FORECASTING CPO PRICE USING ARIMA, ARCH AND GARCH MODELS

RAED HAMEED MAHDI AL-TEMEEME

A Project Report submitted in partial fulfilment of the requirement for the award of the degree of Master of Science (Industrial Statistic)

Faculty of Science, Technology and Human Development
Universiti Tun Hussein Onn Malaysia

JANUARY 2016
This Master’s Project Report has been examined on date 28th December 2015 and is sufficient in fulfilling the scope and quality for the purpose of awarding the degree of Master of Science (Industrial Statistics)

Examiners

Dr. Maria Elena Bt Nor
Faculty of Science, Technology and Human Development
Universiti Tun Hussein Onn Malaysia

Dr. Mohd Saifullah Rusiman
Faculty of Science, Technology and Human Development
Universiti Tun Hussein Onn Malaysia
ACKNOWLEDGMENT

“In the name of ALLAH S. W. T, the Most Gracious and the Most Merciful Lord”,

I want to express my gratitude of thanks to ALLAH S. W. T. for giving me a good health to enable me to complete this report project. I also feel very grateful and appreciate the opportunity being a postgraduate student at Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat, Johor.

My supervisor, Prof Madya Dr. Azme Bin Khamis, whose advice provided me with when most needed and whose assistance made a great impact on my study. The special thank goes to my helpful supervisor, Prof. Madya Dr. Azme Bin Khamis. The supervision and support that he gave truly help the progression and smoothness of the report project. Whose advice provid me with inspiration when most needed and whose assistance made a great impact on my study.

My grateful thanks also go to all my lecturers who had taught me in academic during all semester, which are Dr. Rohayu binti Mohd Salleh, Dr. Siti Noor Asyikin binti Mohd Razali, Dr. Maria Elena binti Nor, Dr. Mohd Saifulah bin Rusiman and Dr. Mohd Asrul Affendi bin Abdullah.

Also, I would like to thank to my family and all my friends in my class because they give moral support and hand in helping me to complete my report project.

Finally, I thank all those whose direct and indirect advice has helped me to complete this project.
ABSTRACT

The oil palm industry in Malaysia directly contributes to the economy through financial returns that enhance the national income. A forecast of crude palm oil (CPO) price is important, especially when the investors will encounter with the increasing risks and uncertainties in the future. Therefore, the applicability of the forecasting approaches in predicting the CPO price is becoming a matter of great concern. The aim of this study is to forecast the price of palm oil in Malaysia for a period of 31 years; 1983 – 2014. The objective of the research is to propose an appropriate model to forecast the CPO price. This study involves three types of model, which are Autoregressive Integrated Moving Average (ARIMA), Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH). Akaike Information Criterion (AIC) and Hannan-Quinn Criterion (H-Q) statistic were used to obtain the best model. It was found that ARIMA (2, 1, 5) performed better compared to ARCH and GARCH models. It is concluded that ARIMA (2, 1, 5) model can be used as an alternative model to forecast the CPO price.
ABSTRAK

CONTENTS

TITLE i
STUDENT'S DECLARATION ii
ACKNOWLEDGEMENT iii
ABSTRACT iv
ABSTRAK v
CONTENTS vi
LIST OF TABLES ix
LIST OF FIGURES x
LIST OF SYMBOLS AND ABBREVIATIONS xiii
LIST OF APPENDICES xiv

CHAPTER 1 INTRODUCTION 1
1.1 Background 1
1.2 Palm Oil 2
1.3 Problem statements 5
1.4 Research Objectives 5
1.5 Scope of Study 5
1.6 Flowchart of Research Process 6
1.7 Organization of the study 8

CHAPTER 2 LITERATURE REVIEW 9
2.1 Introduction 9
2.2 Time series and Forecasting 9
2.3 Box- Jenkins in Time Series Analysis 10
2.4 ARCH and GARCH Model 15
CHAPTER 3 RESEARCH METHODOLOGY

3.1 Introduction
3.2 Stationary Model.
3.3 Non-Stationary Models
   3.3.1 Stationary in Mean
   3.3.2 Stationary in Variance
3.4 Box and Jenkins Models
   3.4.1 A pth-order autoregressive model
   3.4.2 A qth-order moving average model
   3.4.3 Autoregressive moving average models in order \((p, q)\) \((ARMA(p, q))\)
3.5 ARIMA model
   3.5.1 The methodology of ARIMA model
3.6 ARCH model
3.7 GARCH model
3.8 Stationarity For ARCH and GARCH model
3.9 Systematic Representation of model ARCH and GARCH
   3.9.1 Introduction
   3.9.2 Identification
   3.9.3 Estimation
   3.9.4 Diagnostic checking
   3.9.5 Forecasting
3.10 Model selection criterion
   3.10.1 Akaike Information Criterion
   3.10.2 Hannan-Quinn Criterion
3.9 Conclusion

CHAPTER 4 DATA ANALYSIS AND RESULTS

4.1 Introduction
4.2 Summary Statistics and Data Description
4.3 Preliminary Analysis:
4.4 Identification Model
   4.4.1 ARIMA model
   4.4.2 ARCH model
   4.4.3 GARCH model
4.4.4  Most Appropriate Model Selection  52
4.5   Estimation model  53
4.6   Test accuracy model  53
  4.6.1  Test the residuals  53
  4.6.2  Test Ljung-Box Q  54
4.7   Discussion of Results  55
4.8   Forecasting  56

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS  65
  5.1   Introduction  65
  5.2   Conclusion  65
  5.3   Recommendations  66

REFERENCES  67
APPENDICES  74
LIST OF TABLES

3.1 How to determine the model by using ACF and PACF patterns 20
4.1 Descriptive Statistics of price monthly of palm oil (1983 - 2014) 44
4.2 Comparison of suggested ARIMA models with difference statistics. 49
4.3 Comparison of suggested ARCH models with fit statistics. 51
4.4 Comparison of suggested GARCH ($p,q$) models with fit statistics. 52
4.5 Comparison of suggested best models with fit statistics. 52
4.6 Plot forecast price palm oil in Malaysia for the period 2015-2016 57
4.7 Forecast price palm oil and growth rates in Malaysia for the period 2017-2026 59
<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Palm Oil Exports by Country for 2014.</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Palm Oil Imports by Country for 2014.</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Fluctuation in Malaysia Palm Oil Price</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>Flowchart of Research Process</td>
<td>7</td>
</tr>
<tr>
<td>3.1</td>
<td>Box-Jenkins methodology</td>
<td>25</td>
</tr>
<tr>
<td>4.1</td>
<td>Time Series Plot of the of the CPO (January, 1983 to end of December, 2014)</td>
<td>45</td>
</tr>
<tr>
<td>4.2</td>
<td>General Trend plot effect for CPO</td>
<td>45</td>
</tr>
<tr>
<td>4.3</td>
<td>PACF &amp; ACF of the Price of Oil Palm</td>
<td>45</td>
</tr>
<tr>
<td>4.3</td>
<td>Box-Cox Plot of CPO Price</td>
<td>46</td>
</tr>
<tr>
<td>4.4</td>
<td>Time series plot of the first difference of CPO</td>
<td>47</td>
</tr>
<tr>
<td>4.5</td>
<td>The ACF prices after taking first differences</td>
<td>48</td>
</tr>
<tr>
<td>4.6</td>
<td>The PACF prices after taking first differences</td>
<td>48</td>
</tr>
<tr>
<td>4.7</td>
<td>The residual prices from ACF and PACF</td>
<td>54</td>
</tr>
<tr>
<td>4.8</td>
<td>Histogram of the standardized residuals from the ARIMA</td>
<td>55</td>
</tr>
<tr>
<td>4.9</td>
<td>Plot curve forecast value of a series of palm oil price (2015-2016)</td>
<td>58</td>
</tr>
<tr>
<td>4.10</td>
<td>Plot curve forecast value of a series of palm oil price (2015-2016)</td>
<td>64</td>
</tr>
</tbody>
</table>
# LIST OF SYMBOLS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACF</td>
<td>Autocorrelation Function.</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion.</td>
</tr>
<tr>
<td>ARCH</td>
<td>Autoregressive Conditional Heteroscedastic.</td>
</tr>
<tr>
<td>ARI</td>
<td>Autoregressive Integrated</td>
</tr>
<tr>
<td>ARIMA</td>
<td>Autoregressive Integrated Moving Average.</td>
</tr>
<tr>
<td>ARMA</td>
<td>Autoregressive Moving Average.</td>
</tr>
<tr>
<td>B &amp; J</td>
<td>Box and Jenkins Models.</td>
</tr>
<tr>
<td>GARCH</td>
<td>Generalized Autoregressive Conditional Heteroscedastic.</td>
</tr>
<tr>
<td>H-Q</td>
<td>Hannan-Quinn Criterion.</td>
</tr>
<tr>
<td>IMA</td>
<td>Integrated Moving Average.</td>
</tr>
<tr>
<td>MA</td>
<td>Moving Average.</td>
</tr>
<tr>
<td>PACF</td>
<td>Partial Autocorrelation Function.</td>
</tr>
<tr>
<td>Q_{L&amp;B}</td>
<td>Test Ljung and Box.</td>
</tr>
</tbody>
</table>
**LIST OF APPENDICES**

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Output for data PACF &amp; ACF of the Price of Oil Palm</td>
<td>74</td>
</tr>
<tr>
<td>B</td>
<td>Output for data the ACF prices after taking first deference</td>
<td>75</td>
</tr>
<tr>
<td>C</td>
<td>Output for data the R-squared</td>
<td>76</td>
</tr>
<tr>
<td>D</td>
<td>Output for data the Critical value</td>
<td>77</td>
</tr>
<tr>
<td>E</td>
<td>Output for data model ARIMA (2, 1, 5)</td>
<td>78</td>
</tr>
<tr>
<td>F</td>
<td>Output for data model ARCH (1)</td>
<td>79</td>
</tr>
<tr>
<td>G</td>
<td>Output for data model GARCH (3, 1)</td>
<td>80</td>
</tr>
<tr>
<td>F</td>
<td>Output for data Test the residuals from ACF and PACF</td>
<td>81</td>
</tr>
<tr>
<td>I</td>
<td>Output for data Test Ljung-Box Q</td>
<td>82</td>
</tr>
<tr>
<td>J</td>
<td>Output for data errors residuals are distributed</td>
<td>83</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background

Palm oil product is a type of fatty vegetable oil derived from the fruit of the palm tree. It is consumed in both food and non-food products. Palm oil is a highly effective and a yielding source of food and fuel. About 80% of the palm oil is used for food such as cooking oils, margarines, noodles, baked goods, etc (Chuangchid, 2013). In addition, palm oil is utilized as a component in non-edible products such as biofuels, soaps, detergents and pharmaceuticals. With such a wide range of versatile use, the global demand for palm oil is expected to grow further in the future (United States Department of Agriculture, 2012).

Many countries cultivate oil palm trees to produce oil in order to fulfil their local consumption. World trade in palm oil has increased significantly due to increased global demand. Besides that, the world production of palm oil has also increased rapidly in the last 30 years, causing the elimination of large tracts of natural forests which results in fast expansion of palm oil plantation in the Southeast Asian countries. The world production of palm oil was 13.01 million tons in 1992, increasing to 50.26 million tons in 2011, a 286% growth in 19 years (USDA, 2012). The major world producers and exporters of palm oil are Malaysia and Indonesia. In these countries, palm oil production for export purposes has become an important element, as it is one of the agricultural cash crops to replace other traditional crops such as prophylactic. The maintenance of high yields of the palm oil throughout the year is essential to achieve viability for the export market (MPOB, 2010).
1.2 Palm Oil

Palm oil is an edible oil, which is extracted from the pulp of the fruit. The colour of crude palm oil is red, which naturally similar to the pulp colour as it has high inactive vitamin. It is different from kernel oil or coconut oil. Commonly it is combined or mixed with coconut oil to make highly saturated vegetable fat, which is also used for cooking purposes and in manufacturing processed foods, soaps, cosmetics, and the ratio of cholesterol be" zero", making it a very healthy (Behrman et. al, 2005).

Palm oil plays a predominant part in the world vegetable oil export market (MPOC, 2011). It is a vegetable oil derived from the fruit of oil palm trees. They are grown across the region of the equator, mostly in Indonesia and Malaysia. Palm oil is a type of fatty energy crop derived from the fruit of the palm tree. It is practiced for both food and non-food use. It is a highly effective and high yielding source of food and fuel. In addition, palm oil is used as an ingredient in non-edible products such as biofuel, soaps, detergents, and pharmaceuticals. With such a high range of versatile use, the global demand for palm oil is expected to grow further in the future. Palm oil production has increased considerably over the past twenty years as it is easy to produce and cheap to purchase.

The industry employs people around the world and it is estimated that the producers export up to the equivalent of 22 billion Australian dollars of palm oil per annum (USDA, 2014). The palm oil exists in three principal areas Africa, Southeast Asia and South and Central America, where humans have caused the big spreading (Corley and Tinker, 2007). With the lessening of the slave trade, a new commodity was needed, where the oil palm came into the image in the commencement of the 1800s (Henderson and Osborne, 2000). Due to the economic growth and new interventions, such as soaps, candles, margarine and industrial use, the demand for palm oil rose during the 19th century (Corley and Tinker, 2007).

The demand for palm kernel oil has also increased the expansion in using areas of forests. It quickly reached its point where the demand exceeded the supply from natural palm tree plantations. In the early 20th century, the development of the oil palm as an international plantation crop began. The oil palm industry has expanded dramatically during the past 50 years with the efficient plantation management, improved processing and increased marketing. It is currently estimated that six million people are working with palm oil, where many have been put out of
poverty (Gillespie, 2012). Therefore, palm oil producing countries have a great challenge to find a sustainable balance between economic development and environmental protection (OECD/FAO, 2012). Figure 1.1 and 1.2 illustrate the amount of exports and imports ability (1000 TONS).

Figure 1.1: Palm Oil Exports by Country for 2014.

Figure 1.2: Palm Oil Imports by Country for 2014.

Malaysia has played an important role in supporting consumption and remaining competitive in the world’s oils and fats market (World Growth, 2011). The main
consumer and business market for palm oil is the food industry and, for this, the major importers are India, China and the European Union. India is the largest and leading consumer of palm oil worldwide, importing about 7.8 million tons in 2014. The European Union (EU27) is the second biggest importer of palm oil and China become third largest importer of palm oil importing about 5.7 million tons in 2014 (USDA, 2014). The current production of the world palm oil suggests an increased by 32% to nearly 60 million tons by 2020.

Malaysia is one of the largest manufacturers and exporters of palm oil and its products. Currently, the industry is flourishing as it provides employment for its people. As a result of continuous R&D efforts, a wide variety of by-products is produced. Malaysia has been recreating an important role to accomplish the needs and to stay competitive in the world’s oils and fats market (World Growth, 2011). Malaysia aims to boost palm oil industry’s output to the gross domestic product (GDP) to RM 21.9 billion, with RM 69.3 billion in export earnings during the 10th Malaysia Plan period (2011-2015) thus, making palm oil industry as a major contributor to the country’s Gross domestic product (GDP). One of the types of oil produced is crude palm oil (CPO) which can be further refined and fractionated to get a wide range of food and non-food products. The oil palm industry is a contributor to Malaysia’s export revenue. Hence, modelling and forecasting of CPO price is significant so as to hold valuable data relating to the future of CPO price. Now, the Malaysian government is focusing on the use of palm oil for the production of biodiesel to cater to the huge demand from European countries. It has also promoted the construction of more biodiesel plants due to the higher prices of fuel and increasing demand for alternative sources of energy in the Western countries which later pushing the demand and crude palm oil price even higher.

The uncertainty of palm oil price provides many advantages to many countries, not only in Malaysia, but also other developing countries involved in oil palm plantations. The high price of palm oil automatically triggers an influence to the firms or commercial operators in making more capital investments and procures labour recruitments in order to increase the palm oil productions. However, the instability of palm oil price can create significant risks to producers, suppliers, consumers, and other stakeholders. In risky conditions and price instability, an accuracy of palm oil price analysis is very important in supporting policymakers to
make informed decisions and development of their own countries (Chuangchid, 2013).

1.3 Problem Statement

A forecasting of crude palm oil (CPO) price is imperative for the decision makers in order to make the right decisions and also to enhance the managerial decision-making process in several areas of industry, farming, etc. So, forecasting CPO price is very important currently as there are many industries that rely on palm oil raw materials. The high or low CPO price can cause economic catastrophe among dependent countries in all aspects of life. In order to build accurate models with the capacity to be used as predicting price for long periods, these models help to draw the right policies for countries in aiding economic recovery, and can find out the behavior of the export prices of palm oil to reach the solutions can guide the growth of future economic policy. It is very crucial for investors and portfolio managers have a superior forecast of the volatility as this is a good starting point for assessing investment.

1.4 Research Objectives

This study aims to fulfill the following objectives:

i. to apply the ARIMA, ARCH, GARCH model in forecasting CPO.

ii. to find the best model to forecast CPO price.

iii. to forecast palm oil prices for short and long term.

1.5 Scope of The Study

This research was carried out with the data obtained from the Malaysian Palm Oil Board from January 1983 until December 2014. The price was calculated in USD currency, and three models; ARIMA, ARCH and GARCH were applied. The best model was selected to forecast the price of palm oil. The Minitab program and Gretl programs were used to determine the appropriate model. Figure 1.3 presents the fluctuation of palm oil price in Malaysia over the past 31 years (1983 – 2014). The
price was USD 372 per metric ton in January 1983 which increased to USD 731 per metric ton in December 2014.

![Figure 1.3: Fluctuation in Malaysia Palm Oil Price](image)

1.6 Flowchart of Research Process

The flow chart shows the research process in this study. Data were taken from Malaysia Palm Oil Board (MPOB). Then, three models are applied ARIMA, ARCH and GARCH to determine the most applicable model. The model was compared by using AIC and H-Q statistically to identify if the appropriate to be used in forecasting the price. If the selected model is significant, it will be proposed as an alternative model to forecast the CPO price.
Figure 1.4: Flowchart of Research Process
1.7 Organization of The Study

The study is divided into five chapters. Chapter 1 introduces the research study by providing the background to the study, a statement of the problem, objectives, the significance, scope and a brief methodology used in the study. Chapter 2 focuses on the conceptual framework and review related literature that pertains to the study whilst. Chapter 3 presents the detailed methodology used for the study. Chapter 4 covers the data analysis, presentation and discussion of the results. Finally, Chapter 5 encompasses the conclusion as well as recommendations.
CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter will discuss related work which by researcher. There are numerous studies and research on the issue of the use or application of the time series and forecasting phenomena in all areas and the most important studies will be reviewed. The use of time series is not limited to economic studies, but also used in other sciences such as trade, population growth, health care and others. The time series of scientific methods belonged in the phenomena analysis because it gives a clear picture of the relations between the time series data and it can interpret the behavior of this phenomenon through the study of the historical evolution of the time series.

2.2 Time Series and Forecasting

Forecasting is the process of making statements about events whose actual outcomes (typically) have not yet been observed or consists of using prior data to predict future values. Forecasting is a process to assist management in decision making. It is also described as the process of estimation in unknown future situations. In a more general term, it is commonly known as prediction which refers to the estimation of time series or longitudinal type data. The most popular model for this method is the Box-Jenkins model introduced by (Lawless, 2011). A time series is the collection of observations made sequentially over time and methods of analyzing this data of time series which constitute of an important area of statistics (Chatfield, 2005). Several
objectives for a time series data analysis are classified as explanation, description, control and prediction.

A time series is a sequence of data points, typically consisting of successive measurements or observations of quantifiable variables, made over a time interval (Cochrane, 2005). Usually the observations are chronological and taken at regular intervals (days, months, years), but the sampling could also be irregular. It can be defined that a time series is a set of statistics, usually collected at regular intervals. Time series data occur naturally in many application areas. The methods of time series analysis pre-date those of general stochastic processes and Markov Chains. The aims of time series analysis are to describe and summarize time series data, fit low-dimensional models, and make forecasts.

Typical examples of time series include historical data on sales, inventory, customer counts, interest rates, costs, etc. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied sciences and engineering which involves temporal measurements. Time series forecasting is the use of a model to predict future values based on previously observed values.

While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series. However, this type of analysis is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.

2.3 Box-Jenkins in Time Series Analysis

A mathematical model forecasts future economic activity based on past activity. This type of model is called autoregressive. A Box-Jenkins model is so complex that it requires sophisticated specialized software. Box-Jenkins’ ARIMA is widely used to predict the future outcomes for economic or financial purposes. The Box–Jenkins method, named after the statisticians George Box and Gwilym Jenkins, applies autoregressive moving average ARMA or ARIMA models to find the best fit of a time series model for past values of a time series (Gwilym, 1970). ARMA model is
commonly called Box Jenkins models after the US mathematician George Box and Gwilym Jenkins, who popularized them in their 1976 book, *Time Series Analysis Forecasting and Control*. The most general Box-Jenkins model includes difference operators, autoregressive terms, moving average terms, seasonal difference operators, seasonal autoregressive terms, and seasonal moving average terms. As with modeling in general, only necessary terms should be included in the model.

Box-Jenkins approach was by (Areepong, 1999) used to forecast monthly crude palm oil price. In studying to forecasts of palm oil price, the performances of multivariate the univariate model of Autoregressive Integrated Moving Average ARIMA used CPO price which is volatile where the conditional variance of the price series changes between high and low values. The forecasting of agricultural prices has traditionally been carried out by applying econometric models such as Autoregressive Integrated Moving Average (ARIMA), Autoregressive Conditional Heteroscedastic (ARCH) and Generalized Autoregressive Conditional Heteroscedastic models (GARCH) (Christetal, 2010) information criterion AIC is used to assess the goodness of fit and mean absolute percentage error MAPE is used to evaluate the forecasting performance. There are various studies and researches that used time series and models forecast in different phenomenon are, including the prediction of the palm oil prices.

The ARIMA model deals with a univariate time series data and it functions as in auto-regressive (AR) and moving average (MA) model. The process of AR depends on a weighted sum of its past values and a random disturbance term, while the process of MA model depends on a weighted sum of current and lagged random disturbances. If a time series is not stationary, it can be differenced (integrated) once or more to become stationary. Therefore, the stationary process of ARIMA model is a combination of both lagged from past values and random disturbances, as well as a current disturbance term (Pindyck & Rubinfeld, 1998). In terms of short-term forecasting, univariate time series models frequently outperform sophisticated structural models (Harvey & Todd, 1983). Khin (2010) developed multiple forecasting models, included ARIMA, to predict the short-term future prices of palm oil in Malaysia where the short term ex ante and ex post forecasting was being generated. The outcomes showed that forecasting values are satisfactory in term of statistical results. Harvey and Todd (1983) suggested that univariate time series price forecasting was reasonably accurate in predicting the future prices of diesel for short-
term forecasting yet, it was formidable to forecast the long-term price. The results of forecasting and statistical results are somewhat inconsistent when few data points are available and when the forecast horizon is extended (Bailey and Gupta 1999). For accurate forecasting, a lengthy observation is required and it is recommended that at least 50 observations are generally needed to obtain good results (Meyler, et al., 1998).

ARIMA model is a regression model and integrated moving average model and This model consists of three levels, the slope two \( p \) and \( d \), integration level and a moving average rank \( q \) and signified as follows ARIMA \((p, d, q)\).

Wood and Dasgupta (1996) used regression model, ARIMA's model and neural network model to forecast the Capital Market Index of United Stated America (MSCI). They found that the ARIMA model, which was built on the percentage changes in 3-period moving average, is performing better than the ARIMA model build on the index itself.

Lloret, et al. (2000), suggested ARIMA models as the most appropriate to forecast fishery landings in the Hellenic marine waters, since systematic biological time series data sets from explanatory variables are lacking. This methodology has been used to model and forecast the landings and catch per unit effort of many fish and invertebrate species.

Tseng et al. (2001), combined two methods to develop the fuzzy ARIMA model based upon the works of time-series ARIMA \((p, d, q)\) model and fuzzy regression model. He used the new method Fuzzy ARIMA to forecast the foreign market exchange and get the accurate forecasting value in a short time period.

Nochai & Titida Nochai (2006), studied on ARIMA Model for Forecasting oil palm price used ARIMA Model for forecasting by considering the minimum of mean absolute percentage error (MAPE). They found that ARIMA Model for the forecasting farm price of oil palm is an ARIMA \((2,1,0)\), ARIMA model for forecasting wholesale price of oil palm is an ARIMA \((1,0,1)\) or ARMA\((1,1)\), and ARIMA Model for forecasting pure oil price of oil palm is an ARIMA \((3,0,0)\) or AR\((3)\).

Mohd Arshad et al. (1986), studied on crude palm oil price forecasting using Box-Jenkins Approach a univariate ARIMA model developed by Box-Jenkins. They found that \( (0, 2, 1) (0, 1, 1) 6 \) this model indicates that the original crude palm oil series is non-stationary and contains some elements of multiplicity.
Ediger & Akar (2007), used ARIMA model and seasonal ARIMA methods to forecast primary energy demand for fossil fuel in Turkey starts in year 2005 to the year 2020.

Abdullah Lazim (2012), used ARIMA models to forecast the gold bullion coin prices, suggesting that ARIMA (2, 1, 2) is the most suitable model to be used for forecasting gold bullion coin prices. ARIMA model is the most common single variable model to predict the values of economic variables such as prices of commodities and inflation (Ansari, 2001; Meyler, 1998). ARIMA is also used in non-economic variables like the forecast evolution of certain diseases through time (Purohit, 1998).

Karia & Bujang (2011), conducted a study on forecast of crude palm oil price using ARIMA and Artificial Neural Network (ANN). They found that the ARMA family works better with the linear time series data, whereas ANN performs better with the nonlinear time series data.

Nochai Rangsan & Titida Nochai (2006), forecasted the oil palm price of Thailand into three types namely as farm price, wholesale price and pure oil Price using ARIMA Model, by considering the minimum of mean absolute Percentage error (MAPE). They found ARIMA (2,1,0) for the farm price model, ARIMA (1,0,1) for wholesale price, and ARIMA (3,0,0) for a pure oil price.

Md Nor et al. (2014), who conducted a study on forecasting of palm oil price in Malaysia used ARIMA models, neural networks and fuzzy logic systems. They found that the dynamic neural network of NARX and the hybrid system of ANFIS provide higher accuracy than the ARIMA and static neural network for forecasting the palm oil price in Malaysia.

Maizah Hura et al. (2014), studied on Volatility Modeling and Forecasting of Malaysian Crude Palm Oil Prices used and employed Autoregressive Integrated Moving Average (ARIMA) model is first used to fit the series and Generalized Autoregressive Conditional Heteroskedasticity (GARCH). They found that hybrid ARIMA (2, 1, 0)-GARCH (3, 1) model was the most appropriate model for the Malaysian CPO price.
2.4 ARCH and GARCH Model

Engle (1982) presented the ARCH model and revealed that these models were designed to deal with the assumption of non stationarity found in real life financial data. It is research that is based on the ARCH model on the idea that a natural way to update a variable forecast was to average the squared deviation of the rate of return from its mean similar to the principle used in standard deviation. The ARCH process allowed the conditional variance to change over time as a function of past errors leaving the unconditional variance constant. Empirical evidence revealed that the ARCH model required a relatively long lag in the conditional variance equation and to avoid the problems with negative variance parameters, a fixed lag structure was imposed.

Bollerslev (1986), proposed a generalized ARCH (GARCH) to overcome the limitations of the traditional ARCH model of Engle (1982). The GARCH model allowed for both a longer memory and a more flexible lag structure. In the ACRH process, the conditional variance is specified as a linear function of past sample variance only, whereas the GARCH process allows lagged conditional variances to enter in the model as well. Both the ARCH and GARCH models of Engle (1982) and Bollerslev (1986) could not tell how the variance of the return was influenced differently by positive and negative news.

Wang et al. (2005), presented studied testing and modelling autoregressive conditional heteroskedasticity of conventional stream flow models operate under the assumption of constant variance or season-dependent variances. It is shown that the major cause of the ARCH effect is the seasonal variation in variance of the residual Series. The ARMA-GARCH error model combines an ARMA model for modelling the mean behavior and a GARCH model for modelling the variance behavior of the residuals of the ARMA model.

Cheong (2009), studied modeling and forecasting crude oil markets using ARCH-type, modulus study investigates the time-varying volatility of two major crude oil markets, the West Texas Intermediate (WTI) and Europe Brent .The result should that the ARCH provides the best forecasted evaluations for the Brent crude oil data.

Musaddiq (2012), studied modeling and forecasting the volatility of oil futures used ARCH family models. They found that Glosten-Jagannathan-Runkle generalized
autoregressive conditional heteroskedastic (GJR-GARCH) (1, 2) model is best suited to forecast purposes.

Philbertha & Achmad (2014), studied on Volatility Analysis International Crude Palm Oil Price used three models (Arch, Arch-M and GARCH). They found that GARCH (1, 1) is the best model to forecast the international CPO price in foreseeable 10 years.

Zainudin (2013), studied Multi GARCH Approach on Evaluation Hedging Performance in the Crude Palm Oil Futures Market where he employed the dynamic model. They found that a parsimonious model may be appropriate when improvising the hedging performance.

You & Yeap (2014), evaluating the hedging effectiveness in crude palm oil futures markets during financial crises, on crude palm oil (CPO) futures markets used GARCH model. They found that BEKK- GARCH model appears to provide more risk reduction as compared to others.

Alam et al. (2013), presented a study forecasting volatility of stock indices with ARCH model, they investigate the use of ARCH models for forecasting volatility of the DSE20 and DSE general indices by using the daily data. GARCH, exponential generalized autoregressive conditional heteroskedastic (EGARCH), and Threshold Generalized autoregressive conditional heteroskedastic (TARCH) models are used as benchmark models for the study purpose. The GARCH and ARCH models on the other hand are considered as the best performing model jointly for DSE general index returns series.

Jiang (2012), conducted a study using the GARCH model to analyze and predict the different stock markets. The GARCH model, E-GARCH model and GJR-GARCH model analyze the rate of return and under different error distributions. Finally, after comparing the Root Mean Square Error (RMSE), choose the best model is chosen to predict the conditional variance.

Ghosha et al. (2010), studied nonlinear time series modeling and forecasting for periodic and ARCH Effects, modeling and forecasting of monthly rainfall data of Sub-Himalayan West Bengal meteorological subdivision, India. Lagrange multiplier (LM) test for testing presence of Autoregressive conditional heteroscedastic (ARCH) effects is also discussed. Final revealed that the PAR model with AR-GARCH errors has performed better than the Seasonal autoregressive integrated moving average (SARIMA) model for modeling as well as forecasting.
Ramzan et al. (2012), studied modeling and forecasting exchange rate dynamics in Pakistan using an ARCH family of models. The main objective is to provide an exclusive understanding about the theoretical and the empirical working of the GARCH class of models. The data used in the present study consists of monthly exchange rates of Pakistan for the period ranging from July 1981 to May 2010 obtained from State Bank of Pakistan. GARCH (1, 2) is found to be best to remove the persistence in volatility while EGARCH (1, 2) successfully overcome the leverage effect on the exchange rate returns under study.

Zheng et al. (2015), present studied the tuberculosis morbidity in Xinjiang. Using the Box-Jenkins approach, specifically the autoregressive integrated moving average (ARIMA) model, which is typically applied to predict the morbidity of infectious diseases; it can take into account changing trends, periodic changes, and random disturbances in time series. Models are the prevalent tools used to deal with time series heteroscedasticity. Based on the results of this study, the ARIMA (1, 1, 2) (1, 1, 1)12-ARCH (1) model is suggested to give tuberculosis surveillance by providing estimates on tuberculosis is morbidity trends in Xinjiang, China.
CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter deals with the methodology for the study and it has been subdivided into five sections aside the introductory section. Section one looks briefly at time series and its basic concepts like stationary Sections two to four will give a detailed description and explanation of the theory and concept of the ARIMA, ARCH and GARCH models.

3.2 Stationary Model.

Preprocessing Stationary is the behavior of the sum of random variables in a given time and both probabilities does not change with time. This can be achieved stationary in the time series models when the fluctuation observation about the mean of constant and variation contrast, meaning that random variables do not change over time behavior, and be time-series full preprocessing Strictly Stationary and keep the chain characteristics of static over time and realized full stationery the availability of three conditions.

(i) When the mean value Mean fixed $E(\gamma t) = \mu$

(ii) When Variance be fixed $\gamma_0 = \sigma^2 = E(\gamma t - \mu) Var(\gamma t)$

As $\gamma_0$ represent the time series variation and be constant and do not depend on the values of $t$ and estimation.

$$\gamma_0 = \frac{1}{n} \sum_{t=1}^{n}(\gamma_t - \bar{\gamma})^2$$  \hspace{1cm} (3.1)
As $y_0$ represent the time series variation and be constant and do not depend on the values of $t$ and estimation.

$$y_0 = \frac{1}{n} \sum_{t=1}^{n} (y_t - \bar{y})^2$$  \hspace{1cm} (3.1)

and

$$\bar{z} = \frac{\sum z_t}{n}$$  \hspace{1cm} (3.2)

(iii) Function autocovariance is the most common function of time teams.

$$\gamma_2(k) = \text{cov}(y_t, y_{tn}) = E[(y_t - \mu)(y_{t+h} - \mu)] = E[(y_{t+k} - \mu)(y_t - \mu)]$$

$$= r_{t(-k)}$$  \hspace{1cm} (3.3)

that is to say, Joint Distribution function following as

$$F(y_{t1}, \ldots, y_{tn})(y_{t1}, \ldots, y_{tn}) = F(y_{t1+k}, \ldots, y_{tn+k})(y_{t1}, \ldots, y_{tn})$$

As the $k$ represents displacement (Lag) and $\gamma_k$ represent auto covariance for time series and it's an even function and identical around zero. (GwilymM, 2013); (Hamilton, 1994).

3.3 Non - Stationary Models

Many economic and business time series are non-stationary. Non-stationary time series can occur in many different ways. In particular, economic time series usually show time-changing levels. Most of the time series models are a kind of non-stationary Stochastic Models Time Series can be identified through the functions of autocorrelation and partial autocorrelation, were not close value to zero after the second displacement or third, but a large number of displacements values remain, but have the desirable property that is the usability transformation to a stationary time series, and there are two cases of stationary in the time series are (Hamilton, 1994).

3.3.1 Stationary in Mean

The Stationary on average realized when the chain doesn't appear a general trend. The non-Stationary around the mean it means lack of volatility time series about the fixed mean. Can removed by taking appropriate differencing such chains said to be
homogeneous and the number of differences taken is the degree of homogeneity. It can get a stationary time series after taking \( d \) of any differences.

\[
W_t = \nabla^d Y_t \tag{3.4}
\]

As the symbol \( \nabla \) called the backward difference operator, which is known as the following.

\[
\nabla Y_t = (1-B) Y_t = Y_t - Y_{t-1} \tag{3.5}
\]

In general, and to know the differences \((d)\) of the following formula:

\[
\nabla^d Y_t = (1-B)^d Y_t \tag{3.6}
\]

If make up for \( Y_t = \nabla^d Y_t \) we get the new model, can address a certain kind of non-stationary time series, called homogeneous, which called mixed model is non-stationary, And which is symbolized by the (ARIMA) \((p, d, q)\) and the formula:

\[
\phi (B) \nabla^d Y_t = \phi (B) (1-B)^d Y_t = \theta (B) \tag{3.7}
\]

If we substitute a set the value of \((W_t)\), the specimen (ARIMA) \((p, d, q)\) of the series \((Y_t)\) turn into a model \((p, q)\) (ARMA) for the series \((W_t)\) In other words,

\[
\phi (B) W_t = \theta (B) \tag{3.8}
\]

Therefore, all the theoretical concepts of Autoregressive –Moving Average ARMA

Can be applied to models ARIMA \((p, d, q)\) Autoregressive Integrated-Moving Average Models (Anderson, 1971).

### 3.3.2 Stationary in Variance

When a time series is not stationary in variance we need a proper variance stabilizing transformation. It is very common for the variance of a non-stationary process to change as its level changes. Thus, let us assume that the variance of the process is:

\[
V(y_t) = kf(\mu_t) \tag{3.9}
\]

For some positive constant \( k \) and some known function \( f \). The objective is to find a function \( h \) such that the transformed series \( h(y_t) \) has a constant variance. Expanding \( h(y_t) \) in a first-order Taylor series around \( \mu_t \).
\[ h(y_t) \equiv h\mu_t + (y_t - \mu_t)h'(\mu_t) \] (3.10)

Where \( h'(\mu_t) \) isthe first derivative of \( h(y_t) \) evaluated at \( \mu_t \). The variance of \( h(y_t) \) can be approximated as:

\[ V[h(y_t)] \approx V[h\mu_t + (y_t - \mu_t)h'(\mu_t)] \] (3.11)

\[ [h'(\mu_t)]^2 V(y_t) = [h'(\mu_t)]^2 k f(\mu_t) \] (3.12)

Thus, the transformation \( h(y_t) \) must be chosen so that:

\[ h'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}} \] (3.13)

For example, if the standard deviation of a series \( y_t \) is proportional to its level, then \( f(\mu_t) = \mu_t^2 \) and the transformation \( h(u_t) \) has to satisfy \( h'(\mu_t) = \mu_t^{-1} \). This implies that \( h(u_t) = \ln(u_t) \). Hence, a logarithmic transformation of the series will give a constant variance. If the variance of a series is proportional to its level, so that \( f(\mu_t) = u_t \), then a square root transformation of the series, \( \sqrt{y_t} \) will give a constant variance.

More generally, to stabilize the variance, we can use the power transformation introduced by Box and Cox (1964):

\[ y_t^{(\lambda)} = \begin{cases} \frac{y_t^{\lambda-1}}{\lambda} & \text{if } \lambda \neq 0 \\ \ln y_t & \text{if } \lambda = 0 \end{cases} \] (3.14)

Where \( \lambda \) is called the transformation parameter. It should be noted that, frequently, the Box-Cox transformation not only stabilizes the variance but also improves the approximation to normality of process \( y_t \) (Wei, et al., 1990).

### 3.4 Box And Jenkins Models

The study and application of time-series models has received interest from many researchers where it has increased its development and usage in various spheres of life. The Yule in 1927 was one of the first in the use of Autoregressive (AR) models and symbolizes accompanied by two researchers Box and Jenkins, Box & Jenkins in 1970 in that has this development included in the study moving average models Moving Average (MA) symbolizes, followed by mixed models and symbolizes.
(ARMA), which were used in the construction of the specimen, diagnosis and study
the suitability of the studied data is used to predict the future. The Box-Jenkins
methodology refers to the set of procedures for identifying, fitting, and checking
ARIMA models with time series data. Forecasts follow directly from the form of the
fitted model. The Box-Jenkins methodology refers to the set of procedures for
identifying, fitting, and checking ARIMA models with time series data. Forecasts
follow directly from the form of the fitted model (Hanke & Wichern, 2005).

3.4.1 A \textit{p}-th Order Autoregressive Model

AR (p), which has the general form:

\[ Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + \varepsilon_t \quad (3.15) \]

when
\[ Y_t = \text{Response (dependent) variable at time } t \]
\[ Y_{t-1}, Y_{t-2}, Y_{t-p} = \text{Response variable at time lags} \]
\[ t-1, t-2, \ldots, t-p, \text{ respectively} \]
\[ \theta_0, \theta_2, \theta_p = \text{Coefficients to be estimated} \]
\[ \varepsilon_t = \text{Error term at time } t. \]

The ACF of a time series \( y_t \) that is generated by an AR (p) process decays
exponentially with lag \( k \). Thus, if a time series \( y_t \) is generated by an AR (p) process,
then its sample ACF should decrease exponentially with lag \( k \).

3.4.2 A \textit{q}-th Order Moving Average Model

MA (q), which has the general form:

\[ Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \varepsilon_{t-q} \quad (3.16) \]

\[ Y_t = \text{Response (dependent) variable at time } t \]
\[ \mu = \text{Constant mean of the process.} \]
\[ \theta_1, \theta_2, \theta_q = \text{Coefficients to be estimated} \]
\[ \varepsilon_t = \text{Error term at time } t. \]
\[ \varepsilon_{t-1}, \varepsilon_{t-q}, \varepsilon_{t-2} = \text{Errors in previous time periods that are incorporated in the response } Y_t. \]
3.4.3 Autoregressive Moving Average Models In order \((p, q)\) (ARMA \((p, q)\))

The model for the series \(Y_t\) can be an AR \((p)\) model or an MA \((q)\) model or a combination of both the AR \((p)\) and the MA \((q)\) models. The latter model is called an autoregressive moving average of order \((p, q)\), denoted by ARMA \((p, q)\), an ARMA \((p, q)\) model has the general form:

\[
Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}
\]

(3.17)

Which can use the graph of the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to determine the model which processes can be summarized as follows table 3.1. Once a stationary series has been obtained, then identify the form of the model to be used by using the theory in Table 3.1.

<table>
<thead>
<tr>
<th>Model</th>
<th>ACF</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR((p))</td>
<td>Dies down</td>
<td>Cut off after lag (q)</td>
</tr>
<tr>
<td>MA((q))</td>
<td>Cut off after lag (p)</td>
<td>Dies down</td>
</tr>
<tr>
<td>ARMA ((p, q))</td>
<td>Dies down</td>
<td>Dies down</td>
</tr>
</tbody>
</table>

3.5 Autoregressive Integrated Moving Average Models (ARIMA)

ARIMA is a popular time series modeling developed by Box and Jenkins. The model is applied in cases where data show evidence of non-stationarity. Transformations such as differencing is used to remove non-stationarity in the mean of the series while a proper variance stabilizing transformation introduced by Box and Cox can be used to remove non-stationarity in the variance of the series (Contreras et al., 2003).
The model is defined as ARIMA \((p, d, q)\) with the following equation:
\[
\phi(B)(1-B)^d Y_t = \theta(B) \epsilon_t
\]  
(3.18)

And when
\[
(1-B)^d Y_t = W_t = \nabla^d Y_t
\]  
(3.19)

The equation (3.18) become:
\[
\phi(B_1)W_t = \theta(B) \epsilon_t
\]  
(3.20)

that is
\[
(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) W_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \epsilon_t
\]  
(3.21)

The formula represents the latest model ARIMA \((p, d, q)\), where \((p)\) represent the rank AR and \((q)\) represents rank MA and \((d)\) represents the number of differences are taken and both, \((p, q, q)\) take a positive integer number and \(\phi(B)\) represent an Autoregressive Integrated factor stable and which are the roots of the equation \(\phi(B) = 0\) outside the unit circle radius equal to one (Fuller, 1976). The \(\theta(B)\) represents the coefficient of moving averages, which have a reflective property in the sense that the roots of the equation \(\theta(B) = 0\) outside the unit circle radius equal to one.

when \(p = 1, d = 1, q = 1\) in the equation (3.19) we get the model \((1,1,1)\) ARIMA the formula:
\[
(1 - \phi_1 B)(1-B) Y_t = (1 - \theta_1 B) \epsilon_t
\]  
(3.22)

that is
\[
Y_t = Y_{t-1} + \theta_1 Y_{t-2} - \theta_2 Y_{t-2} + \alpha_t - \theta_1 \epsilon_{t-1}; |\phi_1| < 1; |\theta_1| < 1
\]  
(3.23)

when \(q = 0\) in equation (3.18), the Autoregressive Integrated Models, the symbolizes ARI \((p, d, 0)\) writes according to the following formula:

\[
\phi(B)(1-B)^d Y_t = \alpha_t
\]  
(3.24)

or
\[
\phi(B)W_t = \alpha_t
\]  
(3.25)

as the
\[
W_t = (1-B)^d Y_t = \nabla^d Y_t
\]  
(3.26)

that is
\[
(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) \nabla^d Y_t = \epsilon_t
\]  
(3.27)
and \( \emptyset (B) \) it must be located outside the unit circle which radius is equal to one in order to be stationary. And when the value of \( P = 1 \) and \( d = 1 \) in equation (3.25), the model ARIMA \((1, 1, 0)\) or AIR \((1, 1)\) writes according to the following formula:

\[
(1 - \emptyset_1 B)(1 - B)Y_t = \varepsilon_t \quad (3.28)
\]
or

\[
Y_t = Y_{t-1} + \emptyset_1 Y_{t-1} - \emptyset_1 Y_{t-2} = a_t \quad -1 < \emptyset_1 < 1 \quad (3.29)
\]

When \( p = 0 \) in equation (3.18), the Integrated Moving Average Model, which symbolizes IMA \((0, d, q)\) and has the following format:

\[
(1 - B)^dY_t = \theta(B)\varepsilon_t \quad (3.30)
\]
or

\[
W_t = \theta(B)\varepsilon_t \quad (3.31)
\]

As the

\[
W_t = (1 - B)^dY_t = \nabla^d Y_t \quad (3.32)
\]

That is

\[
\nabla^d Y_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - B^q)\varepsilon_t \quad (3.33)
\]

As \( \theta (B) \) check reflectivity as that property located outside the boundaries of the circle which radius is equal to 1 when \( d = 1 \) and \( q = 1 \) in equation (3.28), the model ARIMA \((0, 1, 1)\) or IMA \((1, 1)\) writes the following formula

\[
(1 - B)Y_t = (1 - \theta B)\varepsilon_t \quad (3.34)
\]
or

\[
Y_t = Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad ; \quad -1 < \theta < 1 \quad (3.35)
\]

In 1976, Box and Jenkins, give a methodology (Fig. 3.1) in time series analysis to find the best fit of time series to past values in order to make future forecasts.
REFERENCE


Organization. ISBN 92-5-104859-2. II analyses are carried out using a software called minitab


United States Department of Agriculture (USDA), 2012 Foreign Agricultural Service, World Markets and Trade.

United States Department of Agriculture (USDA), 2014 Foreign Agricultural Service, World Markets and Trade.


