UNIVERSITY OF GHANA

STATISTICAL ANALYSIS OF WATER LEVEL, TEMPERATURE AND HUMIDITY USING COINTEGRATED VECTOR AUTOREGRESSION (VAR) MODELS

BY

APPIAH KOJO ISAAC
[10551074]

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF MPHIL STATISTICS DEGREE

JULY, 2017

DECLARATION

I hereby declare this submission is my own work, except where otherwise stated, under the supervision of Dr. E. N. N. Nortey and Dr. Isaac Baidoo.

Candidate: Appiah Kojo Isaac [10551074]

……
Signature

……
Date

Principal supervisor: Dr. E. N. N. Nortey Co-supervisor: Dr. Isaac Baidoo

…… ……
Signature Signature

…… ……
Date Date

DEDICATION

This project work is dedicated to Agnes Emefa Kudenu

ACKNOWLEDGEMENT

I want to show gratitude to the Almighty God for His knowledge, strength and grace which has sustained me in sailing through to this stage of my life and throughout this study. A profound appreciation to my supervisors Dr. N. N. Nortey and Dr. Isaac Baidoo for all the advice and guidance given to me by putting this work together. I also acknowledge Dr. F. O. Mettle for his assistance and suggestions when I called on him; I say God richly bless you.

Finally, sincere thanks to all friends, course mates and loved ones for their encouragement, support and contributions that helped in the course of this study.

ABSTRACT

The leading climate factors influencing availability of water are; temperature, relative humidity, precipitation, and evaporation. Water and agricultural production cycles are indisputably influenced by temperature and relative humidity. Temperature and relative humidity projections can therefore proficiently be employed in making decisions when optimal usage of water resources is of interest. Thus, the current study explored both the "long-run" and "short-run" impact of both temperature and relative humidity on water level through cointegrated VAR models with specific application to the Akosombo Dam water level. The quarterly averages of the daily Akosombo water level, temperature and humidity of its surrounding was computed from the daily data obtained and it spanned the period January 1980 to December 2014. Seasonal ARIMA models $ARIMA(0,1,1)(0,1,1)_4$, $ARIMA(1,0,1)(1,1,1)_4$ and $ARIMA(2,1,1)(1,1,1)_4$ were estimated using values of AIC, AICc and BIC for Akosombo Dam water level, temperature, and humidity respectively. Also, water level was observed to granger causes both temperature and relative humidity whiles relative humidity also granger causes both water level and relative humidity. In addition, both water level and temperature responded positively to the impulse of humidity. The VAR model outperformed the SARIMA model in forecasting water level, temperature, and relative humidity. Finally, water level and relative humidity were cointegrated whiles a cointeration relation was observed between temperature and relative humidity. The rate of adjustment to equilibrium observed by temperature was very high among the three variables.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.0 Background to the study

"Time series", in general, refers to sequential data of any form. The data may take discrete or continuous values, and form a time series because they are provided with discrete indices indicating an ordering. The index may correspond to the onset of time, with for example price or position data, or have no intuitive meaning with for example the Fibonacci sequence, or genetic sequence data (Bracegirdle, 2013).

The challenges related to establishing patterns in observed sequential data is both well-known and widely studied, transcending fields of research and types of data. There is a very wide variety of problems and correspondingly wide variety of solutions. In general, one is interested in modelling time series in order to understand more about something that occurs over time, in order to make decisions, predictions, or simply to understand more about the world (Bracegirdle, 2013).

A time series with stochastic trends are usually independent of each other and thus exhibits non stationarity among the series. However, a linear combinations of a number of unit root time series is often observed in practice and depict a stationary series. When this occurs, we referred to the process as cointegration. One of the major concept in time series is cointegration. It has been established that the differences between two time series may be more predictable than the individual series itself.

Cointegration is said to exist among a vector of "non-stationary" "time series" processes if a linear combinations of the series which is "stationary" can be established. This concept, first used by Granger (1983), is extensively used by

econometricians for modelling the dynamics of economic systems. Johansen (1988) is one of the major contributors of this technique and his maximum likelihood approach to establishing the cointegration is widely accepted and is arguably the most frequently used method for cointegration analysis. To establish cointegration, a "Vector Autoregressive (VAR) model" for a set of non-stationary variables with normally distributed residuals is first fitted. The VAR model is then converted into vector error correction form. In this form, the "long-run impact matrix" of the VAR serve as the coefficient matrix on the lagged levels, while the variables are also expressed in first differences. This is done without imposing any constraints on the fitted VAR model parameters. At least one cointegration relation can be established. The number of cointegration relation is less by one the number of variables in consideration. The "Vector Error Correction Model (VECM)" has both the "long-run and short-run impact matrices". For a cointegrated process, the long-run impact matrix must have a reduced rank and this rank is equivalent to the number of "cointegration relations". The concept of cointegrated VAR model was deduced in 1987 by Engle and Granger. Several authors for which Johansen cannot be left out have developed rank testing procedures in attempts to find the number of cointegration equations.

Cointegration is a well-accepted technique by econometricians as it consist of both stochastic and deterministic terms. Kaufmann and Stern, in 1997 and 2002, established stochastic trends in climate time series data. However, it is unusual that the concept of cointegration, well advocated and accepted by econometricians is rarely employed in the area of geophysics. In this paper, the cointegration approach

by Johansen and co-authors (Johansen and Juselius, 1990 and Johansen, 1988) were used to fit a "time series model" to explore the long-run relation among the Akosombo Dam water level, temperature and relative humidity of the regions.

1.1 The Akosombo Dam

"Lake Volta", adjudged the world's largest artificial lake (covers 8,502 square kilometers, 3.6% of the land area of Ghana) houses the nations' major source of electricity, the Akosombo Dam. Built close to the range of hills which forms a point of separation between Akwapim and Togo, the project was to serve as the main power supply to an aluminium company in Ghana (Zakhary, 1997). The Akosombo Dam project has seen major phases of development and could generate a power of 1020 megawats (as at 2006) as compared to its initial production rate of 912 megawats.

In spite of the variety of it sources, the Lake Volta conforms to a remarkable regular time-table of rise and fall. The extent of these is always uncertain. Precise records of water levels on the lake have been kept. The river is mostly at its lowest in March each year and at its highest at the end of September or early October (Mason, 1984). The generation process depends highly on the inflows of water that refill the reservoir, thus, the water level is of keen interest to the management of the Akosombo Dam. Establishing the optimal operating scheme of hydro-electric power production installation is one of the major areas of interest as far as the water level is concerned.

To regulate and manage the growth of the project, the Volta River Authority (VRA) through the government of Ghana and the Volta River Development Act was

formed in 1961. Their task was to control the activities of the members of the surrounding communities such as “fishing, lake transportation and communication, the welfare” of the people as well as its primary responsibility of development of the “Volta River Basin” (Fobil, 2003).

In order to carry out its daily operations, the level of the Akosombo Dam water is key importance to the Volta River Authority (VRA). Thus, the current study sought to identify an appropriate model to project the water level in the near future to enable the authorities in the day to day administrative work.

1.2 Problem Statement

For secured electric power supply to the nation Ghana, the activity of the Akosombo Dam which depends mostly on the dynamics of water inflows into the Volta River Basin should be of major concern. Knowledge of approximately what the water level of the Akosombo Dam will be tomorrow, next week or perhaps in a month’s time is very vital to the operations of the Volta River Authority(VRA) as it will enable the authority better manage production and distribution of hydro-electric power to the country and its environs.

For this reason, many studies have tried to fit a model that would help in the prediction of the water level. (Examples include Mensah, 2013; Amfo-Otu, 2010; and Marfo 2009). Mensah (2013) used a step-by-step approach of the Box-Jenkins ARIMA process and arrived at a seasonal model. Amfo-Otu (2010) modeled the water level dynamics of the “Akosombo Dam” through differential equations. However, Marfo (2009) fitted a first order and second order autoregressive models

using least-square fitting method purposely for forecasting future daily dam water levels even though series is random.

The leading climate factors influencing availability of water are; temperature, relative humidity, precipitation, and evaporation. Water and agricultural production cycles are indisputably influenced by temperature and relative humidity. Temperature and relative humidity projections can therefore proficiently be employed in making decisions when optimal usage of water resources is of interest (Sarraf, Vahdat, and Behbahaninia, 2011).

One common trait of all studies in attempt to predicting the Akosombo Dam water level is the failure to explore external factors to the dynamics of the water level. Thus, the current study explored both the "long-run" and "short-run" impact of both temperature and relative humidity on water level through cointegrated VAR models with specific application to the Akosombo Dam water level.

1.3 Objectives of the study

The main objective of the study was to explore the relation amid water level, temperature and relative humidity of the regions using cointegrated VAR models. Specifically, the study sought to;

1. Fit a Seasonal Autoregressive Integrated Moving Average (SARIMA) for the water level, temperature, and humidity of the regions of the Akosombo Dam.
2. Fit VAR model to the water level, temperature, and humidity of the regions data.

3. Establish a VECM for the water level, temperature, and humidity of the regions.
4. Explore the cointegration relation between the water level, temperature and humidity of the regions.
5. Forecast the water level, temperature and humidity of the regions using VAR and SARIMA models.
6. Compare the VAR model to the SARIMA model.

1.4 Hypotheses

a) Temperature granger causes water level and humidity.

b) Humidity granger causes water level and temperature.

c) Water level granger causes temperature and humidity.

1.5 Significance of the study

By using a new method to the prediction of water level, the study will contribute greatly to the body of literature. The study provides time series stationary models for forecasting water levels, temperature and humidity of the regions. This is to help the Volta River authorities manage production and distribution of hydro-electric power to the country and its environs.

Furthermore, this study aside serving as basis for future research in the area of predicting water level will direct the attention of researchers to explore more on this new method of predicting water level as the method has proven much in the world of economics.

1.6 Limitations

The daily data on temperature and relative humidity had a quiet a number of missing observations which were not estimated. Rather, due to the slight differences in the daily occurrences, monthly averages were considered. However, quarterly averages were used to avoid "overfitting". Thus, care should be taken in the interpretation of the results.

1.7 Organization of the study

The rest of the study is organized as follows: Chapter two focus on the pertinent literature to the study. Literature on the time series model such as the SARIMA as well as the historical background to cointegration and its contribution so far is explored. Chapter Three touches on the methods employed in "time series" procedures. The "Box-Jenkins methodology" and the VAR methodology is extensively laid out in this chapter. This is followed by Chapter Four where the researcher looked at the analysis. Finally, Chapter Five presents the discussion of results, conclusion and offers recommendation where needed.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

The review of related literature on the study is presented in this chapter. The literature review provides the framework of both The "Box-Jenkins Seasonal ARIMA" and the VAR methodologies. Pertinent empirical review on the application of these methodologies is also provided.

2.1 Review

Time series forecasting has an especially high utility for predicting "economic and business trends". Many forecasting methods both physical and data driven has been developed in the last decades. These includes; neural networks, the "Box-Jenkins methodology", Vector Autoregressive (VAR) models, Autoregressive Distributed Lag (ArDL) models, among others. "Neural networks" as models for time series has been identified as a field to match the "Box-Jenkins method" for "long and short term memory series" (Rani and Parekh, 2014; Tang and Fishwick, 1993). It has been proven to outperform the Box-Jenkins model for series with short memory. However, the most frequently and well-accepted "time series forecasting methods" in practice is the "Box-Jenkins method" (Tang and Fishwick, 1993). In this study, the "Box-Jenkins Seasonal ARIMA (SARIMA)" and the VAR models were employed.

The Box-Jenkins method was proposed by "George Box and Gwilym Jenkins" in 1970. The Box-Jenkins methodology begins by fitting an ARMA model (stationary process) or an ARIMA model (for a "non-stationary process") to the "time series"

data. The process is referred to as a stochastic model building and that it is an iterative approach that consist of "identification, estimation, and diagnostic checking" as depicted in Figure 2.1 (Box, Jenkins and Reinsel, 2008).

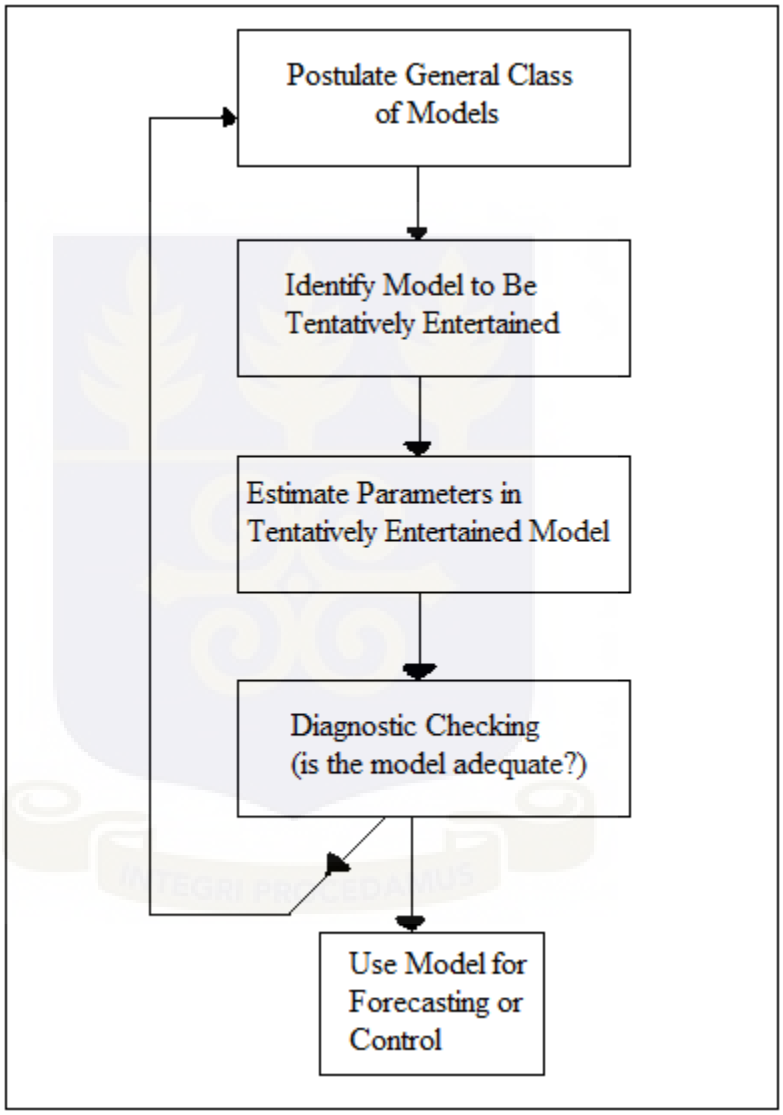

Figure 2.1: Stages in the Iterative Approach to model building

From 2.1, a useful class of models for the purpose at hand is first considered. Secondly, a parsimonious subclass of models are deduced tentatively. The yielded

model from the identification process is then used to estimate of the parameters in the model. The initial estimates obtained serve as a starting point for an iterative process to arrive at a final estimate. Finally, diagnostic checks are perform to assess the "goodness-of-fit" of the model and also to diagnose a cause of "lack of fit". The model is approved for further use if found adequate. However, for lack of fit, the iterative cycle of "identification, estimation, and diagnostic checking" is repeated until a suitable model for the data is found (Box, Jenkins, and Reinsel, 2008). Box-Jenkins methods has been used widely in several fields of study and it provides a comprehensive model in the domain of "time series analysis".

"Time series" analysis has been used in predicting water levels of rivers, lakes and dams, temperature as well as relative humidity by various authors (Nia, Babazadeh and Boustani, 2009; Vahab and Alikhani, 2005; Jahanbakhsh and Baser, 2003; Modarres, 2003 are examples). As a result of the established seasonality in time series data of water level, temperature, and relative humidity, recent studies have resulted to seasonal models. For instance, Tularam and Ilahee in a study to investigate the interactions between rainfall and temperature in coastal catchment in Australia employed time series. Seasonality was observed in both the rainfall and temperature series. Employing Box-Jenkins ARIMA methodology and allowing for seasonality and autocorrelation, the water level was modelled (Tularam and Ilahee, 2010).

Federal agencies in the U.S and Canada in attempt to model and forecast the dynamics of the water level of major Lakes in order to minimize the adverse impact of inconsistency in the levels, employed multiplicative, seasonal ARIMA models

using monthly mean level data. Lakes Erie and Ontario are specifics. $ARIMA(1,0,1)(0,1,1)_{12}$ model was identified for both lakes using data for the period 1900 to 1986 (Eberhardt and Irvine, 1992).

In Ghana, several studies (Mensah, 2013; Amfo-Otu, 2010; and Marfo 2009 are examples) have been carried out in attempt to fit an appropriate model to the fluctuating levels of the Akosmbo Dam water level. Several methodologies have been employed in this exercise. Mensah (2013) however used the Box-Jenkins ARIMA process and arrived at a seasonal model "$ARIMA(1,1,0)(0,1,1)_{12}$", identified as the best model for predicting the Akosombo Dam water level. This study used the mean monthly water level data which span from January, 1980 to December, 2010. The study recommended other methodologies such as the Artificial Neural Network as a forecasting tool.

In another development, the first decade of the 21st century monthly data of maximum and minimum temperatures from the Greek Island of Kefalonia were analyzed using multiplicative seasonal ARIMA model. Seasonal patterns were observed to evolve slowly over time (Mills, 2013). Seasonal patterns was also observed in temperature modeling study by Afrifa-Yamoah. Using 1975 to 2009 monthly data from the department of meteorological and climatology in the Brong Ahafo region of Ghana, an $ARIMA(1,0,0)(0,1,2)_{12}$ was identified as the best model for to forecast the average surface temperature of the region (Afrifa-Yamoah, 2015). In other several noted studies involving modeling temperature, seasonal ARIMA was employed due to the observed seasonality in temperature data (Nury, Koch and Alam, 2013; Hsing and Tsokos, 2008).

One common trait of all these studies to predicting the Akosombo Dam water level is the failure to explore external factors to the dynamics of the water level. Water and agricultural production cycles are indisputably influenced by temperature and relative humidity. Temperature and relative humidity projections can therefore proficiently be employed in making decisions when optimal usage of water resources is of interest (Sarraf, Vahdat, and Behbahaninia, 2011).

To bridge the gap in the studies in attempt to predicting the Akosombo Dam water level, the current study explored the relations of temperature and relative humidity to water level in attempt to measure both the "short and long run" relations of temperature and relative humidity on the dynamics of levels of water through Cointegrated VAR models. One run the risk of spurious regression in attempt to model temperature and relative humidity as result of the established presence of stochastic trends in these "time series" data and this is a justification of the choice of the cointegrated VAR methodology.

VAR, a multivariate technique, is used to explore the linear interdependence amid a vector of "time series" variables and a generalization to the univariate autoregressive (AR) model by making room for at least two variables. The VAR procedure produces an equation for each variable included in the model. An equation for a variable has "lagged values" of its own as well as "lagged values" of the other variables included in the process and a term of innovations.

The popularity of VAR models for analyzing the dynamics of economic systems is due to Sims's (1980) influential work. Sims backed VAR procedures, disapproving the assertions and efficiency of earlier modeling in "macroeconomic econometrics"

(Sims, 1980). It was seen as a statistical field in control theory, being a theory-free procedure to model economic relations, which suffices to be another option to the "incredible identification restrictions" in structural models (Sims, 1980).

The step-by-step approach to modelling VAR are presented in in Figure 2.2 (Luetkepohl, 2011).

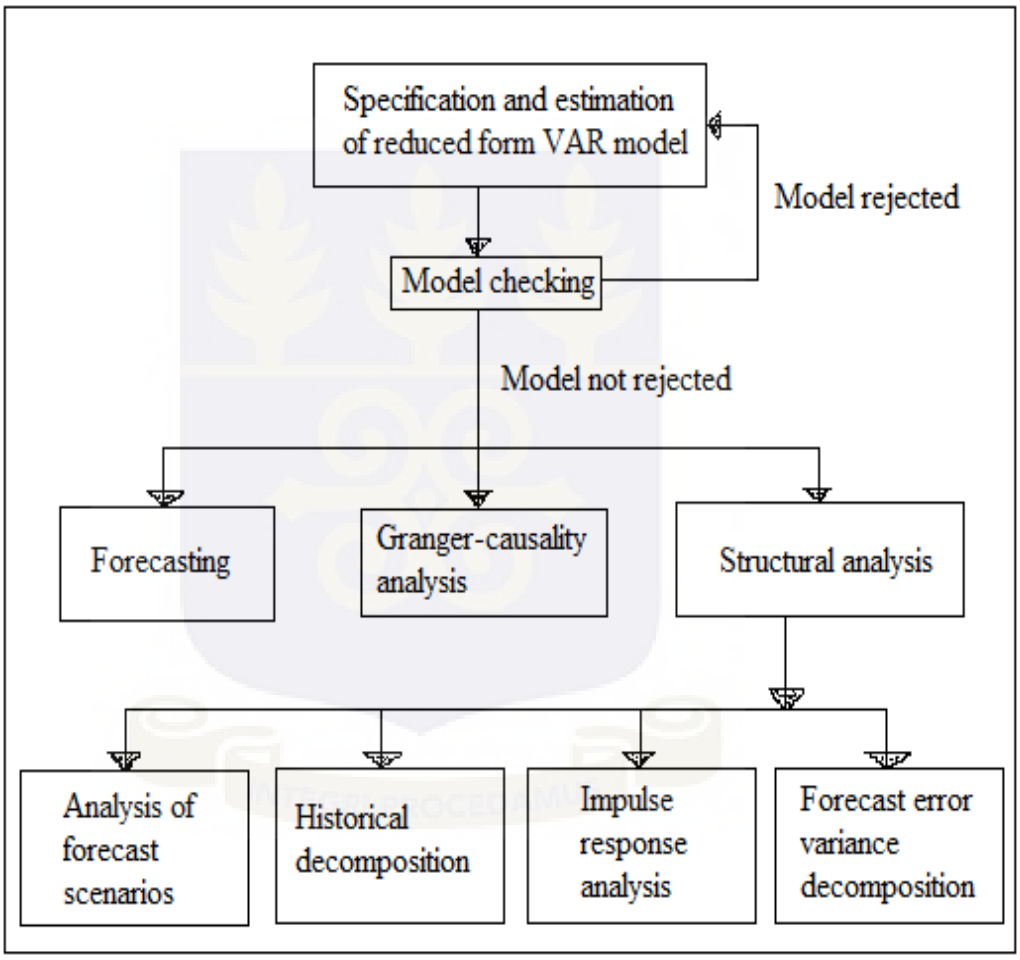

Figure 2.2: Vector Autoregression (VAR) Analysis

In VAR analysis, we first specify and estimate a reduced form of the model the Data Generating Process (DGP) and "goodness-of-fit" assessed. If inadequacy is

detected, the model is modified until an adequate form is obtained. Forecasting, Granger causality is done suing the approved model.

There are three forms of the VAR models. These are; reduced, recursive, and structural forms. The reduced form, which is the basic form of VAR produces a linear equation for each variable considered in the procedure. A linear equation in which the predictors are the ”lagged values” of the response variable, “lagged values” of the other considered variables, and an independent term of innovations. For instance, in the case of the current study, the VAR procedure will produce three equations: Akosombo Dam water level as a linear equation of “lagged values” of water level, temperature and the relative humidity; average temperature as a linear equation of “lagged values” of average temperature, water level and the relative humidity; and similarly for the relative humidity equation. The number of “lagged values” to include in each model can be identified by a number of methodologies.

In a “recursive VAR models”, concomitant values are introduced as regressors. This is done to make the error terms independent to each other. Considering the variables in this study for instance, the first equation will treat the average temperature as the response variable and the covariates being the “lagged values” of temperature, humidity, and water level. Water level becomes the response variable with covariates of lags of temperature, humidity, and water level as well as the present value of the average temperature in the second equation. Relative humidity is the response variable in the third equation also with covariates of lags of temperature, humidity, water level and the present value water level (Blanchard, Olivier and Watson, 1986).

For structural VAR, economic theory is used to categorize the concurrent relations amid the variables (Bernanke, 1986; Blanchard and Watson, 1986; Sims, 1986). An "identifying assumptions" that enables the establishment of specific causal relations is required by the procedure. This crops contributory variables which allows concurrent relations to be modeled using contributory variables regression. The number of "structural VARs" is determined by the ingenuity of the researcher (Stock and Watson, 2001).

The proposed measures of exploring the links among the variables of a VAR model include; "granger causality test", "impulse response analysis", "forecast error variance decompositions", "historical decompositions", and the "analysis of forecast scenarios". Conventionally VAR models were built for for stationary variables without time drifts. Usually, deterministic terms are added to the model to capture trending comportment in the series. The establishment of "stochastic trend" and the concept of cointegration in the 1980s have shown that "stochastic trend" can also be measured by VAR models (Johansen, 1995; Engle and Granger, 1987; Granger, 1981). Most importantly, the "long-run and short-run" dynamics are separated when trend is observed in the variables using an appropriate and well accepted technique. The most expedient frame for this purpose is the VECMs (Luetkepohl, 2011).

In the current study, reduced form of VAR model was considered. Possible cointegration was explored to measure both the "long-run and short-run" impact among the variables considered through cointegrated VAR models. The concept of being able to combine "long-run and short-run equilibrium" processes in the data

through the properties of cointegration is an obvious and the most important reason the choice of VAR model continues to gain the backing and the attention of both econometricians and applied economists as well other researchers in the fields of study (Juselius, 2006).

VAR models are used extensively by econometricians. For instance, Clarida and Friedman (1984) forecasted the United States short-term interest rates from April 1979 to February 1983 through VAR models. Mei, Lui and Jing (2011) by using six economic indicators fitted a multi-factor dynamic system VAR forecast model of the GDP in China. Other studies includes (Chepngtich and Kihoro, 2015; Bayraci, Ari and Yildirim, 2011).

In recent studies, the use of VAR models have been extended to other fields of study such as geophysics. For instance, Schmith, Johansen, and Thejll, (2011) through cointegrated VAR methods observed link amid sea level and temperature. It was established that sea level granger causes temperature and this can be attributed to the enormous heat capacity of the sea. VECM fitted to the data showed that "temperature adjusts to the cointegration equilibrium", while there is "no measureable adjustment of the ocean". Kaufmann and Stern (2002) also used cointegration procedures to fit a time series model of the "link among temperature and the radiative force of solar irradiance, greenhouse gases, and tropospheric sulfates". Their findings confirmed existing simple hypotheses regarding the effect of deviations in radiative forcing on temperature and offered some lines of evidence that indicate human activity to some extent leads to the observed increase in temperature over a century now.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

The methods employed in the analysis were discussed in this chapter. Validity, reliability and generalizability of the research findings are highly dependent on the research methodologies employed (Geoffrey, David and David, 2005). This chapter deals with the data source, Box-Jenkins methodology, maximum likelihood estimation of VAR models, and Johansen's approach to cointegration.

3.2 Data

The study covered data on the Akosombo water level, and the temperature and humidity of its surroundings. The daily Akosombo water level data was obtained from the Electro-Volta House office of the Volta River Authority in Accra and it covered the period of 35 years from January 1980 to December 2014. The data on the temperature and humidity of the regions was obtained from the Ghana Meteorological Agency office in Accra. The data was gathered at their station at Akosombo. Both the daily minimum and maximum temperature and humidity from January 1980 to December 2014 was obtained.

3.3 Time Series

A "time series" is a collection of data observed consecutively over time. These observations, either continuous or discrete are gathered. The continuous and discrete series were conventionally referred to, irrespective of the variable being measured being either continuous or discrete. That is, the interval of observation may either be exact or in an interval. For a discrete time series, the time of recording

is made at distinct or separate points in time whiles continuous time series domain is continuum though the variable itself need not be continuous.

3.3.1 Notational Convention

Time series models involve lagged terms and may involve differenced data to account for trend. The following notations are often used.

Backward shift Operator

The "backward shift" or "lag operator", *B*, is used to operate on the past values of time series. Applying B before the values of the series y_t implies, the observation should be moved back by one period. For instance

$$By_t = y_{t-1}.$$

Moving back a number of time is denoted by the power of the operator. As an example,

$$B^2 y_t = y_{t-2}.$$

y_{t-2} represents y_t two units back in time. $B^k y_t = y_{t-k}$ represents y_t k units back in time. The backshift operator *B* does not operate on coefficients as they are fixed quantities that do not move in time. For example

$$B\theta_1 = \theta_1.$$

Differencing

Differencing is used to deal with nonstationarity as a result of linear or quadratic pattern as well as seasonality. The difference $y_t - y_{t-1}$ can be expressed as

$(1-B)y_t$ using the backward shift operator. An alternative notation for a difference is $\nabla = 1 - B$. Thus,

$$\nabla y_t = (1-B)y_t = y_t - y_{t-1}.$$

a) A subscript defines a difference of a lag equal to the subscript. For instance

$$\nabla_2 y_t = y_t - y_{t-2}.$$

b) A superscript says to repeat the differencing the specified number of times. As an example,

$$\nabla^2 y_t = (1-B)^2 y_t = (1-2B+B^2)y_t = y_t - 2y_{t-1} + y_{t-2}.$$

This is the difference of the first difference. That is

$(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$.

c) For an observed trend pattern of rise and fall in the series, it is appropriate to difference the series ones. A second differencing may be required when a quadratic trend is present.

d) Differencing the series more than necessary can bring about levels of dependency. In practice, the first and second difference is the norm. If stationarity is still not achieved, transformation or smoothing of the data may be opted for.

Definition 3.1

The autocorrelation of a stationary process $\{y_t\}$ at lag k

$$\rho_x(k) = corr(y_t, y_{t-k})$$

denotes the linear relation among y_t and y_{t-k}.

However, the dependency structure among the intermediate process variables $y_s, t-k<s<t,$ is very important. For instance, y_t and y_{t-2} are not directly correlated. The correlation coefficient results indirectly, since y_t is linearly related to y_{t-1} and y_{t-1} is also linearly related to y_{t-2}. The autocorrelation function (ACF) comprises all direct and indirect relationship among y_t and y_{t-k}. For a direct correlation between y_t and y_{t-k}, partial autocorrelation function (PACF) must be considered.

Definition 3.2

Let $\{y_t\}$ be a "stationary process". The "partial autocorrelation" at lag k for $k \geq 0$, is defined as the direct relationship amid y_t and y_{t-k} with the linear dependence between the intermediate variables y_s with $t-k<s<t$ removed.

3.3.2 Unit Root Test

Most "time series" data depict trending or "non-stationarity" in both the mean (μ_t) and variance. This is an undesirable property when modelling an adequate model to the patterns in a time series data. Now, y_t is said to be "stationary" if; $E(y_t)=\mu,\ E(y_t y_{t-k})=\sigma$, and the "autocovariance function" between y_{t_1} and y_{t_2} hinges only on the interval t_1 and t_2.

The existence of "long-run equilibrium" links amid "non-stationary time series" variables has been advocated by "theories of economics and finance". The cointegration test is assumes that the variables are integrated of order zero ($I(0)$) or one ($I(1)$). Determination of the order of integration is an important prerequisite to cointegration procedures as to ascertain that that the variables are not integrated of order two ($I(2)$). This will help avoid spurious outcomes. The cointegration procedure requires that all the variables should be $I(1)$ to be able to observe a long-run equilibrium among the variables. Thus, test for unit roots should always be considered first when it comes to modelling cointegration.

The ADF and KPSS test, were used to assess the "stationarity" of the variables in this study.

3.3.2.1 Augmented Dickey-Fuller (ADF) tests

The ADF test is derived using the model

$$y_t = \boldsymbol{\beta}' D_t + \phi y_{t-1} + \sum_{i=1}^{p} \psi_i \Delta y_{t-i} + \varepsilon_t. \qquad (3.1)$$

Here, Δy_{t-i} are differenced past values of the process, ε_t are sequentially independent and homoscedastic, and D_t contains deterministic terms.

The following hypothesis is tested,

H_o : The series y_t is $I(1)$ (that is $\phi = 1$).

H_1 : The series y_t is $I(0)$ (that is $|\phi| < 1$).

where y_t is $I(1)$ an indication that $\phi = 1$.

The t-statistic and normalized bias statistic are given by

$$ADF_t = t_{\phi=1} = \frac{\hat{\phi} - 1}{SE(\phi)}$$

$$ADF_n = \frac{T(\hat{\phi} - 1)}{1 - \hat{\psi}_1 - \ldots - \hat{\psi}_p}$$

respectively.

3.3.2.2 Kwiatkowski Phillips Schmidt and Shin (KPSS) test

The KPSS test for "stationarity" in a time series data, that is, the null hypothesis that y_t is "integrated" of order zero. It must be noted that the presence of unit root is assessed by the ADF test. KPSS is arguably the most frequently used stationarity test, proposed by Kwiatkowski, Phillips, Schmidt and Shin (1992). The test was derived using the equation;

$$\left.\begin{aligned} y_t &= \beta' D_t + \mu_t + u_t \\ \mu_t &= \mu_{t-1} + \varepsilon_t, \varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right) \end{aligned}\right\} \quad (3.2)$$

where D_t consist of deterministic terms, μ_t is $I(0)$ and may be heteroscedastic. The following hypothesis is tested

H_o : The series y_t is $I(0)$ (that is $\sigma_\varepsilon^2 = 0$).

H_1 : The series y_t is $I(1)$ (that is $\sigma_\varepsilon^2 > 0$).

The KPSS test statistic is given by

$$KPSS = \left(T^{-2} \sum_{t=1}^{T} \hat{S}_t^2 \right) / \hat{\lambda}^2. \quad (3.3)$$

Here, $\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j$, where $\hat{u}_t$ is the residual, and $\hat{\lambda}^2$ is maximum likelihood estimate of μ_t using $\hat{u}_t$.

3.3.3 The Box-Jenkins ARIMA Processes

In time series analysis, the "Box–Jenkins methodology" applies "Autoregressive Processes (AR)", "Moving Average processes (MA)" and "Integrated Processes" to attain the "ARMA or ARIMA models" to model an adequate model that fit a "time series" to "lagged values" of this "time series", for forecasting purposes. The Autoregressive Integrated Moving Averages (ARIMA) class of models has proven to be an efficient tool for forecasting, and has been the basis for most important ideas in "time series" analysis. The different components of this general class of models includes AR and MA. It sometimes referred to as Box-Jenkins models since it was pioneered by Box and Jenkins in 1970.

3.3.3.1 Autoregressive (AR) processes

An AR(p) is given by

$$y_t = \sum_{r=1}^{p} \alpha_r y_{t-1} + z_t \tag{3.4}$$

$$= \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + z_t$$

where $\{\alpha_r\}_1^p$ are fixed constant and $\{z_t\}$ is purely white noise. Using the "backward shift operator", Equation (3.4) is rewritten as

$$\left(1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p\right) y_p = z_t$$

$$\alpha(B)y_t = z_t \tag{3.5}$$

From Equation (3.5),

$$y_t = \alpha(B)^{-1} z_t$$

$$y_t = \left(1 - \alpha_1 B - \alpha_2 B - ... - \alpha_p B^p\right)^{-1} z_t \tag{3.6}$$

From Equation (3.6),

a) $E(y_t) = 0$ since z_t is a random process.

b) $Var(y_t)$ is finite provided $\sum_{i=1}^{\infty} \alpha(B)_i$ converges (a condition for stationarity).

c) Equivalently, the stationarity condition is that the values of B to

$$\alpha(B) = 0$$

must be greater than one.

The simplest form of an AR model is the first order denoted AR(1) and defined as

$$y_t = \alpha y_{t-1} + z_t$$

It is stationary provided $|\alpha| < 1$.

3.3.3.2 Moving Average (MA) processes

An MA(q) is given by

$$y_t = \beta_1 z_{t-1} + \beta_2 z_{t-2} + ... + \beta_q z_{t-q} + z_t$$

$$= \sum_{r=0}^{q} \beta_r z_{t-r}, \beta_0 = 1 \tag{3.7}$$

where $\{\beta_i\}$ are constants and $\{z_t\}$ is purely white noise. Using the "backward shift operator", Equation (3.7) can be rewritten as

$$y_t = \beta(B)z_t.$$

Properties

a) $E(y_t) = 0.$

b) $Var(y_t) = Var\left(\sum_{r=0}^{q} \beta_r z_{t-r}\right) = \sigma^2 \sum_{r=0}^{q} \beta_r^2.$

The basic form of an MA model is MA(1), the first order, defined as

$$y_t = \beta z_{t-1} + z_t .$$

Invertibility of the MA process

From AR(1) process

$$y_t = \alpha y_{t-1} + z_t, \ |\alpha| < 1$$

$$y_t = (1 - \alpha B)^{-1} z_t$$

$$= \left(1 + \alpha B + \alpha^2 B^2 + \ldots\right) z_t$$

$$= z_t + \alpha z_{t-1} + \alpha^2 z_{t-2} + \ldots. \qquad (*)$$

$$= \sum_{r=0}^{\infty} \alpha^r z_{t-r}.$$

Equation (*) is $MA(\infty)$. Thus an AR(1) process has $MA(\infty)$ representation which is stationary.

Consider the MA(1) process

$$y_t = z_t + \beta z_{t-1};\ |\beta| < 1$$

or

$$y_t = z_t - \beta z_{t-1}$$

$$y_t = (1 - \beta B) z_t$$

$$(1 - \beta B)^{-1} y_t = z_t$$

$$(1 + \beta B + \beta^2 B^2 + \ldots) y_t = z_t$$

$$z_t = \sum_{r=0}^{\infty} \beta^r y_{t-r}$$

which is an $AR(\infty)$. Since $|\beta| < 1$, the series converges. MA(1) process has $AR(\infty)$ representation. The process is thus said to be invertible. In other words, we can invert the function taking the z_t sequence to the y_t sequence and recover z_t from present and past values of y_t by a convergent sum.

3.3.3.3 Mixed Models

Let $\{y_t\}$ be expressed as

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + z_t + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \ldots + \beta_q z_{t-q}$$

$$y_t = \sum_{r=0}^{p} \alpha_r y_{t-1} + \sum_{s=0}^{q} \beta_s z_{t-s} \qquad (3.8)$$

where $\sum_{r=0}^{p} \alpha_r y_{t-r}$ is an autoregressive process and $\sum_{s=0}^{q} \beta_s z_{t-s}$ is a moving average process. Equation (3.8) is a mixture of AR(p) and MA(q) referred to as "Autoregressive Moving Average of order (p,q)". It is denoted by $ARMA(p,q)$.

From Equation (3.8),

$$\left(1-\alpha_1 B-\alpha_2 B^2-\ldots-\alpha_p B^2\right) y_t=\left(1+\beta_1 B+\beta_2 B+\ldots+\beta_q B^q\right) z_t$$

$$\alpha(B) y_t=\beta(B) z_t$$

where α and β are "polynomials in terms of *B* of degree *p* and *q"* respectively. For stationarity, the roots of $\alpha(B)=0$, that is, the values of $\{\alpha_i\}$ must lie outside the unit circle. Also, for "invertibility", the values of B to $\beta(B)=0$, that is, the values of $\{\beta_i\}$ must be greater than 1.

3.3.3.4 Integrated Processes

A series which is "non-stationary" may be made "stationary" after differencing *d* times. Such a series is said to be "integrated of order *d*", denoted by *I*(*d*).

Definition 3.3

If d is a non-negative integer, then $\{y_t\}$ is $ARIMA(p,d,q)$ process if $x_t=(1-B)^d y_t$ is a causal $ARMA(p,q)$.

Thus $\{y_t\}$ satisfies a difference equation of the form;

$$\alpha(B)(1-B)^d y_t=\beta(B) z_t$$

where z_t is a "white noise" with mean zero (0) and variance σ^2. $\{y_t\}$ is stationary if and only if $d=0$, that is $ARMA(p,q)$. An $ARIMA(p,d,q)$ model/process is never stationary because of the unit multiplicity.

The Role of the Constant term in ARMA and ARIMA

Let consider an ARMA model with a constant term (Gerolimetto, 2010):

$$\alpha(B)y_t = \beta_o + \beta(B)z_t$$

Applying expectation on both sides gives;

$$\alpha(B)\mu = \beta_o$$

$$\beta_o = \mu(1-\alpha_1-...-\alpha_p)$$

and it is the link between the constant term and the mean value of the process.

The following conclusions can be made;

a) If y_t is MA, that is the AR component does not exist, thus,

$E(y_t) = \mu = \beta_o$.

b) If y_t is only AR, then the above-mentioned relation holds.

If X_t is ARIMA, by including a constant term;

$$\alpha(B)\nabla y_t = \beta_o + \beta(B)z_t$$

This is called ARIMA model with drift. The simplest case is the Random Walk plus drift process:

$$\nabla y_t = \beta_o + z_t.$$

The first difference of y_t yields ∇y_t whose mean is not zero: a drift β_0 adjust for this. The idea is that the constant β_0 was the gradient of a deterministic trend, that after differencing $d = 1$ times, it leaves only a level (β_0) whiles ∇y_t moves with

"stationary" fluctuations. The "random walk" with drift consist of both "stochastic and deterministic trend".

Characteristics of ARIMA Processes

The following scenarios are considered

a) If $d = 0$, we have a "stationary process"

b) If $d = 1$, we have a "nonstationary process". The level has a constant change in time.

c) If $d = 2$, we have a nonstationary process. Both the level and increments are stationary.

3.3.3.5 The Box-Jenkins Seasonal (SARIMA) Model

"Time series", usually depict "seasonal periodic component" which repeats every *m* observations. For instance, with quarterly data where $m = 4$, y_t is expected to depend on the terms such as y_{t-4} and perhaps y_{t-8}, and terms such as $y_{t-1}, y_{t-2}, y_{t-3}, \ldots$ "Box and Jenkins" have generalized the ARIMA to deal with seasonality and defined a general multiplicity "Seasonal ARIMA (SARIMA) model" and denoted as $ARIMA(p,d,q)(P,D,Q)_m$ where *m* is the number of periods per season and defined using the backward shift operator as

$$\alpha(B)\Phi(B^m)\nabla^d\nabla^D y_t = \beta_o + \beta(B)\Theta(B^m)z_t$$

$$\alpha(B)\Phi(B^m)(1-B)^d(1-B^m)^D y_t = \beta_o + \beta(B)\Theta(B^m)z_t$$

where

$$\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p$$

$$\Phi(B^m) = 1 - \phi_1 B^m - \phi_2 B^{2m} - \ldots - \phi_P B^{Pm}$$

$$\beta(B) = 1 - \beta_1 B - \beta_2 B^2 - \ldots - \beta_q B^q$$

$$\Theta(B^m) = 1 - \theta_1 B^m - \theta_2 B^{2m} - \ldots - \theta_Q B^{Qm}$$

Also, *p,d,q* are the orders of "non-seasonal AR", "differencing", and MA respectively whiles *P,D,Q* are the orders of "seasonal AR", "differencing" and MA respectively.

For example, an $ARIMA(1,1,1)(1,1,1)_4$ model (without a constant) is for a quarterly data $(m = 4)$ and can be written as

$$(1 - \alpha_1 B)(1 - \phi_1 B)(1 - B)(1 - B^4) y_t = (1 + \beta_1 B)(1 + \theta_1 B^4) z_t$$

The choice of suitable values for the two orders of differencing (seasonal (*D*) and non-seasonal (*d*)), that addresses seasonality and stationarity in SARIMA models is the foremost thing to consider while fitting SARIMA models. You then proceed by fitting ARMA-type model to the differenced series with the added complication that there may be AR and MA terms at lags which are a multiple of the season length *m*. Values of *d*, *D*, *q* and *Q* need to be assessed by looking at the "autocorrelation function (ACF)" and "partial autocorrelation function (PACF)" of the differenced series and choosing a "SARIMA model" whose ACF and PACF are of similar form.

The estimation of the model involves three steps, namely: "identification", "estimation of parameters" and "diagnostics". The ACF and PACF are used in the

identification process to identify the tentative orders for both the "non-seasonal" and seasonal components of the model.

The "seasonal lags" of the PACF and ACF depicts the seasonal part of an AR or MA model. For example, an $ARIMA(0,0,0)(1,0,0)_4$ model will show exponential decay in the "seasonal lags" of the ACF and a single significant spike at lag 4 in the PACF. Similarly, an $ARIMA(0,0,0)(0,0,1)_4$ model will show a spike at lag 4 in the ACF but no other significant spikes and the PACF will show exponential decay in the "seasonal lags"; that is lags 4, 8, 12, 16,…

The second step involves estimation of the parameters of the tentative models that have been selected. In this study, the model with the minimum values of "Akaike Information Criterion (AIC)", "modified Akaike Information Criterion (AIC))", and "Bayesian Information Criterion (BIC)" is adjudged the best model. The last stage which is the diagnostic stage involves checking whether the selected model adequately represents the Currency in Circulation. An overall check of the model adequacy was made at this stage using the "Ljung-Box" test and ARCH-LM test. These tests were performed to check for higher order autocorrelation and homoscedasticity respectively.

Definition 3.4

The "seasonal difference" of a "time series" is the series of changes from one series to the next.

Let consider a quarterly data which there are 4 periods in a season for instance. The "seasonal difference" of *y* at a period *t* is $y_t - y_{t-4}$. Seasonal differencing is usually employed to remove seasonality whiles addressing most of the trend. It is a simple procedure of additive seasonal modification. It is attained by subtracting an index (the value that was observed in the same season one year earlier) from each value of the "time series".

Definition 3.5

The first difference of the "seasonal difference" of a monthly "time series" *y* at a period *t* is equal to $(y_t - y_{t-12}) - (y_{t-1} - y_{t-13})$ which can also be expressed as $(y_t - y_{t-1}) - (y_{t-12} - y_{t-13})$.

First difference of the "seasonal difference" is the amount by which the change from the previous period to the current period is different from the change that was observed exactly one year earlier.

The following information criteria were used in best model selection

Akaike Information Criteria (AIC) by Akaike (1978)

$$AIC = \frac{Log\hat{\sigma}_k^2 + n + 2k}{n}$$

where

$$\hat{\sigma}_k^2 = \frac{SSE_k}{n}$$

denotes the MLE for the error variance. K is the number of "seasonal and non-seasonal AR and MA" parameters to be estimated in the model. Wei (1990) says $k = p + q + P + q + 1$ and n is the number of observations.

AICc is AIC with a correction for the finite sample sizes. It given as (Burnham and Anderson, 2004);

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

where n and k are as defined.

AICc is basically AIC with a superior penalty for extra parameters. When n is not as large as compared to k^2, AIC increases the chance of selecting a model with too many parameters, that is, overfitting. However, if n is many times larger than k^2, then the correction will be negligible; hence, there will be negligible penalty in using AIC other than AICc.

Bayesian Information Criteria (BIC)

$$BIC = -2\ln(L) + k\ln(n)$$

where L is the likelihood function and k is the number of parameters (Stoffer and Dhumway, 2010).

Parameter Estimation

The next thing in Box-Jenkins methodology after an adequate model has been tentatively obtained is to estimate the parameters of the model. The Conditional Least Estimation (CLS) and Yule-Walker are two frequently used techniques in the parameter estimation process. The CLS estimates are conditional on the assumption

that the unobserved errors equal to zero (0). The series y_t can be represented in terms of the previous observations, as

$$y_t = z_t + \sum_{i=1}^{\infty} \pi_i B^i.$$

The weights (π) are computed from the ratio of the α and β polynomials as

$$\frac{\alpha(B)}{\beta(B)} = 1 - \sum_{i=1}^{\infty} \pi_i B^i.$$

The CLS method produces estimates minimizing

$$\sum_{i=1}^{n} \hat{z}_t^2 = \sum_{i=1}^{n} \left(y_t - \sum_{i=1}^{\infty} \pi_i y_t - i \right)^2$$

where the unobserved past values of X_t are set to zero (0) and $\hat{\pi}_i$ are computed from the estimates of α and β at each iteration. The *k*-step forecast of X_{t+k} is computed as

$$\hat{y}_{t+k} = \sum_{i=1}^{k-1} \hat{\pi}_i \hat{y}_{t+k-i} + \sum_{i=k}^{\infty} \hat{\pi}_i \hat{y}_{t+k-1}.$$

Model adequacy and Diagnostic Check

A conventional procedure in ARIMA modelling is to tentatively fit more than one model for a data and adequately select the best by avoiding overfitting. The parameters of the accepted model is then estimated and the goodness-of-fit assessed through diagnostic checks. Forecasting is then performed after the model has been validated.

In particular, the following has to be performed:

a) A study of the residual series obtained after fitting the model to the data to see if any pattern remains accounted for. The ACF and PACF plots of the residual series help in detecting any unaccounted pattern.

b) A study of the sampling statistics of the current optimum solution to check if any further simplification of the model is possible. The residuals left over after fitting an ARIMA model should ideally be just random noise (white noise) with zero mean and constant variance.

For model adequacy and diagnostic check, the ACF and PACF plots of the residuals as well as the Ljung-Box Chi-Square test was performed in this study.

ACF and PACF plots of the residuals

For an adequate model, the both the ACF PACF should not have any significant spikes (an indication of randomness and proper fit) at any lag order. However, in practice, there maybe a few spikes which are close to significance. A significant lag in every 20 lags is considered to by chance and thus an acceptable outcome.

Ljung-Box Chi-Square test

The most accepted statistical test for checking randomness of residuals is the Ljung-Box test which follows the Chi-Square distribution. The null hypothesis is that the set of "autocorrelations" for residuals is white noise is tested in this case. The test statistics is given by

$$\chi_m^2 = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-2}$$

where n is the size of sample, r_k is the sample "autocorrelation" at lag k, and m is the number of lags being tested.

3.3.4 Box-Cox Transformation

Heteroscedasticity affect the performance of an ARIMA model. This issue can persist even after differencing (which does not address non constancy of variance in the series). The issue of "non-stationarity" in variance is usually handled by transformation of the time series. A general form of transformation is the Box-Cox transformation, a well-accepted procedure that handles non constancy of variance in time series.

The original Box-Cox transform is given by

$$z_t(\lambda) = \begin{cases} \dfrac{y_t^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \log(y_t), & \lambda = 0 \end{cases}$$

The Box-Cox transformation allows reading a good level of symmetry, independence, and stable variance of random effects.

3.4 Vector Autoregressive Process (VAR)

Let $\{y_t\}_{t=1}^{T}$ be a realization of the k – dimensional vector autoregressive (VAR) process of lag length p:

$$\boldsymbol{y}_t = \phi_0 + \sum_{i=1}^{p} \boldsymbol{\Pi}_i \boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t, \qquad (3.10)$$

where $\boldsymbol{\varepsilon}_t \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, ϕ_0 is a k-dimensional constant vector and $\boldsymbol{\Pi}_i$ are $k \times k$ matrices for $i > 0, \boldsymbol{\Pi}_p \neq 0$.

Using the "backward shift operator", Equation (3.10) can be rewritten as

$$\phi(B)\boldsymbol{y}_t = \phi_0 + \boldsymbol{\varepsilon}_t \tag{3.11}$$

where $\phi(B) = \boldsymbol{I}_k - \sum_{i=1}^{p} \Pi_i B^i$ with $\Pi_p \neq 0$.

3.4.1 A VAR(1) Representation

Define $\boldsymbol{Y}_t = (\tilde{y}_t', \tilde{y}_{t-1}', ..., \tilde{y}_{t-p+1}')'$, which is a pk – dimensional time series. The VAR(p) model in Equation (3.11) can be written (without the constant) as (Tsay, 2014);

$$\begin{bmatrix} \tilde{y}_t' \\ \tilde{y}_{t-1}' \\ \tilde{y}_{t-2}' \\ \vdots \\ \tilde{y}_{t-p+1}' \end{bmatrix} = \begin{bmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_{p-1} & \Pi_p \\ \boldsymbol{I}_k & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_k & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_k & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \tilde{y}_{t-1}' \\ \tilde{y}_{t-2}' \\ \tilde{y}_{t-3}' \\ \vdots \\ \tilde{y}_{t-p}' \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}$$

$$\boldsymbol{Y_t} = \boldsymbol{\Phi Y_{t-1}} + \boldsymbol{u_t} \tag{3.12}$$

where $\boldsymbol{u}_t = \left(\boldsymbol{\varepsilon}_t', \boldsymbol{0}'\right)'$. It must be noted that the **0** is a zero vector with dimension $k(p-1)$, and $\boldsymbol{\Phi}$ known conventionally as the companion matrix is $kp \times kp$-dimensional matrix given by

$$\boldsymbol{\Phi} = \begin{bmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_{p-1} & \Pi_p \\ \boldsymbol{I}_k & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_k & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_k & \boldsymbol{0} \end{bmatrix}.$$

Here, $\boldsymbol{I}_k$ is the identity matrix whiles $\mathbf{0}$ is zero matrix, with a dimension of $k \times k$ each. Also, $\boldsymbol{u}_t$ has a covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, with all its element zero except those in the upper-left corner.

3.4.2 Stationarity Condition

From Equation 3.12, $\boldsymbol{Y}_t$ follows a VAR(1) model and is stationary if the roots of the determinant equation

$$\left|\boldsymbol{I}_{kp} - \boldsymbol{\Phi B}\right| = 0$$

are greater than 1 in absolute value.

Lemma 3.1

For the $k \times k$ matrix polynomial $\phi(B) = \boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\Pi}_i B^i$,

$$\left|\boldsymbol{I}_{kp} - \boldsymbol{\Phi B}\right| = \left|\boldsymbol{I}_k - \boldsymbol{\Pi}_1 B - \cdots \boldsymbol{\Pi}_p B^p\right|$$

holds, where $\boldsymbol{\Phi}$ is defined in Equation (3.12).

Proof of Lemma (Tsay, 2014)

$$\left|\boldsymbol{I} - \boldsymbol{\Phi B}\right| = \begin{bmatrix} \boldsymbol{I_k} - \boldsymbol{\Pi_1 B} & -\boldsymbol{\Pi_2 B} & \cdots & -\boldsymbol{\Pi}_{p-1}\boldsymbol{B} & -\boldsymbol{\Pi}_p \boldsymbol{B} \\ -\boldsymbol{I_k B} & \boldsymbol{I_k} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{I_k B} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & -\boldsymbol{I_k B} & \boldsymbol{I_k} \end{bmatrix}$$

If the last, pth, column is multiplied by B and the result added to the preceding column, that is, the $(p-1)th$ column,

$$|\boldsymbol{I}-\boldsymbol{\Phi}B|=\begin{bmatrix} \boldsymbol{I_k}-\boldsymbol{\Pi}_1 B & -\boldsymbol{\Pi}_2 B & \cdots & -\boldsymbol{\Pi}_{p-1}\boldsymbol{B}-\boldsymbol{\Pi}_p B^2 & -\boldsymbol{\Pi}_p B \\ -\boldsymbol{I_k}B & \boldsymbol{I_k} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I_k}B & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I_k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I_k} \end{bmatrix}$$

Also, the $(p-1)th$ column multiplied by B and added to the $(p-2)th$ column, gives $-\boldsymbol{\Pi}_{p-2}B-\boldsymbol{\Pi}_{p-1}B^2-\boldsymbol{\Pi}_p B^3$ at the (1, $p-2$) block and **0** at ($p-1$, $p-2$) block. Recursively, we obtain

$$|\boldsymbol{I}-\boldsymbol{\Phi}B|=\begin{bmatrix} \boldsymbol{\phi}(B) & -\sum_{j=2}^{p}\boldsymbol{\Pi}_j B^{j-1} & \cdots & -\boldsymbol{\Pi}_{p-1}B-\boldsymbol{\Pi}_p B^2 & -\boldsymbol{\Pi}_p B \\ \boldsymbol{0} & \boldsymbol{I_k} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I_k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I_k} \end{bmatrix}=|\boldsymbol{\phi}(B)|.$$

By Lemma, we have $|\boldsymbol{I}_{kp}-\boldsymbol{\Phi}(B)|=|\boldsymbol{\phi}(B)|$ for the VAR(p) time series. Thus, the VAR(p) is stable if

$$|\boldsymbol{I}_k-\boldsymbol{\Pi}_1 B-\ldots-\boldsymbol{\Pi}_p B^p|=0 \qquad (3.13)$$

in modules has all solutions greater than 1.

The "Wold moving average (MA)" representation is given as

$$\boldsymbol{y}_t=\Phi_o\boldsymbol{\varepsilon}_t+\Phi_1\boldsymbol{\varepsilon}_{t-1}+\Phi_2\boldsymbol{\varepsilon}_{t-2}+\ldots$$

with $\Phi_o=\boldsymbol{I}_k$, and the Φ_s matrices can be computed recursively according to

$$\Phi_s = \sum_{j=1}^{s} \Phi_{s-j}\Pi_j \quad \text{for } s = 1, 2, ...,$$

where $\Phi_o = \boldsymbol{I}_k$ and $\Pi_j = 0$ for $j > p$.

3.4.3 Periodic VAR Processes

A cointegrated VAR models though having time invariant coefficients results in a "non-stationary processes" with possibly time varying first and second moments. It is appropriate to model the "non-stationarity" in the framework of time varying coefficient methods. If a significant seasonality is observed in the time series data to be modelled, a VAR(*p*) model with a seasonal intercept can be estimated for each season (Lütkepohl, 2005):

$$y_t = d_t + \sum_{i=1}^{p} \Pi_i y_{t-i} + \boldsymbol{\varepsilon}_t \tag{3.14}$$

Here, d_i associated with the *ith* season has a dimension of $(k \times 1)$. In other words, in Equation (3.14), *t* is the time index assumed to be associated with the *ith* season of the year. Clearly, the process has a possibly different expectation for a season (for instance, a quarter) of the year.

Assuming *m* seasons, Equation (3.14) could be written alternatively as

$$y_t = n_{1t}d_1 + ... + n_{mt}d_m + \sum_{i=1}^{p} \Pi_i y_{t-i} + \boldsymbol{\varepsilon}_t,$$

where

$$n_{it} = 0 \text{ or } 1 \text{ and } \sum_{i=1}^{m} n_{it} = 1.$$

In this case, n_{it} is a seasonal dummy variable.

3.4.4 The VAR(*p*) Model with Time Varying Coefficients

For a k-dimensional VAR(*p*) model with "time varying coefficients" (Lütkepohl, 2005);

$$y_t = d_t + \sum_{i=1}^{p} \Pi_{it} y_{t-i} + \varepsilon_t,\ t \in \mathbb{Z} \tag{3.15}$$

3.4.4.1 Maximum Likelihood (ML) Estimation

Equation (3.15) can be rewritten as (Lütkepohl, 2005);

$$y_t = \boldsymbol{B}_t \boldsymbol{Z}_{t-1} + \boldsymbol{u}_t \tag{3.16}$$

where $\boldsymbol{B}_t = (d_t, \Pi_{1t}, ..., \Pi_{pt})$, $\boldsymbol{Z}_{t-1} = (1, Y'_{t-1})'$. Here, $\boldsymbol{B}_t$ is $(k \times (kp+1))$ dimensional and a function of $\boldsymbol{\gamma}$, that is, $\boldsymbol{B}_t = \boldsymbol{B}_t(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a $(N \times 1)$ dimensional vector of fixed, time invariant parameters. In addition, Σ_t is also a function of $\boldsymbol{\sigma}$, an $(M \times 1)$ dimensional vector of fixed parameters. It must also be noted that $\boldsymbol{\sigma}$ is disjoint of and unrelated with $\boldsymbol{\gamma}$.

Assuming that $\boldsymbol{u}_t \sim N(\boldsymbol{0}, \Sigma_t)$, the log-likelihood function of Equation 3.15 is

$$\ln l(\boldsymbol{\gamma}, \boldsymbol{\sigma}) = -\frac{kT}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^{T} \ln |\boldsymbol{\Sigma}_t| - \frac{1}{2} \sum_{t=1}^{T} \boldsymbol{u}_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t, \tag{3.17}$$

for all initial condition terms not considered. The normal equations are

$$0 = \frac{\partial \ln l}{\partial \boldsymbol{\gamma}} = -\sum_{t=1}^{T} \frac{\partial \boldsymbol{u}_t'}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t$$

$$= -\sum_{t=1}^{T} \frac{\partial (y_t - \boldsymbol{B}_t \boldsymbol{Z}_{t-1})'}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t$$

$$= \sum_{t=1}^{T} \frac{\partial vec(\boldsymbol{B}_t)'}{\partial \boldsymbol{\gamma}} (\boldsymbol{Z}_{t-1} \otimes \boldsymbol{I}_k) \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t$$

$$= \sum_{t=1}^{T} \frac{\partial vec(\boldsymbol{B}_t)'}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t \boldsymbol{Z}_{t-1}' \tag{3.18}$$

and

$$0 = \frac{\partial \ln l}{\partial \boldsymbol{\sigma}} = -\frac{1}{2} \sum_t \left[\frac{\partial vec(\boldsymbol{\Sigma}_t)'}{\partial \boldsymbol{\sigma}} \frac{\partial \ln |\boldsymbol{\Sigma}_t|}{\partial vec(\boldsymbol{\Sigma}_t)} \right] - \frac{1}{2} \sum_t \left[\frac{\partial vec(\boldsymbol{\Sigma}_t)'}{\partial \boldsymbol{\sigma}} \frac{\partial u_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t}{\partial vec(\boldsymbol{\Sigma}_t)} \right]$$

$$= -\frac{1}{2} \sum_t \left[\frac{\partial vec(\boldsymbol{\Sigma}_t)'}{\partial \boldsymbol{\sigma}} vec(\boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t \boldsymbol{u}_t' \boldsymbol{\Sigma}_t^{-1} \right]. \tag{3.19}$$

The second partial derivatives of Equation (3.17) with respect to $\boldsymbol{\gamma}$ gives

$$\frac{\partial^2 \ln l}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = -\sum_t \left[\frac{\partial vec(\boldsymbol{B}_t)'}{\partial \boldsymbol{\gamma}} (\boldsymbol{Z}_{t-1} \otimes \boldsymbol{I}_k) \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{Z}_{t-1}' \otimes \boldsymbol{I}_k) \frac{\partial vec(\boldsymbol{B}_t)}{\partial \boldsymbol{\gamma}'} \right] + \text{terms with mean zero}$$

$$\frac{\partial^2 \ln l}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = -\sum_t \left[\frac{\partial vec(\boldsymbol{B}_t)'}{\partial \boldsymbol{\gamma}} (\boldsymbol{Z}_{t-1} \boldsymbol{Z}_{t-1}' \otimes \boldsymbol{\Sigma}_t^{-1}) \frac{\partial vec(\boldsymbol{B}_t)}{\partial \boldsymbol{\gamma}'} \right] + \text{terms with mean zero} \tag{3.20}$$

The $\partial vec(\boldsymbol{\Sigma}_t)' / \partial \boldsymbol{\sigma}$ is assumed to be independent of $\boldsymbol{\sigma}$, hence, a second partial derivatives of $\boldsymbol{\Sigma}_t$ wit respect to $\boldsymbol{\sigma}$, gives

$$\frac{\partial^2 \ln l}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'} = -\frac{1}{2}\sum_t \left(\frac{\partial vec(\boldsymbol{\Sigma}_t)'}{\partial \boldsymbol{\sigma}} \left[\frac{\partial vec(\boldsymbol{\Sigma}_t^{-1})}{\partial \boldsymbol{\sigma}'} - (\boldsymbol{I}_k \otimes \boldsymbol{\Sigma}_t^{-1}\boldsymbol{u}_t\boldsymbol{u}_t') \frac{\partial vec(\boldsymbol{\Sigma}_t^{-1})}{\partial \boldsymbol{\sigma}'} - (\boldsymbol{\Sigma}_t^{-1}\boldsymbol{u}_t\boldsymbol{u}' \otimes_t \boldsymbol{I}_k) \frac{\partial vec(\boldsymbol{\Sigma}_t^{-1})}{\partial \boldsymbol{\sigma}'} \right] \right)$$

$$= \frac{1}{2}\sum_t \left[\frac{\partial vec(\boldsymbol{\Sigma}_t)'}{\partial \boldsymbol{\sigma}} (\boldsymbol{\Sigma}_t^{-1} \otimes \boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1} \otimes \boldsymbol{\Sigma}_t^{-1}\boldsymbol{u}_t\boldsymbol{u}_t'\boldsymbol{\Sigma}_t^{-1} \right.$$

$$\left. -\Sigma_t^{-1}u_t u_t'\Sigma_t^{-1} \otimes \Sigma_t^{-1}) \frac{\partial vec(\boldsymbol{\Sigma}_t)}{\partial \boldsymbol{\sigma}'} \right]. \qquad (3.21)$$

Under the present assumptions, it can be observed that

$$E[\partial^2 \ln l / \partial \boldsymbol{\gamma} \partial \boldsymbol{\sigma}'] = 0.$$

Also, the "information matrix" becomes

$$I(\boldsymbol{\gamma}, \boldsymbol{\sigma}) = E\left[\frac{\partial^2(-\ln l)}{\partial \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\sigma} \end{bmatrix} \partial(\boldsymbol{\gamma}', \boldsymbol{\sigma}')} \right] = -E\left[\frac{\partial^2(\ln l)}{\partial \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\sigma} \end{bmatrix} \partial(\boldsymbol{\gamma}', \boldsymbol{\sigma}')} \right]$$

$$= \begin{bmatrix} \sum_t \left[\frac{\partial vec(\boldsymbol{B}_t)'}{\partial \boldsymbol{\gamma}} \left[E(\boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}') \otimes \boldsymbol{\Sigma}_t^{-1} \right] \frac{\partial vec(\boldsymbol{B}_t)}{\partial \boldsymbol{\gamma}'} \right] & 0 \\ 0 & \frac{1}{2}\sum_t \left[\frac{\partial vec(\boldsymbol{\Sigma}_t)'}{\partial \boldsymbol{\sigma}} (\boldsymbol{\Sigma}_t^{-1} \otimes \boldsymbol{\Sigma}_t^{-1}) \frac{\partial vec(\boldsymbol{\Sigma}_t)}{\partial \boldsymbol{\sigma}'} \right] \end{bmatrix}. \qquad (3.22)$$

3.4.5 Maximum Likelihood Estimation and Testing for Time Varying Coefficients

In this section, various forms of coefficients of VAR(p) model is addressed. Considering a "periodic VAR model", the "normal equations" given in Equation (3.18) reduces to

$$0 = \frac{\partial \ln l}{\partial \boldsymbol{\gamma}} = \sum_{t=1}^{T} \frac{\partial vec(n_{1t}\boldsymbol{B}_1 + ... + n_{mt}\boldsymbol{B}_m)'}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t \boldsymbol{Z}_{t-1}'$$

$$= \sum_{i=1}^{m} \sum_{t=1}^{T} n_{it} \frac{\partial vec(\boldsymbol{B}_i)'}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{u}_t \boldsymbol{Z}_{t-1}', \tag{3.23}$$

where $\boldsymbol{B}_i = \left[d_i, \boldsymbol{\Pi}_{1i}, ..., \boldsymbol{\Pi}_{pi} \right] = \left[d_i, \boldsymbol{\Pi}_i \right]$, $i = 1, ..., m$, and

$$\boldsymbol{\Sigma}_t^{-1} = \left(\sum_{i=1}^{m} n_{it} \boldsymbol{\Sigma}_i \right)^{-1} = \sum_{t} n_{it} \boldsymbol{\Sigma}_i^{-1}$$

Has been used. Moreover, Equation (3.19) reduces to

$$0 = \frac{\partial \ln l}{\partial \sigma} = -\frac{1}{2} \sum_{i=1}^{m} \sum_{t} n_{it} \left[\frac{\partial vec(\boldsymbol{\Sigma}_i)'}{\partial \boldsymbol{\sigma}} vec(\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \boldsymbol{u}_t \boldsymbol{u}_t' \boldsymbol{\Sigma}_i^{-1} \right]. \tag{3.24}$$

3.4.5.1 All Coefficients Time Varying

We consider a scenario where all coefficients of a periodic "VAR(p) model" are "time varying", that is

$$H_1 : \boldsymbol{B}_t = [d_t, \boldsymbol{\Pi}_t] = \sum_{i=1}^{m} n_{it} \boldsymbol{B}_i, \ \boldsymbol{\Sigma}_t = \sum_{i=1}^{m} n_{it} \boldsymbol{\Sigma}_i.$$

For this case,

$$\boldsymbol{\gamma} = vec[\boldsymbol{B}_1, ..., \boldsymbol{B}_m]$$

and

$$\boldsymbol{\sigma} = [vech(\boldsymbol{\Sigma}_1)', ..., vech(\boldsymbol{\Sigma}_m)'].$$

The "ML estimators" can be obtained from Equations (3.23) and (3.24) as

$$\tilde{\boldsymbol{B}}_i^{(1)} = \left(\sum_{t=1}^{T} n_{it} y_t \boldsymbol{Z}'_{t-1} \right) \left(\sum_{t=1}^{T} n_{it} \boldsymbol{Z}_{t-1} \boldsymbol{Z}'_{t-1} \right)^{-1} \tag{3.25}$$

and

$$\tilde{\boldsymbol{\Sigma}}_i^{(1)} = \sum_i n_{it} (y_t - \tilde{\boldsymbol{B}}_i^{(1)} \boldsymbol{Z}_{t-1})(y_t - \tilde{\boldsymbol{B}}_i^{(1)} \boldsymbol{Z}_{t-1})' / T\bar{n}_i, \tag{3.26}$$

for $i = 1,...,m.$ Here, $\bar{n}_i = \sum_{t=1}^{T} n_{it} / T$. The corresponding maximum of the log-likelihood function is

$$\lambda_1 = -\frac{1}{2} \sum_t \ln \left| \tilde{\boldsymbol{\Sigma}}_t^{(1)} \right| = -\frac{1}{2} T (\bar{n}_1 \ln \left| \tilde{\boldsymbol{\Sigma}}_1^{(1)} \right| + ... + \bar{n}_m \ln \left| \tilde{\boldsymbol{\Sigma}}_m^{(1)} \right|) \tag{3.27}$$

without an additive constant.

3.4.5.2 All Coefficients Time Invariant

For a basic stationary VAR(p) model, where all the coefficients are time invariant;

$$H_2 : \boldsymbol{B}_i = \boldsymbol{B}_1, \ \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_1, \quad i = 2,...,m.$$

For this case, the "ML estimators" are

$$\tilde{\boldsymbol{B}}_1^{(2)} = \left(\sum_{t=1}^{T} y_t \boldsymbol{Z}'_{t-1} \right) \left(\sum_{t=1}^{T} \boldsymbol{Z}_{t-1} \boldsymbol{Z}'_{t-1} \right)^{-1} \tag{3.28}$$

and

$$\tilde{\boldsymbol{\Sigma}}_1^{(2)} = \sum_i (y_t - \tilde{\boldsymbol{B}}_1^{(2)} \boldsymbol{Z}_{t-1})(y_t - \tilde{\boldsymbol{B}}_1^{(2)} \boldsymbol{Z}_{t-1})' / T. \tag{3.29}$$

Also, without an additive constant, the corresponding maximum log-likelihood is

$$\lambda_2 = -\frac{1}{2} T \ln \left| \tilde{\Sigma}_1^{(2)} \right|. \tag{3.30}$$

The LR test statistic is given as

$$\lambda_{LR} = 2[\ln l(\tilde{\delta}) - \ln l(\tilde{\delta}_r)],$$

where $\tilde{\delta}_r$ and $\tilde{\delta}$ are the restricted and unrestricted ML estimators respectively.

The LR statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of linearly independent restrictions.

In testing H_2 (null hypothesis) versus H_1 (alternative hypothesis), the LR statistic is given as $2(\lambda_1 - \lambda_2)$ with degrees of freedom $(m-1)k[k(p+\frac{1}{2})+\frac{3}{2}]$.

3.4.6 Selecting VAR Model Order

VAR model order selection is a key requirement when fitting VAR model. The appropriate order needs to be selected for model adequacy whiles avoiding overfitting. In this study, the following "information criterion" were used in the order selection of the VAR model (Lütkepohl, 2005).

The "Akaike Information Criterion (AIC)" by Akaike (1973),

$$AIC(p) = \ln \left| \hat{\Sigma}_{a,p} \right| + \frac{2}{T} pk^2;$$

the "Bayesian Information Criterion (BIC)" or "Schwarz criterion (SC)" by Schwarz (1978),

$$BIC(p) = SC(p) = \ln\left|\hat{\Sigma}_{a,p}\right| + \frac{\ln(T)}{T} pk^2 ;$$

the Hanan and Quinn (HQ) criterion by Hanan and Quinn (1979) and Quinn (1980) $HQ(p) = \ln\left|\hat{\Sigma}_{a,p}\right| + \frac{2\ln[\ln(T)]}{T} pk^2 ;$

and the "Final prediction Error (FPE)" criterion

$$FPE(p) = \left[\frac{T + pk + 1}{T - pk - 1}\right]^k \left|\hat{\Sigma}_{a,p}\right|.$$

Here, $\hat{\Sigma}_{a,p}$ is the ML estimate of Σ_a and T is the sample size.

3.4.7 Model Diagnostics

When a model, such as the vector autoregressive model, is considered, it adequacy and goodness-of-fit is of foremost importance. Any one, or several, of the features of the model, such as "serial correlation", constancy of error variance, normality of the error terms, may not be appropriate for the particular data at hand. After the model is validated, an application of it is then considered.

3.4.7.1 Test of Constancy of Variance

The study adopted the multivariate ARCH test for the tests for heteroscedasticity in the in the error terms (Engle, 1982, and Hamilton, 1994).

The Multivariate ARCH – LM Test

The test is based on the following regression

$$vech(\hat{\varepsilon}_t\hat{\varepsilon}_t') = \beta_o + B_1 vech(\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-1}') + ... + B_q vech(\hat{\varepsilon}_{t-q}\hat{\varepsilon}_{t-q}') + v_t$$

Here, β_0 and B_i (for $i = 1,...,q$) have dimensions $\frac{1}{2}k(k+1)$ and $\frac{1}{2}k(k+1)\times\frac{1}{2}k(k+1)$ respectively where v_t assigns a "spherical error process". The following hypothesis is tested:

$$H_o : B_1 = B_2 = ... = B_q \neq 0$$

$$H_1 : B_1 \neq 0 \cap B_2 \neq 0 \cap ... \cap B_q \neq 0$$

The test statistics is defined as;

$$VARCH_{LM}(q) = \frac{1}{2}TK(K+1)R_m^2$$

with

$$R_m^2 = 1 - \frac{2}{k(k+1)} tr(\hat{\boldsymbol{\Omega}}\boldsymbol{\Omega}_o^{-1}),$$

where $\hat{\boldsymbol{\Omega}}$ is the covariance matrix. The test statistic follows the Chi-Square distribution with degrees of freedom $qK^2(K+1)^2/4$ (Pfaff and Taunus, 2008).

3.4.7.2 Test of Normality

In this study, the residuals of the VAR(*p*) model is tested for normality, skewness and kurtosis by adopting Jarque-Bera normality test proposed by Bera and Jarque, (1980 and 1987).

The test statistics are defined as:

$$JB_{mv} = s_3^2 + s_4^2,$$

where

$$s_3^2 = \frac{T\boldsymbol{b}_1'\boldsymbol{b}_1}{6}$$

$$s_4^2 = \frac{T(\boldsymbol{b}_2 - 3k)'(\boldsymbol{b}_2 - 3k)}{24}.$$

Here, $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are respectively the third and fourth non-central moment vectors of the standardized residuals

$$\hat{\boldsymbol{\varepsilon}}_t^s = \tilde{P} - (\hat{\boldsymbol{\varepsilon}}_t - \bar{\hat{\boldsymbol{\varepsilon}}}_t).$$

The test statistics JB_{mv}, s_3^2, and s_4^2 follows the Chi-Square distribution with $2k$, k, and k degrees of freedom respectively (Pfaff and Taunus, 2008).

3.4.7.3 Test of Serial Correlation

"Serial correlation" is an undesired property in a multivariate time series data analysis. "Multivariate Portmanteau test" (Ljung and Box, 1978), and the LM test proposed by Breusch and Godfrey (1987) were employed to test the residuals of a VAR(p) model for the existence of serial serial correlation.

The Multivariate Portmanteau Test

The Portmanteau statistic is defined as;

$$Q_\ell = T\sum_{j=1}^{\ell} tr\left(\hat{C}_j' C_o^{-1} \hat{C}_j \hat{C}_o^{-1}\right),$$

with

$$\hat{C}_i = \frac{1}{T}\sum_{t=i+1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_{t-i}'.$$

which follows the χ^2 distribution with $k^2\ell - n$ degrees of freedom. Here, *n* and is the number of coefficients of the VAR(*p*) model excluding the deterministic terms. The small sample properties of the test statistic is

$$Q_\ell^* = T^2 \sum_{j=1}^{\ell} \frac{1}{T-j} tr\left(\hat{C}_j' C_o^{-1} \hat{C}_j \hat{C}_o^{-1}\right).$$

Breusch – Godfrey LM Test

The test (Breusch, 1978, and Godfrey, 1978), is based upon the auxiliary regression:

$$\hat{\varepsilon}_t = A_1 y_{t-1} + ... + A_p y_{t-p} + CD_t + B_1 \hat{\varepsilon}_{t-1} + ... + B_\ell \hat{\varepsilon}_{t-\ell} + u_t$$

The following hypothesis is tested:

$H_o : B_1 = ... = B_\ell = 0$

$H_1 : B_i \neq 0$ for $i = 1, 2, ..., \ell.$

The test statistic is given as

$$LM_\ell = T(k - tr(\tilde{\Sigma}_R^{-1} \tilde{\Sigma}_e)),$$

where $\tilde{\Sigma}_R$ and $\tilde{\Sigma}_e$ are the restricted and unrestricted model residual covariance matrix respectively. The test statistics is follows χ^2 with ℓk^2 degrees of freedom.

The sample correction is (Edgerton and Shukur, 1999);

$$LMF_\ell = \frac{1-(1-R_r^2)^{1/r}}{(1-R_r^2)^{1/r}} \frac{N_r - q}{km},$$

with

$$R_r^2 = 1 - \frac{\left|\tilde{\Sigma}_e\right|}{\left|\tilde{\Sigma}_R\right|},\ r = \sqrt{\frac{(k^2m^2-4)}{(k^2m^2-5)}},\ q = \frac{1}{2}km - 1,\ \text{and } N = T - k - m - \frac{1}{2}(k-m+1),\ \text{and } m = kh.$$

where *n* denote the predictors and LMF_ℓ is distributed as $F(\ell k^2, \text{int}(N_r - q))$.

"Structural stability" can also be assessed by exploring the empirical fluctuation process (Zeileis, Leisch, Kleiber and Hornik, 2005; Kuan and Hornik (1995).

3.4.8 Forecasting

One key application of a validated model is forecasting. However, detection of the dynamic relationship amid variables is a basic purpose for the adoption VAR analysis. For a given empirical VAR, forecasts can be calculated recursively according to

$$\boldsymbol{y}_{T+h|T} = Cd_t + \Pi_1 \boldsymbol{y}_{T+h-1|T} + \ldots + \Pi_p \boldsymbol{y}_{T+h-p|T}$$

for $h = 1, 2, \ldots, n.$ The forecast error covariance matrix is given as

$$\text{cov}\begin{pmatrix} \boldsymbol{y}_{T+1} & \boldsymbol{y}_{T+1|T} \\ & \vdots \\ \boldsymbol{y}_{T+h} & \boldsymbol{y}_{T+h|T} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & 0 & \cdots & 0 \\ \Phi_1 & \boldsymbol{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \cdots & \boldsymbol{I} \end{pmatrix} (\Sigma_\varepsilon \otimes \boldsymbol{I}_h) \begin{pmatrix} \boldsymbol{I} & 0 & \cdots & 0 \\ \Phi_1 & \boldsymbol{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \cdots & \boldsymbol{I} \end{pmatrix},$$

where Φ_i are matrices of coefficient matrices of the Wold moving average representation of a stable VAR(*p*) process. The estimates of the confidence interval is obtained as

$$\left[y_{k,T+h|T} - c_{1-\gamma/2}\sigma_k(h), y_{k,T+h|T} + c_{1-\gamma/2}\sigma_k(h) \right],$$

where $\sigma_k(h)$ is the standard deviation of the kth variable h steps ahead and $c_{1-\gamma/2}$ indicates the $(1-\gamma/2)\%$ point of the normal distribution.

3.4.9 Causality Analysis

"Granger causality" test, Granger, 1969), is the frequently used test for exploring the causality among variables included in a VAR analysis. It is primary purpose for fitting a VAR model and the interest of researchers to establish the causality between the variables. When causality is detected, care must be taken in it interpretation thereof. We say that variable *y* granger-causes variable *x* if variable *y* helps in the prediction of the variable *x*. According to Lütkepohl, (2005), the test for non-zero correlation between the error processes of the cause and effect variables is carried out in the process. Furthermore, the vector of endogenous variables y_t is usually, split into two sub-vectors y_{1t} $(k_1 \times 1)$ and y_{2t} $k = k_1 + k_2$. Thus, the VAR(p) is rewritten as

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \Pi_{11,i} & \Pi_{12,i} \\ \Pi_{21,i} & \Pi_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + \boldsymbol{C} d_t + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \tag{3.31}$$

The following hypothesis is then tested;

$H_0 : \Pi_{21,i} = 0$ for $i = 1, 2, ..., p$ (that is y_{1t} does not Granger-cause y_{2t}).

$H_1 : \Pi_{21,i} \neq 0$ for $i = 1, 2, ..., p$.

The test statistic is distributed as $F(pk_1k_2, kT - n^*)$, where n^* is the total number of parameters in the model in Equation (3.31).

Also, the non-instantaneous causality is tested using the following hypothesis

$H_0 : C\sigma = 0$ versus

$H_1 : C\sigma \neq 0$

Here, *C* is a matrix of dimension $(N \times k(k+1)/2)$ rank *N*.

The Wald statistic is defined as

$$\lambda_w = T\tilde{\boldsymbol{\sigma}}\boldsymbol{C}\left[2\boldsymbol{C}\boldsymbol{D}_k^{+}(\tilde{\boldsymbol{\Sigma}}_{\varepsilon} \otimes \tilde{\boldsymbol{\Sigma}}_{\varepsilon})\boldsymbol{D}_k^{+\prime}\boldsymbol{C}'\right]^{-1}\boldsymbol{C}\tilde{\boldsymbol{\sigma}},$$

where the Moore-Penrose inverse of the duplication matrix $\boldsymbol{D}_k$ is assigned by $\boldsymbol{D}_k^{+}$ and $\tilde{\boldsymbol{\Sigma}}_{\varepsilon} = \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'$. The duplication matrix $\boldsymbol{D}_k$ has dimension $(k^2 \times \frac{1}{2}k(k+1))$ and is defined such that, for any symmetry $(k \times k)$ matrix $\boldsymbol{A}$, $vec(\boldsymbol{A}) = \boldsymbol{D}_k vech(\boldsymbol{A})$ holds. The test statistic λ_w is asymptotically distributed as $\chi^2(N)$ (Pfaff and Taunus, 2008).

3.4.10 Impulse Response Functions

In the causality analysis, the impact of the impulse variable on the response variable over time cannot be quantified. This is addressed by impulse response function. The impulse response analysis, based upon the Wold moving average representation of a VAR(*p*)-process, is used to explore these kinds of dynamic relations amid the endogenous variables.

The impulse response are the coefficients of the matrices Φ_s. The expected response of variable $y_{i,t+s}$ for a unit change in y_{jt} is measured by the $(i, j)th$ coefficients Φ_s. The effect of a variable on the other is usually cumulated through time, that is, $s = 1, 2, \ldots,$ (for quarterly periods for example).

The "Choleski decomposition" of the error variance-covariance matrix, that is, $\Sigma_\varepsilon = PP'$, is the orthogonal impulse responses. Here, the matrix P is lower triangular. Its MA depiction can then be transformed to

$$y_t = \Psi_o \tau_t + \Psi_1 \tau_{t-1} + \ldots \text{ for } i = 0,1,2,\ldots, \tau_i = P^{-1}\varepsilon_t, \Psi_i = \Phi_i P, \text{ and } \Psi_0 = P.$$

3.4.11 Forecast Error Variance Decomposition

The "forecast error variance decomposition (FEVD)" is based upon the orthogonal impulse response coefficient Ψ_n matrices. The FEVD measures contribution of variable *j* to the *h*-step forecast error variance of variable *k*. Formally, the FEVD for $y_{k,T+h} - Y_{k,T+h|T}$ is defined as

$$\sigma_k^2(h) = \sum_{n=0}^{h-1} (\psi_{k1,n}^2 + \ldots + \psi_{kK,n}^2),$$

which can be written as

$$\sigma_k^2(h) = \sum_{j=1}^{K} (\psi_{kj,0}^2 + \ldots + \psi_{kj,h-1}^2). \qquad (**)$$

Dividing through (**) by $\sigma_k^2(h)$ yields the "forecast error variance decompositions" in percentage terms:

$$\omega_{kj}(h) = (\psi_{kj,0}^2 + \ldots + \psi_{kj,h-1}^2) / \sigma_k^2(h).$$

3.5 The Vector Error Correction Model (VECM)

If Equation (3.13) has a unit root, then some or all of the variables in y_t are $I(1)$ and they may also be cointegrated. Equation (3.10) can then be written in VECM form as:

$$\Delta y_t = \phi_0 + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t \tag{3.32}$$

where

$$\Pi = \Pi_1 + \ldots + \Pi_p - I_k$$

and

$$\Gamma_\tau = -\sum_{j=\tau+1}^{p} \Pi_j, \quad \tau = 1, \ldots, p-1.$$

The Π and Γ_τ matrices are the "long-run and short-run" impact matrices respectively. From Equation (3.32), the VAR parameters, Π_i, can be estimated as;

$$\left.\begin{aligned} \Pi_1 &= \Gamma_1 + \Pi + I_k \\ \Pi_\tau &= \Gamma_\tau - \Gamma_{\tau-1}, \tau = 2, \ldots, p \end{aligned}\right\} \tag{3.33}$$

In the Equation (3.32), the term Δy_t and its past values are stationary, that is, $I(0)$). However, (1*)

$$\Pi y_{t-1} \tag{1*}$$

is the only term that is likely to have variables that are non-stationary, that is, $I(1)$. For us to have a stationary model (that is, Δy_t being $I(0)$), the term (1*) must be $I(0)$. It therefore suffices that, if cointegration relation exist, it is contained in (1*).

If the unit root test in Equation (3.13) indicates the presence of non-stationarity (that is, if the there is unit root in the VAR model), then the matrix, Π is a singular and thus has a reduced (that is, rank(Π) = $r < n$). When this happens, the following case are possible;

a) If $rank(\Pi) = 0$, it is an indication that the matrix, $\Pi = 0$. Thus, y_t, though $I(1)$ is not cointegrated. Thus Equation (3.32) becomes

$$\Delta y_t = \phi_0 + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t .$$

That is, a first differenced VAR($p - 1$).

b) If $0 < rank(\Pi) = r < k$, (which is the desire of every researcher), y_t is cointegrated and the number of linearly independent cointegration relations is r $k - r$ common stochastic trends.

Based on the presence of cointegration relations, the matrix, Π is usually rewritten as the product

$$\underset{(k\times k)}{\Pi} = \underset{(k\times r)}{\alpha}\ \underset{(r\times k)}{\beta'} \qquad (2^*)$$

where α and β are (kxr) matrices with $rank(\alpha) = rank(\beta) = r$.

The rows of β' form a basis for the r "cointegrating vectors" whiles α distribute the "long-run impact" of the "cointegrating vectors" to the variations of Δy_t. Equation (3.32) becomes

$$\Delta y_t = \phi_0 + \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t \qquad (3.34)$$

where $\beta' y_{t-1} \sim I(0)$ since β' is a matrix of "cointegrating vectors".

It is important to note that the factorization (2*) is not unique since for any rxr nonsingular matrix **H** we have

$$\alpha\beta' = \alpha \boldsymbol{H}\boldsymbol{H}^{-1}\beta' = (\alpha \boldsymbol{H})(\beta \boldsymbol{H}^{-1\prime})' = \alpha^{*}\beta^{*\prime}.$$

Hence the factorization (2*) only identifies the space spanned by the "cointegrating vectors". To obtain unique values of α and β' requires further restrictions on the model. A proposed normalization and maximum likelihood estimation is proposed by Johansen, (1995) and Hamilton (1994) and is presented in Appendix C. The deterministic terms specification of Equation (3.34) are also detailed.

3.5.1 Likelihood Ratio Test for the Number of Cointegration Relations

In this test, the VECM (Equation (3.34) is denoted H(r). It must be noted that the cointregrated VECM is unrestricted. The *I*(1) model H(r) can be formulated as the condition that $rank(\Pi) \leq r$. The following nested hypotheses is then tested

$$H(0) \subset H(1) \subset \ldots \subset H(r) \subset \ldots \subset H(n).$$

Here, $H(0)$ is such that there is no cointegration which implies that $\Pi = 0$ whiles $H(n)$ represent an unrestricted stationary VAR(*p*) model. This nested formula creation is expedient for developing a sequential procedure to test for the number *r* of cointegrating relations.

The likelihood ratio (LR) statistics proposed by Johansen determines the rank of the "long-run matrix" Π, which is equivalent to the number of cointegration relations in y_t. These statistics are based on the estimated eigenvalues

$$\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_n$$

of the matrix Π. These eigenvalues also happen to equal the squared canonical correlation between $\Delta \boldsymbol{y}_t$ and $\Delta \boldsymbol{y}_{t-1}$ corrected for lagged $\Delta \boldsymbol{y}_t$ and d_t and so lie between 0 and 1.

3.5.1.1 Johansen's Trace Statistics

The LR statistic is

$$LR_{trace}(r_o) = -T \sum_{i=r_o+1}^{n} \ln(1-\hat{\lambda}_i) .$$

The following nested hypotheses is tested;

$$H_o(r) : r = r_o$$

$$H_1(r) : r > r_o$$

If $rank(\Pi) = r_o$ then $LR_{trace}(r_o)$ is small for values of $\hat{\lambda}_{r_o+1}, \ldots, \hat{\lambda}_n$ close to zero. However, $LR_{trace}(r_o)$ is large for $rank(\Pi) > r_o$ and some of the values of $\hat{\lambda}_{r_o+1}, \ldots, \hat{\lambda}_n$ are nonzero but less than 1. The asymptotic null distribution of the test statistic is multivariate version of the "Dickey – Fuller Unit root" distribution which depends on the dimensions $n - r_o$ and the specification of the "deterministic terms".

CHAPTER FOUR

RESULT AND DISCUSSION

4.0 Introduction

In this chapter of the study, the analysis of the data is presented and it begins with basic statistical analysis involving graphical tools and basic statistics. The Box-Jenkins methodology is used to estimate "Seasonal Autoregressive Integrated Moving Average (SARIMA) model" to the univariate data. Furthermore, the cointegration among the water level, temperature and humidity of the regions was explored through cointegrated vector autoregressive (VAR) models. Johansen's classical approach was employed in the estimation of the VECM to establish both "long-run and short-run" effect of temperature and humidity on the Akosombo water level. The software used in the analysis is R.

4.1 Exploration of Data

The quarterly averages of the daily Akosombo water level, temperature and humidity of its surrounding was computed from the data obtained from January 1980 to December 2014. This section of the study explores both the pattern as well as the univariate "stationarity" of the "time series" data.

Figure 4.1 is the "time series plot" of the quarterly Akosombo water level (in feet), temperature (in Degrees Celsius) and the relative humidity.

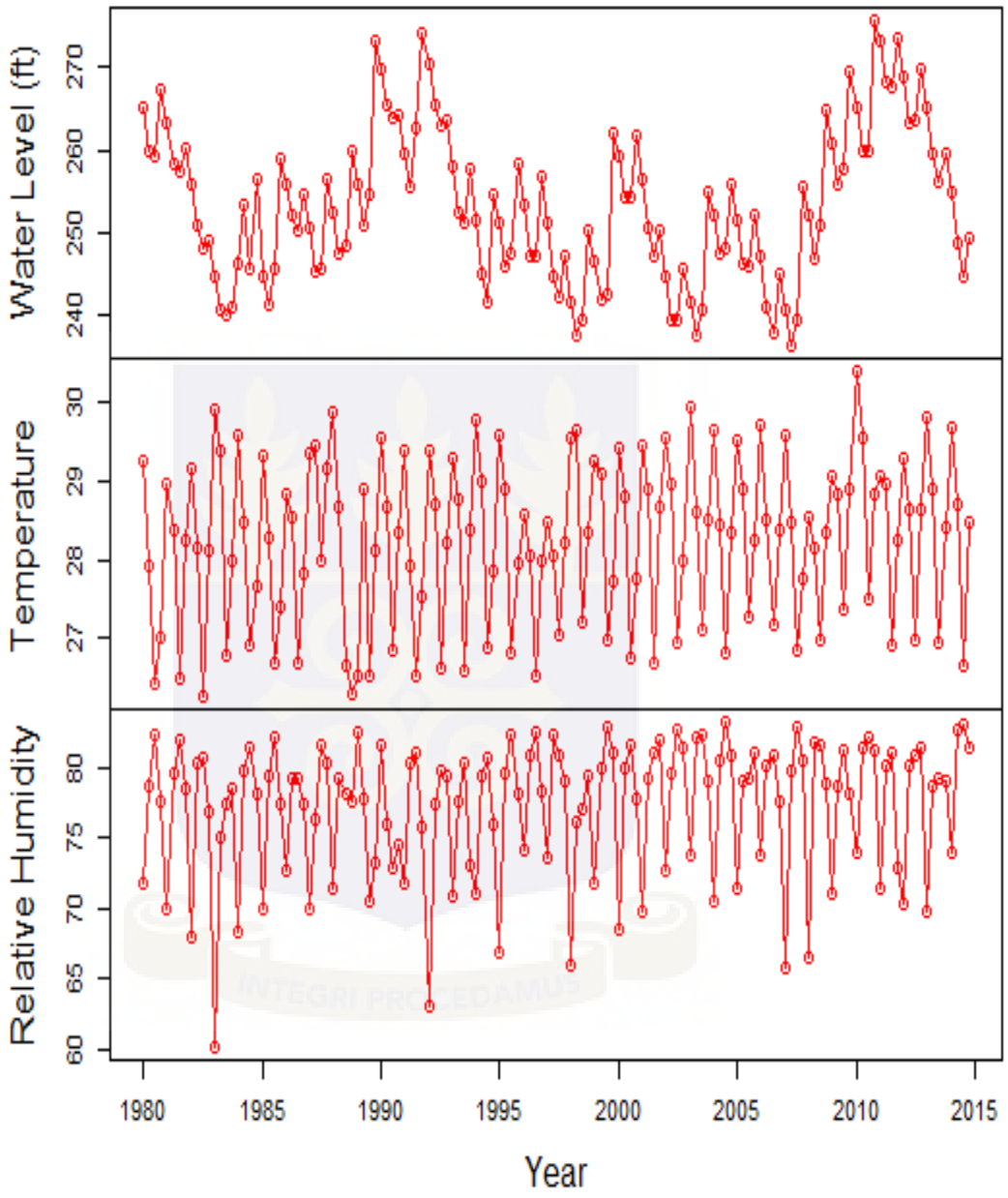

Figure 4.1: Time series plot of quarterly Akosombo water level, temperature and relative humidity

From Figure 4.1, the series for water level, temperature and relative did not exhibit any patterns of trend. However, there as a consistent upward and downward patterns in all three series with temperature and relative exhibiting high variability. The presence of seasonality is observed in both the temperature and relative humidity series. Interestingly, the series move in a parallel pattern which is an indication of possible correlation among the variable. Figure 4.2 depicts the plots for the differenced series.

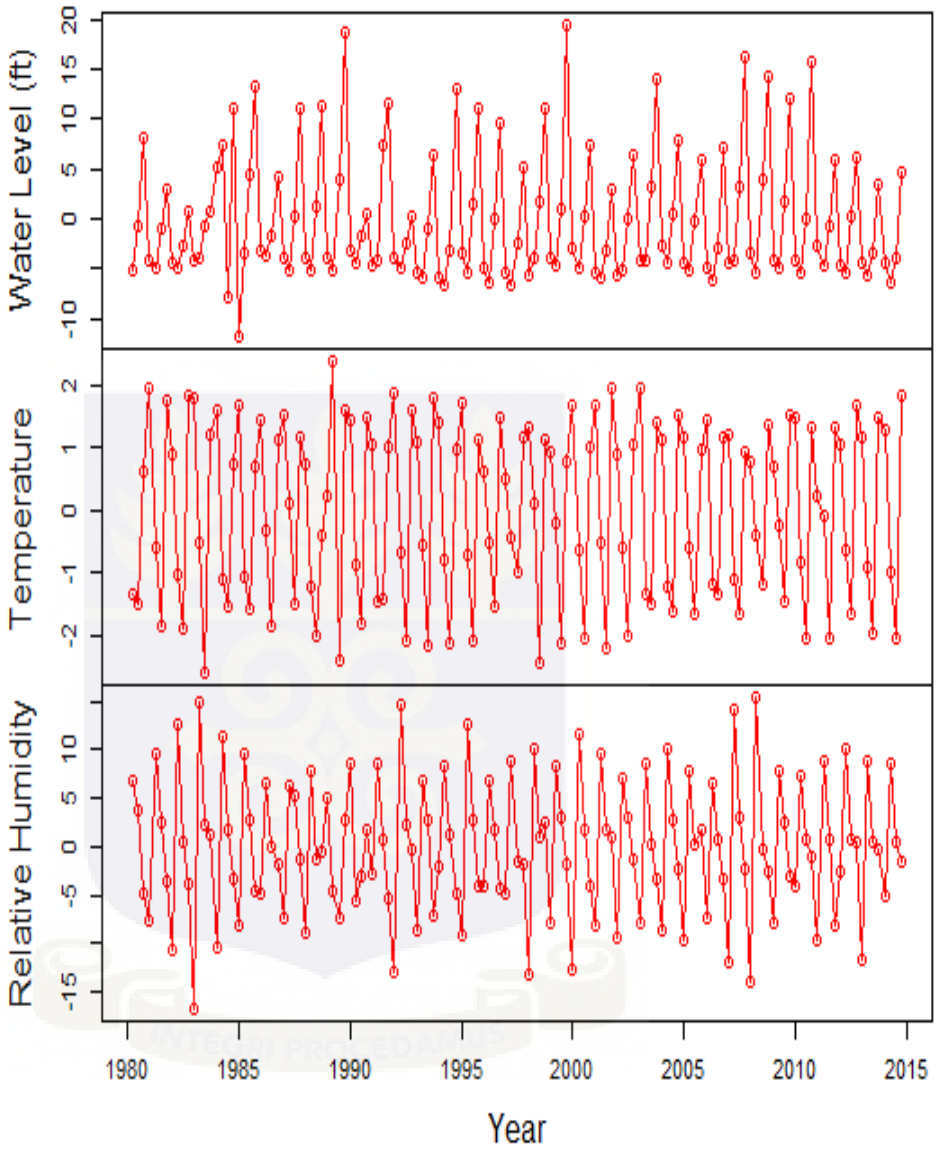

Figure 4.2: Time series plot of first differenced quarterly Akosombo water level, temperature and relative humidity

After first difference of the water level series, the data appears stationary with respect to mean. However, the series still exhibited variation across time span. There is seasonality in all series with high volatility (Figure 4.2).

4.1.1 Unit Root Test

The ADF and KPSS tests for unit root and stationarity were adopted to ascertain the order of integration of all variables. Both the levels of the variables and the "first difference" of the levels of the variables were tested. Table 4.1 presents the unit root test with the associated test statistics and critical values at 5%.we failed

Table 4.1: Unit root test on variables

Variable	ADF test		KPSS test	
	t – stat	Critical value at 5%	t – stat	Critical value at 10%
Water Level (ft)	-0.2949	-1.95	0.3871	0.347
Temperature	-0.2443	-1.95	0.4041	0.347
Relative Humidity	-0.3077	-1.95	0.6482	0.347

From the ADF test, we failed to reject the null hypothesis of nonstationarity was for all series since in each case the value of the test statistic was greater than the critical value of -1.95 at 5%. From the KPSS test, it was observed that all series were not $I(0)$. Since, the value of the test statistic for all the variables were greater than the critical value (0.347) at 10%, the null hypothesis that the variables were stationary was rejected in each case. Thus the study boldly concluded that the series were nonstationary (Table 4.1).

In order to establish the order of integration of the variables, the "unit root test" was performed on the first difference of the variables. This is discussed in Table 4.2.

Table 4.2: Unit root test on first difference of levels variables

Variable	ADF test		KPSS test	
	t – stat	Critical value at 5%	t – stat	Critical value at 5%
Water Level (ft)	-14.9639	-1.95	0.0461	0.463
Temperature	-26.1482	-1.95	0.0242	0.463
Relative Humidity	-16.2863	-1.95	0.0197	0.463

After the first difference, the ADF test indicated that the qaurterly water level, temperature and relative humidity, were stationary at the 5% critical value. This was confirmed by the KPSS test, all series were stationary after first difference. Thus, all series are integrated of order one ($I(1)$) (Table 4.2).

4.2 Fitting the SARIMA Model

In this section of the study, the Box-Jenkins methodology was employed to fit SARIMA model to the Akosombo water level, temperature and relative humidity data. The evidence of seasonality was assessed using the *ets* function in *forecast* package in *r*. The function fits an Error, Trend, Seasonality (ETS) model and the significance of the seasonal terms is assessed. The seasonality terms were observed to be significant in this study.

4.2.1 The Akosombo Water Level

The estimate of the SARIMA model for the Akosmbo Dam water level was fitted in this section. The plot of ACF and PACF of the undifferenced water level data is depicted in Figure 4.3.

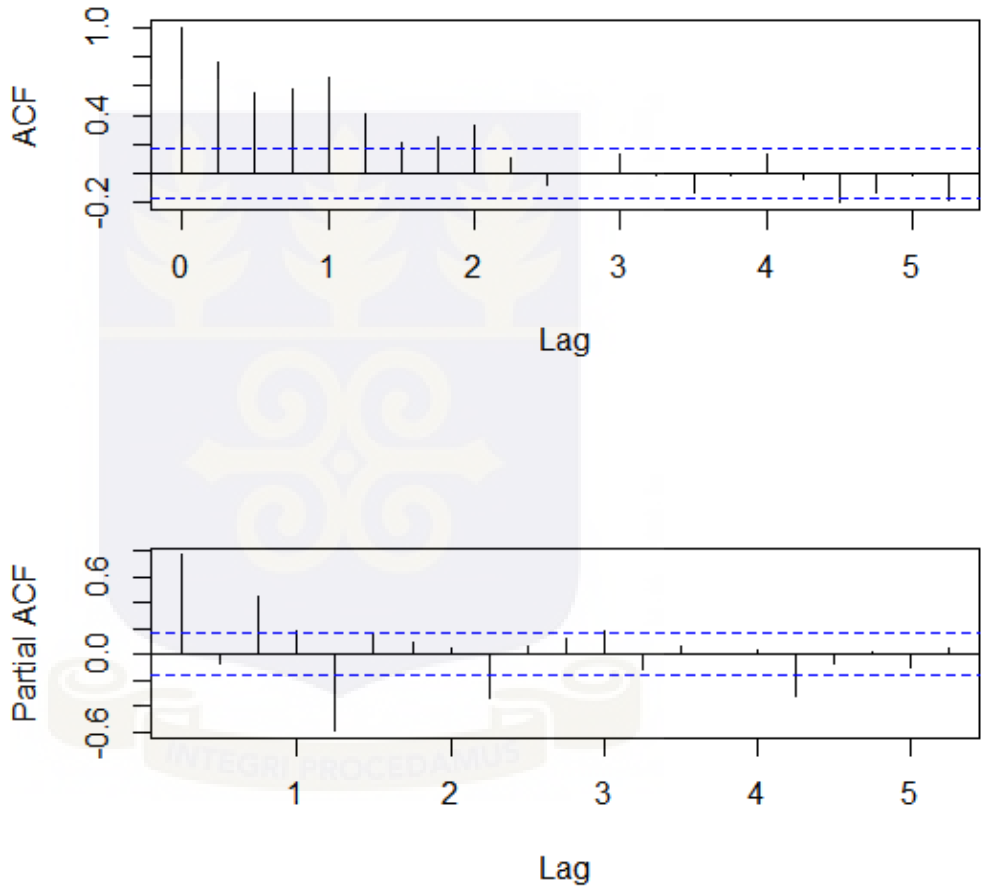

Figure 4.3: Correlogram (ACF and PACF) of Akosombo Water level data

It can be observed that there is one positive spike which cuts off after lag 1 as depicted by the PACF graph. However, the graph had other significant spikes at the other lags. Furthermore, its ACF is in sinusoidal waves and it gradually tails off. A

possible model of $ARIMA(1,0,0)(1,0,0)_4$ can be observed. However, the data is not stationary. Thus after differencing, an $ARIMA(1,1,0)(1,1,0)_4$ can be thought of (Figure 4.3).

Due to seasonality observed in the plot of the data (Figure 4.1) and the result of the seasoanlity test, a seasonal difference was obtained. The study further performed a difference of the seasonal difference to attain stationarity. Figure 4.4 is the ACF and PACF plot of the seasonally first differenced water level data.

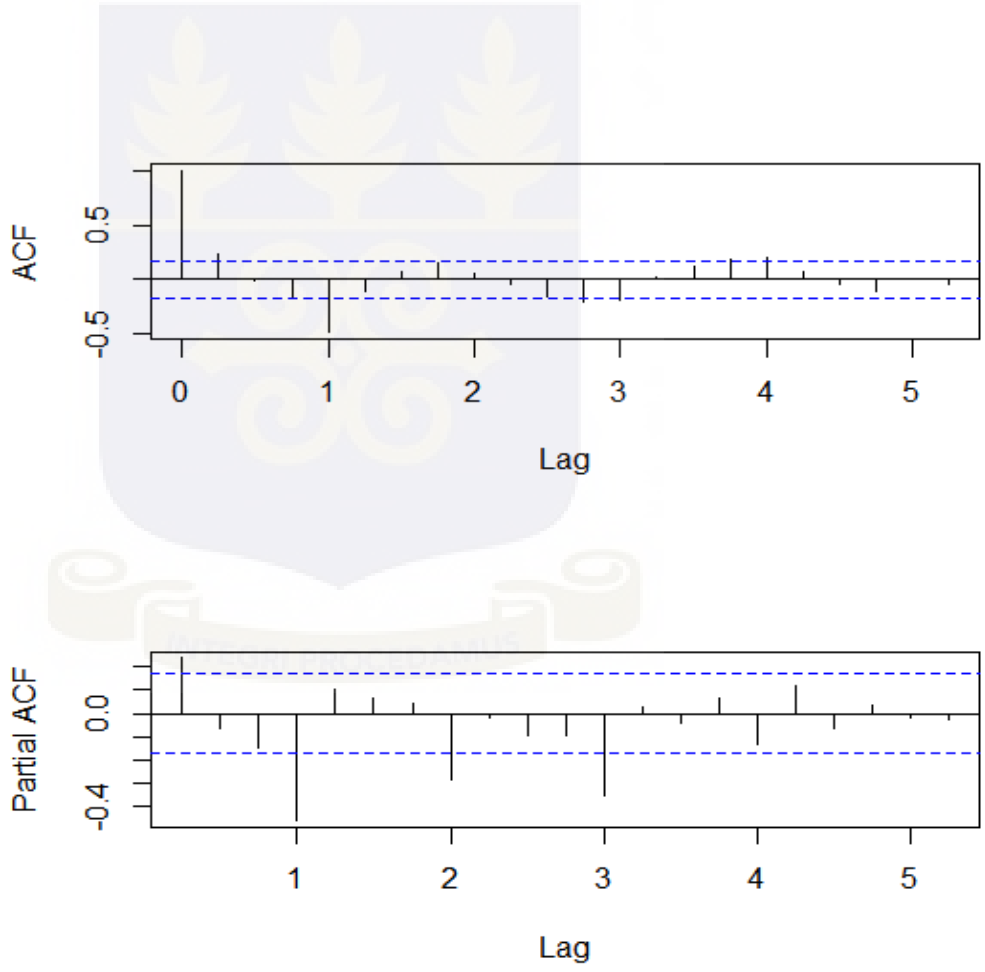

Figure 4.4: Correlogram (ACF and PACF) of seasonally first differenced Water level Data

The ACF plot depicts a sinusoidal waves and it gradually tails off. It have significant spike at the non-seasonal lag 1 and seasonal lag 4, with significant spikes at other non-seasonal lags. The PACF plot also has significant spikes at the non-seasonal lag 1 and seasonal lags 4, 8 and 12. It can be observed that, the positive significant spike at lag (1) of the ACF cut off after lag 1 and this is depicted in the PACF plot. Therefore, it suffices that we may have a seasonal autoregression with period one. Furthermore, there is a difference of one (Figure 4.4). Based on the non-seasonal significant lags of both the ACF and PACF, tentative models were identified for the Akosombo water level (Table 4.3).

Table 4.3: Tentative SARIMA Models for Water Level Data

Model	AIC	AICc	BIC
$ARIMA(1,1,1)(1,1,1)_4$	727.25	727.71	741.77
$ARIMA(1,0,1)(1,1,1)_4$	724.97	725.96	739.54
$ARIMA(1,1,0)(1,1,0)_4$	764.85	765.04	773.57
$ARIMA(0,1,1)(0,1,1)_4^*$	**723.58***	**723.76***	**732.29***
$ARIMA(1,1,1)(0,1,1)_4$	725.5	725.81	737.12
$ARIMA(0,1,1)(1,1,1)_4$	725.33	725.64	736.95
$ARIMA(0,1,2)(0,1,2)_4$	727.24	727.71	741.77

*: *Means best based on the selection criteria*

From Table 4.3, using the values of AIC, AICc, and BIC, the model $ARIMA(0,1,1)(0,1,1)_4$ was the most appropriate model that project the water level. The parameters of the model $ARIMA(0,1,1)(0,1,1)_4$ was then estimated and presented in Table 4.4.

Table 4.4: Estimates of Parameters for $ARIMA(0,1,1)(0,1,1)_4$

Variable	Coefficient	Standard Error	t-statistic	p - value
β_1	0.2536	0.0852	2.9777	0.0034
θ_1	-1.00	0.1040	-9.6190	0.0000

Let us recall the general form of the SARIMA model given as;

$$\alpha(B)\Phi(B^m)(1-B)^d(1-B^m)^D y_t = \beta(B)\Theta(B^m)z_t$$

From Table 4.4, the SARIMA model for the Akosombo water level can mathematically be presented as;

$$(1-B)(1-B^4)y_t = (1+\beta_1 B)(1+\theta_1 B^4)z_t$$

$$(1-B-B^4+B^5)y_t = (1+\beta_1 B+\theta_1 B^4+\beta_1\theta B^5)z_t$$

$$y_t - By_t - B^4 y_t + B^5 y_t = z_t + \beta_1 Bz_t + \theta_1 B^4 z_t + \beta_1\theta_1 B^5 z_t$$

$$y_t - y_{t-1} - y_{t-4} + y_{t-5} = z_t + \beta_1 z_{t-1} + \theta_1 z_{t-4} + \beta_1\theta_1 z_{t-5}$$

$$y_t = y_{t-1} + y_{t-4} - y_{t-5} + z_t + \beta_1 z_{t-1} + \theta_1 z_{t-4} + \beta_1\theta_1 z_{t-5}$$

To assess the fitness of the model, the plot of the residuals, its ACF, quantile plot as well the plot of p values associated with Ljung-Box statistic was examined. This is presented in Figure 4.5.

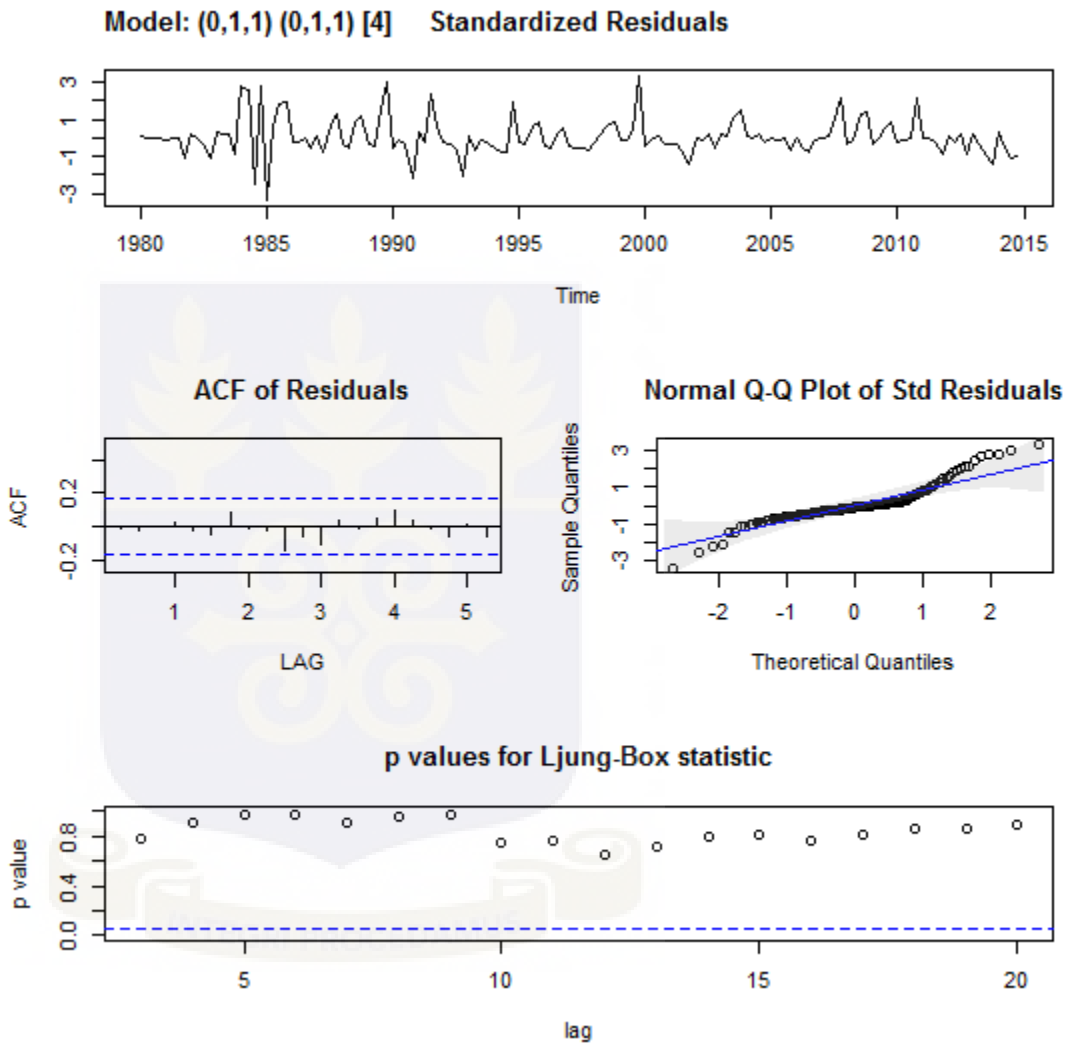

Figure 4.5: Diagnostic plot of $ARIMA(0,1,1)(0,1,1)_4^*$

The time series plot of residuals shows some cluster of volatility. However, the "autocorrelation functions" of the residuals from this estimated model include no significant values indicating the goodness of fit of the model to the series. The, "Ljung-Box" is a test of "autocorrelation" that confirms or otherwise the presence

of "autocorrelations" in the residuals of a time series model. From the Ljung-Box plot, the p-values are greater than 0.05. It suffices that the "autocorrelation" is different from zero (0). Thus, the selected model is an appropriate one for the Akosombo water level. Finally, from the "Normal Q-Q plot", the residuals are approximately normal (Figure 4.5).

From the residual plot in Figure 4.5, there is an observed cluster of volatility. Therefore, the ARCH-LM test for constancy of variance was performed and presented in Table 4.5.

Table 4.5: ARCH-LM Test for $ARIMA(0,1,1)(0,1,1)_4$ Residuals

Lag	Test Statistic	df	p - value
12	23.155	12	0.0264
24	25.356	24	0.3866
36	35.783	36	0.4788

To explore the cluster of volatility depicted in the residual plot in Figure 4.5, the ARCH-LM test result shown in Table 4.5, failed to reject the null hypothesis of no ARCH effect in the residuals of the selected model. Hence, the selected model satisfies all the assumptions and it can be concluded that $ARIMA(0,1,1)(0,1,1)_4$ model provides adequate representation of the Akosombo water level.

Having confirmed the model through residual diagnostics, the model was used to forecast the Akosmbo water level for the next five years. The forecasted Akosombo water level with its prediction interval is presented in Table A1 (Appendix A) and depicted in Figure 4.6.

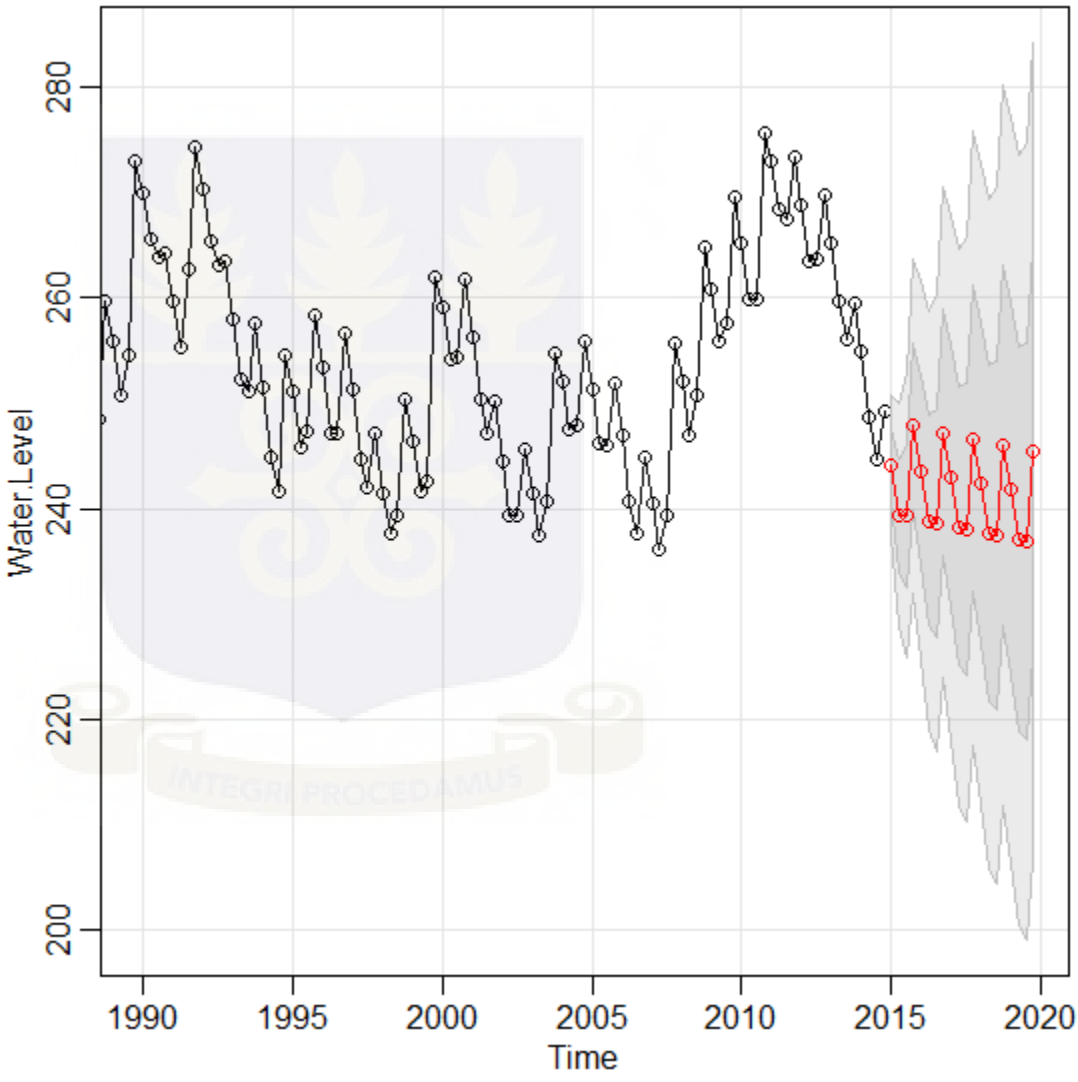

Figure 4.6: Plot of Forecast of water level using $ARIMA(0,1,1)(0,1,1)_4$

4.2.2 Temperature

This section of the study estimated the SARIMA model for the temperature of the Akosombo Dam environs. Figure 4.7 is the correlogram (ACF and PACF) of the undifferenced Temperature data.

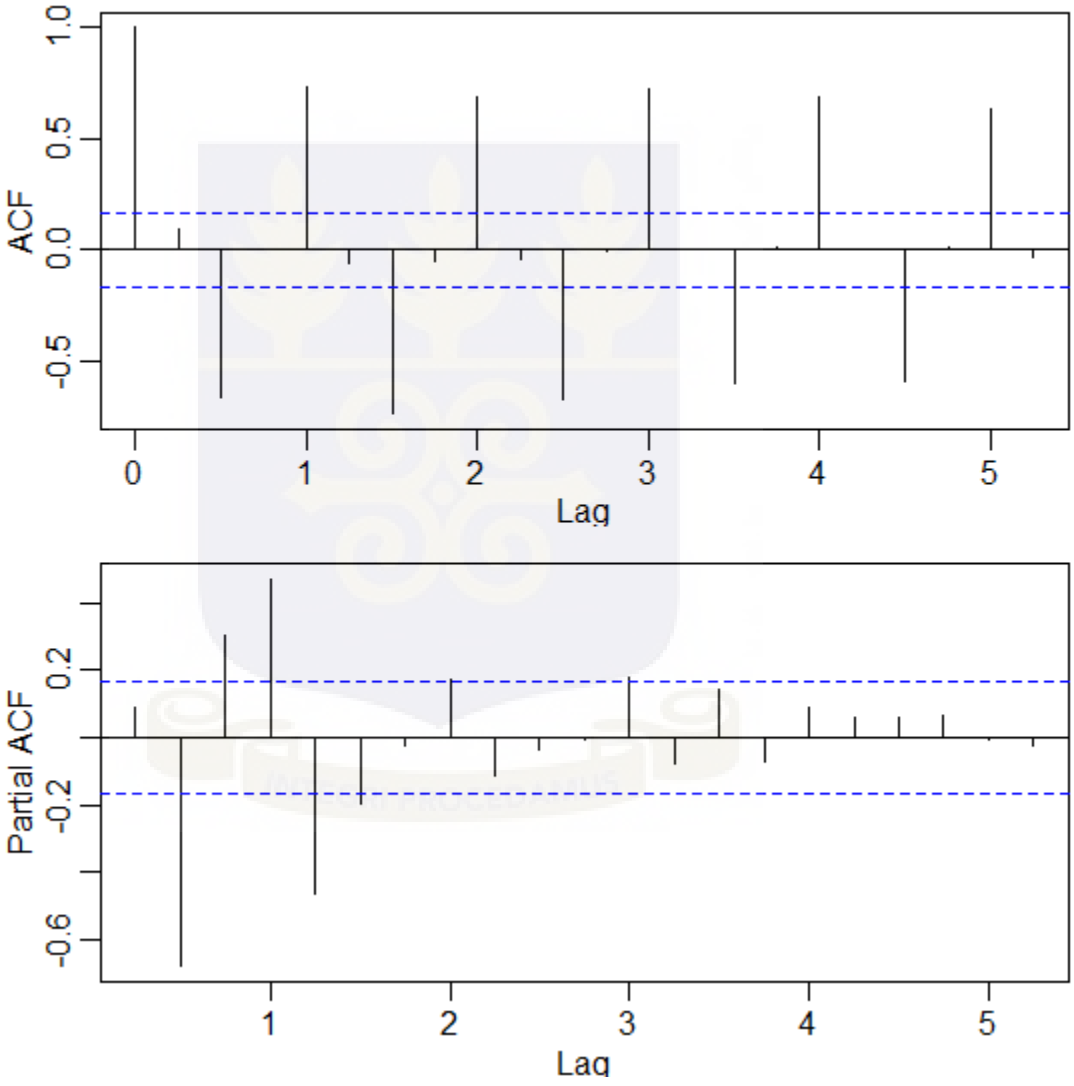

Figure 4.7: Correlogram (ACF and PACF) of Temperature data

PACF graph has both one negative and positive non-seasonal spikes which gradually tails off. However, the graph had other significant spikes at the other lags

with one significant seasonal lag. Furthermore, its ACF exhibits several significant spikes at both the seasonal and non-seasonal lags.

The study further performed a difference of the seasonal difference to attain stationarity. Figure 4.8 is the ACF and PACF plot of the seasonally first differenced water level data.

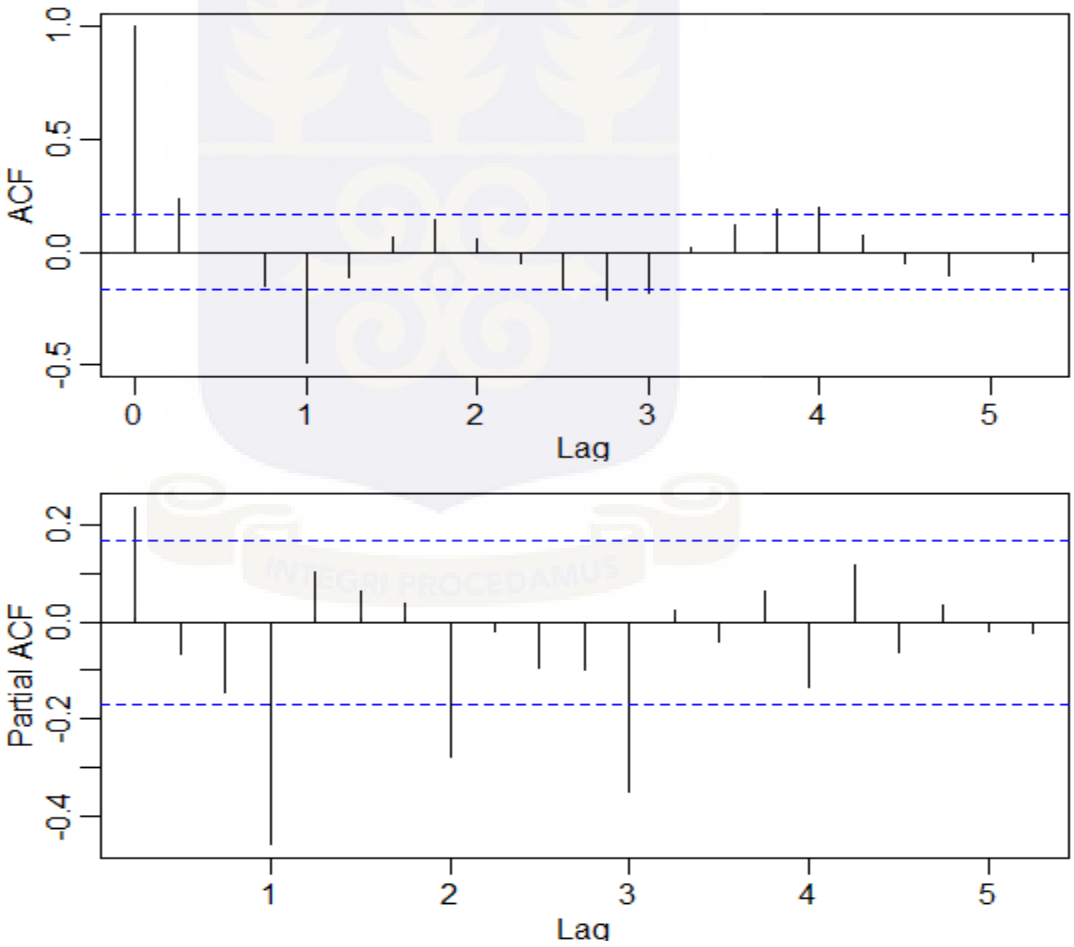

Figure 4.8: Correlogram (ACF and PACF) of seasonally first differenced Temperature data

From Figure 4.8, the ACF plot is sinusoidal and it gradually tails off. It have a positive significant spike at the non-seasonal lag 1 and seasonal lag 4, with significant spikes at other non-seasonal lags. The PACF plot also has a positive significant spikes at the non-seasonal lag 1 and seasonal lags 4, 8 and 12. It can be observed that, the positive significant spike at lag (1) of the ACF cut off after lag 1 and this is depicted in the PACF plot (Figure 4.8). Based on the non-seasonal significant lags of both the ACF and PACF, tentative models were identified for the Temperature data and this is presented in (Table 4.6).

Table 4.6: Tentative SARIMA Models for the Temperature Data

Model	AIC	AICc	BIC
$ARIMA(1,1,1)(1,1,1)_4$	184.38	184.85	198.91
$ARIMA(1,0,1)(1,1,1)_4^*$	**180.33***	**180.79***	**194.89***
$ARIMA(1,1,0)(1,1,0)_4$	251.28	251.46	260
$ARIMA(0,1,1)(0,1,1)_4$	202.35	202.54	211.07
$ARIMA(1,1,1)(0,1,1)_4$	182.53	182.83	194.15
$ARIMA(0,1,1)(1,1,1)_4$	203.59	203.90	215.21
$ARIMA(0,1,2)(0,1,2)_4$	188.32	188.78	202.84

*: *Means best based on the selection criteria*

From Table 4.6, the model $ARIMA(1,0,1)(1,1,1)_4$ was chosen as the appropriate model that fit the data well because it has the minimum values of AIC, AICc and BIC compared to other models. The estimates of the parameters of the model $ARIMA(1,0,1)(1,1,1)_4$ are presented in Table 4.7.

Table 4.7: Estimates of Parameters for $ARIMA(1,0,1)(1,1,1)_4$

Variable	Coefficient	Standard Error	t-statistic	p - value
α_1	0.8908	0.0487	18.2909	0.0000
β_1	0.2904	0.0879	3.3041	0.0012
ϕ_1	0.0754	0.0920	0.8196	0.4139
θ_1	-1.0000	0.0608	-16.4519	0.0000
β_o	-0.0351	0.0735	-0.4770	0.6341

Recalling the general form of the SARIMA model;

$$\alpha(B)\Phi(B^m)(1-B)^d(1-B^m)^D y_t = \beta_o + \beta(B)\Theta(B^m)z_t$$

The SARIMA model for the Temperature can mathematically be presented as;

$$(1-\alpha_1 B)(1-\phi_1 B)(1-B^4)y_t = \beta_o + (1+\beta_1 B)(1+\theta_1 B^4)z_t$$

$$(1-B^4-\phi_1 B+\phi_1 B^5-\alpha_1 B+\alpha_1 B^5+\alpha_1\phi_1 B^2-\alpha_1\phi_1 B^6)y_t = \beta_o + (1+\theta_1 B^4+\beta_1 B+\beta_1\theta_1 B^5)z_t$$

$$y_t = \beta_o + (\alpha_1 + \phi_1)y_{t-1} - \alpha_1\phi_1 y_{t-2} + y_{t-4} - (\alpha_1 + \phi_1)y_{t-5} + \alpha_1\phi_1 y_{t-6} + \beta_1 z_{t-1} + \theta_1 z_{t-4} + \beta_1\theta_1 z_{t-5} + z_t$$

Figure 4.9 is the diagnostic plot of $ARIMA(1,0,1)(1,1,1)_4$.

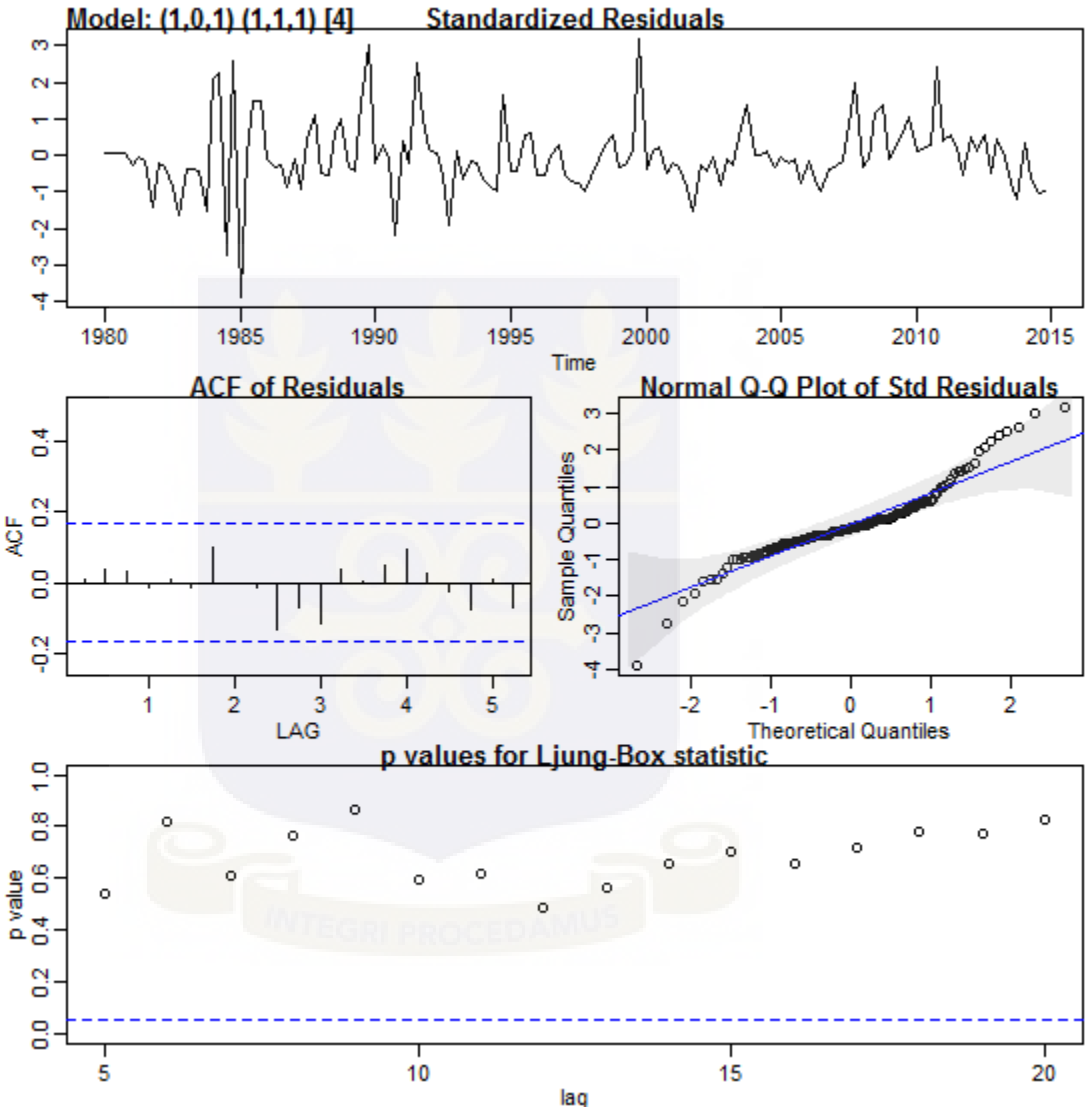

Figure 4.9: Diagnostic plot of $ARIMA(1,0,1)(1,1,1)_4$

There appears to be non-constancy of variance in the residual plot. However, the "autocorrelation functions" of the residuals from this estimated model include no

significant values indicating the goodness of fit of the model to the series. From the "Ljung-Box plot" and "Normal Q-Q plot", the selected model adequately fit the Temperature of the region. Also, from the the residuals are approximately normal (Figure 4.9). The ARCH-LM test for constancy of variance was performed to assess the volatility depicted in the residual plot in Figure 4.9. The result is presented in Table 4.8.

Table 4.8: ARCH-LM Test for $ARIMA(1,0,1)(1,1,1)_4$ Residuals

Lag	Test Statistic	df	p - value
12	58.939	12	0.0000
24	56.89	24	0.0002
36	64.649	36	0.0024

The ARCH-LM test result shown in Table 4.8, rejected the null hypothesis of no ARCH effect in the residuals of the selected model. Hence, the selected model satisfies all the assumptions but not free of volatility. To address the volatility, the Box- Cox transform was employed. However, the volatility was not quiet addressed. The study concluded that $ARIMA(1,0,1)(1,1,1)_4$ model provides adequate representation of the Temperature of the Akosombo region. Therefore, the model was used to forecast the Temperature for the next five years. These are presented in Table A2 (Appendix A) and depicted in Figure 4.10.

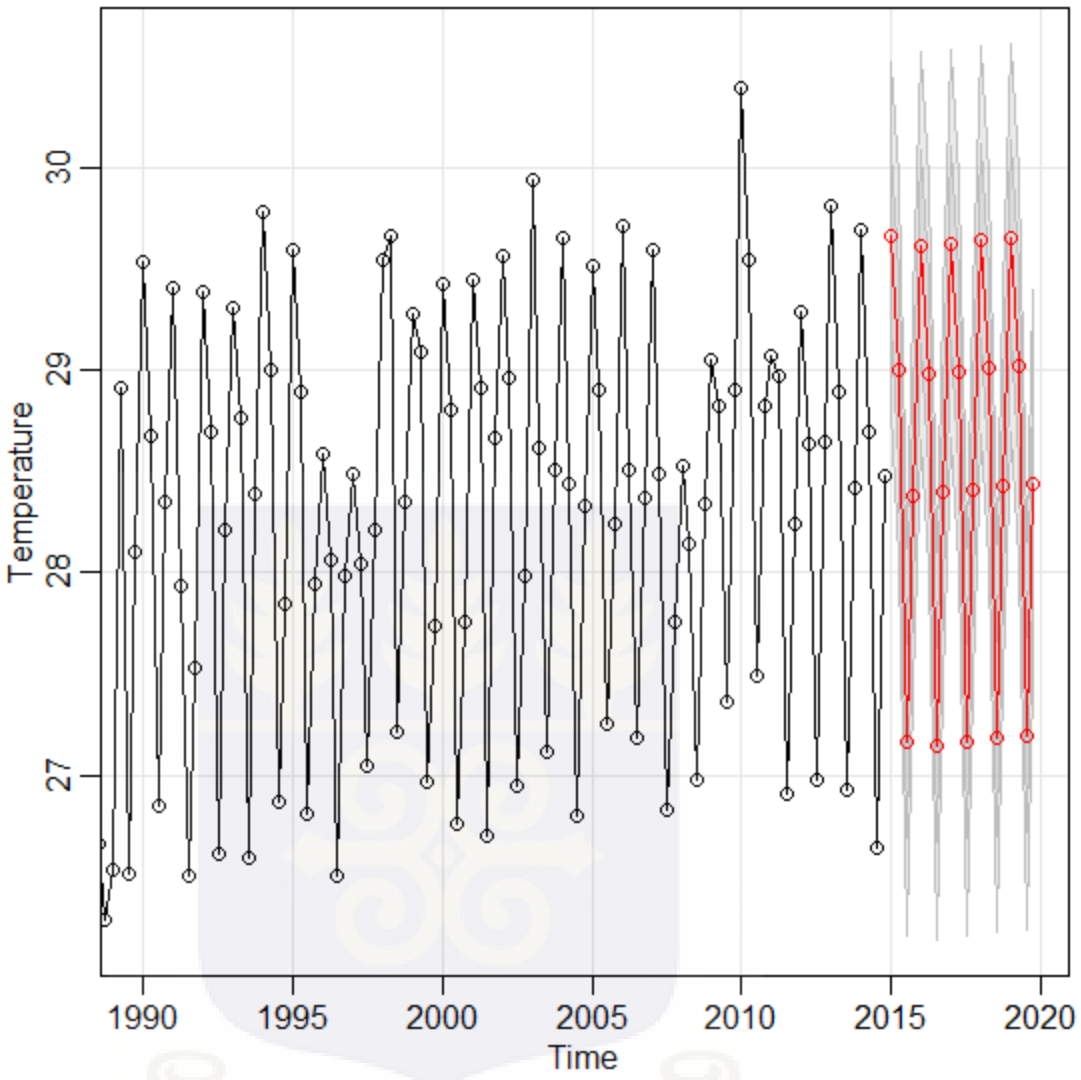

Figure 4.10: Plot of forecast of Temperature using $ARIMA(1,0,1)(1,1,1)_4$

4.2.3 Relative Humidity

In this section, the SARIMA model for relative humidity was estimated. The plot of ACF and PACF of the undifferenced relative humidity data is presented in Figure 4.11.

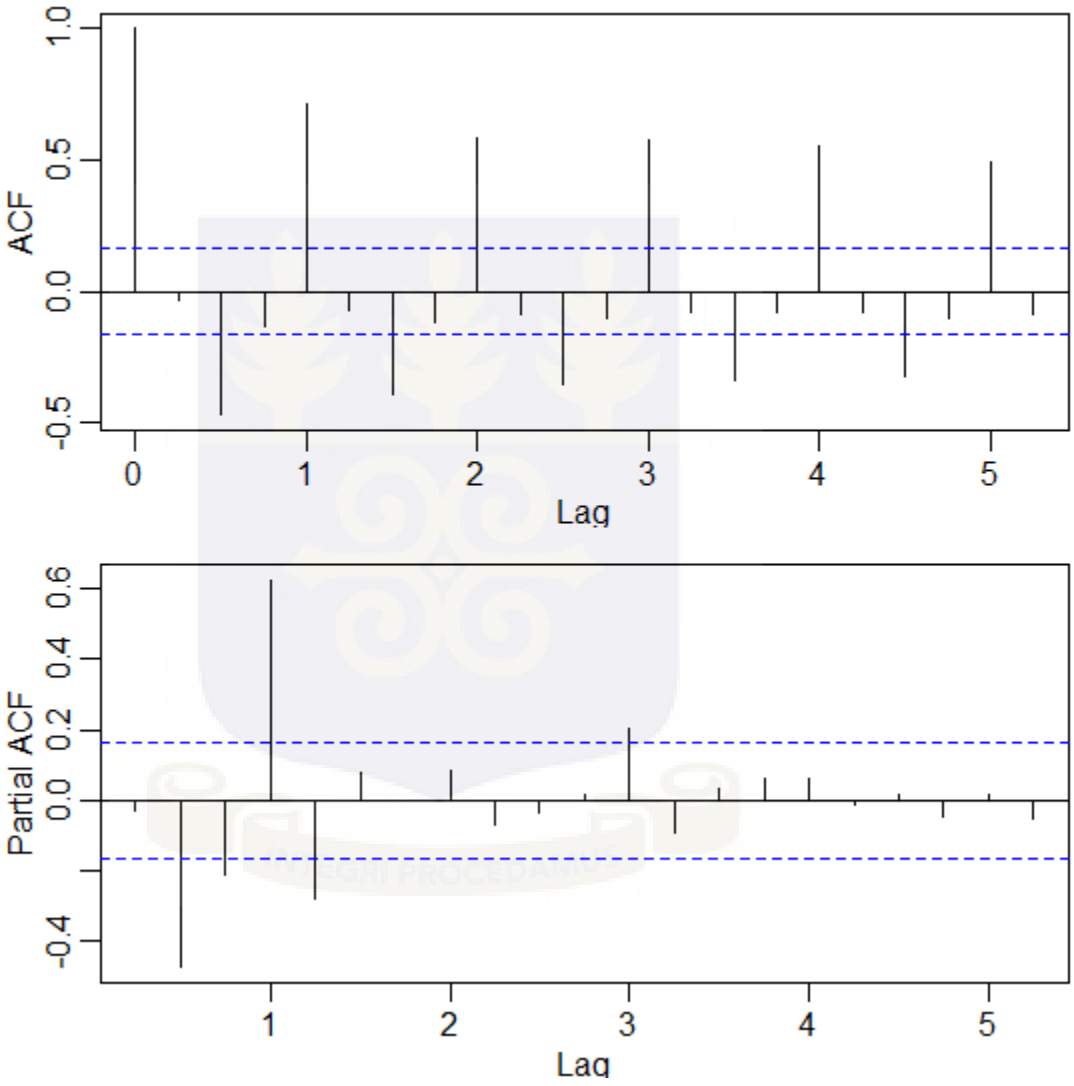

Figure 4.11: Correlogram (ACF and PACF) of Relative Humidity data

The ACF has one negative significant non-seasonal spike which cut off after lag 1. It however has other significant spikes at other lags. The PACF plot also has two

negative significant non-seasonal spikes at lags 2 and 3. There are also significant seasonal spikes at lags 4 and 12 (Figure 4.11). Figure 4.12 is the ACF and PACF plot of the seasonally first differenced relative humidity data.

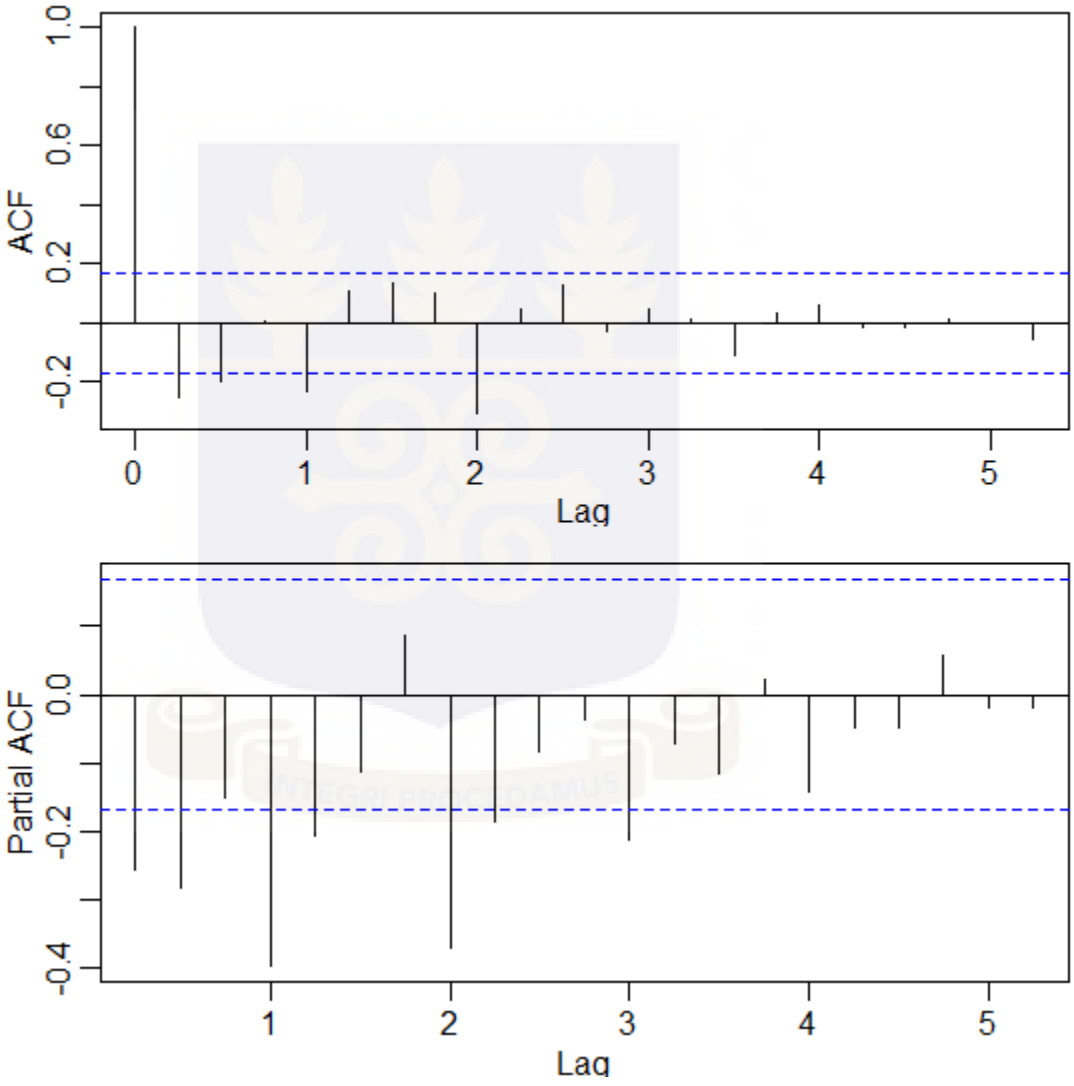

Figure 4.12: Correlogram (ACF and PACF) of seasonally first differenced Relative Humidity Data

The ACF have two negative non-seasonal significant spike at lags 1 and 2 with seasonal significant spikes at lags 4 and 8. Furthermore, plot depicts a sinusoidal waves and it gradually tails off. The PACF plot also has two negative non-seasonal significant spikes at lags 1 and 2 1 and significant seasonal spikes at lags 4, 8 and 12 (Figure 4.12). Tentative models were identified for the relative humidity (Table 4.9).

Table 4.9: Tentative SARIMA Models for Relative Humidity Data

Model	AIC	AICc	BIC
$ARIMA(1,1,1)(1,1,1)_4$	684.75	685.21	699.27
$ARIMA(1,0,1)(1,1,1)_4$	678.82	679.28	693.39
$ARIMA(1,1,0)(1,1,0)_4$	769.08	769.26	777.79
$ARIMA(0,1,1)(0,1,1)_4$	696.34	696.52	705.05
$ARIMA(1,1,1)(0,1,1)_4$	694.17	694.47	705.79
$ARIMA(0,1,1)(1,1,1)_4$	688.89	689.19	700.51
$ARIMA(0,1,2)(0,1,2)_4$	678.82	679.28	693.34
$ARIMA(2,1,1)(1,1,1)_4^*$	**678.76***	**679.41***	**696.19***

*: *Means best based on the selection criteria*

From Table 4.9, the model $ARIMA(2,1,1)(1,1,1)_4$ was chosen as the appropriate model that fit the data well because it has the minimum values of AIC, AICc and

BIC compared to other models. The estimates of the parameters of the model $ARIMA(2,1,1)(1,1,1)_4$ are presented in Table 4.10.

Table 4.10: Estimates of Parameters for $ARIMA(2,1,1)(1,1,1)_4$

Variable	Coefficient	Standard Error	t-statistic	p - value
α_1	0.2528	0.0840	3.0084	0.0031
α_2	-0.2449	0.0851	-2.8769	0.0047
β_1	-1.0000	0.0427	-23.4243	0.0000
ϕ_1	0.2533	0.0891	2.8434	0.0052
θ_1	-1.0000	0.0568	-17.6061	0.0000

Recalling the general form of the SARIMA model;

$$\alpha(B)\Phi(B^m)(1-B)^d(1-B^m)^D y_t = \beta(B)\Theta(B^m)z_t$$

The SARIMA model for the relative humidity can mathematically be presented as;

$$(1-\alpha_1 B-\alpha_2 B^2)(1-\phi_1 B)(1-B)(1-B^4)y_t = (1+\beta_1 B)(1+\theta_1 B^4)z_t$$

Figure 4.13 is the residual diagnostics of the selected model. It present the plot of the residuals, its ACF, Q-Q plot as well the plot of p values associated with Ljung-Box statistic.

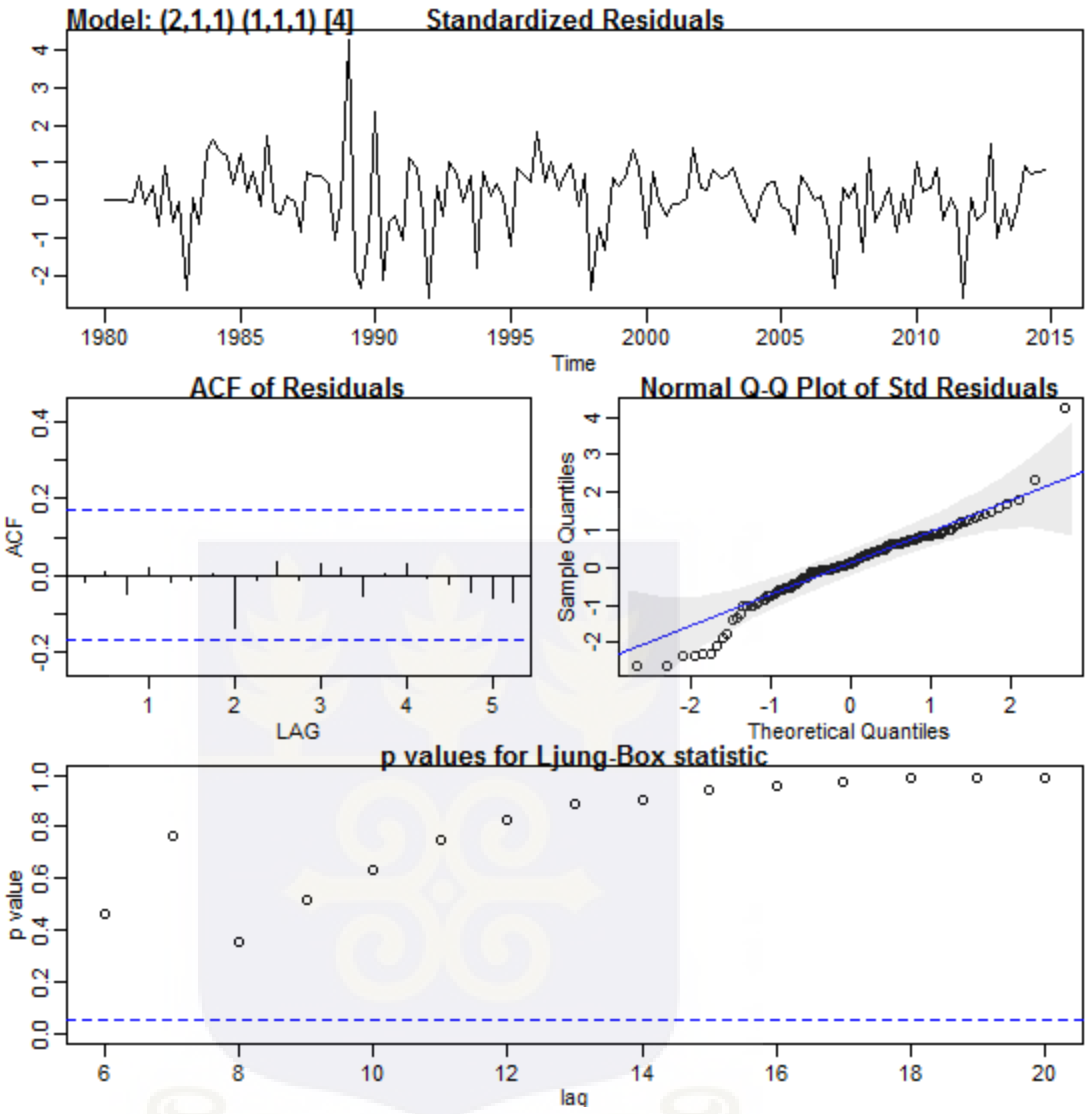

Figure 4.13: Diagnostic plot of $ARIMA(2,1,1)(1,1,1)_4$

From Figure 4.13, here also, the adequacy of the model is confirmed by the ACF plot, Ljung-Box plot, and the Normal Q-Q plot. Thus, the selected model is an appropriate one for the Temperature of the region. In addition, the ARCH-LM test for constancy of variance was performed to assess the volatility depicted in the residual plot in Figure 4.13. The result is presented in Table 4.11.

Table 4.11: ARCH-LM Test for $ARIMA(2,1,1)(1,1,1)_4$ Residuals

Lag	Test Statistic	df	p - value
12	15.697	12	0.2055
24	29.566	24	0.1996
36	28.037	36	0.8259

The ARCH-LM test result shown in Table 4.11, failed to reject the null hypothesis of no ARCH effect in the residuals of the selected model. Hence, the selected model satisfies all the assumptions and it can be concluded that $ARIMA(2,1,1)(1,1,1)_4$ model provides adequate representation of the relative humidity. The model was then used to forecast relative humidity of the regions for the next five years. The forecasted relative humidity with its prediction interval is presented in Table A3 (Appendix A) and depicted in Figure 4.14.

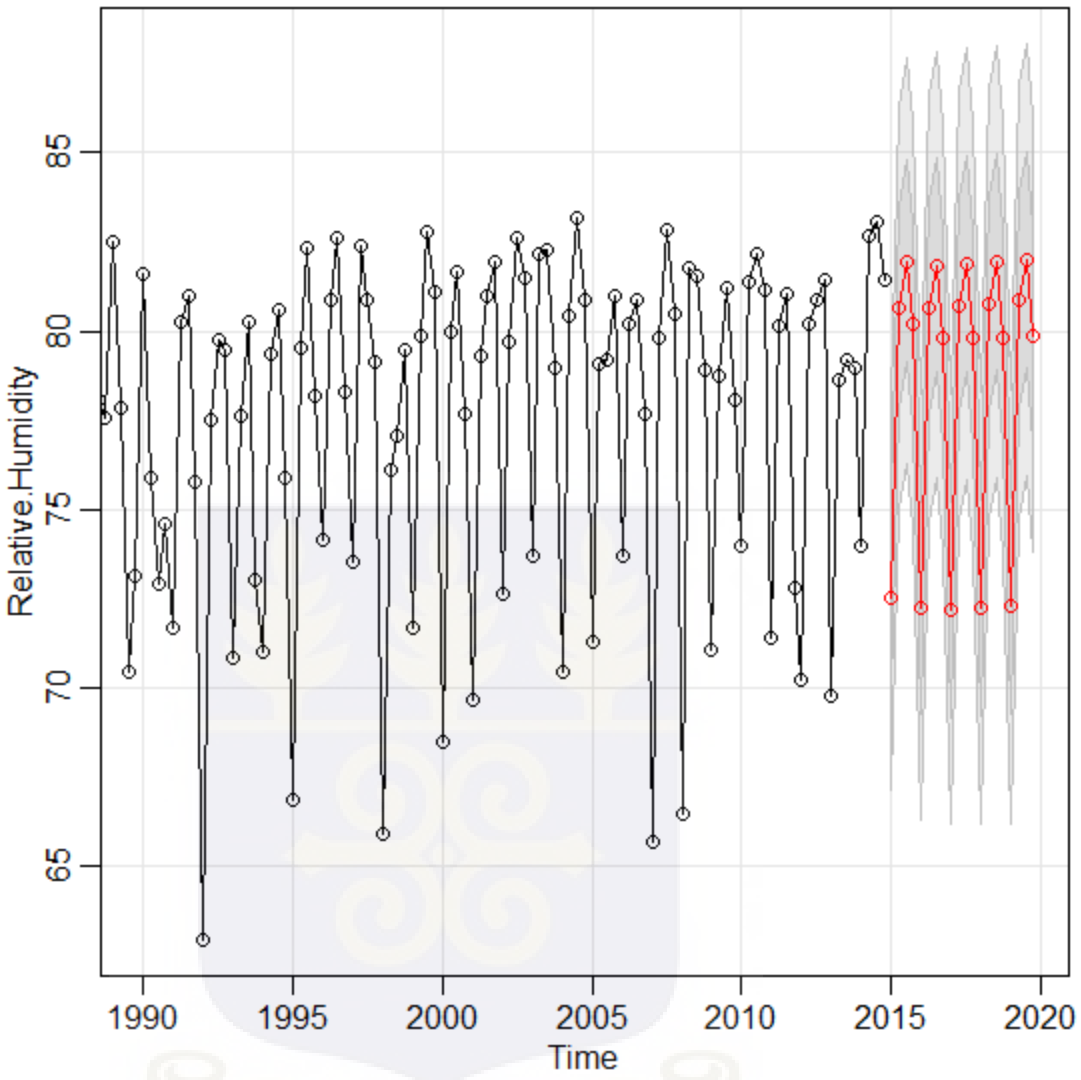

Figure 4.14: Plot of forecast of Relative Humidity using

ARIMA(2,1,1)(1,1,1) $_4$

4.3 Fitting the VAR Model

In this section of the study, a vector autoregression (VAR) model was fitted using the water level, temperature and the relative humidity. The causality analysis was performed. In addition, the "impulse response analysis" as well as the "forecast error variance decomposition" was explored. Furthermore, five years forecast was estimated for each variable. Table 4.12 present the results of the various information criteria employed in the order selection of the VAR model.

Table 4.12: Order Selection

Lag	AIC(n)	HQ(n)	SC(n)	FPE(n)
1	2.9242	3.1426	3.4617	18.6285
2	2.8129	3.1132	3.5519	16.6798
3	2.8998	3.2820	3.8404	18.2201
4	2.8742	3.3382	4.0164	17.7973
5	2.9733	3.4733	4.2710	18.8264

From Table 4.12, with the exception of the Bayesian Information Criterion (SC) which suggested a VAR order of one (1), the rest of the criteria considered opted for a VAR order of two (2) as the computed statistic begun to increase after lag 2 in each case.

After, the VAR order was appropriately determined, the VAR model was estimated. The estimated coefficients for both lags 1 and 2 for the three variables considered with the associated significance is discussed in Table 4.13.

Table 4.13: Estimate of VAR Model

Equation	Variables	Coefficient	Standard Error	Z statistics	p – value
Water Level	WaterLevelL1	1.2047	0.0851	14.156	0.0000
	TemperatureL1	1.2701	0.6060	2.096	0.0380
	HumidityL1	0.0492	0.0931	0.529	0.5980
	WaterLevelL2	-0.2882	0.0848	-3.399	0.0009
	TemperatureL2	-0.7604	0.6052	-1.256	0.2112
	HumidityL2	0.0369	0.0937	0.394	0.6943
	sd1	-17.4992	2.1410	-8.174	0.0000
	sd2	-14.8044	2.0811	-7.114	0.0000
	sd3	-8.8882	1.5355	-5.789	0.0000
Temperature	WaterLevelL1	-0.0130	0.0120	-1.078	0.2829
	TemperatureL1	0.6462	0.0856	7.548	0.0000
	HumidityL1	0.0274	0.0132	2.079	0.0396
	WaterLevelL2	0.0213	0.0120	1.779	0.07764
	TemperatureL2	0.1511	0.0855	1.767	0.0796
	HumidityL2	0.0195	0.0132	1.473	0.1431
	sd1	0.8538	0.3025	2.823	0.0055

	sd2	-0.7298	0.2940	-2.482	0.0144
	sd3	-2.4222	0.2169	-11.166	0.0000
Humidity	WaterLevelL1	0.0969	0.0747	1.298	0.1967
	TemperatureL1	-0.1562	0.5318	-0.294	0.7694
	HumidityL1	0.3256	0.0817	3.985	0.0001
	WaterLevelL2	-0.0786	0.0744	-1.056	0.2930
	TemperatureL2	2.1712	0.5311	4.088	0.0000
	HumidityL2	-0.1224	0.0822	-1.489	0.1390
	sd1	3.4821	1.8789	-1.853	0.0661
	sd2	6.0040	1.8264	3.287	0.0013
	sd3	0.7247	1.3475	0.538	0.5916

From Table 4.13, there are three models considered here. In the first model, Akosombo water level is the response variable with first two lags of it as well as that of temperature and relative humidity being the predictors. It was observed that the past two quarters of data on water level significantly predict the present water level (p – value < 0.05). Again, the first lag of temperature significantly influenced the present water level data. However, relative humidity was not a significant predictor to the water level. Furthermore, the seasonal dummies considered in the model were all significant (the fourth quarter was treated as reference).

Also, in the second model, temperature was the response variable with first two lags of it as well as that of water level and relative humidity being the predictors.

Interestingly, the first lag of water level did not significantly predict the current temperature. It second lag, however, could influence the current temperature if we relax the level of significance to 10%. The first lags of temperature and humidity were observed to significantly influence the level of temperature. The seasonal dummies were significant contributors in this model as well.

In the model involving the relative humidity as the response variable, the first and second lags of humidity and temperature were the observed significant predictors respectively.

From the result discussed in Table 4.13, it suffices the existence of causality among the variables. To ascertain the presence of causality, the Granger causality test was performed and is presented in Table 4.14. In the test, pairwise causality as well as group causality was assessed.

Table 4.14: Granger Causality Test

Equation	Excluded	F statistic	p – value
Water Level	Temperature		0.0380
	Humidity		0.5980
	All	2.0204	0.0909
Temperature	Water Level		0.0776
	Humidity		0.0396
	All	13.969	0.0000
Humidity	Water Level		0.1967
	Temperature		0.0000
	All	2.836	0.0243

From Table 4.14, water level causes temperature but not relative humidity. It however, may cause both temperature and relative humidity. Interestingly, temperature was observed to cause both water level and humidity but more significantly to relative humidity. Furthermore, relative humidity causes both temperature and water level.

Table 4.15 present the roots of the VAR model.

Table 4.15: Roots of VAR model

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
0.999905	0.806647	0.496448	0.496448	0.374036	0.253493

Clearly, the VAR model is stationary since all of its roots are less that unity. However, the first root is almost on the unit circle. In this case there could be possibility of cointegration among the variable. This is addressed in Section 4.4 of this chapter of the study.

The stationarity of the VAR model is confirmed by the OLS-CUSUM plot in Figure 4.15 since each model lies within the bound.

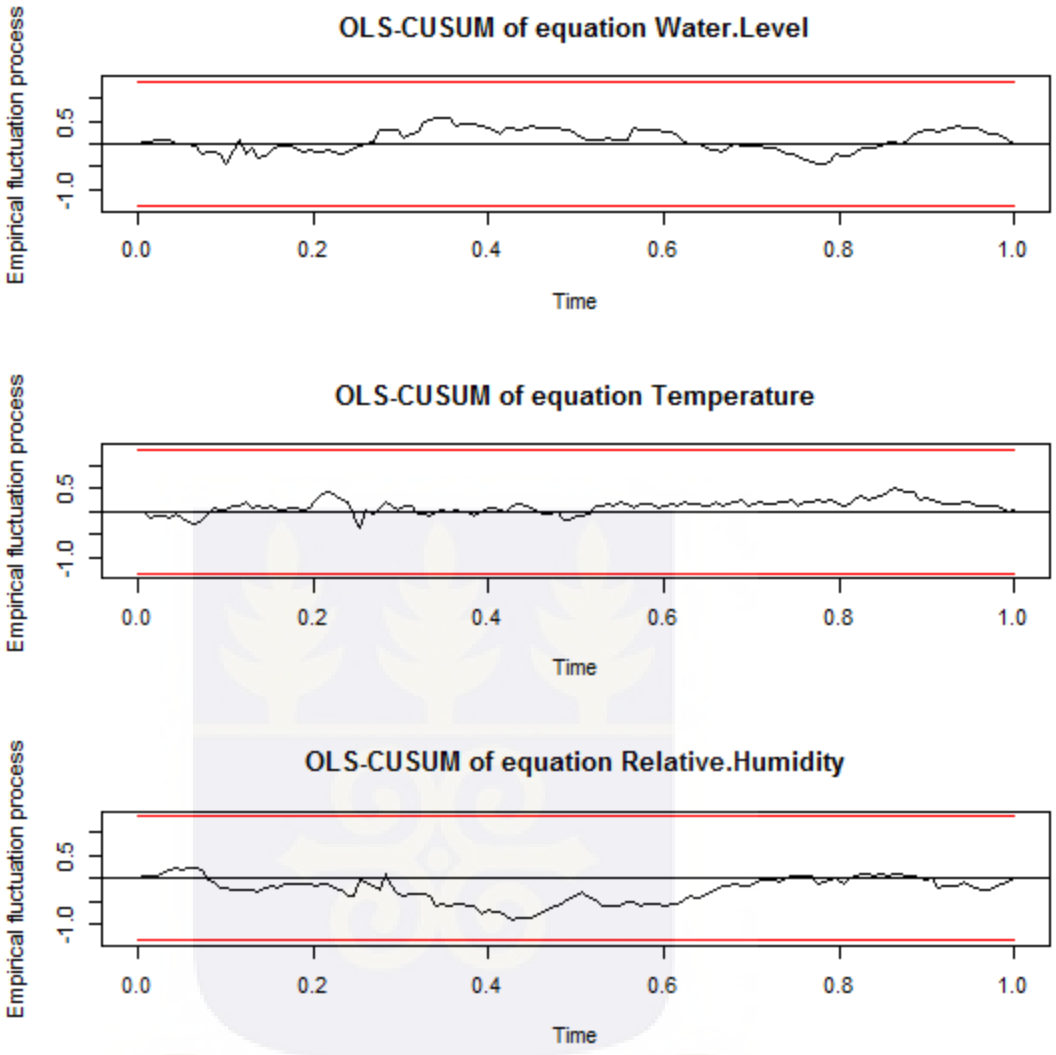

Figure 4.15: OLS-CUSUM plot of VAR (2) equations

Figure 4.16 is five year forecast using the VAR model for water level, temperature and relative humidity. The forecast values for water level, temperature and relative humidity with the associated confidence intervals are presented in Tables B1, B2, and B3 respectively (Appendix B).

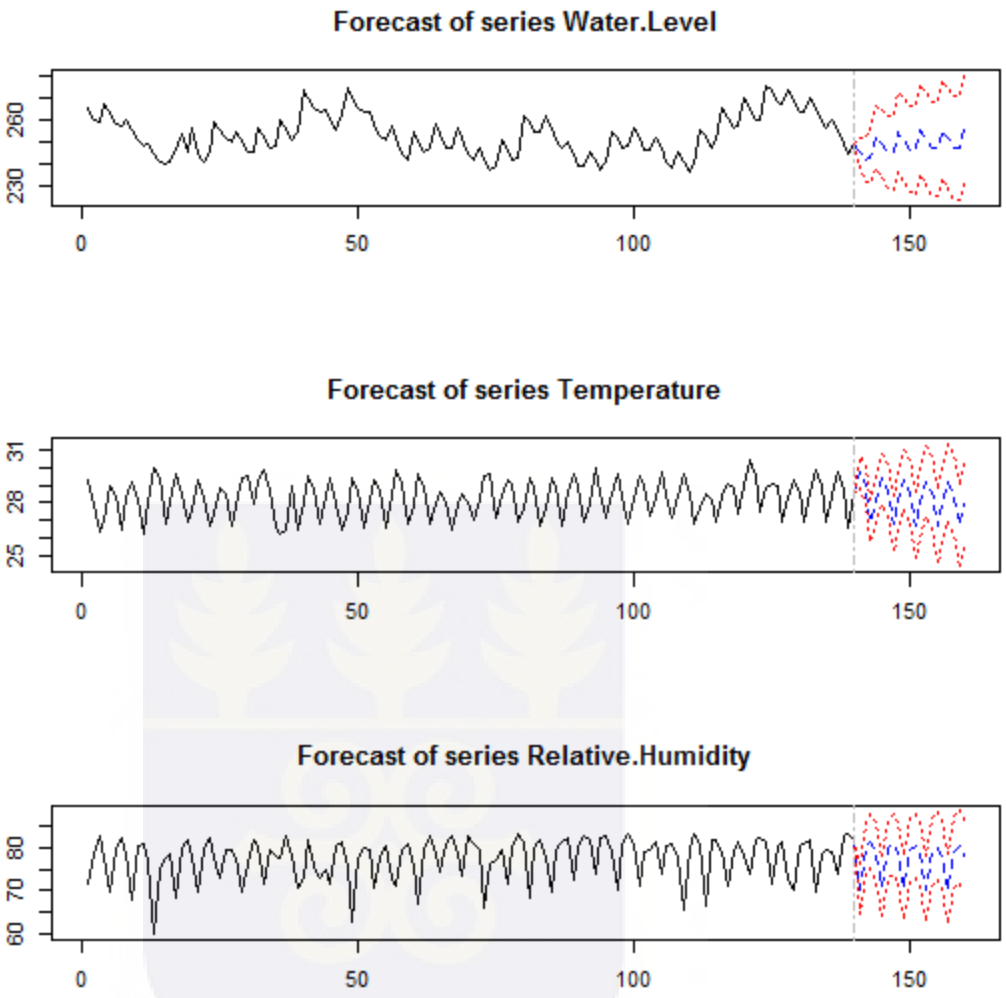

Figure 4.16: Plot of Forecast of VAR(2) equations

The “impulse response function” of VAR is to analyze the dynamic effects of the system when the model receives the impulse. The impulse response among the water level, temperature, and humidity was explored. Figures 4.17 is the response of Akosombo dam water level to temperature and relative humidity.

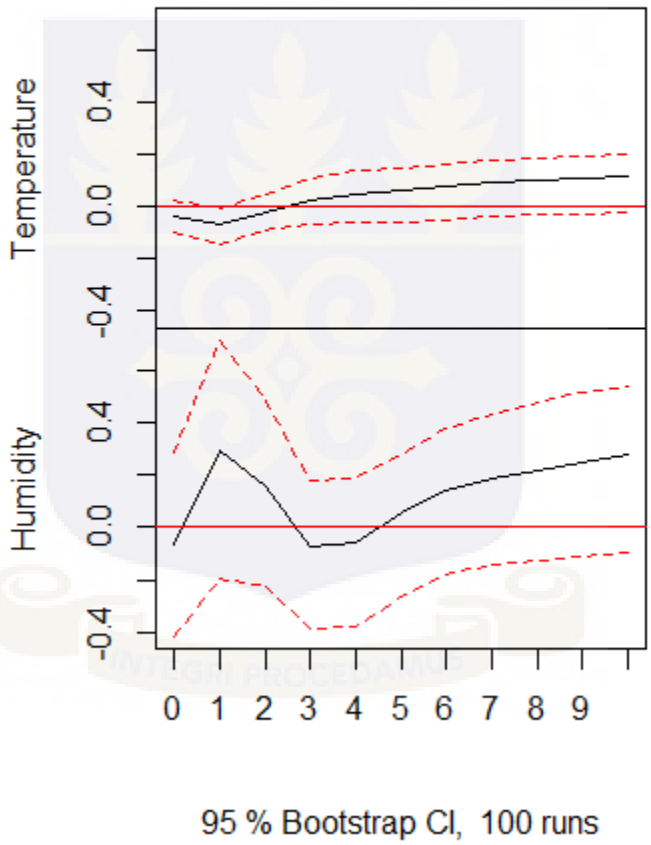

Figure 4.17: Impulse Response Analysis of Water Level

From Figure 4.17, when the impulse is water level, a unit change in water level will lead to a change of about -0.2 unit in temperature after the first quarter. However, the response of temperature remained positive after the third quarter. Also,

humidity change by 0.3 for a unit change in water level after the first quarter and by -0.4 after the second quarter. It however, had a positive smooth response after the fourth quarter. The response of temperature to Akosombo dam water level and relative humidity was also explored. This is presented in Figure 4.18.

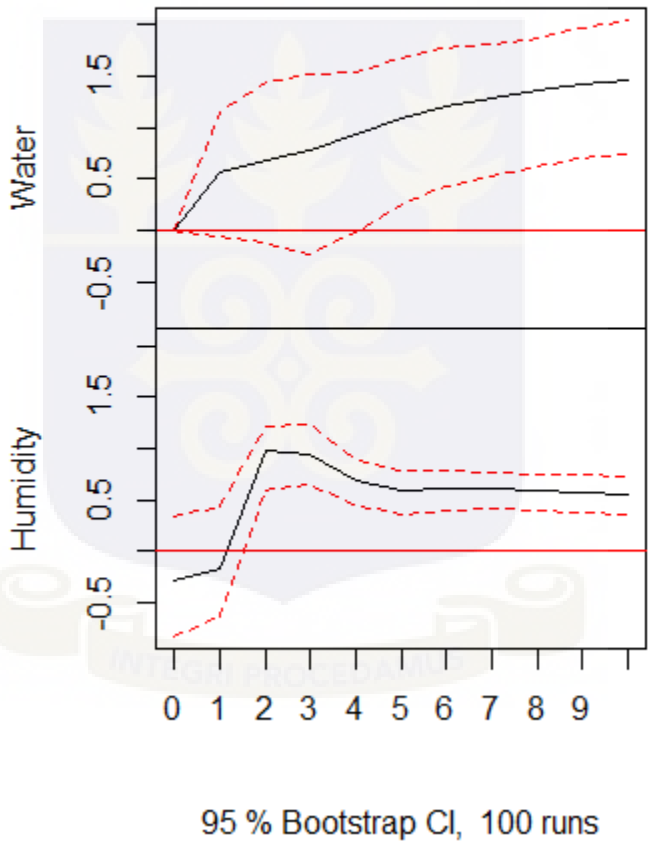

Figure 4.18: Impulse Response Analysis of Temperature

From Figure 4.18, when the impulse is temperature, the response of water level is positive at each time responsive period. The response increased consistently. Also, humidity had a positive response for each quarter. For a unit change in temperature,

there is a change of 1.2 in humidity after the second quarter. However, the response in humidity remained constant after the fourth quarter. The impulse response of humidity to water level and temperature is presented in Figure 4.19.

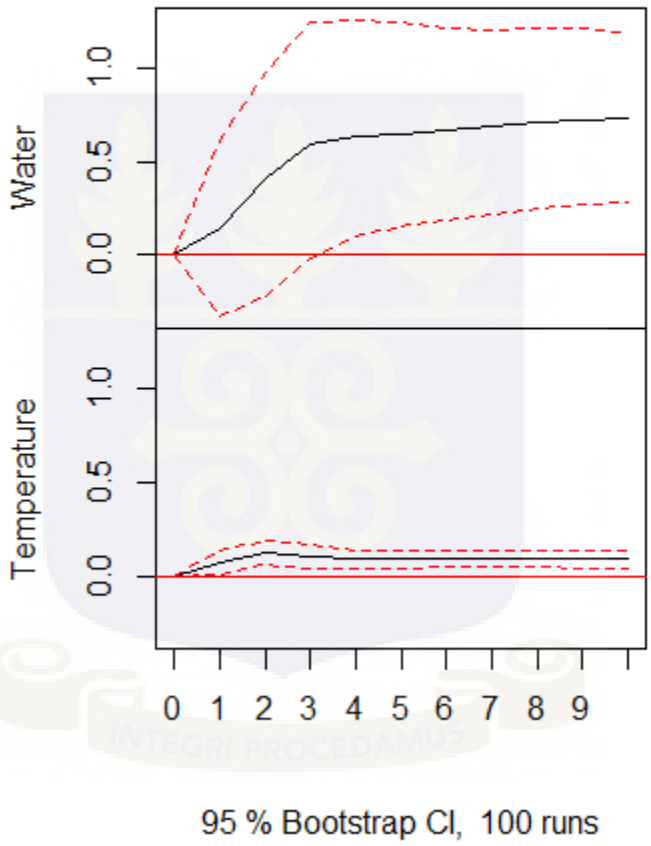

Figure 4.19: Impulse Response Analysis of Humidity

From Figure 4.19, the response of water level is positive at each time responsive period. There was a consistent increase in response. Temperature also experienced a positive response throughout the responsive period.

The variability in other variables explained by each variable in the autoregression is captured by the variance decomposition. The "forecast error variance decomposition" for water level is depicted in Table 4.16.

Table 4.16: Forecast Error Variance Decomposition for Water Level

Period	Water Level	Temperature	Humidity
1	1.0000	0.0000	0.0000
2	0.9868	0.0125	0.0008
3	0.9755	0.0197	0.0048
4	0.9626	0.0269	0.0105
5	0.9483	0.0364	0.0154
6	0.9322	0.0484	0.0194
7	0.9149	0.0620	0.0231
8	0.8968	0.0764	0.0268
9	0.8786	0.0911	0.0304
10	0.8603	0.1059	0.0338

From Table 4.16, it can be observed that variation in water level is completely explained for all forecast. However, a fraction of the variation is significantly explained by temperature and humidity. The "forecast error variance decomposition" for temperature is also presented in Table 4.17.

Table 4.17: Forecast Error Variance Decomposition for Temperature

Period	Water Level	Temperature	Humidity
1	0.0054	0.9946	0.0000
2	0.0018	0.9622	0.0198
3	0.0157	0.9244	0.0599
4	0.0145	0.970	0.0785
5	0.0169	0.8978	0.0853
6	0.0220	0.8880	0.0901
7	0.0293	0.8758	0.0949
8	0.0387	0.8624	0.0989
9	0.0493	0.8488	0.1019
10	0.0605	0.8354	0.1041

From Table 4.17, the forecast for temperature was completely explained by temperature. A fraction of variation in latter forecast was better explained by humidity as compared to water level. In addition, "forecast error variance decomposition" for humidity was explored and this is presented in Table 4.18.

Table 4.18: Forecast Error Variance Decomposition for Humidity

Period	Water Level	Temperature	Humidity
1	0.0005	0.0098	0.9898
2	0.0103	0.0116	0.9781
3	0.0119	0.1066	0.8815

4	0.0113	0.1799	0.8088
5	0.0110	0.2123	0.7766
6	0.0109	0.2330	0.7560
7	0.0121	0.2537	0.7341
8	0.0144	0.2734	0.7122
9	0.0175	0.2901	0.6924
10	0.0215	0.3040	0.6744

From Figure 4.18, the early forecast were completely explained by humidity. However, a significant fraction of the latter forecast were explained greatly by temperature as compared to humidity.

From the “impulse response analysis” and the “forecast variance decomposition”, it can be observed that both temperature play a key role in the dynamics of the water level. Both the SARIMA model and the VAR models can be concluded to be significant in forecasting the Akosombo dam water level, and the temperature and humidity of its environs. The forecast by the SARIMA methodology is presented in Figures A1, A2, and A3 in Appendix A whiles that of VAR methods is presented in Figures B1, B2, and B3 in Appendix B. The length of the confidence intervals for the VAR forecast are smaller as compared to that of SARIMA. Thus, the VAR model appears to improve upon the forecast and hence a better method for forecasting the water level and for that matter the Akosombo water level.

The “short-run and long-run” impact of temperature and humidity on the dynamics of the water level was further explored in section 4.3.

4.4 Fitting Vector Error Correction Model

In this section of the study, the vector error correction from the VAR model fitted in Section 4.3 was estimated using the Johansen's classical approach. Johansen's trace statistic was performed to test the number of "cointegration relations" among the variables. This is presented in Table 4.19.

Table 4.19: Rank Test

	Test	10%	5%	1%
$r \leq 2$	9.11	10.49	12.25	16.26
$r \leq 1$	44.86	22.76	25.32	30.45
$r = 0$	127.93	39.06	42.44	48.45

From Table 4.19, the null hypothesis of two cointegration relations was retained rejecting the null hypotheses of no and 1 cointegrated relation. The VECM was then estimated and presented in Table 4.20. "Error correction models (ECMs)" estimate the rate at which a response variable returns to equilibrium after a change in other variables. Error correction models may also arise because one variable forecasts another (Campbell and Shiller, 1988). The rate of adjustment for water level, temperature and relative humidity were explored.

Table 4.20: The Estimated VECM Model

Coefficient	Δ Water level	Δ Temperature	Δ Humidity
ect1	0.000625	-0.001616	-0.035612
ect2	0.115513	-0.568692	-0.133101
constant	-4.208815	16.384240	95.493768
sd1	-18.064733	0.762075	-4.034924
sd2	-15.408038	-0.591994	6.694692
sd3	-8.985525	-2.176930	2.096685
Δ Water levelL1	0.248580	-0.016859	0.100865
Δ TemperatureL1	0.961759	0.033425	-1.086509
Δ HumidityL1	0.002448	0.003725	0.264065

From Table 4.20, the estimated impact of the "cointegrating vectors" to the evolution of Δy_t measured by α is given below

$$\hat{\alpha} = \begin{pmatrix} 0.000625 & 0.115513 \\ -0.001616 & -0.568692 \\ -0.035612 & -0.133101 \end{pmatrix}$$

The adjustment parameters in $\hat{\alpha}$ have correct signs and imply rapid adjustment toward equilibrium in the long-run. The negative signs of temperature and humidity is an indication that the system reverts to a long-run equilibrium (that is, when the system overshoots, it is adjusted downwards) whiles the positive sign of the water level is an indication of sustainability of deviations from it (that is when the system undershoots it is corrected upwards). The rate of "adjustment to equilibrium" after

a shock every quarter of the year is low for water level (0.063%), average temperature (0.16%) and relative humidity (3.6%) in the first "error correction model". However, in the second "error correction model", the rate of reverting to equilibrium observed by temperature was very high (56.9%) as compared to adjustment rates of 11.6% and 13.3% observed by water level and relative humidity respectively.

It must be noted that when the predictions from the cointegrating equations are positive, water level is above equilibrium value. This is due to the positive coefficient of water level in the cointegration equation. When the Akosombo dam water is low, it gradually rises in response to temperature and relative humidity due to the positive adjustment coefficient of 0.000625. The estimates of the adjustment coefficients for temperature and relative humidity in the first cointegration relation are -0.0016 and -0.035. Thus, when the Akosombo dam water level is high, the average temperature and relative humidity quickly adjusts towards the water level at the same time that the water level is adjusting. This same observation was made in the second cointegration relation. However, water level adjust quickly in this case.

Furthermore, the components of Δ Water levelL1, Δ TemperatureL1, and Δ HumidityL1 are the short-run adjustment to the previous water level, temperature, and relative humidity respectively. Also, sd1, sd2, and sd3 are seasonal dummies with the fourth quarter being the reference point.

The cointegration relation is estimated as follows:

$$\hat{\beta}' = \begin{pmatrix} 1.0000 & 0.0000 & 30.8837 & -0.5416 \\ 0.0000 & 1.0000 & -0.0866 & -0.00214 \end{pmatrix}$$

Thus, the equilibrium relationship between water level and relative humidity is given as

$$y_{1t} = 0.5416t - 30.8837 y_{3t}$$

$$Water\ Level = 0.5416t - 30.8837 Humidity$$

and the equilibrium relationship between temperature and relative humidity is also given as

$$y_{2t} = 0.00214t + 0.0866 y_{3t}$$

$$Temperature = 0.00214t + 0.0866 Humidity$$

Thus, the long-run effect of relative humidity is 30.8837, that is, for a unit change in relative humidity in the lon-run, there is a change of 30.8837 ft in the Akosombo Dam water level. Also, a unit change in relative humidity in the long-run will trigger a change of 0.0866 in average temperature.

Furthermore, $0.5416t$ and $0.00214t$ are the deterministic trends. It accounts for the deterministic trends whiles the stochastic trends are catered by the cointegration process.

CHAPTER FIVE

DISCUSSION, CONCLUSSION AND RECOMMENDATION

5.1 Introduction

This chapter concludes the study by discussing the findings. Summary and conclusions of the findings are also drawn. In addition, recommendations and few issues for further investigations are raised. In this study, the following objectives were addressed:

1. Fit a SARIMA model for the water level, temperature, and humidity of the regions of the Akosombo Dam.
2. Fit a VAR model for the water level, temperature, and humidity of the regions.
3. Establish a VECM for the water level, temperature, and humidity of the regions.
4. Explore the cointegration relation between the water level, and temperature and humidity of the regions.
5. Forecast the water level, temperature and humidity of the regions using VAR and SARIMA models.
6. Compare the VAR model to the SARIMA model.

5.2: Discussion

The study explored the relationship between water level, temperature and humidity of the regions using cointegrated VAR models. The quarterly averages of the daily Akosombo water level, temperature and humidity of its surrounding was computed from the daily data obtained and it spanned the period January 1980 to December 2014. Presence of seasonality was observed and was significant.

According to the seasonal properties in the parameters of water level, temperature, and relative humidity, most recent studies have resulted to seasonal models (Eberhardt and Irvine, 1992; Tularam and Ilahee, 2010; Mensah, 2013). In the current study, seasonal ARIMA models $ARIMA(0,1,1)(0,1,1)_4$, $ARIMA(1,0,1)(1,1,1)_4$ and $ARIMA(2,1,1)(1,1,1)_4$ were tentatively arrived at using values of AIC, AICc and BIC for Akosombo Dam water level, temperature, and humidity respectively.

The autocorrelation functions of the residuals from these estimated models had no significant values indicating the "goodness-of-fit" of the models to the series. Also, the Ljung-Box plot indicated the absence of autocorrelation in each case. Five year forecast for Akosombo Dam water, temperature and humidity were obtained using the respective certified models.

Furthermore, the VAR model was fitted. It was observed that the past two quarters of data on water level significantly predict the present water level. Also, the first lag of temperature significantly influenced the present water level data. However, relative humidity was not a significant predictor to the water level. The first lags of temperature and humidity were observed to significantly influence the level of temperature. In the model involving the relative humidity as the response variable, the first and second lags of humidity and temperature were the observed significant predictors respectively.

The granger-causality test revealed that water level granger-causes temperature but not relative humidity. It however, granger-causes both temperature and relative humidity. Interestingly, temperature was observed to granger-cause both water

level and humidity but more significantly to relative humidity. Furthermore, relative humidity granger-causes both temperature and water level. The response of water level to the impulse of temperature was positive at each time responsive period. Humidity also responded positively and consistently to the impulse of temperature. In addition, both water level and temperature responded positively to the impulse of humidity.

The influence of temperature and humidity on the dynamics of water level appropriate the use of the VAR model (Kaufmann and Stern, 2002; Schmith, Johansen and Thejll, 2011). The smaller confidence intervals for forecast by VAR model compared to that of SARIMA indicates the supremacy of VAR models in the prediction of water level over that of the SARIMA model.

Finally, the speed of "adjustment to equilibrium" after a shock every quarter of the year is low for water level (0.63%), average temperature (0.16%) and relative humidity (0.36%) in the first "error correction model". However, in the second "error correction model", the speed of adjustment to equilibrium observed by temperature was very high (56.9%) as compared to adjustment rates of 11.6% and 13.3% observed by water level and relative humidity respectively. The two cointegraion relations among the variables were;

$$Water\ Level = 0.5416t - 30.8837Humidity$$

and

$$Temperature = 0.00214t + 0.0866Humidity$$

It suffices that, the long-run effect of relative humidity is 30.8837 and 0.0866 in the Akosombo Dam water level and temperature for a unit change in relative humidity respectively.

5.3 Conclusion

Seasonal ARIMA models $ARIMA(0,1,1)(0,1,1)_4$, $ARIMA(1,0,1)(1,1,1)_4$ and $ARIMA(2,1,1)(1,1,1)_4$ were estimated using values of AIC, AICc and BIC for Akosombo Dam water level, temperature, and humidity respectively. Also, water level was observed to granger causes both temperature and relative humidity whiles relative humidity also granger causes both water level and relative humidity. In addition, both water level and temperature responded positively to the impulse of humidity. The VAR model outperformed the SARIMA model in forecasting water level, temperature, and relative humidity. Finally, water level and relative humidity were cointegrated whiles a cointeration relation was observed between temperature and relative humidity. The rate of adjustment to equilibrium observed by temperature was very high among the three variables.

As a result of the findings of this study, the following recommendations were made;

1. As established that cointegration relation exist among water level, temperature and relative humidity, cultivation along river banks should be discouraged and measures should be taken to improve upon the climate conditions in the regions of our river bodies, especially, the Akosombo Dam.

2. VRA should adopt this technique to explore other conditions that necessitate the availability of water and put in measures to reduce their impact on the Akosombo Dam water level.
3. VRA should also, as part of their managerial role, employ all necessary legislative policies to help manage the issues of climate conditions in the regions.
4. The Bayesian approach to cointegration has gain grounds over the past two decades. This estimation approach should be adopted to investigate the "long-run equilibrium relations" among these variables in a future study.

REFERENCES

Afrifa-Yamoah E., (2015). Application of ARIMA Models in Forecasting Monthly Average Surface Temperature of Brong Ahafo Region of Ghana. *International Journal of Statistics and Applications*. 5(5):237-246.

Akaike, H., (1973). Information theory and an extension of the maximum likelihood principle, in B. N. Petrov & F. Cs´aki (eds), 2nd International Symposium on Information Theory, Acad´emiai Kiad´o, Budapest, pp. 267–281.

Akaike, H., (1978). On the likelihood of a time series model. *The Statistician*, 27, 217-235.

Amfo-Otu, W., (2010). Predicting the Level of Akosombo Dam using Mathematical modeling. Master's Thesis, Kwame Nkrumah University of Science and Tecnology, Kumasi, Ghana.

Bayraci S., Ari Y., and Yildirim Y., (2011). A Vector Aut-Regressive (VAR) Model for the Turkish Financial Markets. MPRA Paper 30475, University Library of Munich, Germany.

Bernanke, B. S., (1986). "Alternative Explorations of the Money-Income Correlation." Carnegie-Rochester Conference Series on Public Policy. 25, pp. 49 – 99.

Blanchard, Olivier J., and Watson, M. W., (1986). "Are Business Cycles All Alike?" in The American Business Cycle. R.J. Gordon, editor. Chicago: University of Chicago Press.

Blanchard, O. J. and Watson, M. W., (1986). Are Business Cycles All Alike?, in Gordon, R. (ed.), The American Business Cycle: Continuity and Change, University of Chicago Press, Chicago, pp. 123-56.

Box, G. and Jenkins, G., (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Box, G. E. P. and Cox, D. R. (1964). An Analysis of transformations. *Journal of the Royal Statistical Society*. 26(2): 211-252.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008). *Time Series Analysis. Forecasting and Control*. 4th Edition, Wiley & Sons Inc., New Jersey.

Bracegirdle, C. I., (2013). Inference in Bayesian Time-series Models. PhD Thesis, University College London.

Breusch, T. S., (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*. 17: 334–355.

Burnham, K. P. and Anderson, D. R., (2004). Multimodel inference: understanding AIC and BIC in Model Selection. *Sociological Methods and Research*, 33: 261-304.

Campbell, J. Y. and Shiller, R. J., (1988). Interpreting cointegrated models. *Journal of Economic Dynamics and Control*. 12(2-3): 505-522.

Chepngtich M. and Kihoro J., (2015). Application of Vector Autoregressive (VAR) Process in Modelling Reshaped Seasonal Univariate Time Series. *Science Jouranl of Applied Mathematics and Statistics. 3(3): 124-135.*

Clarida, R. H. and Friedman, B. M., (1984). The behavior of U.S. Short-term interest rates since October 1979. *The Journal of finance, 39(3), 671-682.*

Eberhardt A. K. and Irvine K. N., (1992). Multiplicative seasonal ARIMA models for Lake Erie and Lake Ontario water levels. *Water Res Bull.* 28(3):385-396.

Edgerton, D. and Shukur, G., (1999). Testing autocorrelation in a system perspective. *Econometric Reviews.* 18: 343–386.

Engle, R. F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica.* 50: 987–1007.

Engle, R. E. and Granger, C. W. J., (1987). Cointegration and error-correction: representation, estimation, and testing. *Econometrica*. 55: 251-276.

Fobil, J.N., (2003). Remediation of the environmental impacts of the Akosombo and Kpong dams. Attaquayefio, D.K. and Volta Basin Research Project [VBRP]. HORIZON Solutions Site: Public Health. Yale University Department of Biology: HORIZON International.

Gerolimetto, M., (2010). ARIMA and SARIMA models. Available at: www.dst.unive.it/~margherita/TSLecturenotess6.pdf (accessed 2nd May, 2017).

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*. 46: 1303–1313.

Granger, C. W. J., (1981). Some properties of time series data and their use in econometric model specification, Journal of Econometrics 16: 121–130.

Granger, C. W. J., (1983). 'Co-Integrated Variables and Error-Correction Models', UCSD Discussion Paper.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Hannan, E. J. and Quinn, B. G., (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*. B41: 190–195.

Hsing, S. and Tsokos C. P., (2008). A temperature forecasting model for the continental United States. *Neural, Parallel and Scientific Computations*. 16(1): 59-72.

Jahanbakhsh, S. and Baser, E. A. B., (2003). Studying and predicting the mean temperature of Tabriz using ARIMA model. *Geographic Researches Magazine. 18*(3): 34-46.

Jarque, C. M. and Bera, A. K., (1980). Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals. *Economic Letters*. 12, 255-259.

Jarque, C. M. and Bera, A. K., (1987). A test for normality of observations and regression residuals. *International Statistical Review*. 55, 163-172.

Johansen, S., (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. and Juselius, K., (1990). Maximum likelihood estimation and inference on cointegration with application to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52, 169-209.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

Juselius, K., (2006). *The Cointegrated VAR Approach: Methodology and Applications, Advanced Texts in Econometrics*. Oxford University Press, Oxford.

Kaufmann R. K. and Stern, D. I., (1997). Evidence for human influence on climate from hemispheric temperature relations. *Nature*. 388: 39-44.

Kaufmann, R.K. and Stern, D. I., (2002). Cointegration Analysis of Hemisphere Temperature relations. *Journal of Geophysical Research 107(D2)*, article No. 4012.

Kuan, C. M. and Hornik, K., (1995). The generalized fluctuation test: A unifying view. *Econometric Reviews*. 14(2): 135-161.

Kwiatkowski, D., Phillips P. C. B., Schmidt, P. and Shin, Y., (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics.* 54, 159–178.

Ljung, G. M. and Box, G. E. P., (1978). On a measure of lack of fit in time series models. *Biometrika*. 65: 297–303.

Ljung, G. M. and Box, G. E. P., (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*. 65(2): 297-303.

Lütkepohl, H., (2005). *New Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin.

Luetkepohl H., (2011). *Vector Autoregressive Models*. European University Institute.

Marfo, J., (2009). Inventory Control Modeling for Water Level Scheduling: A Case study of Akosombo Dam, Ghana. Master's Thesis, Kwame Nkrumah University of Science and Tecnology, Kumasi, Ghana.

Mason, J., (1984). *BK-Volta greatest lake, the story of Ghana's Akosombo dam.* Frankfurt: Natural History Museum.

Mensah, D., (2013). Application of Time Series in Predicting the Water Levels of the Akosombo Dam. Master's Thesis, University of Ghana, Department of Statistics, Legon, Ghana.

Mei, Q., Liu, Y. and Jing, X., (2011). Forecast the GDP of shanghai based on the multi-factors VAR model. *Journal of Hubei University of Technology, 26(3).*

Mills, T. C., (2013). Time series modelling of temperatures: an example from Kefalonia. *Meteorological Applications*. 21(3): 578-584.

Modarres, R., (2003). Box-Jenkins time series modeling for monthly precipitation in Ghale Shahrokh Station, third conference for Numerical Prediction of Weather.

Nia, S., Babazadeh, S. A. H. and Boustani, F., (2009). "Predicting the precipitation and average monthly temperature of Shiraz using ARIMA model and its influence on drought", first national conference of agriculture and sustainable development, opportunities and challenges. Islamic Azad University of Shiraz.

Nury, A.H., Koch M. and Alam, M.J.B., (2013). Modeling Temperatures in the Sylhet Division of Bangladesh using Time Series Analysis, In: Proceedings

of the 4th *International Conference on Environmental Aspects of Bangladesh (ICEAB).* Fukuoka, Japan, August 24-26, 2013.

Pfaff, B. and Taunus, K. I., (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, 27(4).

Rani, S. and Parekh, F., (2014). Predicting reservoir water level using artificial neural network. *International Journal Innovation Research Sci. Eng. Technol.* 3: 14489-14496.

Sarraf, A., Vahdat, S. F. and Behbahaninia, A., ((2011). Relative humidity and mean monthly temperature forecasts in Ahwaz Station with ARIMA model in time Series Analysis. IPCBEE, 12, IACSIT Press, Singapore: 149-153.

Schmith, T., Johansen, S. and Thejll, P., (2011). Statistical analysis of global surface air temperature and sea level using cointegration methods. University of Copenhagen, Department of Economics.

Schwarz, G., (1978). Estimating the Dimension of a Model. *Annals of Statistics*. 6(2): 461-464.

Sims, C. A., (1980). "Macroeconomics and Reality," *Econometrica.* 48: 1-48.

Sims, C. A., (1986). "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis Quarterly Review. *Winter.* 2 – 16.

Stock, J. H. and Watson, M. W., (2001). Vector Autoregression. *Journal of Economics Perspectives.* 15(4): 101-115.

Stoffer, D. S. and Dhumway, R. H., (2010). *Time Series Analysis and its Applications.* 3rd Edith, Springer, New York.

Tang, Z. and Fishwick, P. A., (1993). Feed-forward Neural Nets as Models for Time Series Forecasting. *ORSA Journal on Computing*. 5(4): 374-385.

Tsay, R. S., (2014). *Multivariate Time Series Analysis With R and Financial Applications*. John Wiley and Sons, Inc., Hoboken, New Jersey.

Tularam, G. A. and Ilahee M., (2010). Time Series of Rainfall and Temperature Interactions in Coastal Catchments. *Journal of Mathematics and Statistics*. 6(3): 372-380.

Vahab, B. A. and Alikhani, B., (2005). Studying the drought and wet year and predicting the climate changes in Birjand using statistical models. *Geographics researches magazine*. 37(52): 37-46.

Wei, W. S., (1990). Time Series Analysis: Univariate and Multivariate Methods.

Zakhary K., (1997). Factors affecting the prevalence of schistosomiasis in the Volta region of Ghana. *McGill Journal of Medicine* 3: 93-101.

Zeileis, A., Leisch, F., Kleiber, C. and Hornik, K., (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*. 20(1): 99-121.

Appendix A

Table A1: Forecast of Water level using $ARIMA(0,1,1)(0,1,1)_4$

Year	Quarter	Forecast	Lower	Upper	CI
2015	1	244.1636	237.6536	250.6736	13.0199
2015	2	239.4259	228.9890	249.8628	20.8737
2015	3	239.2496	226.0029	252.4963	26.4935
2015	4	247.8210	232.2639	263.3781	31.1143
2016	1	243.5657	225.9130	261.2184	35.3055
2016	2	238.8280	219.2866	258.3694	39.0827
2016	3	238.6517	217.3888	259.9146	42.5258
2016	4	247.2231	224.3680	270.0782	45.7102
2017	1	242.9678	218.5606	267.3750	48.8144
2017	2	238.2301	212.3524	264.1078	51.7554
2017	3	238.0538	210.7848	265.3228	54.5380
2017	4	246.6252	218.0325	275.2179	57.1853
2018	1	242.3699	212.4591	272.2807	59.8217
2018	2	237.6322	206.4505	268.8139	62.3634
2018	3	237.4559	205.0532	269.8586	64.8055
2018	4	246.0273	212.4479	279.6067	67.1588
2019	1	241.7720	207.0091	276.5349	69.5258
2019	2	237.0343	201.1206	272.9480	71.8274
2019	3	236.8580	199.8292	273.8868	74.0576
2019	4	245.4294	207.3181	283.5407	76.2225

Table A2: Forecast of Temperature using $ARIMA(1,0,1)(1,1,1)_4$

Year	Quarter	Forecast	Lower	Upper	CI
2015	1	29.66250	28.8241	30.5009	1.6768
2015	2	29.00376	28.0791	29.9284	1.8494
2015	3	27.16696	26.2301	28.1038	1.8737
2015	4	28.38093	27.4424	29.3195	1.8772
2016	1	29.61112	28.6728	30.5495	1.8767
2016	2	28.97934	28.0410	29.9177	1.8767
2016	3	27.14948	26.2112	28.0878	1.8766
2016	4	28.41165	27.4734	29.3499	1.8765
2017	1	29.62745	28.6892	30.5657	1.8766
2017	2	28.99542	28.0571	29.9338	1.8767
2017	3	27.16562	26.2273	28.1040	1.8767
2017	4	28.41165	27.4734	29.3500	1.8766
2018	1	29.64230	28.7118	30.5728	1.8609
2018	2	29.01030	28.0720	29.9487	1.8767
2018	3	27.18050	26.2422	28.1189	1.8767
2018	4	28.42660	27.4883	29.3649	1.8766
2019	1	29.65724	28.7189	30.5955	1.8766
2019	2	29.02524	28.0869	29.9636	1.8767
2019	3	27.19545	26.2571	28.1338	1.8767
2019	4	28.44154	27.5032	29.3798	1.8766

Table A3: Forecast of Relative Humidity using $ARIMA(2,1,1)(1,1,1)_4$

Year	Quarter	Forecast	Lower	Upper	CI
2015	1	72.53983	67.2322	77.8474	10.6152
2015	2	80.63474	75.1338	86.1357	11.0020
2015	3	81.95486	76.3950	87.5148	11.1198
2015	4	80.17816	74.6004	85.7559	11.1555
2016	1	72.26246	66.4338	78.0912	11.6574
2016	2	80.65615	74.7900	86.5223	11.7322
2016	3	81.84304	75.9790	87.7071	11.7282
2016	4	79.82148	73.9580	85.6849	11.7269
2017	1	72.19158	66.2882	78.0949	11.8067
2017	2	80.72008	74.8096	86.6306	11.8210
2017	3	81.87881	75.9711	87.7866	11.8155
2017	4	79.78215	73.8746	85.6897	11.8151
2018	1	72.21996	66.2954	78.1445	11.8491
2018	2	80.85040	74.9234	86.7774	11.8540
2018	3	81.93788	76.0131	87.8627	11.8496
2018	4	79.82213	73.8969	85.7474	11.8505
2019	1	72.27666	66.3382	78.2152	11.8770
2019	2	80.85040	74.9104	86.7904	11.8800
2019	3	82.00235	76.0645	87.9403	11.8758
2019	4	79.88180	73.9432	85.8204	11.8772

Appendix B

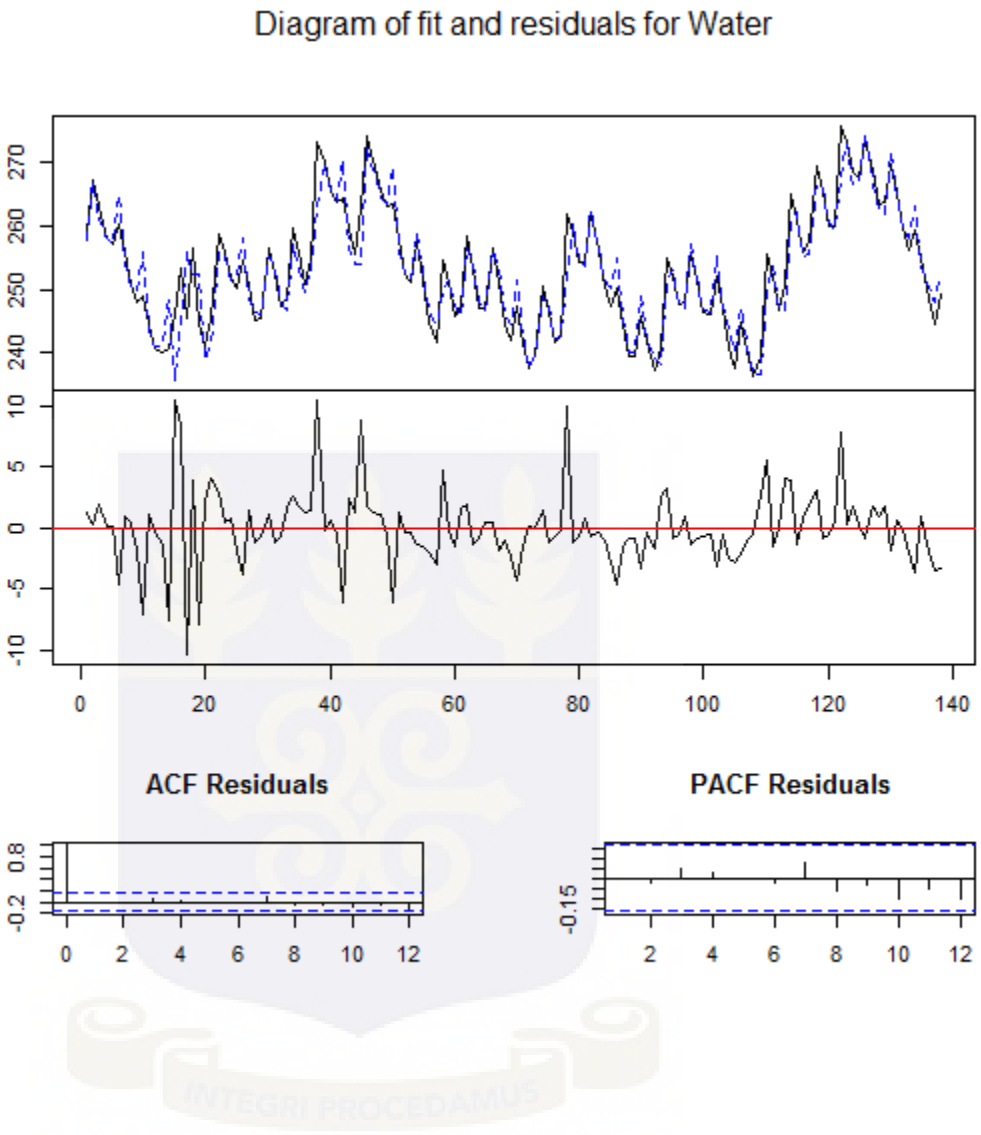

Figure B1: Diagnostic plot of VAR(2) Water Level

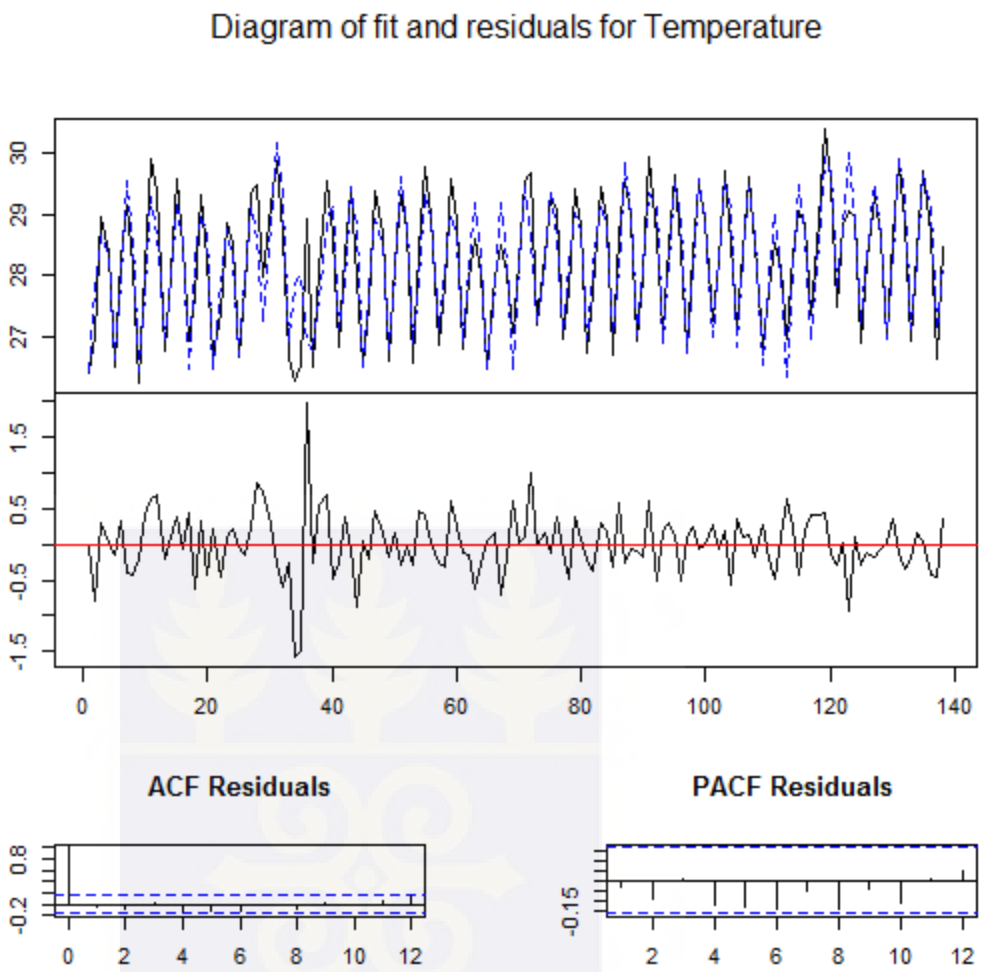

Figure B2: Diagnostic plot of VAR(2) Temperature

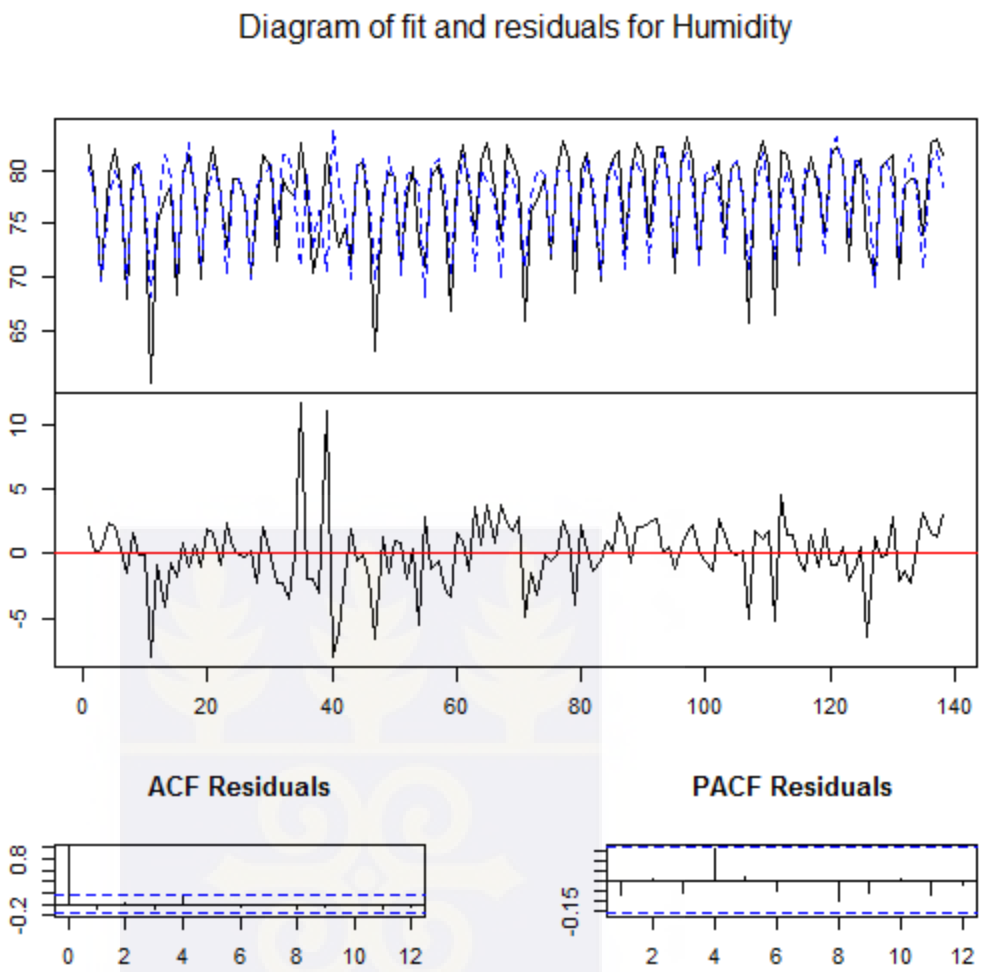

Figure B3: Diagnostic plot of VAR(2) Humidity

Table B1: Forecast of Water Level

Year	Quarter	Forecast	Lower	Upper	CI
2015	1	245.5025	239.1846	251.8205	6.3180
2015	2	241.9412	232.0528	251.8295	9.8883
2015	3	242.8531	230.5507	255.1556	12.3024
2015	4	252.3308	238.2715	266.3901	14.0593
2016	1	248.9389	233.5355	264.3424	15.4034
2016	2	244.9562	228.4826	261.4299	16.4737
2016	3	245.4398	228.0806	262.7980	17.3587
2016	4	254.5266	236.4081	272.6450	18.1184
2017	1	250.7312	231.9392	269.5232	18.7920
2017	2	246.3821	226.9780	265.7863	19.4041
2017	3	246.5682	226.5972	266.5393	19.9710
2017	4	255.4214	234.9179	275.9248	20.5035
2018	1	251.4370	230.4281	272.4459	21.0089
2018	2	246.9335	225.4407	268.4263	21.4928
2018	3	246.9944	225.0355	268.9532	21.9588
2018	4	255.7466	233.3368	278.1565	22.4099
2019	1	251.6811	228.8331	274.5292	22.8481
2019	2	247.1122	223.8372	270.3872	23.2750
2019	3	247.1202	223.4283	270.8120	23.6919
2019	4	255.8298	231.7301	279.9296	24.1000

Table B2: Forecast for Temperature

Year	Quarter	Forecast	Lower	Upper	CI
2015	1	29.6820	28.7894	30.5746	0.8926
2015	2	28.9667	27.8944	30.0389	1.0723
2015	3	26.9966	25.8014	28.1919	1.1952
2015	4	28.1757	26.8851	29.4664	1.2906
2016	1	29.3649	27.9841	30.7457	1.3808
2016	2	28.7126	27.2486	30.1765	1.4640
2016	3	26.8324	25.2917	28.3730	1.5406
2016	4	28.0423	26.4293	29.6552	1.6130
2017	1	29.2462	27.5637	30.9286	1.6824
2017	2	28.6115	26.8620	30.3610	1.7495
2017	3	26.7503	24.9359	28.5647	1.8144
2017	4	27.9751	26.0977	29.8525	1.8774
2018	1	29.1898	27.2512	31.1285	1.9387
2018	2	28.5638	26.5654	30.5621	1.9983
2018	3	26.7096	24.6532	28.7661	2.0565
2018	4	27.9403	25.8271	30.0535	2.1132
2019	1	29.1598	26.9911	31.3284	2.1686
2019	2	28.5374	26.3147	30.7602	2.2228
2019	3	26.6864	24.4106	28.9621	2.2757
2019	4	27.9195	25.5919	30.2471	2.3276

Table B3: Forecast for Relative Humidity

Year	Quarter	Forecast	Lower	Upper	CI
2015	1	70.4006	64.8560	75.9453	5.5447
2015	2	79.5581	73.6922	85.4240	5.8659
2015	3	81.2824	75.1029	87.4619	6.1795
2015	4	79.1214	72.6691	85.5737	6.4523
2016	1	71.1100	64.4899	77.7300	6.6200
2016	2	79.5529	72.8022	86.3035	6.7506
2016	3	80.5674	73.6896	87.4451	6.8778
2016	4	78.3770	71.3722	85.3817	7.0048
2017	1	70.6288	63.5014	77.7562	7.1274
2017	2	79.2172	71.9709	86.4635	7.2463
2017	3	80.2723	72.9090	87.6356	7.3634
2017	4	78.1128	71.3310	85.5922	7.4794
2018	1	70.4092	62.8149	78.0035	7.5943
2018	2	79.0391	71.3310	86.7472	7.7081
2018	3	80.1243	72.3036	87.9451	7.8207
2018	4	77.9870	70.0547	85.9193	7.9323
2019	1	70.3016	62.2589	78.3444	8.0428
2019	2	78.9467	70.7947	87.0988	8.1521
2019	3	80.0444	71.7842	88.3046	8.2602
2019	4	77.9169	69.5497	86.2842	8.3672

APPENDIX C

Maximum Likelihood Estimation of the Cointegrated VECM

The ML estimation of Equation (3.34) without the deterministic term as detailed by Johansen (1995) and Hamilton (1994):

Begin by estimating the following regression equations

$$\Delta \boldsymbol{y}_t = \hat{\boldsymbol{\Gamma}}_1 \Delta \boldsymbol{y}_{t-1} + \ldots + \hat{\boldsymbol{\Gamma}}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \hat{\boldsymbol{u}}_t$$

$$\boldsymbol{y}_t = \hat{\boldsymbol{\phi}}_1 \Delta \boldsymbol{y}_{t-1} + \ldots + \hat{\boldsymbol{\phi}}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \hat{\boldsymbol{v}}_t$$

Then solve the eigenvalue problem

$$\left| \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} \right| = 0$$

for $\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_n$ and associated eigenvectors $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n$ that satisfy

$$\hat{\lambda}_i S_{11} \hat{v}_i = S_{10} S_{00}^{-1} S_{01}, \; i = 1, 2, \ldots, n$$

$$\hat{V} S_{11} \hat{V} = \boldsymbol{I}_n$$

where $\hat{V} = \left[\hat{v}_1, \ldots, \hat{v}_n \right]$, and

$$S_{00} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{u}}_t \hat{\boldsymbol{u}}_t', \; S_{00} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{u}}_t \hat{\boldsymbol{v}}_t', \; S_{11} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{v}}_t \hat{\boldsymbol{v}}_t'.$$

The unnormalized ML estimate for $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{mle} = (\hat{v}_1, \ldots, \hat{v}_r).$$

"By imposing an appropriate normalizing and identifying restrictions, the normalized ML estimate for $\boldsymbol{\alpha}$ is given by"

$$\hat{\alpha}_{c,mle} = S_{01}\hat{\beta}_{c,mle}.$$

At this juncture, use the "multivariate least squares of the VECM" to estimate the remaining parameters. The estimates for α and β must be substituted as shown below;

$$\Delta y_t = \alpha_c \hat{\beta}'_{c,mle} y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

From

$$L_{\max}^{-2/T} \alpha |S_{00}| \prod_{i=1}^{r} (1 - \hat{\lambda}_i),$$

"the maximum value for the likelihood function for r cointegrating vectors used in the construction of LR tests for the number of cointegrating vectors is, the estimates of the orthogonal complements of α_c and β_c are given by"

$$\hat{\alpha}_{c,\perp} = S_{00}^{-1} S_{11}(\hat{v}_{r+1}, \ldots, \hat{v}_n)$$

$$\hat{\beta}_{c,\perp} = S_{11}(\hat{v}_{r+1}, \ldots, \hat{v}_n).$$

"Let $\boldsymbol{C}$ be any $(n \times r)$ matrix such that $\hat{\beta}_c$ has full rank. Then β may be normalized as $\beta_c = \beta(\boldsymbol{C}'\beta)^{-1}$ satisfying $\boldsymbol{C}'\beta_c = \boldsymbol{I}_r$ provided $|\boldsymbol{C}'\beta| \neq 0$. Johansen suggests setting $\boldsymbol{C} = (\boldsymbol{I}_r : \boldsymbol{0})'$. This choice of $\boldsymbol{C}$ corresponds to solving the cointegrating relations $\beta' y_t$ for the first r variables. To see this, let $y_t = (y'_{1t}, y'_{2t})'$ where y_{1t} contains the first r variables and y_{2t} contains the remaining $n - r$

variables and let $\beta' = (-\beta_1 \vdots \beta_2)$ where β_1 is $(r \times r)$ and β_2 is $r \times (k-r)$". Then $\beta_c' = -\beta_1$ and

$$\beta_c = \begin{pmatrix} I_r \\ -\beta_1^{-1}\beta_2 \end{pmatrix}$$

provided β_1 has full rank *r*.

R Codes

```
##PAckages##
library(fUnitRoots)
library(urca)
library(MTS)
library(vars)
library(MASS)
library(tseries)
library(forecast)
library(astsa)
library(FinTS)
##Data###
read.csv("Data")
##Functions##
```

plot.ts: function used to obtain time series plot.

urdfTest: "ADF test of unit root"

urkpssTest: "KPSS test of stationarity"

tsdisplay(diff(data,4)): Plot of quarterly Seasonal difference

diff(diff(data,4)):Difference of a quarterly seasonal difference

acf: Plot ACF (Correlogram)

pacf: Plot of PACF (Correlogram)

Arima: Used in identification and estimation of SARIMA model

sarima: Used in estimation of SARIMA model

tsdisplay: For diagnostic plot

Box.test: Ljung-Box test for autocorrelation

ArchTest: Constancy of variance test

sarima.for: Forecasting with SARIMA model

VARselect: Selecting the VAR order

VAR: Estimating the VAR model

plot: Obtaining diagnostic plots for the fitted VAR models

roots: Computes roots of VAR model

arch: Multivariate test of Heteroscedasticity

normality: Jarque- Bera Multivariate test of Normality, Skewness and Kurtosis

serial: Multivariate Portmanteau and Breusch – Godfrey LM test of Serial correlation

stability: OLS-CUSUM Stability test

causality: Granger causality test

predict: Predicting with VAR model

plot: Plot of forecast with actual values

irf: “Impulse response analysis”

fevd: “Forecast error variance decomposition”

ca.jo: Johansen trace statistic for identifying number of cointegration relations

cajorls: Normalized Vector error correction model

APPENDIX D

Table 1C: Quarterly Water level, Temperature and Relative Humidity data

Year	Quarter	Water Level	Temperature	Relative Humidity
1980	1	265.02	29.2599	71.72228
1980	2	259.87	27.91552	78.6147
1980	3	259.11	26.40731	82.33889
1980	4	267.2	27.02138	77.61989
1981	1	263.1267	28.96272	69.92608
1981	2	258.1433	28.37091	79.51487
1981	3	257.2533	26.49925	82.03907
1981	4	260.2333	28.26342	78.56918
1982	1	255.7367	29.15943	67.875
1982	2	250.83	28.13948	80.37832
1982	3	248.1233	26.25229	80.74677
1982	4	248.82	28.10756	76.81971
1983	1	244.63	29.91617	60.02942
1983	2	240.6233	29.38848	74.99301
1983	3	239.9833	26.78276	77.33746
1983	4	240.83	28.00294	78.49176
1984	1	246.08	29.58808	68.23582
1984	2	253.3287	28.46482	79.68262
1984	3	245.42	26.92348	81.45394
1984	4	256.3967	27.64808	78.09892
1985	1	244.61	29.32377	69.86905
1985	2	241.07	28.27063	79.34283
1985	3	245.48	26.68686	82.12652
1985	4	258.8367	27.38591	77.44337
1986	1	255.7333	28.84983	72.73675
1986	2	251.9533	28.54504	79.25448
1986	3	250.2367	26.68719	79.27348

Year	Quarter	Water Level	Temperature	Relative Humidity
1986	4	254.4433	27.82124	77.37957
1987	1	250.3967	29.36607	69.95123
1987	2	245.1533	29.46749	76.35323
1987	3	245.42	27.97894	81.54498
1987	4	256.3967	29.14697	80.30538
1988	1	252.4067	29.88697	71.405
1988	2	247.27	28.67401	79.23833
1988	3	248.46	26.66572	78.06111
1988	4	259.78	26.28523	77.53833
1989	1	255.8	26.52947	82.50833
1989	2	250.69	28.91344	77.855
1989	3	254.5167	26.51254	70.45497
1989	4	273.1033	28.1052	73.14228
1990	1	269.97	29.5381	81.58011
1990	2	265.5033	28.67213	75.89833
1990	3	263.88	26.84539	72.89367
1990	4	264.36	28.35099	74.57333
1991	1	259.6133	29.40259	71.69163
1991	2	255.3933	27.93172	80.25108
1991	3	262.6967	26.50591	80.96398
1991	4	274.28	27.5262	75.75143
1992	1	270.29	29.38775	62.9657
1992	2	265.49	28.69975	77.48477
1992	3	263.12	26.61382	79.75
1992	4	263.4433	28.21029	79.47832
1993	1	258.0467	29.30749	70.85
1993	2	252.2633	28.76032	77.5948
1993	3	251.1867	26.59246	80.24427
1993	4	257.5367	28.38937	73.04333
1994	1	251.59	29.78652	71.00326

Year	Quarter	Water Level	Temperature	Relative Humidity
1994	2	244.8433	29.0026	79.34785
1994	3	241.6	26.87039	80.60735
1994	4	254.5433	27.85131	75.91022
1995	1	251.2133	29.59111	66.84178
1995	2	245.7267	28.89392	79.53513
1995	3	247.26	26.80618	82.35108
1995	4	258.3	27.94703	78.1957
1996	1	253.35	28.58615	74.12792
1996	2	247.07	28.06041	80.84964
1996	3	247.0467	26.50554	82.59086
1996	4	256.6133	27.98228	78.29749
1997	1	251.2667	28.48969	73.53629
1997	2	244.59	28.04586	82.40502
1997	3	242.0567	27.05097	80.87724
1997	4	247.21	28.20699	79.1319
1998	1	241.5167	29.54184	65.91782
1998	2	237.56	29.66204	76.09731
1998	3	239.3867	27.21233	77.04803
1998	4	250.35	28.35235	79.47957
1999	1	246.3833	29.27467	71.66656
1999	2	241.64	29.08754	79.8871
1999	3	242.5467	26.96466	82.79068
1999	4	261.9633	27.7409	81.08244
2000	1	259.13	29.42258	68.49852
2000	2	254.1267	28.7995	79.98244
2000	3	254.32	26.75566	81.66774
2000	4	261.8033	27.75618	77.69444
2001	1	256.3033	29.44384	69.6682
2001	2	250.42	28.9129	79.31039
2001	3	247.2033	26.69763	81.00699

Years	Quarter	Water Level	Temperature	Relative Humidity
2001	4	250.1267	28.66213	81.91523
2002	1	244.4467	29.56463	72.61233
2002	2	239.3233	28.964	79.6681
2002	3	239.3333	26.94337	82.63387
2002	4	245.6267	27.98939	81.46129
2003	1	241.51	29.94215	73.68011
2003	2	237.3633	28.6152	82.16157
2003	3	240.6733	27.11185	82.28692
2003	4	254.7933	28.50522	78.95466
2004	1	252.0167	29.65536	70.4251
2004	2	247.53	28.43887	80.43781
2004	3	247.99	26.80394	83.16738
2004	4	255.8333	28.33054	80.87168
2005	1	251.3433	29.51229	71.30357
2005	2	246.2733	28.89839	79.0733
2005	3	245.95	27.25869	79.20735
2005	4	251.96	28.24557	80.96828
2006	1	246.99	29.71033	73.72139
2006	2	240.7467	28.51201	80.21237
2006	3	237.7033	27.18674	80.86703
2006	4	244.9267	28.37303	77.67043
2007	1	240.49	29.59562	65.6922
2007	2	236.1933	28.48884	79.83065
2007	3	239.4267	26.83099	82.85538
2007	4	255.6067	27.759	80.50269
2008	1	252.1367	28.52859	66.47256
2008	2	246.8533	28.14495	81.76487
2008	3	250.6933	26.97982	81.54229
2008	4	264.9267	28.33842	78.90394
2009	1	260.79	29.05346	71.08257

Year	Quarter	Water Level	Temperature	Relative Humidity
2009	2	255.9167	28.82061	78.76183
2009	3	257.5233	27.36516	81.20538
2009	4	269.58	28.89835	78.05233
2010	1	265.3	30.39555	73.96448
2010	2	259.9633	29.54857	81.37599
2010	3	259.9233	27.49027	82.13477
2010	4	275.75	28.82683	81.13763
2011	1	273.0933	29.07143	71.42723
2011	2	268.3567	28.96801	80.16254
2011	3	267.5633	26.91052	81.01462
2011	4	273.4867	28.24108	72.79715
2012	1	268.8833	29.28547	70.25195
2012	2	263.4133	28.63663	80.21649
2012	3	263.7367	26.98256	80.88349
2012	4	269.79	28.64864	81.4043
2013	1	265.2967	29.81398	69.79071
2013	2	259.6067	28.89625	78.64032
2013	3	256.1333	26.93127	79.1948
2013	4	259.48	28.41649	78.99104
2014	1	255.0167	29.69364	73.95065
2014	2	248.61	28.69384	82.63517
2014	3	244.58	26.64615	83.04014
2014	4	249.2033	28.47781	81.43278