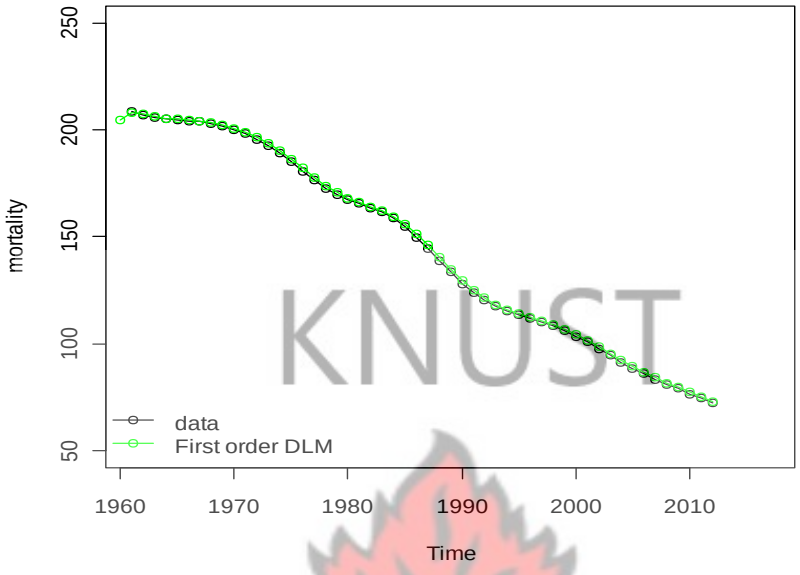


Figure 4.7: A time plot of the observed series and the updated (DLM) Model Values

Table 4.4: Updated values for the time series (Under – Five Mortality Rates for Ghana from (1961-2013) by the Dynamic Linear (DLM) Model

Time	**Forecast Distribution**		**Adaptive Coefficient**	**Observations**	**Error**	**Posterior Information**	
Year	f_t	Q_t	A_t	Y_t	e_t	m_t	C_t
1960						**205**	**10000**
1961	205	17000	0.88	208.7	3.7	**208.3**	1760
1962	208.3	8760	0.77	207.1	-1.2	**207.4**	1540
1963	207.4	8540	0.77	205.9	-1.5	**206.2**	1540
1964	206.2	8540	0.77	205.3	-0.9	**205.5**	1540
1965	205.5	8540	0.77	204.9	-0.6	**205**	1540

1966	205	8540	0.77	204.4	-0.6	**204.5**	1540
1967	204.5	8540	0.77	204	-0.5	**204.1**	1540
1968	204.1	8540	0.77	203.2	-0.9	**203.4**	1540
1969	203.4	8540	0.77	202	-1.4	**202.3**	1540
1970	202.3	8540	0.77	200.4	-1.9	**200.8**	1540
1971	200.8	8540	0.77	198.3	-2.5	**198.8**	1540
1972	198.8	8540	0.77	195.8	-3	**196.5**	1540
1973	196.5	8540	0.77	192.9	-3.6	**193.7**	1540
1974	193.7	8540	0.77	189.3	-4.4	**190.3**	1540
1975	190.3	8540	0.77	185.5	-4.8	**186.6**	1540
1976	186.6	8540	0.77	180.9	-5.7	**182.2**	1540
1977	182.2	8540	0.77	176.6	-5.6	**177.9**	1540
1978	177.9	8540	0.77	172.7	-5.2	**173.9**	1540
1979	173.9	8540	0.77	169.8	-4.1	**170.7**	1540
1980	170.7	8540	0.77	167.5	-3.2	**168.2**	1540
1981	168.2	8540	0.77	165.5	-2.7	**166.1**	1540
1982	166.1	8540	0.77	163.7	-2.4	**164.3**	1540
1983	164.3	8540	0.77	161.7	-2.6	**162.3**	1540
1984	162.3	8540	0.77	158.8	-3.5	**159.6**	1540
1985	159.6	8540	0.77	154.7	-4.9	**155.8**	1540
1986	155.8	8540	0.77	149.9	-5.9	**151.3**	1540
1987	151.3	8540	0.77	144.4	-6.9	**145.9**	1540
1988	145.9	8540	0.77	138.8	-7.1	**140.4**	1540
1989	140.4	8540	0.77	133.3	-7.1	**134.9**	1540

1990	134.9	8540	.0.77	128.1	-6.8	**129.7**	1540
1991	129.7	8540	0.77	123.8	-5.9	**125.5**	1540
1992	125.5	8540	0.77	120.1	-5.1	**121.3**	1540
1993	121.3	8540	0.77	117.3	-4	**118.2**	1540
1994	118.2	8540	0.77	115	-3.2	**115.7**	1540
1995	115.7	8540	0.77	113.3	-2.4	**113.9**	1540
1996	113.9	8540	0.77	111.6	-2.3	**112.1**	1540
1997	112.1	8540	0.77	110.1	-2	**110.6**	1540
1998	110.6	8540	0.77	108.1	-2.5	**108.7**	1540
1999	108.7	8540	0.77	106	-2.7	**106.6**	1540
2000	106.6	8540	0.77	103.4	-3.2	**104.1**	1540
2001	**104.1**						

REFERENCES

Ambalavanan, N., Carlo, W.A., Bobashev, G., Liu, B., Poole, K., Fanaroff, A. A.,Stoll, B.J., Ehrenkranz, R. and Wright, L. L. (2005): Prediction of Death for Extremely Low Birth Weight Neonates.

Amabalavanan, N., and Carlo, W.A. (2001). Comparison of the prediction of extremely low birth weight neonatal mortality regression analysis and by neural networks.

Amouzou, A., Richard, S. A., Friberg, I.K., Bryce, J., Baqui, A. H. and El Arifeen. (2010). "How well does LiST capture mortality by wealth quintile? A comparison of measured versus modelled mortality rates among children under-five in Bangladesh." *International journal of Epidemiology (2010)39(suppl)*

Bisgaard, S., Kulachi, M. (2011).Time Series Analysis and Forecasting by Example. John Wiley & sons, Inc. Hoboken, New Jersey.

Bitwe, R., Dramaix, M. and Hennart, P. (2006). "Simplified prognostic model of overall intrahospital mortality of children in central Africa." *Tropical medicine of International health.2006 Jan; 11(1):73-80.*

Blot, S. Brusselaers, N., Monstrey, S., Vandewounde, K., De Waele, J.J., Colpaert, K., Decruyenaere, J.,Malbrain, M., Lafaire C., Fauville, J.P., Jennes, S., Casaer, M. P., Muller, J., Jacquemin, D., De Bacquer, D. and Hoste, E.(2009). "Development and validation of a model for prediction of mortality in patients with acute burn injury. *The British journal of surgery.2009 Jan; 96(1):1117*

Boschi-Pinto, C., Velebit, L. and Shibuya K. (2008).”Estimating Child Mortality due to diarrhea in developing in developing countries.” *Bulletin of the World Health Organization 2008;86: 710-717.*

Broughton, S.J., Berry, A., Jacobe, S.,Cheeseman, P. and Tarnow-Mordi, W.O. (2004). “The Mortality Index for Neonatal Transportation Score.” A new Mortality Prediction Model for Retrieved Neonates.

Chartfield, C. (2000). Time Series Forecasting

Chinamu, K. and Chikobvu, D. (2010). Random Walk or mean reversion? Empirical evidence from the Crude oil price market.

Choi, K.M.S., Ng D.K.K., Wong S.F., Kwok, K.L., Chow, P.Y., Chan, C.H. and Ho, J.C.S. (2005). “Assessment of the Paediatric Index of Mortality (PIM) and the Paediatric Risk of Mortality (PRISM) III Score for prediction of Mortality in a paediatric intensive care unit in Hong Kong.” *Hong Kong Medical journal.*

Clancy, P. R., Mc Gaurn, S.A., Wernovsky G., Spray, T.L., Norwood, W.I., Jacobs, M.L., Murphy, J.D., Gaynor J.W. and Goin, J.E.(2000). “Preoperative risk-of-death prediction model in heart surgery with deep hypothermic circulatory arrest in the neonate.” *The journal of thoracic and Cardiovascular Surgery.*

Ellington Jr, M., Richardson D.K., Gray J.E., PursleY, D.M., Gortmaker, S., Goldman D. and Mc Cormick, M.(1997). “Mortality Prediction in very Low Birth Weight Infants At Days 3 and 14.”*The American Pediatric Society and the society for Pediatric Research.(1997)41,195-195*

Fowlie, P.W., Tarnow-Mordi, W.O., Gould, C.R., and Strang, D.(1997).Predicting outcome in very low birthweight infants using an objective measure of illness severity and cranial ultrasound scanning.

Fry-Johnson, Y.W., Levine, R., Rowley, D. and Agboto, V. (2010). "United States Black: White Infant Mortality Disparities are not Inevitable: Identification of Community Resilience Independent of Socioeconomic Status." *Journal of Ethnicity & Disease, volume 20.*

Horbor, J.D., Onstad, L. and Wright, E.(1993).Predicting mortality risk for infants weighing 501 to 1500 grams at birth: a National Institutes of Health Neonatal Research Network report.

Hussein, M.A.(1993). "Time Series analysis for forecasting both fertility and mortality levels in Egypt until year 2010." *The Egyptian population and family planning review 1993 Dec;27(2):67-81.*

Kayode, G.A., Adekanmbi, V. T. and Uthman, O. A. (2012). Risk factors and predictive model for under-five mortality in Nigeria: evidence from Nigeria demographic and health survey.

http:// www.biomedcentral.com/1471-2393/12/10

Klose, C., Pircher, M. , Sharma, S. (2004). Univariate Time Series Forecasting "UK Okonometrische Prognose"

Kumar, D. N. and Maity, R. (2008). Bayesian dynamic modelling for nonstationary hydroclimatic time series forecasting along with uncertainty quantification.

Liu, L., Johnson, H.L.,Cousens, S., Perin, J., Scott, S., Lawn, J.E., Rudan, I., Campbell, H., Cibulskis, R., Li, M., Mathers, C. and Black, R.E.(2012). Global. regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000.

Marshall, G., Tapia, J.L., D'Apremont I., Grandi, C., Alegria, A., Standen, J., Panniza, R., Roldan, L., Musante G.,Bancalari, A., Bambaren, E., Lacarruba, J., Hubner, M.E., Fabres, J., Decaro, M., Mariani, G., Kurlat, I. and Gonzalez, A. (2005). " A new score for predicting neonatal very low birth weight mortality risk in the NEOCOSUR South American Network." *Journal of perinatology.2005Sep;25(9):577-82.*

Migon, H. S., Gamerman, D.,Lopes, H.F. and Ferreira, M.A.R. (2005). Dynamic Models.

Mojekwu, J.N., and Ajijola, L.A. (2011). Developing a model for estimating infant mortality rate of Nigeria.

Mohamed, I. E. (2008). Time Series Analysis Using SAS, Part 1, The Augmented Dickey-Fuller (ADF) Test.

Nakwan, N., Nakwan, N., Wannaro, J.,(2011). "Predicting mortality in infants with persistent pulmonary hypertension of the new born with the Score for Neonatal Acute Physiology-version II (SNAP-II) in Thai neonates." *Journal of perinatal Medicine.Volume39, issue 3, pages 311-315.*

O'Neill, S.M., Fitzgerald, A., Briend, A. and Van den Broeck, J. (2012). "Child mortality as predicted by nutritional status and recent weight velocity in children under two in rural Africa." *Journal of nutrition 2012.Mar;142(3):520-5.*

Pearson, G.A., Stickley, J. and Shann, F.(2001). "Calibration of the paediatric index of mortality in UK paediatric intensive care units." *Archives of disease in childhood.2001, Feb; 84(2):125-8.*

Pedroza, C.(2006). A Bayesian forecasting model: Predicting U.S. male mortality. *Oxford journals, volume 7 ,issue 4. Pp 530-550*

Pepler, F.T., Uys, D.W. and Nel, D.G. (2012). "Predicting mortality and length-of-stay for neonatal admissions to private hospital neonatal intensive care units: a Southern African retrospective study." *African Health Sciences vol 12,No 2 (2012)*

Purwanto, Eswaran, C. and Logeswaran, R. (2010). "A Comparison of ARIMA, Neural Network and Linear Regression Models for predicting of Infant Mortality Rate." *Ams, pp 34-39, 2010. Fourth Asia International Conference on Mathematical/ Analytical Modelling and Computer Simulation, 2010.*

Pollack, M.M., Koch, M.A., Bartel, D.A., Rapoport, I., Dhanireddy, R., El-Mohandes, A.A.E., Harkavy, K., and Subramanian, K.N.S.(2000): A Comparison of Neonatal Mortality Risk Prediction Models in very Low Birth Weight Infants.

Rao, C., Adair, T., and Kinfu, Y. (2010). Using Historical Vital Statistics to Predict the Distribution of Under-Five Mortality by cause.

Rastogi, P.K., Sreenivas, V. and Kumar, N. (2009). Validation of CRIB II for prediction of Mortality in Premature Babies.

Sankrithi, U., Emanuel, I., and Van Belle, G. (1991). "Comparison of linear and exponential multivariate models for explaining national infant and child mortality." *International journal of epidemiology.1991 Jun;20(2):565-70.*

Sergio, I.V., and Leon, A.C. (2009). "Analysis of mortality from diarrheic diseases in under-five children in Brazilian cities with more than 150,000 inhabitants.

Siu-Hang Li, J., Hardy, M.R. and Tank, S. (2009). "Uncertainty in Mortality Forecasting: An Extension to the classical Lee-Carter Approach." *The journal of the International Actuarial Association; volume 39, issue 01.*

Skarsgard, E. D., MacNab, Y.C., Qui, Z., Little, R. and Lee, S.K. (2005). "SNAP-II predicts mortality among infants with congenital diaphragmatic hernia." *Journal of perinatology.2005 /May;25(5):315-9.*

Slater, A., and Shann, F. (2004). "The Suitability of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand." *Journal of the Pediatric critical care medicine (2004) Sep;5(5) 447-54.*

Terra de Souza, A.C., Cuffino, E., Peterson, K. E., Gardner, J, Vasconcelos do Amaral, M. I. and Ascherio, A. (1999). "Variations in infant mortality rates among municipalities in the state of Ceara, Northwest Brazil: an ecological analysis. *International journal of Epidemiology.1999 Apr,28(2):267-75.*

West, M. and Harrison, J. (1997). Bayesian Forecasting and Dynamic Models, 2nd ed. Springer Series in Statistics. Springer-Verlag New- York, Inc.

Wong, S.L., Yu, I.T., Wong, T.W., and Lloyd, O.L. (1997). “Trends in infant mortality in Hong Kong (1956- 90) and evaluation of preventable infant deaths.” *Journal of paediatric and child health.*

Zernikow, B., Holtmannspoetter, K., Michel, E., Hornschuh, F., Westermann, A., and Hennecke, K. (1998). “Artificial neural network for risk assessment in pretern neonates. *Archives of Disease in Childhood: Fetal &Neonatal.*