ANALYSIS OF THE RELATIONSHIP BETWEEN EXCHANGE RATE, INFLATION RATE AND GOLD PRICE OF SRI LANKA: CO-INTEGRATION APPROACH

K.M.E.M. KARUNAWARDANA

(138378F)

Degree of Master of Science

Department of Mathematics

University of Moratuwa

Sri Lanka

MAY 2018

ANALYSIS OF THE RELATIONSHIP BETWEEN EXCHANGE RATE, INFLATION RATE AND GOLD PRICE OF SRI LANKA: CO-INTEGRATION APPROACH

K.M.E.M. KARUNAWARDANA

(138378F)

Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Operational Research

Department of Mathematics

University of Moratuwa

Sri Lanka

MAY 2018

DECLARATION

I do hereby declare that the work reported in this thesis was exclusively carried out by me under the supervision of Mr. T.M.J.A. Cooray. It describes the results of my own independent research except where due reference has been made in the text. No part of this thesis has been submitted earlier or concurrently for the same or nay other degree.

Date:………………………..   ……………………………...

Signature of the Candidate

Certified by:

1. Supervisor(Name):………………………..   Date:………………………...

   Signature:…………………………..

To my beloved parents

ACKNOWLEDGEMENT

Completing this master work has been a wonderful chance I got in my life. Due to the limited time frame it became a challenging and a real learning experience. First I would like to thank my mother because of her affection. I greatly thank to my supervisor Mr T.M.J.A.Cooray to his patience and willingness to instruct in this research work. Also I would like to thank Head of the Department of Mathematics, University of Moratuwa for giving me a chance to carry out my work at the department. Unless support of my friends my research and thesis would have not be materialized. Their words and suggestions often boosted my courage and determination to write this thesis. Finally, I would like to thank everyone who encouraged me to finish this exciting endeavor.

Thank you.

K.M.E.M. Karunawardana.

ABSTRACT

Recent records show that the price of gold has been rising at a higher rate than in the past. This has been shown to be true for Sri Lankan gold prices as well. In this study an attempt has been made to develop a forecasting model for gold price and to examine the relationship between selected factors, that is the inflation rate, exchange rate and gold price. The data was mined from the World Gold Council and the Central Bank of Sri Lanka. The sample data of gold price were gathered from 2007 January to 2016 March in the currency of US dollars per troy ounce. It was converted into Sri Lankan rupees per 22 carat. Data until December 2015 were used to build the ARIMA model and the VEC model remainder was used to forecast the gold price and to check the accuracy of the model. Box-Jenkins, Auto Regressive Integrated Moving Average methodology (ARIMA) has been used to developed the model $D[Ln[GOLD\ PRICE]]$; with terms AR (3) and MA(3) and to forecast the future gold price. The MAPE value of fitted data in the appropriate model is 9.4%. To identify the relationship with gold price, inflation rate and exchange rate, quarter value data of all three factors were used. Two models were developed by based on the minimum AIC and the minimum SIC values. Firstly, the stationarity of the data is checked through the Augmented Dickey Fuller test and then the Johansen co-integration test and the Vector error correction model (VECM) are employed for analysis. The results of the Johansen co-integration test revealed that exchange and inflation rates are co-integrated with the gold price that led to run VECM. The VEC model developed for minimum AIC value provides evidence for the existence of long run and short run relationships between the gold prices, the exchange rate and the inflation rate and the model developed for minimum SIC value as well. The model developed based on minimum SIC value is rejected since the existence of serial correlation. The speed of adjustment to equilibrium is 12.1%, the model explains the gold price of the current quarter as 69.3% of the gold price of the previous quarter, and the exchange and inflation rates in the VEC model developed based on minimum AIC value. The MAPE value of fitted data from appropriate VEC model is 6.36%. When forecasting time period is increasing the percentage error in ARIMA model is higher than the percentage error increasing in appropriate VEC model. According to the mean absolute percentage error as forecasting accuracy measure the study concluded that the VEC model is more appropriate fitted model to forecast the gold price in Sri Lanka than the fitted ARIMA model.

Key Words: Auto Regressive Integrated Moving Average (ARIMA), Regression Analysis, Co-integration, Vector Error Correction Model, Granger Causality

TABLE OF CONTENTS

		Page
DECLARATION		i
DEDICATION		ii
ACKNOWLEDGEMENT		iii
ABSTRACT		iv
TABLE OF CONTENTS		v
LIST OF FIGURES		xi
LIST OF TABLES		xiii
ABBRIVATIONS		xiv
1.	**INTRODUCTION**	**01**
1.1	History of Gold	01
1.2	Background of the Study	02
1.3	Determining Price of Gold	04
1.4	Significance of the Study	05
1.5	Objectives of the Study	06
1.6	Data Collection of the Study	06
1.7	Limitations of the Study	07
1.8	Period of the Study	07
1.9	Forecasting	07
1.10	Content of Thesis	08

2. LITERATURE REVIEW **09**

2.1 Background 09

2.2 Exertion in India 09

2.3 Exertion in Malaysia 10

2.4 Gold Price in the World 11

2.5 Statistical Methods and Models Used to Analysis the Gold Price 12

3. METHODLOGY **15**

3.1 Definition of Time Series 15

3.2 Time Series Analysis 15

3.3 Components of a Time Series 16

3.4 Stationary and Non Stationary 17

3.4.1 Transforming Non -Stationary to Stationary 18

3.4.2 Unit Root Test 19

3.4.2.1 ADF test and PP test 19

3.5. Kruskal- Wallis Test for Seasonality 20

3.6 The Covariance 20

3.6.1 The Autocorrelation Coefficient 21

3.6.2 The Partial Autocorrelation Function 22

3.7 Autoregressive Model of Order p-AR(p) 22

3.8 Moving Average Model of order q-MA(q) 22

3.9 Autoregressive Moving Average Model of Order p, q-ARMA(p, q) 23

3.10 Autoregressive Integrating Moving Average Model (ARIMA) 23

3.10.1 Assumptions 24

3.10.2 Limitations 24

3.10.3 Identify the Order of Differencing ARIMA Models 25

3.11 Error Forecasting 25

3.11.1 Mean Absolute Percentage Error 25

3.12 Time Series Model Building 25

3.13 Model Selection Criteria 26

3.13.1 Akaike's Information Criteria (AIC) 26

3.13.2 Schwarz Information Criteria (SIC) 27

3.13.3 Coefficient of Determination 27

3.14 Model Diagnostic Checking 27

3.14.1 Normality of Residuals 27

3.14.2 Lagrange's Multiplier Test 28

3.14.3 White's General Test 28

3.14.4 Anderson- Darling Test 28

3.15 Forecasting Accuracy 29

3.16 Regression Analysis 29

3.16.1 Correlation 29

3.16.2 Pearson Correlation Coefficient 30

3.16.3 Linear Regression Model 31

3.17 Spurious Regression 31

3.17.1 Durbin-Watson Statistic 32

3.18 Granger Causality 32

3.18.1 Best Lag Selection for Testing Linear Granger Causality 34

3.19 Integrated Order p or $I(p)$ 35

3.20 Co-integrated Order(d,p) or $CI(d,p)$ 35

3.21 Background for Co-integration 35

3.22 Definition for Co-integration 36

3.22.1 Importance of Co-integration 36

3.22.2 Testing for Co-integration 36

3.22.3 Pre Condition to Johansen Test 37

3.22.4 Johansen Test 37

3.22.5 Johansen Maximum Likelihood Co-integration Test 38

3.22.5.1 Trace Test 38

3.22.5.2 Maximum Eigen Value Test 39

3.23 VAR Models and Co-integration 39

3.24 Error Correction Models 40

3.25 Vector Error Correction Model 40

3.26 Procedures for Testing Co-integration and VECM 41

3.27 Wald Test 42

3.27.1 Hypothesis Check to Joint Significance of Parameters 42

4. TIME SERIES MODELS TO FORECAST GOLD PRICE 43

4.1 Preliminary Analysis 43

4.2 First Difference Series of Monthly Gold Price 45

4.3 Model Identification and Coefficient Estimation $D[GOLD\ PRICE]$ 47

4.4 Box-Cox Transformation 48

4.5 Model Identification and Coefficient Estimation $D[Ln\ [GOLD\ PRICE]]$ 51

4.6 Residual Analysis $D[Ln[GOLD\ PRICE]]$ −Model 2 53

4.7 Forecasting and Accuracy of the Model 57

4.8 Summary of Findings 58

5. INFLUENCE OF INLATION RATE AND EXCHANGE RATE ON THE GOLD PRICE 59

5.1 Introduction 59

5.2 Descriptive Analysis 59

5.2.1 Regression Model 60

5.3 Co-integration Equation 62

5.3.1 Lag Selection Criteria 62

5.3.2 Pre condition for Johansen Co-integration Test 63

5.4 The Development of a VEC Model Based on Minimum AIC value 64

5.4.1 Johansen Co-integration Test 64

5.4.2 Vector Error Correction Model (VECM) 65

5.4.3 Wald Test Results for Short Run Causalities 67

5.4.4 Residual Analysis for the Model 69

5.5 The Development of a VEC Model Based on Minimum SIC Value 72

5.5.1 Johansen Co-integration Test 72

5.5.2 Vector Error Correction Model (VECM) 73

5.5.3 Wald Test Results for Short Run Causalities 75

5.5.4 Residual Analysis for the Model 77

5.6 Appropriate Final VEC Model 78

5.7 Forecasting and Accuracy of the Model 79

5.8 Summary of Findings 79

6. DISCUSION **81**

6.1 Summary in Chapter 4 81

6.2 Summary in Chapter 5 82

6.3 Model Comparison 84

7. CONCLUSIONS AND RECOMMENDATIONS **85**

7.1 Conclusions 85

7.2 Recommendations 86

LIST OF REFERENCES 87

APPENDIX A 91

APPENDIX B 94

APPENDIX C 96

List of Figures

Figure 1.1	Percentages of gold use in different areas (pie chart)	02
Figure 1.2	The Plot of world gold price	04
Figure 3.1	Granger causes effect between two variables	33
Figure 4.1	Time series plot of monthly gold price	43
Figure 4.2	The correlogram of monthly gold price	44
Figure 4.3	The graph of first difference of gold price	46
Figure 4.4	The correlogram of first difference of gold price	47
Figure 4.5	Box-Cox transformation plot	49
Figure 4.6	The time series plot of $D[Ln[GOLD\ PRICE]]$	49
Figure 4.7	The correlogram of $D[Ln[GOLD\ PRICE]]$	51
Figure 4.8	The *E-views* output for AR(2) and MA(2) terms	52
Figure 4.9	The *E-views* output for AR(3) and MA(3) terms	52
Figure 4.10	The correlogram of residuals of model 2 $D[Ln[GOLD\ PRICE]]$	53
Figure 4.11	The histogram of residuals of model 2 $D[Ln[GOLD\ PRICE]]$	54
Figure 4.12	The normal probability plot of residuals of model 2 $D[Ln[GOLD\ PRICE]]$	55
Figure 4.13	The serial correlation of residuals of model 2 $D[Ln[GOLD\ PRICE]]$	55
Figure 4.14	The heteroscedasticity of residuals of model 2 $D[Ln[GOLD\ PRICE]]$	56
Figure 5.1	Lag selection criteria for VAR model	62
Figure 5.2	Johansen co-integration test for minimum AIC value	64

Figure 5.3	Co-integration output by *E-views*	65
Figure 5.4	The probability values of coefficients of model	66
Figure 5.5	Wald test results for coefficients of C(2), C(3) and C(4)	67
Figure 5.6	Wald test results for coefficients for C(5), C(6) and C(7)	68
Figure 5.7	Wald test results for coefficients for C(8), C(9) and C(10)	68
Figure 5.8	The correlogram of the residuals of the model	70
Figure 5.9	The histogram of the residuals of the model	70
Figure 5.10	The serial correlation of the residuals of the model	71
Figure 5.11	The heteroscedasticity of the residuals of the model	71
Figure 5.12	Comparison of the actual and fitted values of *GOLD*	72
Figure 5.13	Johansen co-integration test for minimum SIC value	72
Figure 5.14	The co-integration output by *E-views*	73
Figure 5.15	The probability values of coefficients of model	74
Figure 5.16	Wald test results for coefficients of C(2) and C(3)	75
Figure 5.17	Wald test results for coefficients of C(4) and C(5)	76
Figure 5.18	Wald test results for coefficients of C(6) and C(7)	76
Figure 5.19	The correlogram of the residuals of the model	77
Figure 5.20	The histogram of the residuals of the model	77
Figure 5.21	The serial correlation of the residuals of the model	78
Figure 5.22	The heteroscedasticity of the residuals of the model	78
Figure 6.1	Granger causality test	82

List of Tables

Table 1.1	The sample daily gold prices in Sri Lanka	05
Table 3.1	Transformations for lambda value	18
Table 3.2	Theoretical ACF and PACF	23
Table 3.3	Interpretation of r values	30
Table 4.1	Kruskal- Waillis test results for monthly gold price	45
Table 4.2	P-values for the Unit Root test	45
Table 4.3	ADF and PP test values for first difference	46
Table 4.4	Coefficient and probability values for AR(p) and MA(q) terms	48
Table 4.5	AIC, R^2 and S.E. regression values $D[GOLD\ PRICE]$	48
Table 4.6	ADF and PP test values for $D[Ln[GOLD\ PRICE]]$	50
Table 4.7	Possible combinations and probability values $D[Ln[GOLD\ PRICE]]$	52
Table 4.8	R^2, S.E. regression and AIC values for model 1 and model 2	52
Table 4.9	The forecast gold prices for appropriate model	58
Table 5.1	Correlation values	60
Table 5.2	Regression analysis	61
Table 5.3	Model summary	61
Table 5.4	ADF test results for three variables on level	63
Table 5.5	ADF test results for first difference of three variables	63
Table 5.6:	Forecasting gold price for VEC model	79
Table 6.1	Forecast gold prices for appropriate ARIMA model	84
Table 6.2	Forecast gold prices for appropriate VEC model	84

ABBREVIATIONS

ACF	Auto Correlation Function
AD	Anderson Darling
ADF	Augmented Dickey Fuller
AIC	Akaike Information Criteria
AR	Autoregressive
ARMA	Autoregressive and Moving Average
ARIMA	Autoregressive Integrated Moving Average
ECM	Error Correction Model
LM	Lagrange’s Multiplier
Ln	Logarithm
MAPE	Mean Absolute Percentage Error
PACF	Partial Auto Correction Function
PP	Phillips-Perron
SIC	Swartz Information Criteria
VACM	Vector Error Correction Model
VAR	Vector Auto Regressive

CHAPTER 1

INTRODUCTION

This chapter discusses the background of the study, method of data collection, objectives, limitations of the study, an introduction for forecasting and outline of the thesis.

1.1 History of Gold

When studying about the gold it can be recognized that it has a long history as a valuable metal. For instance it has been identified that gold articles came into usage more than 500 years ago in Egypt (*Kirkemo et al (2005-2017))* and the capstones on the Pyramids of Giza were made from solid gold. According to a study by *(Deshmukh et al (2017))*, the gold has been recognized as one of the most expensive metal but it is not clear as to why it was known as exceptionally luxurious metal throughout the human history. However, studies show that people used it to show off their wealth and power to society (*Deshmukh et al (2017)).* It has also been reported that possession of gold as a symbol of personal power and the governments use gold as a relative standard for currency equivalents (*Tharmmaphornphilas et al (2012)).* The first use of gold as money occurred around 700 B.C. The ancient records said the first known currency exchange ratio, which mandated the correct ratio of gold to silver has been produced by the Egyptians. One piece of gold was recognized to be equal to two and a half parts of silver. Therefore, it is believed that humans almost intuitively place a high value on gold. Pure gold has a bright yellow colour and traditionally it is considered as very attractive (*Prakash et al (2014)*). Further *(Prakash et al (2014))* also states that gold was known as one of the most eye-catching metals for thousands of years because of its rarity, luster and natural beauty. Hence, due to these reasons not only gold has been popular as jewelry but also its vitality spread in large areas *(Prakash et al (2014)).*

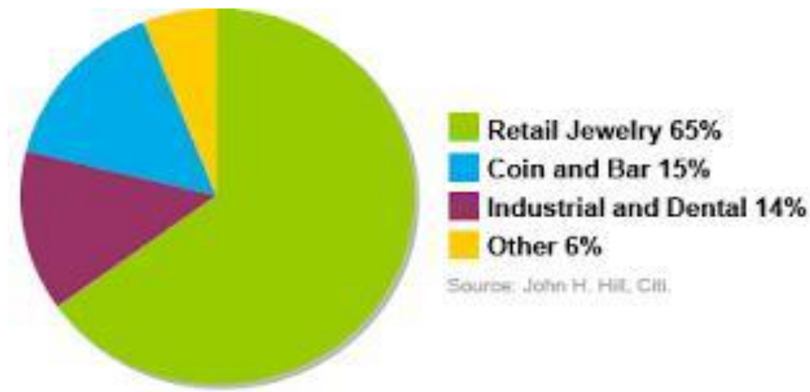


Figure 1.1: Pie chart describes the approximate percentage of gold use in different areas (http://www.google.image).

1.2 Background of the Study

In accordance with the study carried out by (*Ismail et al (2009)),* gold is one of the most important commodities in the world. Thus, day by day the trend of the market for gold is increasing and is accumulated over centuries. In 2008 and early 2009 prices of the most metals fell and the global economy was in recession. Consequently it is arguable as to whether there will be growing demand in gold usage in the future (*Tharmmaphornpilas et al, (2012)).* For example, in year 1998, the total world supply of gold is 125000, metric tons (*Ismail et al (2009)).* Even though, supply and the demand are main factors for determining the gold price (*Simakova (2011)* & (*Ismail et al (2009))* & (*Tharmmaphornpils et al (2012))*; there are many factors that must be taken into account which understanding the causes result in rising price of gold. Therefore it is understood that the price of gold is fluctuated due to various reasons which we cannot control. But it can be measured and forecast for future decisions *(Sepanek et al (2014)).* Approximately 60% of global corporate gold production is owned by the World Gold Council, and the price of gold is in many terms influenced by government and central banks, monitory policy performed by governments, changes in interest rates, inflationary policies. *(Simakova (2011))* states that in terms of the gold market, the dollar exchange rates are very important and there is a negative correlation between the gold price and dollar exchange rates. Accordingly, strong dollar keeps the price of gold controlled and low. In addition, oil price and inflation rate are two other major macro economic variables that influence the gold market. A study by (*Shafee et al (2010))* confirms the above pointing that there is a positive correlation between gold and crude oil prices, and the correlation between gold price and inflation rates are negative with a value of -9%.

Further the study explains that there are factors contributing to short-term and long-term gold price escalations. In the short-term, there are two main reasons why gold prices dramatically increase. First reason can be recognized as a period in which the global financial market crashes and the global economy is in a recession. Second reason is the devaluation of the US dollar versus other currencies, and international inflation with high oil prices (*Shafiee et al (2010)) & (Sindhu (2012)) & (Poonam et al (2014))*. This is justified by (*Bishnoi et al (2014)) & (Tripathi et al (2014))* in their studies. Consequently, it implies that gold trading will set the potential movement of real value in the short term market against US dollar oscillations and inflation (*Shafiee et al (2010)).*

The long-term there are three major reasons for increasing gold prices reveled by (*Shafiee et al (2010))*. One reason is that mine production has gradually reduced in recent years. As a result if the gold price increases, the profit of mining companies expected to rise and as a result of that, share price may increase (*Prakash et al (2014))*. Second reason is that the institutional and retail investment has rational expectations when markets are uncertain and therefore price can be increased. Investing in gold compared to other financial market scan is recognized as the third reason. There seemed to have been two significant gold price jumps in the historical trends. The first was in early January 1980 and the second historical jump in price has been started in 2008 and is continually increasing. Gold price dramatically increased in 1980. With the rapid rising of gold prices it was encouraged to explore to new gold mines and consequently new mines were opened in 1980. The sharp declines in consumption in 1974 and 1980 were resulted from the reduced demand for jewelry and investment products, which in turn create a rapid price increase in those years. The world gold demand from 1998 to

2007 decreased while net retail investment, exchanging traded funds (ETEs) and industrial demand increased (*Shafiee et al (2010))*. Due to recent financial crisis in 2008, the gold price was increased by 6%, while many key mineral prices fell and other equalities dropped by around 40% (*Shafiee et al (2010))*. But people continued to buy and sell gold for a wide variety of reasons i.e. as an investment, money and insurance etc (*Prakash et al (2014))*.

Following this trend, in the end the price of gold has grown fairly and steadily over the past few decades and many experts predict that it will continue its gradual climb over the next few years.

Price development in the years 2000 to 2007 is illustrated in Figure 1.2.

Figure 1.2: The plot of world gold price (http://www.kitco.com).

1.3 Determining Price of Gold

Gold is traditionally weighted by Per Troy Ounces and by grammes; 1 troy ounce equal to 31.1035 grammes. The above Figure 1.2 depicts the collected data that are in US dollars per troy ounce and it has been converted to Sri Lankan LKR as follows;

$$Gold\ Price\ LKR, 22k = \frac{Gold\ Price\ LKR}{31.1\ g} \times 8g, \qquad where\ 22k = 8g$$

Then it is converted to 22 k gold price in Sri Lanka nominal market. The gold price determining procedure is known as the gold Fixing in London which provides a daily benchmark price to the industry.

1.4 Significance of the Study

The following table shows the sample daily gold price data in Sri Lanka.

Table 1.1: The sample daily gold prices in Sri Lanka.

Date	Gold Price 22 Carat(LKR)
2017-01-27	46, 200.00
2017-01-30	46, 600.00
2017-01-31	46, 750.00
2017-02-01	47, 250.00
2017-02-02	47, 400.00
2017-02-03	47, 400.00
2017-02-06	47, 850.00
2017-02-07	48, 200.00
2017-02-08	48, 300.00
2017-02-09	48, 500.00
2017-02-13	48, 500.00
2017-02-14	48, 150.00
2017-02-15	48, 000.00
2017-02-16	48, 350.00
2017-02-17	48, 400.00
2017-02-20	48, 300.00
2017-02-21	48, 400.00
2017-02-22	48, 500.00
2017-02-23	48, 500.00
2017-02-27	49, 300.00

According to Table 1.1 the daily gold price in Sri Lanka has been fluctuating. Thus, it is necessary to determine a proper model to forecast gold price in Sri Lanka to understand a clear picture as to what will happen in the future gold market, what are the reasons for these fluctuations and to recognize whether there are any other factors that influence on changes of gold price.

1.5 Objectives of the Study

- In the gold market, it is important to undertake the various fluctuations of gold price because of sudden upwards and downwards trend appears recently. One of the objectives of this study is to recognize an appropriate time series model to forecast future gold price in Sri Lanka. Thus, the ARIMA methodology is employed to build the model and Mean Absolute Percentage Error is used to test the forecasting accuracy of the appropriate fitted model.

- To identify the factors which affected gold price in Sri Lanka and to determine the influence of inflation rate and exchange rate on gold price. Regression Analysis and Johansen co-integration test have been employed and depending on the statistical analysis Vector Error Correction Model (VECM) has developed. Mean Absolute Percentage Error is calculated to test the forecasting accuracy of the appropriate fitted model.

1.6 Data Collection of the Study

The study is based absolutely on secondary data obtained from various data sources. The monthly gold price data are available on the site www.kitco.com and world gold council website www.worldgold.com are used as the main sources. Daily gold price is taken from www.ideabeam.com. In addition, US dollar exchange rates for Sri Lankan rupees were taken from Sri Lanka Central Bank website and inflation rates were captured from the site www.goldpinflation.com.

1.7 Limitations of the Study

The following major limitation was met when conducting the study.

- The analysis of gold price mainly based on historical data
- All relevant secondary data were collected from websites and study based on monthly wise data.
- The finding of the study may be useful for the investors. But the investment decisions of the investors may be depend on their level of personal expectations.

1.8 Period of the Study

The study covers a period of 8 years spanning from 2007 January to 2016 March. The data captured from that period January 2007 to December 2015 is used to build the model and remain data is used to check the accuracy of the ARIMA model. Quarterly value gold price, inflation rate and exchange rate data from 2007, January to 2015, December are used to develop the VEC model.

1.9 Forecasting

A forecasting is a prediction of some future events. It is a study of understanding and keeping an accurate view point of past and current data to recognize how it will affect the future and to minimize the future failures. Forecasting is used in many practical fields such as Business, Economics, Finance, Science etc to ensure the ease of taking future decisions and planning. However, without a proper preparation, there is a high possibility that forecasting may fail. Therefore, it is important to have an idea as to what kind of method or model is going to use for forecasting, what kind of data we have and how we are going to handle these data to make accurate forecasts. There are two broad types of forecasting techniques: Quantitative methods and Qualitative methods (*Montgomery et al, (2015)).*

Quantitative forecasts are often used in situations where there is little or no historical data available to make the forecast. For instance, when a new product is going to introduce to the market, which would not have past data, qualitative forecasting techniques are often used (*Montgomery et al (2015))*. Sometimes qualitative forecasting methods make use of marketing tests, surveys of potential customers, and experience with the sales performance of other products. Quantitative forecasting techniques make formal use of historical data and forecasting model. Depending on the historical data it is important to develop a model first and then to use it to project the patterns in the data into the future. In other words, the forecasting model is used to extrapolate past and current behavior into the future. Among several types of forecasting models the three most widely used are regression models, smoothing models and general time series models (*Montgomery et al (2015)).*

This study uses general time series models to analysis the data and for forecasting purposes. Further in order to understand the relationship between the gold price and the factors which effect gold price, regression models multivariable time series models have been used.

1.10 Content of Thesis

This thesis contains of six chapters that are organized as follows;

Chapter 1: An Introduction
Chapter 2: Literature Review
Chapter 3: The Methodology
Chapter 4: Time Series Models to Forecast Gold Price
Chapter 5: Influence of Inflation Rate and Exchange Rate on Gold Price.
Chapter 6: The Discussion
Chapter 7: Conclusions and Recommendations

CHAPTER 2

Literature Review

This chapter briefly analyses the earlier works that past researchers have done related to the topic.

2.1 Background

Inflation of the gold price has a long history which goes beyond 1900. According to (*Tharmmaphomphilas et al (2009)),* in early stage, gold has been used for pricing rather than as a commodity. In Indian history, it is recorded that different parts of the country have different monetary system and only gold was treated as a common exchange commodity (*Guha et al (2016))*. Thus, it can be noted that gold has an authority and strength among other commodities such Silver, Platinum and etc. The situation has widened all over the world (*Prakash et al (2014))*. Gold has become a significant icon to control the world currency inflation. For instance, when foreign nations that hold billions of dollars in US debt start buying gold as they fear the value of the dollars will go down, the rising price of gold becomes more than a novelty (*Khan(2016))*. Moreover, an allocation to gold provides investors with the confidence to invest in a wider range of strategies including alternative assets (*Sindhu(2013))*.

2.2 Exertion in India

Price of gold cannot be controlled. To further analyzed this a number of studies has been carried out, especially in India to find out what factors mostly impact to alter the gold price. A study by (*Sindhu(2013))* states that the prices of the gold are increasing and is affected by the various factors such as Exchange Rate of US dollar with INR, Crude Oil prices, Repo Rate and Inflation Rate. Gold prices vs. each of the factors were studied and conclusions were made. Accordingly, there is an inverse relationship between the US dollars and gold prices; the crude oil prices have an impact on the gold prices. Eventhough gold prices and repo rates are interdependent, gold price and inflation rates are dependent but positively correlated. Some of the conclusions that were made by (*Sindhu(2013))* in his study is confirmed by an early study done by (*Sherman et al (1983))* and another study by (*Moore(1990*)); stating that the gold price has a significant positive relationship with unexpected inflation. In contrast to that (*Adrangi(2003))* says that there exists a positive relationship with expected inflation, but no relationship between the gold price and unexpected inflation.

In addition to factors that were considered by (*Sindhu(2013))*; US dollar long-term interest rate and US real GDP also have an influence on gold price *(Bishnoi et al (2014))*. Further (*Sindhu (2013))* concluded gold prices and inflation rate are dependent and positively correlated where as (*Bishnoi et al (2014))* concluded that there is a negative relationship with inflation, US long term interest rate and US real GDP.

Gold is priced in dollars. If the dollar loses its values the nominal gold price tends to rise, thus preserving the real value of gold. It implies that currency exchange rates for US dollar affecting gold price. However, (*Poonam et al (2014))* in his study reveled that there is an inverse relationship between the exchange rate and gold price. This fact was supported by (*Sindhu(2013))* and further studied by (*Capie(2015))* with a conclusion that gold as a hedge against the US dollars.

2.3 Exertion in Malaysia

Another study has been done by (*Ibrahim et al (2014))* to analyze factors that affect the prices of gold in Malaysia. Accordingly, there is an inverse relationship between inflation rates and exchange rates on the gold price. The meaning of the inverse relationship is that when one goes up other factors go down while crude oil prices and gold price exhibit positive relationship. These findings were strongly supported by the study carried out by (*Sindhu(2013)).* Crude oil and gold price relationship discussed in (*Simakova(2011))* reveals the existence of a long-term relationship between them.

2.4 Gold Price in the World

The world gold market and the historical trend of gold prices were discussed on a study by (*Shafiee et al (2010)).* Two conclusions were made in his paper; there is a high correlation between gold price and oil price which was around 85% and no significant relationship between the gold price and global inflation.

In addition to gold, silver seems to have a significant position among other commodities. (*Prakash et al (2014))* had done a study to examine the relationship between gold and silver prices over the period 2001-20013 and found that there is a significant relationship between gold and silver price.

To identify the factors that influence on gold price, further to the the regression analysis (*Gangopadhyay et al (2015))* developed a VECM to explain and forecast gold prices in India. Five key determinants were considered; stock market index, oil prices, exchange rate, interest rate and consumer price index and showed that gold price has a long-term relationship with these factors.

Another study reported to have identify the long run and the short run relationship between Indian stock and gold price. Accordingly, daily closing price data were collected for the period of ten years and VECM was employed for analysis. The results proved that there is a long run relationship between stock and gold price yet, no short run relationship exists between them. On the other hand, neither of them exist vice versa (*Banumathy et al (2015)).*

Instead of employing VECM to determine the relationship between the gold price and global factors in India (*Tripathi et al (2014))* attempted to investigate the existence of a causal relationship. It was identified that all global factors do not make Granger Cause on gold price in India. However exchange rate and crude oil price had such relationship in the period under the study.

2.5 Statistical Methods and Models Used to Analyses the Gold Price

Most of studies has been carried out using Auto Regressive Integrated Moving Average ARIMA methodology (*Guha et al (2016)) & (Khan(2013))*. The sample data of gold price were taken from January 2003 to March 2012. Data till January 02, 2012 were used to build the model while the remaining data used to forecast the gold price and to check the accuracy of the model in the study by (*Khan (2013))*. Both authors used ACF and PACF graphs to check the stationarity of the data set and (*Khan (2013))* used the "Augmented Dickey Fuller" and "Philips Perron" unit root tests used to statistically justify the stationarity of series.

To test the suitability of data set for time series analysis (*Guha et al (2016))* checked the Durbin-Watson statistic. He further highilighted that if DW values between 0 to 1.5 or between 2.5 to 4 the data is longitudinal which means it is good for time series analysis.

Based on Akaike Information Criteria (AIC) and Schwartz Information Criteria (SIC), the best model selected to forecast gold price.

To measure accuracy of forecast, Root Mean Square Error (RMSE), Mean Absolute Error(MAE) and Mean Absolute percentage Error(MAPE) had been considered and selected the least value on MAE and RMSE indicated model (*Khan(2013))*. (*Guha et al (2016))* used the same statistics and Lungs Box Q statistics. In addition to ARIMA models, (*Jadhav et al (2015))* used artificial neural network and ARIMA models to study Indian stock market indices. ARIMA (1, 0, 1) was concluded the best model, that was chosen.

(Tharmmaphomphilas et al (2012)) applied system dynamics to predict monthly gold prices from January 2010 to June 2011. Having ability to take into account of qualitative factors particularly political chaos and economic crisis events, the developed model had reduced prediction error than Holt-Winter Exponential Smoothing and Box-Jenkins method.

Due to the rising demands for gold, researchers preferred to study the reasons behind and to recognize what would be the factors that are most likely to make this influence and what level of relationship should there be between gold price and other factors. Most of studies were based on regression analysis. *(Sindhu(2013))* has done a quantitative data analysis through regression, trend analysis, standard deviation and correlation. Similar to that (*Poonam et al (2014))* attempt to test the relationship between the factors and the gold prices using statistical tools, such as T-Test and Trend analysis.

However, (*Ismail et al (2009))* had used the multiple linear regression method to understand of gold prices. For his study; gold price was used as a single dependent variable and several econometric factors had been selected as independent variables. Mean Square Error (MSE) through Statistical Packages for Social Science (SPSS) with was identified as the fitness function and had used to determine the parameter estimation and forecast accuracy receptively.

(Nair et al (2015)) had used Johansen co-integration test to check the long term association between exchange rate of US dollar to INR and gold prices in India. Further, the Granger causality test was used to check the lead lag relationship between the variables.

Similar to (*Nair et al (2015))*, a study had done by (*Tripathi et al (2014))* investigated the relationship between gold price in India and other selected factors report. The data set covered the period of nine years from April, 2004 to March, 2013 and unit root test, co-integration test and Granger Causality test have been used with *Eviews 5* statistical software packages (*Tripathi et al (2014)).* Granger Causality between gold price and oil price had been checked in a study by (*Simakova(2011))* and made a conclusion that there is no strong causality between two of. Yet, co-integration test revealed long- term relationship and proportional analysis confirmed that gold price and oil ratio moving during the period of 1970 - 2010 on its long-term values.

(Gangopadhyay et al (2015)) carried a study to determine how was the stock market index, oil prices, exchange rate, interest rate and consumer price index influence the gold price in India. The study was based on collected monthly data of each variable for a long time period starting from April, 1970 to August, 2013. All of them were analyzed in co-integrating framework therefore variables were non stationary. For this purpose ADF test were carried out and optimal lag length was found by the procedures suggested by Ng and Perron, (1995). All data are integrated of order 1 and they have one co-integrating relationship among them and VECM model were fitted.

Another VECM was done by *(Banumathy et al (2015))* for gold price and Indian stock price. Daily closing price data, for the period of ten year ranging from 2004 April to 2014 March were used in this study. ADF test was employed to check the stationarity. All data series were non stationary at level but the same becomes stationary at first difference. Based on Johansen co-integration test results, VECM was carried out. To identify the short run causality, Wald test has been employed. In addition to VECM (*Tripathi et al (2014))* used a Granger causality test to discuss long run integration between gold price and variables. However, results were not significant enough to fit a VAR model. (*Simakova(2011))* fitted a VECM by the positive results of variables and it was done for *log* of variables since they were stationary on. Whereas VAR model was not fitted.

Chapter 3

Methodology

This chapter consists of Statistical formulations which are necessary to understand the mathematical and statistical simulations in Time Series Analysis, Regression, Granger Causality and Co-integration.

3.1 Definition of Time Series

The variable or the observation $X(t)$ is arranged in a proper chronological order of time intervals. It is defined mathematically as a set of vectors X(t); t = 0,1,2,… where t represents the time elapsed. This time interval can be considered daily, weekly, monthly, quarterly or annually. The collected data set or observations have the following property, then it is regarded as a time series.

- The observations are recorded after successive equal intervals of any unit of time.

Examples of time series are the daily stock prices, daily temperature, daily oil index and weekly or monthly records of babies born at a hospital etc. The basic graphical display of time series data is the time series plot. This is just a graph versus the time t, for $t = 1,2,3, \dots, T$. Time series observations can be continuous or discrete. Observations are taken continuously in time in continuous time series. For example maximum daily temperature. In discrete time series measurements are taken only at specific points in time. Accumulations of rainfall measuring would be an example.

3.2 Time Series Analysis

There are many reasons for an analysis of time series data, Eg: to study the past behavior of data, to make better forecasts of future values and to understand the present situation better. After collecting the data set the most important task is to analyze it in order to develop a suitable mathematical model which can be used to understand the behaviour of past events as well as to forecast future events. It is necessary to have an appropriate mathematical tool to develop an adequate model. The selection of a suitable mathematical tool depends on the nature of the data set, the observations and the purpose of the analysis. While there is no minimum sample size for observations, sufficient observations should be available for better analysis. Time series analysis model is usually classified as either a time domain model or a frequency domain model. In a time domain model, mathematical functions are usually used to study the data with respect to time while in a frequency domain model, frequency is considered rather than time.

3.3 Components of a Time Series

In general, time-ordered data can exhibit different patterns over time. This pattern consists of several different components. These are Secular trend or simply, trend (T), Cyclic variation(C), Seasonal Variation(S) and Random or Irregular Variation(R). Every data set may not be characterized by all these components. The trend (T) is a long-term growth or decline pattern of a time series. This can exist as linear or non-linear. If it does not exhibit any trend it is called stagnate in the mean or it is said that the time series data set is stationary in the mean. By studying the trend which is exhibited in the plotted graph one can get an idea about past and future directions of the data set. Cyclical component shows swings around either side of the trend line and has no regular pattern. However, it repeats regularly after a certain interval of time which can be 3 to 10 years long. Seasonal Variation(S) or short-term variation occurs over a short period of time such as days, weeks months or quarters due to seasonal factors. This pattern is repetitive and predictable around the trend line. Irregular component is unpredictable and occurs due to random factors. It does not follow any regular pattern and can be obtained as a residue after elimination of the effects of other factors.

These components combine to provide a value for $X(t)$. The presence of random error occurs in both models underlying the pattern seen in the additive model and the multiplicative model.

The additive model can be written as follows;

$$X(t) = T(t) + C(t) + S(t) + I(t) + e(t).$$

The multiplicative model can be written as follows;

$$X(t) = T(t) * C(t) * S(t) * I(t) * e(t);$$

where $e(t)$ is the residual of the time series. While the multiplicative model is highly affected by its components this does not happen in the additive model.

3.4 Stationary and Non-Stationary

Basically this can be identified by plotting given time series data. Stationary implies a type of statistical equilibrium or stability in the data which produces a constant variance and mean. In general a stationary time series does not have a predictable pattern in the long term where the plots are roughly horizontal with a constant variance. Even though most of the time series encountered are non-stationary there can be stationary time series as well. There are several kinds of stationary time series. A series is said to be stationary in the wide sense, weak sense or second order if it has a fixed mean and a constant variance. A series could be non-stationary because of random walk, drift or trend.

A time series is said to be strictly stationary if its statistical properties have not changed over time. If the joint probability distribution of any two observations, say, X_t and X_{t+k}, is the same for any two time periods t and $t + k$ that are separated by the same interval k then the time series is strictly stationary. In other words, if the distributions of the observations are normally distributed, the series is said to possess strict stationarity.

If the mean is equal to zero, then the time series $X_t : t \epsilon T$ is called a white noise time series. Thus, a white noise time series consists of independent, identically distributed, random variables with mean zero and common variance σ^2. It is important that the time series data set should be stationary. Then its properties are not affected by a change in the time origin. It means that time series with trends or seasonality are not stationary.

3.4.1 Transforming Non-Stationary to Stationary

Most of the time series data exhibit seasonal and trend variations which causes the data series to become non-stationary. This can be identified by plotting the time series values $X_1, X_2, X_3, \dots, X_n$ with respect to the time. Therefore, it is necessary to make them stationary by using proper statistical work. When a non linear relationship exists it can be transformed into a linear one prior to modelling, by either a natural log transformation of the dependent variable or, if the series is non-stationary with respect to non-constant variance, then one can use Box-Cox variance stability transformation test which will stabilize the variance of the data. A very popular type of data transformation to deal with non constant variance is the powerful family of transformation, reported in *(Montgomery et al(2015))*

$$y^{\lambda} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda}}, & \lambda \neq 0 \\ \dot{y} \ln \lambda, & \lambda = 0. \end{cases}$$

Where $\dot{y} = \exp(\frac{1}{T}) \sum_{t=1}^{T} \ln y_t$ is the geometric mean of the observations.

The transformation for the various λ values is given in the following Table 3.1 (*Montgomery et al(2015)).*

Table 3.1: Different transformations for lambda values.

λ	-1	-0.5	0	0.5	1
Transformation	$\frac{1}{x}$	$\frac{1}{\sqrt{x}}$	$\ln x$	$\sqrt{x}$	x

The most applications, the first differencing $(d = 1)$ and occasionally the second differencing $(d = 2)$ would be enough to achieve stationary. However, sometimes transformations other than differencing are needed in reducing a non stationary time series to a stationary one.

The log transformation is used frequently in situations where the variability in the original time series increases with the average level of the series.

Difference method is another technique that makes the data series stationary. It helps to stabilize the mean of a time series by removing changes in the level of a time series. If the plot of a time series values $X_1, X_2, X_3, \ldots, X_n$ indicates that these values are non stationary; by differencing the data that are applies the difference operator ∇ the original time series will produce a new time series that may be stationary. The first difference of the time series values $X_1, X_2, X_3, \ldots, X_n$ is defined as follows;

$$Y_t = X_t - X_{t-1} = \nabla X_t; \text{ where } t = 2, 3, 4, \ldots, n.$$

Where ∇ is the (backward) difference operator. In general, the backward difference operator is defined as ∇^d , where d is referred to as lag.

3.4.2 Unit Root Test

One way to determine whether more differencing is required is the Unit Root test. These are statistical hypotheses tests of stationarity that are designed for determining whether differentiating is required. A number of unit root tests are available. One of the most popular tests is the Augmented Dickey Fuller (ADF) and the results can be justified by Phillips-Perron(PP) test as well.

3.4.2.1 Augmented Dickey Fuller Test (ADF) and Phillips-Perron Test (PP)

These statistical tests are employed to test whether the time series data set is stationary or non-stationary. Both tests define the null hypothesis as one where the unit root is present in the time series sample. Here the hypothesis H_0: Data are not stationary (unit root exists) Vs H_1: Data are stationary (unit root does not exist). If ADF statistics exceed critical value, then H_0 can be rejected. Hence, H_1 is accepted which means the data are stationary. A similar argument applies to Phillips-Perron test.

3.5 Kruskal-Wallis Test for Seasonality

Even though a plot of data over the time period can be used to check the seasonality variation that should be proved statistically. Kruskal-Wallis one way analysis of variance test determines whether there exists a seasonality or not. H_0: Seasonality does not exist in the series Vs. H_1: Seasonality exists in the series is the hypothesis.

The H_0 can be accepted if the $p-$ value of the series is greater than 5% significance level and rejected if the $p-$ value is less than 5% significance level. This can be accomplished by computing the statistics.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{n} \frac{R_i^2}{n_i} - 3(N+1)$$

where

$N =$ the total number of rankings $\qquad H =$ Kruskal − Wallis teststatistic

R_i = the sum of the rankings in the i^{th} season and n_i is the number of rankings in a i^{th} season.

3.6 The Covariance

The covariance between X_t and its value at another time period, say X_{t+k} is called the autocovariance at $lag\ k$, and is defined as follows;

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

The collection of the values of γ_k, k = 0,1, 2,… is called the autocovariance function. Note that the autocovariance at $lag\ k = 0$ is the variance of the time series; that is, $\gamma_0 = \sigma_x^2$, which is constant for a stationary time series. The autocovariance measures the linear dependence between two points on the same series observed at different times.

3.6.1 The Autocorrelation Coefficient

The autocorrelation coefficient at $lag\ k$ for a stationary time series is defined as follows;

$$\rho(k) = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sqrt{E[(X_t - \mu)^2 E[(X_{t+k} - \mu)]^2]}} = \frac{COV(X_t, X_{t+k})}{Var(X_t)} = \frac{\gamma_k}{\gamma_0}$$

Where $-1 \leq \rho(k) \leq 1$and $\rho_0 = 1$.

Here X_t is the current observation and X_{t+k} is the observation after k time period. The collection of the values of γ_k where $k = 0, 1, 2, \ldots$ is called the autocorrelation function(ACF). The ACF plot is also useful for identifying non-stationary time series. For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decrease slowly. In other words a strong and slowly dying ACF suggests deviations from stationary. This quantity measures the linear relationship between time series observations separated by a lag of k time units. Characteristics of the ACF are; it is a dimensionless quantity, $\rho_k = \rho_{-k}$ that is the ACF is symmetric around zero.

Sample auto correlation is defined because it is necessary to estimate ACFs from a time series of finite length which is defined as follows;

$$\gamma_k = \hat{\rho} = \frac{\hat{\gamma}_k}{\gamma_0} \quad k = 0, 1, 2, \ldots, k$$

A good general rule of thumb is that at least 50 observations are required to give a reliable estimate of the ACF, and the individual sample autocorrelation should be calculated up to lag k where k is about T/4.

3.6.2 The Partial Autocorrelation Function

The partial autocorrelation function between X_t and X_{t+k} is the autocorrelation between X_t and X_{t-k} after adjusting for $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$(*Montgomery et al(2015)).* The partial autocorrelation is a conditional correlation of X_{t+k} with X_t.

3.7 Autoregressive Model of Order p - AR(p)

An autoregressive(AR) model is a representation of a type of random process which describes certain time varying processes in nature, economics etc. Suppose that current data value X_t is based on its past values that are X_{t-p}, which means X_t can be expressed as a linear combination of $X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p}$ then the model is called auto regressive model of order p. The model equation can be written as follows;

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + e_t$$

Where X_t is a response variable at time t, and $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_p$ are constants or parameters to be determined by giving observations and e_t is an error term at time, $e_t \sim iid(0, \sigma^2)$.

3.8 Moving Average Model of Order q- MA(q)

Suppose that current X_t depends on only past residuals or error values that are $e_{t-1}, e_{t-2}, \dots, e_{t-q}$ then such models are defined as a moving average model. Order of MA model is defined by how many error points are to be considered in the model. The model equation can be written as follows;

$$X_t = \beta_0 + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \cdots + B_q e_{t-q}$$

Where X_t is a response variable at time t, $\beta_1, \beta_2, \dots, \beta_q$ are constants or parameters to be determined by given observations.

3.9 Autoregressive Moving Average Model of Order p, q-ARMA(p, q)

An Autoregressive Moving Average(ARMA) model is a combination of autoregressive terms and moving average terms. Therefore, the autoregressive moving average of order p and q that is ARMA(p, q) the response variable is a linear combination of p past responses and q past errors. The model equation can be written as follows;

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \cdots + B_q e_{t-q}$$

The stationarity of ARMA process relates to the AR and MA components in the model. Thus, the ACF and PACF of ARMA process are determined by the AR and MA components respectively. The theoretical values of the ACF and PACF for stationary time series are summarized in Table 3.2.

Table 3.2: Behavior of theoretical ACF and PACF for stationary process.

Model	**ACF**	**PACF**
MA(q)	Cuts off after lag q	Exponential decay and or damped sinusoid
AR(p)	Exponential decay and or damped sinusoid	Cuts of after lag p
ARMA(p, q)	Exponential decay and or damped sinusoid	Exponential decay and or damped sinusoid

3.10 Autoregressive Integrating Moving Average Model (ARIMA)

The method, called Box-Jenkins(ARIMA) method differences the series to achieve stationarity and then combines the moving averages with autoregressive parameters to yield a comprehensive model an enable to forecast. This non-seasonal model is taken by combining an autoregressive and a moving average model. Following assumptions and limitations have to be considered when apply Box-Jenkins(ARIMA) method .

Let X_t' be an autoregressive integrated moving average ARIMA process of orders p, d and q, that is, ARIMA (p, d, q), where d is the degree of first differentiating involved. The term integrated is used $d = 1$, we can write X_t'. Hence, an ARIMA (p, d, q) can be written as follows (*Montgomery et al(2015)).*

$$X_t' = c + \alpha_1 X_{t-1}' + \alpha_2 X_{t-2}' + \cdots + \alpha_p X_{t-p}' + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t,$$

Where
$X_t' =$ the first difference series, $p =$ order of the autogregressive part, $c =$ constant
$q =$ order of the moving average part, $d =$ degree of the difference involved

To determine the values of p and q, ACF plot and PACF plot can be used. The sationarity of an ARIMA process relates to the AR component in the model.

3.10.1 Assumptions

The Box-Jenkins method requires that the discrete time series data be equally spaced over time and that there be no missing values in the series.

The series also needs to be stationary in the second or weak sense, i.e. series must be stationary in mean, variance and autocovariance.

The mean, variance and autocovariance structure between the same amount of time lags should be constant.

3.10.2 Limitations

There are a few limitations to the Box-Jenkins models. If there are not enough data, they may be no better at forecasting than the decomposition or exponential smoothing techniques. These models are better at formulating incremental rather than structural change. They presume weak stationary, equally spaced intervals of observations and at least 30 to 50 observations.

3.10.3 Identify the Order of Differencing of ARIMA Models

It is important to identify the order of differencing when the data is set transformed to get stationary ARIMA models. Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose autocorrelation function(ACF) plot decays. *(Montgomery et al(2015))* states that for model identification ACF and PACF should have very distinct and indicative properties for MA and AR models. Therefore, (*Montgomery et al(2015))* strongly recommends that both the samples, ACF and PACF be used simultaneously.

3.11 Error Forecasting

Error for the time series can be written as follows;

Time Series = Pattern – Error

3.11.1 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) measures the size of the error in terms. It is calculated as the average of the unsigned percentage error.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_i-\widehat{X_i}}{X_i}\right| \times 100 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.11.1)$$

Where X_i is the actual value and $\widehat{X}_i$ is the forecasted value and n is the number of observations. Generally if MAPE is less than 10%, it is taken as an acceptable model.

3.12 Time Series Model Building

When an ARIMA model is being built there are three iterative steps that need to be pursued *(Montgomery et al(2015))* & *(Yaffee et al (2003))*.

- First a tentative model of the ARIMA class is identified through analysis of historical data.

- Second, the unknown parameters of the model are estimated.

- Third, through residual analysis, diagnostic checks are performed to determine the adequacy of the model or to indicate potential improvements.

- Fourth, the forecasting accuracy is checked after getting the forecasted results from the fitted model.

3.13 Model Selection Criteria

There are several models that can be used for forecasting a particular time series. Consequently, selecting an appropriate forecasting model is of considerable practical importance. Thus, various criteria for model assumption have been introduced in the literature for model selection.

3.13.1 Akaike's Information Criteria (AIC)

To assess the quality of model fitting, Akaike introduced an information criteria for a sample of n observations, AIC is defined as follows (*Montgomery et al(2015));*

$$\begin{aligned} \mathrm{AIC} &= -\,2\ln(\text{maximum likelihood}) + \,2\mathrm{k}\ln(\mathrm{n}) \\ &\approx \mathrm{n}\ln(\hat{\sigma}_a^2) + \mathrm{k}\ln(\mathrm{n}) \end{aligned}$$

Where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2 where k is the number of parameters estimated in the model including constant term. The model with the smallest AIC is given preference page 402 *in (Baguley T.(2012))* said,

"The best model from the set of plausible models being considered is therefore the one with the smallest AIC value(the least information loss relative to the true model)…Negative AIC indicates less information loss than a positive AIC and therefore a better model".

3.13.2 Schwarz Information Criteria (SIC)

The SIC criterion is defined as

$$\text{SIC} = n^{k/n} \sum \frac{\hat{U}}{n} = n^{k/n} \frac{\text{RSS}}{n} \text{ or}$$

$$\text{SIC} = \frac{k}{n} \ln n + \ln(\frac{\text{RSS}}{n})$$

Where $[\left(\frac{k}{n}\right) \ln n]$ is the penalty factor. Like AIC the lower the value of SIC, the better the model.

3.13.3 Coefficient of Determination

The coefficient of determination R^2 describes strength of the dependent variable described by the independent variable. In other words, it is used as a measure to fit and to evaluate the correctness of the fitted model. The best model gives the largest of R^2 value.

3.14 Model Diagnostic Checking

Once the model has been fitted to the data, it is necessary to carry out number of diagnostics checks. If the model fits well, the residuals should essentially behave like white noise, that is, normalized residuals are uncorrelated random shock with zero mean and constant variance. Hence model diagnostic checking is accomplished through careful analysis of residual series. In this study, the following diagnostic checking has been done to find the behaviour of residuals.

3.14.1 Normality of Residuals

To determine whether the errors are normally distributedthe histogram of residuals is checked. If the residuals are normally distributed, they should all lie more or less on a straight upward sloping line. A skewness closer to 0 and the kurtosis closer to3 also suggest that the residuals follow anormal distribution.

H_0: Residuals are normally distributed

H_1: Residuals are not normally distributed

3.14.2 Lagrange's Multiplier Test

When error terms from different time periods are correlated, it is said that the error term is serially correlated. There are several statistical tests to check whether the serial correlation exists or not. The Lagrange Multiplier (LM) test has provided a standard means of testing parametric restrictions for a variety of models. It is one of the general tests that can be used to test the serial correlation. All the hypotheses are generally checked using 5% of significance level.

H_0: There is no serial correlation of any order up to p

H_1: There is a serial correlation of some order up to p

3.14.3 White's General Test

This test is used to check constant variance of residuals. It appears that there is more difference among Y values when X is larger. If non constant variance of residuals exists, residuals showed heteroscedasticity. If homoscedastic, the variance of the residuals is the same for all observations in the sample. The heteroscedasticity can be checked by setting hypothesis for the residuals for 5% of significance level.

H_0: There is homoscedasity among the residuals of variances

H_1: There is heteroscedasticity among the residuals of variances

3.14.4 Anderson -Darling Test

The Anderson-Darling test is used to test if a sample of data comes from a population with a specific distribution. It is a modification of the Kolmogorov Simirnov test. This test is considered as an effective method of determining whether the distribution of data in a sample departs from a normal distribution. The Anderson-Darling test statistic is defined as

$$A^2 = -N - S$$

Where $S = \sum_{i=1}^{N} \frac{2i-1}{N} [\ln F(X_i) + \ln(1 - F(X_N + 1 - i))]$

F is the cumulative distribution of the specified distribution and X_i are the ordered data. If the probability value of the Anderson-Darling statistic is less than 5% said that the distribution is not normal and if otherwise that it is normally distributed.

H_0: The residuals follow a normal distribution

H_1: The residuals do not follow a normal distribution

3.15 Forecasting Accuracy

After forecasting by using the fitted model the accuracy of forecasting has to be checked. There are several methods to measure accuracy and compare one forecasting method with another. The selected model has minimum Mean Absolute Percentage Error which is regarded as showing highest accuracy. This is calculated as described in (3.11.1). The MAPE is employed to check the forecasting accuracy of the model.

3.16 Regression Analysis

Regression analysis is a statistical technique to measure and identify a relationship between an independent variable and an explanatory variable. This method is useful to obtain an idea about future behaviour of the data and the nature of the relationship between variables.

3.16.1 Correlation

Correlation is a statistical technique which shows whether two or more variables or processes are related and if so how strongly (*Simakova(2011)).*

The following hypothesis has considered for correlation matrix for 5% of significance level.

H_0: There is no correlation between dependent variable and independent variable
H_1: There is a correlation between dependent variable and independent variable

3.16.2 Pearson Correlation Coefficient

It measures the strength of the linear relationship between an independent variable and a dependent variable is defined as follows;

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad -1 \leq r \leq 1$$

Table 3.3: Interpretation of r values

Value of r	Interpretation
$r = +1$	Perfect positive correlation exists between the variables
$r = -1$	Perfect negative correlation exists between the variables
$r = 0$	No relation exists between the variables
$+0.78 \leq r < +1$	The high positive correlation exists between the variables
$-0.75 \geq r > -1$	The high negative correlation exists between the variables
$+0.50 \leq r < 0.75$	The moderate positive correlation exists between the variables
$-0.50 \geq r > -0.75$	The moderate negative correlation exists between the variables
$r < +0.50$	The low positive correlation exists between the variables
$r > -0.50$	The low negative correlation exists between the variables

3.16.3 Linear Regression Model

A first-order regression model is hypothesized to be as follows;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \text{……………..(3.16.3)}$$

where Y denotes the response variable, X_1 and X_2 are independent variables and the ϵ denotes the error terms in the model. Here β_0 and β_1 are called partial regression coefficients because they convey information about the effect on Y of the predictor that they multiply.

3.17 Spurious Regression

In time series analysis, we often encounter situations where we wish to model one non stationary time series Y_t as a linear combination of other non stationary time series (X_{1t}, $X_{2t}, \dots, X_{kt}$). The regression equation can be written as follows;

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \epsilon_t$$

Where $\beta_i,\ i = 1,2,3,\dots.,k$ and$\epsilon_t \sim iid(0, \sigma^2)$.

In general when analyzed data series are non stationary, the regression equation by which they can be modeled is inadequate or **spurious**(nonsense) as it shows illogical correlation between series. This type of relationship is due to the presence of trends in the data series, the processes not necessarily having the same causal phenomena. If the series are not stationary, the regression seems to be statistically significant, even though the only thing, period correlation and not causal relations between the data is presented. The statistical tests validate the regression coefficients, whenever the series contains trends. When a time series contains unit roots, statistical tests estimate the dependency between the variables, and the estimators are doubtful. The typical statistical symptoms of existence of spurious regression is

$$\text{High}R^2, \quad t-\text{values}, \quad F-\text{values but } \textbf{low } \boldsymbol{DW}$$

The problem of spurious regression can be eliminated by differencing the data, but this implies the loss of long run information content in the data. Long run information content can be retained through co-integration.

3.17.1 Durbin-Watson Statistic

The popular test for serial correlation is the Durbin-Watson statistic defined as follows;

$$d = DW = \frac{\sum_{t=2}^{T}(\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^{T}\hat{\epsilon}_t^2}$$

Where T is the number of time periods and ϵ_t is the error term in the model at time period t. If successive values of $\hat{\epsilon}_t$ are close to each other, the DW statistic will be low, indicating the presence of a positive correlation. The d lies between lower and upper bounds, say, d_L and d_U, such that if d is outside these limits, the decision procedure is as follows for hypothesis:

H_0: no autocorrelation exists
H_1: autocorrelation exists

If $0 < d_L$ reject H_0. Evidence of positive auto correlation
$d_L > d_U$ Indecision
$d_U \leq 2 \leq 4 - d_U$ Do not reject H_0 or H_1 or both
$4 - d_U < 4 - d_L$ Zone of indecision
$4 - d_L < 4$ reject H_0. Evidence of negative auto correlation

3.18 Granger Causality

Granger Causality (1969) test is a statistical hypothesis test for determining whether one time series is useful in forecasting another. In other words the current value of one variable is caused by the prior value of other variables (*Simakova(2011))*. The evidence of Granger causality is a good indicator that a VAR model rather than a univariate model, is needed. Even though in the ordinary regression, the correlation and regression illustrate the relation between two variables at the same time they do not imply causation. Granger Causality is used to provide such type of information about casual relations (*Simakova(2011))*.

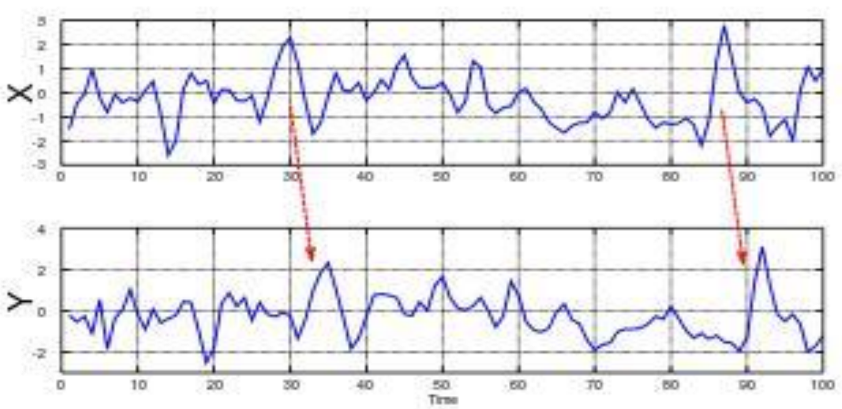


Figure 3.1: When time series X Granger-causes time series Y , the pattern in X are approximately repeated in Y after some time lag. Thus, past values of X can be used for the prediction of future values of Y (www.wikipedia.org).

Consider the following equations;

$$Y_t = \alpha_1 + \sum_{j=1}^{k_1} b_{ij}\, X_{i-j} + \sum_{i=1}^{m_1} C_{1i} Y_{t-i} + u_t$$

And

$$X_t = \alpha_2 + \sum_{j=1}^{k_2} b_{2j}\, X_{i-j} + \sum_{i=1}^{m_2} C_{2i} Y_{t-i} + v_t$$

where $k_1,\ k_2,\ m_1,\ m_2$ are all positive integers and u_1, v_t are the random disturbances of model 1 and 2 respectively.

For model 1

Null hypothesis and alternative hypothesis for model at 5% of significance level

H_0: X does not Granger cause Y at level of significance

H_1: X does Granger cause Y at level of significance.

For model 2

Can be used to test the Granger causality of Y to X. If a positive integer q exists, and for all k_1,m_1, there are $k_1 < q$ and $m_1 < q$, X does not Granger cause Y at some level of significance.

3.18.1 Best Lag Selection for Testing Linear Granger Causality

The best lag for testing linear Granger causality is equal to

$$\max\{n \mid min_{n \leq q}\{|AIC_n|\}\} \quad \text{or} \quad \max\{n \mid min_{n \leq q}\{|SIC_n|\}\}$$

When AIC_n and SIC_n respectively denote the AIC and SIC of the corresponding regression model with lag n.

This section is devoted to cointegration test and construction of final Vector Error Correction (VEC) model that is used to confirm the evidence of the long-term and short- term relationship between price of gold and values of all other two variables. Engle and Granger (1987) pointed out that a linear combination of two or more non-stationary series may be stationary. If such stationary linear combination exists, the non-stationary time series are said to be co-integrated.

The theory of co-integration analysis states that there may be co-integration relation-ship between the variables involved if they are integrated of the same order (*Nair et. al(2015)).*

3.19 Integrated Order p or $I(p)$

A stochastic process is said to be integrated of order p, abbreviated as $I(p)$, if it needs to be difference p times in order to achieve stationarity.

3.20 Co-integrated Order (d, p) or $CI(d, p)$

Let X_{1t} and X_{2t} be two non stationary time series. Both are said to be co-integrated of order $CI(d, p)$ if X_{1t} and X_{2t} are both integrated of order d.

3.21 Background for Co-integration

The regression of a nonstationary time series on another nonstationary time series may produce a spurious regression. Let X_{1t} and X_{2t} be two non stationary time series. Subjecting these time series individually to unit root analysis, both are $I(1)$; that is they contain a unit root. Suppose we regress X_{1t} on X_{2t} as follows;

$$X_{1t} = \beta_1 + \beta_2 X_{2t} + \epsilon_t$$

Where β_1 and β_2 are parameters and ϵ_ta is residual term.

This can be written as follows;

$$\epsilon_t = X_{1t} - \beta_1 - \beta_2 X_{2t} \dots \dots \dots \dots \dots .. (1)$$

If ϵ_t is subjected to unit root analysis it is seen that it is stationary; that is $I(0)$ eventhough individual series are not stationary. Their linear combination cancels out the stochastic trends in the two series and make it $I(0)$. In a case like this , it is said that the two variables are co-integrated. Economically two variables will be co-integrated if they have a long- term or an equilibrium relationship between them. In the short run there may be disequilibrium. Therefore the error term in (1) is named the "equilibrium error". This error term is to tie the short run behavior of X_{1t} to its long run value. Economic theory often implies equilibrium relationships between the levels of time series variables that are best described as being $I(1)$.

3.22 Definition for Co-integration

If two or more time series are non stationary, but a linear combination of these variables are $I((1)$ stationary, then we say that both series are co-integrated. Suppose X_{1t} and X_{2t} have random walk.

Define $\beta = (1 - \beta)'$ and consider the linear combination

$$Z_t = \beta' X_t = (1 - \beta_2)\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = X_{1t} - \beta_2 X_{2t}$$

where Z_t is stationary. Then we say that X_{2t} co-integrate with co-integration vector β. Where $I(1)$ means the difference of the series and β is co-integrating vector. It is obvious that this co-integration matrix is not unique.

3.22.1 Importance of Co-integration

(1) Co-integration implies existence of long run equilibrium.
(2) Co-integration implies common stochastic trend.
(3) With co-integration we can separate short and long run relationships among variables.
(4) Co-integration can be used to improve long run forecast accuracy.

3.22.2 Testing for Co-integration

The basic idea had been given by Engle and Granger 1987.

- If ($y_{1,t}, y_{2,t}, \ldots, y_{m,t}$) are co-integrated, the true equilibrium error process ϵ_t must be $I(0)$.
- If they are not co-integrated, then ϵ_t must be series $I(1)$.
- Test the null hypothesis of no co-integration against the alternative of co-integration by performing a unit root test on the equilibrium error process ϵ_t.

A number of methods for testing co-integration have been developed.

(1) A DF(Dickey Fuller) or ADF(Augmented Dickey Fuller) unit root test on the residuals estimated from the co-integrating regression.
(2) The co-integrating regression Durbin Watson (CRDW) test.
(3) Johansen Maximum Likelihood co-integration test.

3.22.3 Pre Condition to Johansen Test

Before performing Johansen's co- integration test as per precondition, all variables must be non-stationary at the level and stationarity at their first difference; that is, integrated of same order (*Simakova(2011)).* This means that the series needs to be stationary at the same lag value.

3.22.4 Johansen Test

Co-integration assumes the presence of common non-stationary (i.e. $I(1)$) processes underlying the input time series variables.

$$\begin{aligned}
Y_{1,t} &= \alpha_1 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \cdots + \beta_p X_{p,t} + \epsilon_{1,t} \\
Y_{2,t} &= \alpha_2 + \gamma_1 X_{1,t} + \gamma_2 X_{2,t} + \gamma_3 X_{3,t} + \cdots + \beta_p X_{p,t} + \epsilon_{2,t} \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
Y_{m,t} &= \alpha_m + \phi_1 X_{1,t} + \phi_2 X_{2,t} + \phi_3 X_{3,t} + \cdots + \phi_p X_{p,t} + \epsilon_{m,t}
\end{aligned}$$

The number of independent linear combinations (k) is related to the assumed number of common non- stationary underlying process(p) as follows;

$$p = m - k$$

Consider three outcomes:

(1) $k = 0,\ p = m$ in this case, time series variables are not co-integrated.
(2) $0 < k < m, 1 < p < m$. In this case, the time series variables are co-integrated.
(3) $k = m,\ p = 0$. All time series variables are stationary $I(0)$ to start with and co-integration is not relevant here.

Johansen test has two forms; the trace test and the maximum eigenvalue test. Both test address the co-integration presence in the hypothesis.

3.22.5 Johansen Maximum Likelihood Co-integration Test

This method follows the same principles as the Engle-Granger approach to co-integration. The order of integration of the variables are first assessed which is called pre- condition to Johansen test. The Johansen Maxium Likelihood(ML) procedure can then be used to determine whether a stable long run relationship exists between the variables. The results of that can be used to determine the number of co-integrating vectors present. There will be possible $m-1$ vectors, where m is the number of variables included in the model. Based on this result, the long-run coefficients can then be determined and the resultant error correction model can be produced.

The technique produces two statistics the likelihood ratio test based on maximal eigenvalue of the stochastic matrix and the trace of the stochastic matrix. These statistics are then used to determine the number of co-integrating vectors.

3.22.5.1 Trace Test

The trace statistic investigates the null hypothesis of r co-integrating relations against the alternative of r co-integrating relations, where n is the number of variables in the system for $r = 0,1,2,\ldots,n-1$. It can be written as follows;

$$LR_{tr}(r / n) = -T \sum_{i=r+1}^{n} \log(1 - \hat{\lambda}_i)$$

Where $\hat{\lambda}_i$ is the estimated value for the i^{th} ordered eigenvalue and T is the sample size. Hypotheses for the trace statistics can be written as follows;

H_0: There is no co-integration among the variables, r=0

H_1: There is a co-integration among the variables, r>0

The null hypothesis can be rejected if p- value of the trace statistic is less than 5% of significance level.

3.22.5.2 Maximum Eigenvalue Test

The maximum eigenvalue statistics test the null hypothesies of r co-integrating relations against the alternative of $r+1$ co-integrating relations for $r = 0, 1, 2, \dots, n-1$. The maximum eigenvalue test statistic can be written as follows;

$$LR_{max} = (r/n+1) = -T \times \log(1-\hat{\lambda})$$

Where λ is the maximum eigenvalue and T is the sample size. Where $\hat{\lambda}_i$ is the estimated value for the i^{th} ordered eigenvalue from the π matrix.

$\mathrm{H_0}$: There is no co-integration among the variables

$\mathrm{H_1}$: There is a co-integration among the variables

If p- value of the maximum eigenvalue is less than 5% of significance level we can reject null hypothesis. In some cases, trace and maximum eigenvalue statistics may yield different results and in this case the results of the trace test should be preferred.

In this section we are working with the statistical concept of co-integration that is required to make sense of Vector Autoregression(VAR) models with $I(1)$ data. The connection between VAR models and co-integration is made and Johansen's maximum likelihood methodology for co-integration is facilitated in this section.

3.23 VAR Models and Co-integration

The Granger representation theorem states that two non stationary time series X_{1t} and X_{2t} are co-integrated if and only if there exists an error correction model for either X_{1t}, X_{2t} or for both.

3.24 Error Correction Models

These are theoretical driven approach useful for estimating both short term and long term effects of one time series on another. The error correction term is the residual from the co-integrating relationship. The error correction term relates to the fact that last periods deviation from a long run equilibrium and the error influences its short run dynamics. Thus, ECMS directly estimate the speed at which a dependent variable returns to equilibrium after changing in another variable. The coefficient of the AR(1) is negatively signed, indicating a move back towards equilibrium and the model is stable. A positive sign indicates movement away from equilibrium. The coefficient should be lying between 0 and 1, 0 suggesting no adjustment one time period later and 1 indicating full adjustment.

3.25 Vector Error Correction Model

Consider the levels VAR(p) model for the $(n \times 1)$ vector X_t

$$X_t = \emptyset D_t + \pi_1 X_{t-1} + \cdots + \pi_p X_{t-p} + \epsilon_t \quad t = 1,2,3,\dots,T$$

where D_t contains deterministic terms (constant, trend, seasonal etc) the VAR(p) model is not stable and X_t is co-integrated. Then VAR representation is not the most suitable representation for analysis because the co-integration relations are not explicitly apparent. It will be apparent if the level VAR is transformed to the vector error correction model.

Vector error correction model (VECM) can be written as follows;

$$\Delta X_t = \emptyset D_t + \pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{\mathrm{p-1}} \Delta X_{t-p+1} + \epsilon_t \;\dots\dots\dots\dots (2)$$

where $\pi = \pi_1 + \cdots + \pi_p - \mathrm{I}_n$ and $\Gamma_k = - \sum_{j=k+1}^{p} \pi_j \quad k = 1,2,\dots,p-1$

The matrix π is called the long run impact (coefficient) matrix Γ_k are the short run impact matrices. As π is a singular matrix that has a reduced rank; $rank(\pi) = r < n$.

Case I: $rank(\pi) = 0$ implies that $\pi = 0$ and X_t is $I(1)$ and not co-integrated.

Case II: $0 < rank\ (\pi) = r < n$ implies that X_t is $I(1)$ with r linearly independent cointegrating vectors and $n - r$ common stochastic trends. Since π has rank r which can be written as the product

$$\pi_{(n\times n)} = \alpha_{(n\times r)}\beta'_{(r\times n)}$$

Where α and β' are $(n \times r)$ matrices with $rank\ (\alpha) = rank\ (\beta') = r,\ \beta'$ gives the co-integrating vectors α gives the amount of each co-integrating vector entering each equation of VECM, also known as the "adjustment parameter" or "speed of adjustment towards equilibrium" and β' is a matrix of co-integrating vectors.

3.26 Procedures for Testing Co-integration and VECM.

- Determine order of VAR(p). Suggest choosing the minimal p such that the residuals behave like vector white noise.

- Employ the unit root test to check that all the variables are non stationary at level and stationary at same level which is called pre-condition of co- integration test.

- Use trace or λ_{max} tests to determine number of unit roots and to check whether the co- integration equations exist.

- If there is evidence of co-integration, use the residual to form the error correction term in the corresponding ECM.

- Wald test to examine existence of the short run causalities.

- Perform diagnostic checking of residuals.

3.27 Wald Test

The Wald test is a method to find whether the explanatory variables in a model are significant. Significant means that they add something to the model. Variables that add nothing can be deleted without affecting the model in any meaningful way.

The Wald test statistic is

$$W_T = \frac{[\hat{\theta} - \theta_0]^2}{1/I_n(\hat{\theta})} = I_n(\hat{\theta})[\hat{\theta} - \theta_0]^2$$

Where $\hat{\theta}$= Maximum Likelihood Estimator(MLE)

$I_n(\hat{\theta})$ = Expected Fisher information (Evaluated at MLE)

3.27.1 Hypothesis Check for Join Significance of Parameters.

H_0: Parameters equal to zero.

The null hypothesis cannot be rejected if the Chi-squared probability is greater than 5%.

H_1: Parameters not equal to zero.

The null hypothesis can be rejected if the Chi-squared probability is less than 5% and the alternative hypothesis can be accepted.

If the Wald test shows the parameters for certain explanatory variables are zero, those variables can be removed from the model. It implies that those variables are not jointly influenced. If the test shows that the parameters are not zero the relevant variables are kept in the model, indicating that those variables are jointly influenced.

CHAPTER 4

TIME SERIES MODELS TO FORECAST GOLD PRICE

This chapter includes the fitted time series models to forecast gold price. Preliminary analysis has been carried out to check the validity of the data series for the time series analysis and to understand the behavior of the data series. Five models were discussed under Auto Regressive Integrated Moving Average model building techniques.

4.1 Preliminary Analysis

After collecting data it is necessary to test for its suitability to carry on stationary time series analysis.

Figure 4.1 shows the time series plot for monthly gold price in Sri Lanka which consists of 108 observations.

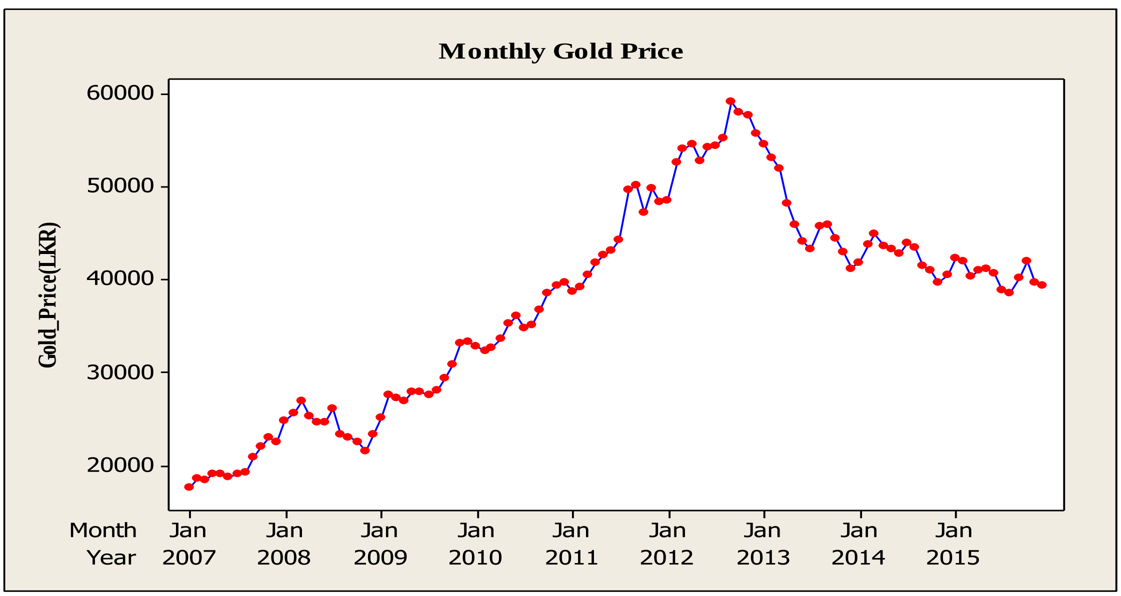
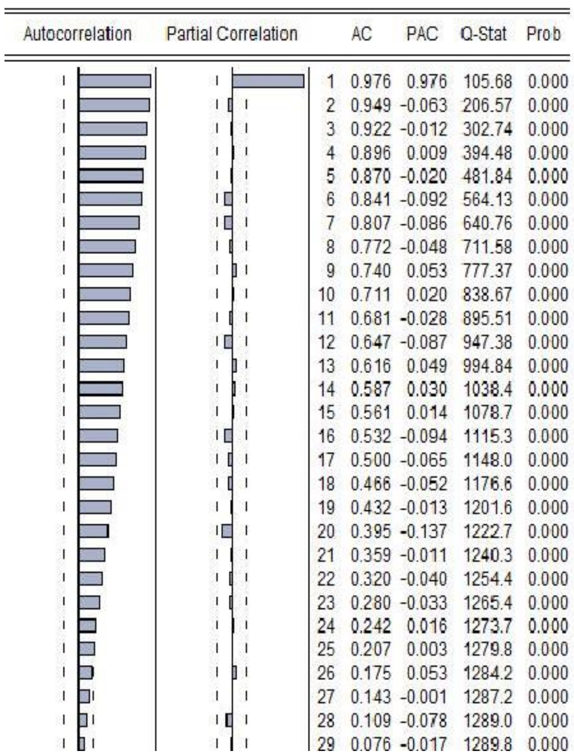


Figure 4.1: The time series plot of monthly gold price in Sri Lanka from Jan. 2007 to Dec. 2015.

Figure 4.1 illustrates change of monthly gold price from 2007, Janurary to 2015, December. First, data are positively autocorrelated. Second, the monthly gold price exhibits a rapid increase, which maximized around year 2012. It has shown a gradual decrease from 2012 to 2015. There are no large swings can be seen yet to exhibit small fluctuations. However, gold price showed an upward trend and then a gradual decrease which implies that data are not stationary.

The non stationarity of data series can be justified by gridding autocorrelation (ACF) and partial autocorrelaiton (PACF) graphs as well.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.976	0.976	105.68	0.000
		2	0.949	-0.063	206.57	0.000
		3	0.922	-0.012	302.74	0.000
		4	0.896	0.009	394.48	0.000
		5	0.870	-0.020	481.84	0.000
		6	0.841	-0.092	564.13	0.000
		7	0.807	-0.086	640.76	0.000
		8	0.772	-0.048	711.58	0.000
		9	0.740	0.053	777.37	0.000
		10	0.711	0.020	838.67	0.000
		11	0.681	-0.028	895.51	0.000
		12	0.647	-0.087	947.38	0.000
		13	0.616	0.049	994.84	0.000
		14	0.587	0.030	1038.4	0.000
		15	0.561	0.014	1078.7	0.000
		16	0.532	-0.094	1115.3	0.000
		17	0.500	-0.065	1148.0	0.000
		18	0.466	-0.052	1176.6	0.000
		19	0.432	-0.013	1201.6	0.000
		20	0.395	-0.137	1222.7	0.000
		21	0.359	-0.011	1240.3	0.000
		22	0.320	-0.040	1254.4	0.000
		23	0.280	-0.033	1265.4	0.000
		24	0.242	0.016	1273.7	0.000
		25	0.207	0.003	1279.8	0.000
		26	0.175	0.053	1284.2	0.000
		27	0.143	-0.001	1287.2	0.000
		28	0.109	-0.078	1289.0	0.000
		29	0.076	-0.017	1289.8	0.000

Figure 4.2: The correlogram for monthly gold price

Sample ACF and PACF are given in Figure 4.2. The sample ACF dies out very slowly, while the sample PACF is only significant at the first lag. This evidence suggests that the process is not stationary. Moreover, that there is no special pattern in the sample ACF graph means that we can conclude the absence of seasonality in the gold price data.

Statistical evidence to prove the absence of seasonality is provided by the Kruskal Wallis test. The p-value of Kruskal Wallis test which is higher than 5% of significance level strongly rejects the seasonality presented in the time series. Thus, it is concluded that seasonality is not present in this data series.

Table 4.1: Kruskal-Wallis test results for seasonality in monthly gold price.

Original Series	
Kruskall-Wallis Statistic Probability Value	$\mathrm{H} = 0.28$ 1.0000

Table 4.2 shows the test results of Unit Root test, Augmented Dickey-Fuller (ADF) and Phillips-Perron(PP) tests with intercept to prove data are not stationary.

Table 4.2: P- values for the Unit Root test statistics.

Original Series	ADF t- statistic	PP Adj. t-statistic
 Probability Value	-1.866564 0.3469	-1.866564 0.3469

Table 4.2, p- values are given for test critical value of 5% significance level. Depending on results of $p-$ values for ADF and PP tests the null hypothesis H_0 can be accepted and reject the alternative hypothesis H_1 which states that the data series has a unit root implied that the time series data are not stationary. Thus, to make them stationary the behavior of first difference of data series was checked.

4.2 First Difference Series of Monthly Gold Price

Depending on the preliminary analysis, it is clear that the data series is non-stationary. The first difference series of gold price data graph was examined and implies that the trend has been removed although high variability of data between 2011, Janurary and 2103, January can be seen.

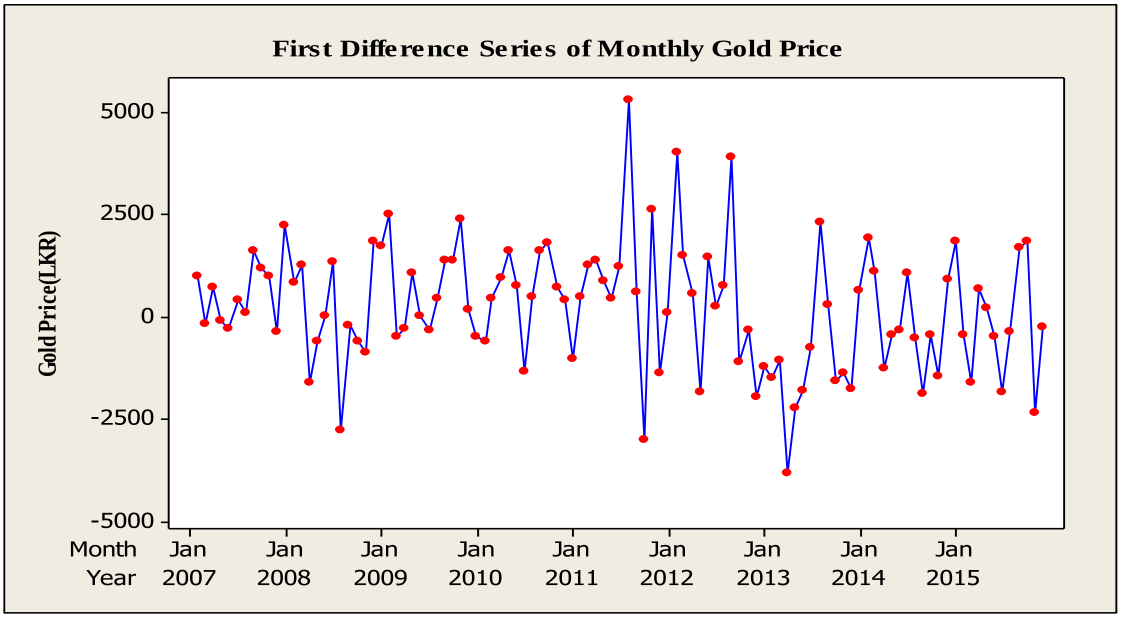


Figure 4.3: The graph of first difference of gold price from 2007, January to 2015, December.

To confirm the stationarity of the first difference series Unit Root test, ADF test and PP tests have been carried out and the results for 5% of significance level are shown in the Table 4.3.

Table 4.3: ADF and PP test values for the first difference series.

$D[GOLD\ PRICE]$	ADF t- statistic	PP Adj. t-statistic
	-8.36265	-8.288831
Pr**o**bability Value	0.0000	0.0000

Depending on the two tests results of ADF and PP tests it is concluded that the data series is stationary at the 1st difference as p −value is 0.0000 which is less than the 5% significance level. The null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted which means that the time series is stationary.

	AC	PAC	Q-Stat	Prob
1	0.197	0.197	4.2491	0.039
2	-0.051	-0.094	4.5416	0.103
3	-0.007	0.023	4.5477	0.208
4	-0.003	-0.011	4.5486	0.337
5	0.162	0.174	7.5663	0.182
6	0.205	0.144	12.437	0.053
7	0.076	0.034	13.112	0.069
8	-0.079	-0.085	13.844	0.086
9	-0.037	0.001	14.011	0.122
10	0.080	0.062	14.772	0.141
11	0.163	0.099	18.012	0.081
12	0.051	-0.038	18.337	0.106
13	-0.092	-0.089	19.393	0.111
14	-0.064	-0.006	19.906	0.133
15	0.011	0.021	19.920	0.175
16	0.086	0.029	20.875	0.183
17	0.100	0.027	22.164	0.178
18	0.061	0.063	22.654	0.204
19	-0.049	-0.009	22.967	0.239
20	-0.073	-0.040	23.682	0.257
21	0.013	-0.015	23.705	0.308
22	0.005	-0.059	23.709	0.363
23	0.011	-0.005	23.728	0.419
24	0.028	0.043	23.841	0.471
25	-0.035	-0.017	24.012	0.519
26	-0.075	-0.062	24.812	0.530
27	0.062	0.077	25.368	0.554
28	0.016	-0.030	25.406	0.606
29	-0.014	-0.006	25.434	0.656

Figure 4.4: The correlogram of first difference series of gold price

Figure 4.4 the ACF and PACF graph of $\boldsymbol{D[GOLD\ PRICE]}$(first difference series) in correlogram do not show any significant pattern. Differences in the original data cancel out the autocorrelation which means that no trend and no evidences can be seen for the existence of seasonality. Kruskall- Walli test which gives $p-$value 0.807 is greater than 5% of significance which implies the rejection of the null hypothesis. Therefore, we can conclude with 95% confidence that seasonality does not exist in $\boldsymbol{D[GOLD\ PRICE]}$. Therefore, the model identification and coefficient estimation are carried out.

4.3 Model Identification and Coefficient Estimation $D[GOLD\ PRICE]$

E-views 7 software has been used to find the significant AR and MA terms for the model. Further the significance of the coefficients of the model is examined and the relevant coefficient and probability values have been taken for AR (p) and MA (q) terms given in Table 4.4 and Table 4.5 shows the AIC, R^2 and standared error of regression values for possible combinations of AR (p) and MA (q) terms in $\boldsymbol{D}[\boldsymbol{GOLD\ PRICE}]$.

Table 4.4: Coefficient and probability values of AR(p) and MA(q) terms.

Combinations	Coefficients	Probability values for AR(p) and MA(q)
AR(1)	0.2108	0.0289
MA(1)	0.2441	0.0111
AR(3) MA(3)	0.9271 -0.9370	0.0000 0.0000

Table 4.5: AIC, R^2 and standard error of regression values for different AR(p) and MA(q) terms in $\boldsymbol{D}[\boldsymbol{GOLD\ PRICE}]$.

Combinations	AIC values	R^2	**S.E. of Regression**
AR(1)	17.4492	0.0281	**1481.748**
MA(1)	17.4369	0.0332	**1472.731**
AR(3) MA(3)	17.4935	0.0210	**1507.663**

It can be concluded with 95% confidence that the coefficients are significance in all terms, Table 4.4. Yet, the standard error of regression which is highlighted in Table 4.5 reports a very high value; this indicates that the fitted models are not accurate enough.

Therefore, it is decided to check the Box-Cox transformation and consider another model.

4.4 Box- Cox Transformation

The Box Cox transformation has been carried out to identify which transformation should be taken into account.

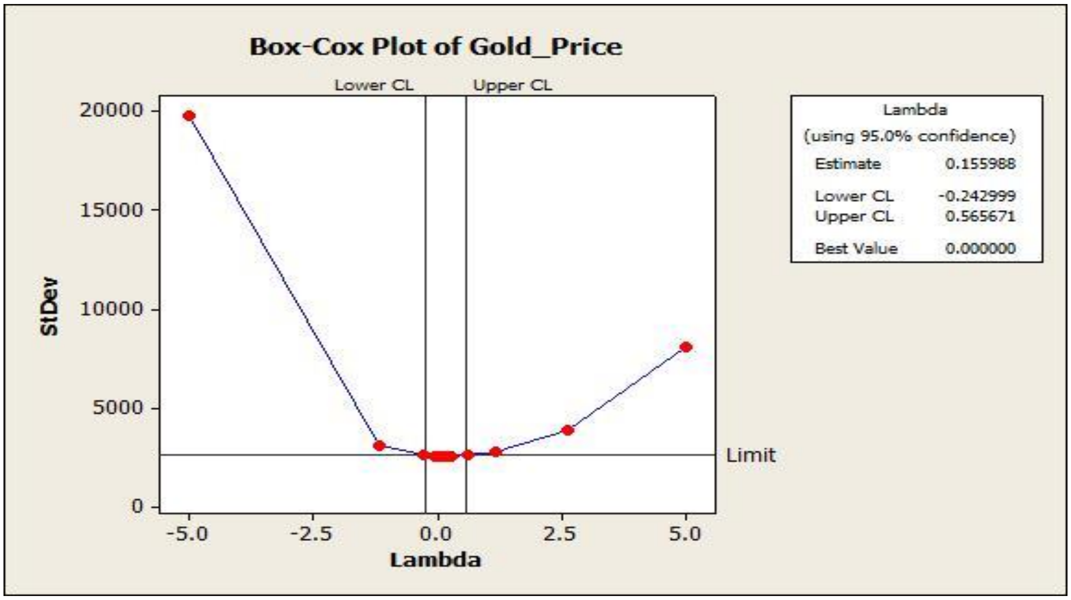


Figure 4.5: Box -Cox transformation plot

Figure 4.5 indicates the best value for λ is 0. $\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]$ time series plot is failed to satisfy the property of sationarity (refer APPENDIX A) thus, the first diference of Ln transformation of the gold price has been taken.

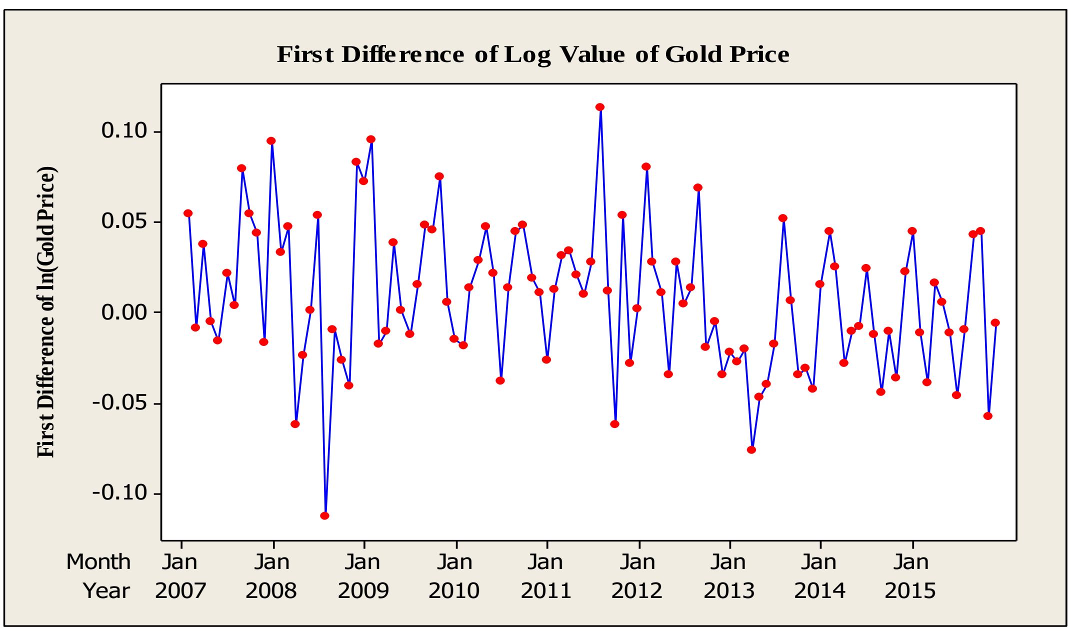


Figure 4.6: The time series plot of the $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$

Since the trend is not visualized in Figure 4.6 it has to be proved statistically with the support of both the ADF test and the PP test which is used to confirm the result given in ADF test.

To confirm the stationarity of the $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$ (first difference series of Ln gold price) we carried on the Unit Root test, Augmented Dickey Fuller and Pillips-Perron tests. The results of 5% of significance level are shown in the Table 4.6.

Table 4.6: ADF and PP test values for $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$.

$D[Ln(GOLDPRICE)]$	ADF t- statistic	PP Adj. t-statistic
	-8.47685	-8.43253
Probability Value	0.0000	0.0000

The probability values 0.0000 given by both test statistics are less than 5% of significance level and indicate that the first difference of the Ln gold price series is stationary. This supports the fact that no trend appears in the time series. This is justified from the correlogram of the $\boldsymbol{D}[\boldsymbol{Ln}(\boldsymbol{GOLD\ PRICE}]]$ in Figure 4.7.

The ACF and PACF graph of the first difference of Ln values of gold price in Figure 4.7 do not show any non significance values. It suggests that the difference in the Ln values in the original data cancel out the autocorrelation. No trend or seasonal pattern exists in the ACF graph. Therefore, it was decided to carry out model identification and coefficient estimation for $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$.

The following Figure 4.7 shows the correlogram for $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$.

Sample: 1 108
Included observations: 107

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.188	0.188	3.9073	0.048
		2	-0.019	-0.057	3.9484	0.139
		3	-0.064	-0.051	4.4031	0.221
		4	0.022	0.045	4.4568	0.348
		5	0.126	0.115	6.2854	0.279
		6	0.144	0.102	8.6939	0.192
		7	0.006	-0.031	8.6988	0.275
		8	-0.055	-0.034	9.0551	0.338
		9	-0.007	0.019	9.0604	0.432
		10	0.068	0.047	9.6095	0.475
		11	0.202	0.161	14.564	0.203
		12	0.014	-0.064	14.587	0.265
		13	-0.101	-0.076	15.844	0.258
		14	-0.048	0.009	16.135	0.305
		15	0.099	0.093	17.367	0.297
		16	0.128	0.046	19.464	0.245
		17	0.115	0.056	21.185	0.218
		18	-0.019	-0.015	21.233	0.268
		19	-0.052	-0.005	21.591	0.305
		20	0.019	0.021	21.640	0.360
		21	0.055	-0.009	22.046	0.397
		22	0.084	0.017	23.020	0.401
		23	0.020	0.012	23.075	0.456
		24	0.030	0.081	23.205	0.508
		25	-0.033	-0.047	23.358	0.557
		26	-0.009	-0.057	23.369	0.612
		27	0.089	0.068	24.534	0.601
		28	0.004	-0.046	24.537	0.653
		29	-0.018	0.010	24.586	0.700

Figure 4.7: The correlogram of $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$

4.5 Model Identification and Coefficient Estimation $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$

Estimate the significance of AR and MA terms for model development *E-views software 7* is employed and the significant terms are given in Table 4.7. AIC, and R^2 values for each model is mentioned in Table 4.8.

Table 4.7 shows the probability values for significant AR(p) and MA(q) terms that produced for each model in $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$**.**

Table 4.7: Possible combinations and probability values for significant AR(p) and MA(q) terms in $\boldsymbol{D[Ln[GOLD\ PRICE]]}$.

Possible combinations	Probability values for AR(p) and MA(q) terms	
AR(2) MA(2)	**0.0000**	**0.0000**
AR(3) MA(3)	**0.0000**	**0.0000**

The probability values of variables of AR(2), MA(2), and AR(3), MA(3) combinations are equal to 0.0000 which is less than 5% of significance level. Therefore, it was concluded that 95% confidence the coefficients are significant in selected combinations.

The following figures, Figure 4.8, and Figure 4.9 show the *E-views* outputs for each significant term.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.924440	0.032625	28.33512	0.0000
MA(2)	-0.958497	0.030808	-31.11204	0.0000

Figure 4.8: *E-views* output for AR(2) and MA(2) $\boldsymbol{D[Ln[GOLD\ PRICE]]}$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(3)	0.908611	0.046321	19.61546	0.0000
MA(3)	-0.944345	0.044700	-21.12612	0.0000

Figure 4.9: *E-views* output for AR(3) and MA(3) $\boldsymbol{D[Ln[GOLD\ PRICE]]}$

The model consists of terms **AR(2) and MA(2) named as model 1** and the model consists of terms **AR(3) and MA(3) named as model 2**. Compared the values given in Table 4.8. to select a better model to forecast fututer gold price.

Table 4.8: R^2, standard error of regression and AIC values for model 1 and model 2.

	R-Squared	**S.E. of Regression**	AIC
AR(2) MA(2)	0.065	**0.0389**	-3.634
AR(3) MA(3)	0.043	**0.0394**	-3.606

R-squared in model 1 is greater than R-squared in model 2. Further AIC value in model 2 is greater than the AIC value in model 1. Yet, the standard error of regression in model 1 and the standard error of regression in model 2 are almost the same. Therefore, we decided to check the correlerogram of residuals in both models and on that evidence model 2 is selected to carried out for residual analysis. Correlerogram of residuals in model 1 is given in APPENDIX A.

4.6 Residual Analysis-$D[Ln[GOLD\ PRICE]]$ −Model 2

After fitting the tentative model to the data, adequacy was examined by analyzing its residuals.

Included observations: 104
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.162	0.162	2.8251	
		2	-0.069	-0.098	3.3418	
		3	-0.079	-0.053	4.0262	0.045
		4	-0.015	0.001	4.0515	0.132
		5	0.074	0.068	4.6594	0.199
		6	0.145	0.120	7.0222	0.135
		7	-0.046	-0.085	7.2615	0.202
		8	-0.096	-0.051	8.3289	0.215
		9	-0.029	0.004	8.4295	0.296
		10	0.060	0.049	8.8574	0.354
		11	0.160	0.124	11.879	0.220
		12	0.024	-0.032	11.947	0.289
		13	-0.148	-0.111	14.597	0.202
		14	-0.063	0.010	15.087	0.237
		15	0.103	0.101	16.394	0.229
		16	0.116	0.045	18.078	0.203
		17	0.053	-0.003	18.429	0.241
		18	0.001	0.038	18.429	0.299
		19	-0.103	-0.050	19.810	0.284
		20	-0.035	-0.017	19.972	0.334
		21	0.058	0.005	20.423	0.370
		22	0.019	-0.037	20.473	0.429
		23	-0.022	0.000	20.540	0.487
		24	-0.005	0.053	20.544	0.549
		25	-0.074	-0.063	21.301	0.563
		26	-0.037	-0.071	21.494	0.609
		27	0.072	0.046	22.240	0.622
		28	-0.038	-0.049	22.449	0.664
		29	-0.054	-0.017	22.873	0.692
		30	0.014	0.043	22.902	0.738

Figure 4.10: The correlogram of residuals of model 2 in $D[Ln[GOLD\ PRICE]]$

The ACF and PACF graphs in Figure 4.10 do not indicate the presence of any structure and all the values lie between 95% of confidence interval and their probability values are greater than 5% of significance level .

The ACF and PACF graphs illustrate that there is no correlation between residual values.

In Figure 4.11, the histogram of residuals suggests that they are normally distributed because of its heavy body and thin tails but outliers can be seen. Further more it is noted that the kurtosis is more than 3 and skewness is closer to 0. Data do not provide evidence against to H_0.

H_0: Residuals are normally distributed

H_1: Residuals are not normally distributed

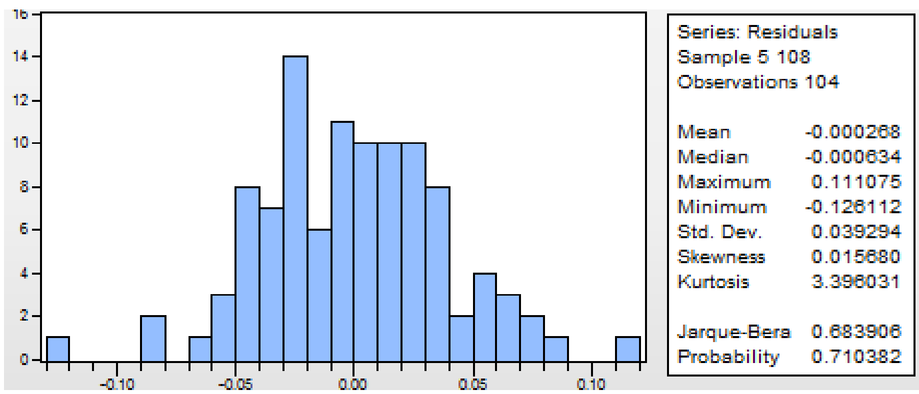


Figure 4.11: The histogram of residuals of model 2 in $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$

Further more the Jarque-Bera statistic has a p-value of 0.710382, implying that the data are consistent with a normal distribution and residuals have constant variance. This can be proved by the evidence of the probability plot of residuals which is shown in Figure 4.12.

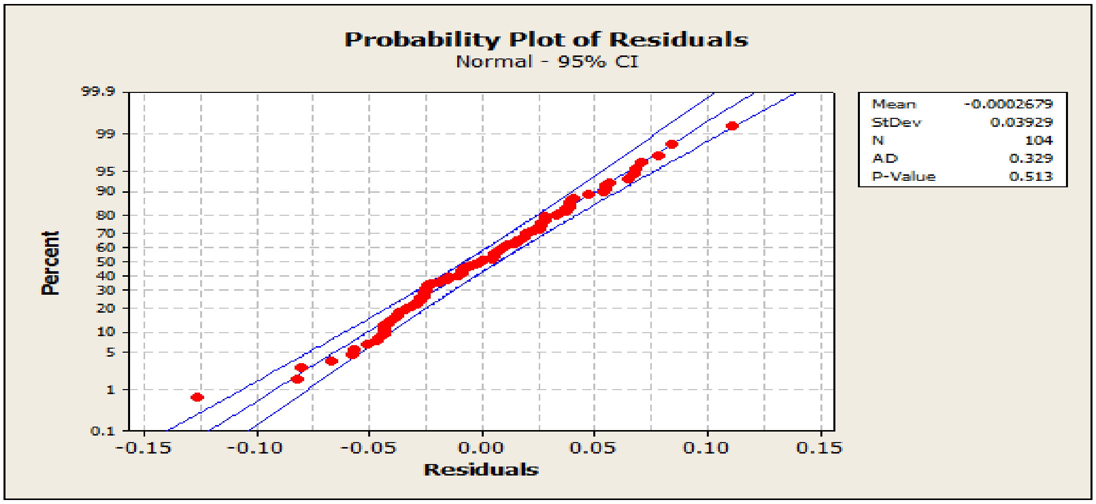


Figure 4.12: The normal probability plot of the residuals of model 2 in $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$.

Figure 4.12 indicates the normal probability plot of residuals. The probability plot of residuals appears to be approximately linear and one outlier exists. The probability value of Anderson Darling test in Figure 4.12 suggests that the residuals follow normality.

H_0: The residuals follow a normal distribution

H_1: The residuals do not follow a normal distribution

Thus, the null hypothesis can be accepted and it can be concluded that the data follows normal distribution with 95% confidence.

The existence of serial correlation among residuals in the considered model is given in Figure 4.13 as follows;

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.867922	Prob. F(2,100)	0.1598
Obs*R-squared	3.740652	Prob. Chi-Square(2)	0.1541

Figure 4.13: The serial correlation of residuals of model 2 in $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$

The Lagrange Multiplier(LM) test has provided a standard means of testing parametric restrictions for a variety of models. It is used to test the independency of residuals. The probability of $\chi^2 -$ statistic(0.1541) greater than 0.05 indicates that no serial correlation occurs in the time series when there are errors associated with a given time period.

H_0: There is no serial correlation of any order up to p

H_1: There is a serial correlation of any order up to p

Therefore the null hypothesis H_0 is accepted and reject the alternative hypothesis H_1.

The presence of heteroscedasticity can be validated by statistical tests. Whit's general heteroscedasticity test (White Test) is used to test the presence of heteroscedaticity among residuals. The probability values are greater than the 5% of significance level which means that the null hypothesis is not rejected and the alternative hypothesis accepted. In other words no heteroscedasticity appears among residuals.

H_0: There is homoscedasity among residuals of variances

H_1: There is heteroscedasticity among residuals of variances

Heteroskedasticity Test: White

F-statistic	0.094252	Prob. F(3,101)	0.9631
Obs*R-squared	0.293134	Prob. Chi-Square(3)	0.9613
Scaled explained SS	0.340140	Prob. Chi-Square(3)	0.9523

Figure 4.14: The heteroscedasticity of residuals of model 2 in $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$

Here all the requirements for the residual diagnostics in time seriesanalysis were tested.

The evidences of residual analysis suggests that the appropriate model to forecast the gold price is $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$.

Therefore the appropriate model can be written as follows;

$$D(1-\alpha_3 B^3)\, Ln\, X_t = \beta_1 e_{t-3}$$

$$D\, Ln\, X_t - \alpha\, DX_{t-3} = \beta_1 e_{t-3}$$

$$\boldsymbol{D}\big[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]\big]_{\boldsymbol{t}} = \alpha DX_{t-3} - \beta_1 e_{t-3}, \text{ where } X \text{ is the } \textit{GOLD PRICE}$$

$$\boldsymbol{D\big[Ln[GOLD\ PRICE]\big]_t} = 0.9086\ D(LnX)_{t-3} - 0.9443\ e_{t-3} \ldots\ldots\ldots (1)$$

Where AR terms written as $D\ (Ln\,X)_{t-3}$, MA terms written as e_{t-3}.

4.7 Forecasting and Accuracy of the Model

Recall the model equation (1)

$$\boldsymbol{D\big[Ln[GOLDPRICE]\big]_t} = 0.9086\ D\ (Ln\,X)_{t-3} - 0.9443\ e_{t-3}$$

$$Ln\,X_t - Ln\,X_{t-1} = 0.9086\ D\ (Ln\,X)_{t-3} - 0.9443\ e_{t-3}$$

$$Ln\ X_t = Ln\,X_{t-1} + 0.9086\ D\ (Ln\,X)_{t-3} - 0.9443\ e_{t-3} \ldots\ldots\ldots\ldots..(2)$$

Substituting values for $Ln\,X$ and $D\ (L\,n\,X)_{t-3}$ in the equation (2) that are derived from *E-views* the future gold price for three months ahead can be found. Estimated future gold price for $t = 109$, $t = 110$ and $t = 111$ is given by equation (2) as follows;

$$Ln\ X_{109} = Ln\,X_{108} + 0.9086\ D\ (Ln\,X)_{106} - 0.9443\ e_{106}$$

$$Ln\ X_{109} = 10.5819 + 0.9086\ \times\ 0.0448 - 0.9443 \times 0.0388$$

$$X_{109} = 39576.86$$

$$Ln\ X_{110} = ln\,y_{109} + 0.9086\ D\ (Ln\,X)_{107} - 0.9443\ e_{107}$$

$$Ln\ X_{110} = 10.5860 + 0.9086\ \times -0.0575 - 0.9443 \times -0.0571$$

$$X_{110} = 39509.64$$

$$Ln\ X_{111} = Ln\,X_{110} + 0.9086\ D\ (Ln\,X)_{108} - 0.9443\ e_{108}$$

$$Ln\ X_{111} = 10.5843 + 0.9086 \times -0.0062 - 0.9443\ \times -0.0080$$

$$X_{111} = 39434.64$$

Table 4.9: The forecast gold prices and error percentages for the model 2 in $D[Ln[GOLD\ PRICE]]$

Index	Month	Predicted	Actual	Error	Error %
109	January 2016	39576.86	40631.69	1054.83	2.59
110	February 2016	39509.64	44424.71	4915.07	11.06
111	March 2016	39434.64	46153.65	6719.01	14.55

From the equation (3.11.1) the calculated Mean Absolute Percentage Error (MAPE) is equal to 9.4 % for $D[Ln[GOLD\ PRICE]]$.

4.8 Summary of Findings

This study recommends that in the time series approach ARIMA model with terms AR(3) and MA(3) was the appropriate model to forecast monthly gold prices in Sri Lanka. Depending the study monthly gold prices in Sri Lanka is increasing.

MAPE value of fitted data for the model is 9.4% which is less than 10%. Therefore, the model is taken as an acceptable model to forecast future gold price.

Chapter 5

Influence of Inflation Rate and Exchange Rate on the Gold Price

5.1 Introduction

This chapter consists of understanding the relationship between inflation rate and exchange rate with the gold price. Regression analysis and the co-integration test have been employed for this purpose.

Depending on the Figure 4.1, it can be concluded that the price of gold in Sri Lanka is increasing and it is probably affected by various factors. Based on a literature review *(Pitigalaarachchi et al (2015)),* the inflation rate and the exchange rate of US dollar to Sri Lankan rupee were selected as the explanatory variables to build the VEC model. In order to examine the relationship among these variables, regression analysis and co-integration test based on Vector Auto-regression (VAR) were carried out on *E views 7* statistical software package.

The models that were built and developed were based on quarterly data of each factor from 2007, January to 2015, December.

5.2 Descriptive Analysis

The time series plots of quarterly gold price, inflation rate and exchange rate are given in APPENDIX B.

The correlation values among gold price, US dollar exchange rate to Sri Lankan rupees and inflation rate have been checked and the results are reported in Table 5.1 as follows;

Table 5.1: Correlation values among quarter values of gold price, inflation rate and exchange rate

	Gold Price	Inflation Rate
Inflation Rate	-0.529 (0.0000)	
Exchange Rate	0.618 (0.0000)	-0.593 (0.0000)

Cell contents: Pearson correlation

(P- value)

The Pearson correlation statistics ($p-$values are in parenthesis) given in Table 5.1 indicate that there is a moderate positive co-relation between gold price and exchange rate (refer Table 3.3) and a negative correlation between gold price and inflation rate. Results based on Table 5.1 give sufficient statistical evidence at the 5% significance level to conclude the existence of a linear relationship between the gold price and exchange rate which means that there is no reason to reject H_0. Hence a regression model was developed in order to determine the relationship between the gold price as the dependent variable with the exchange rates and the inflation rates the independent variables.

5.2.1 Regression Model

The general linear regression model on a non stationary time series Y_t as a combination of another non stationary time series $X_{1t},\ X_{2t}, \ldots ,\ X_{kt}$ can be written as follows;

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \epsilon_t$$

where $\epsilon_t \sim iid(0, \sigma^2)$ and β_i are parameters.

Table 5.1 statistically justifies the existence of a correlation between three variables. Thus, the regression analysis has been carried out to determine the behavior of three variables. The result of the regression analysis is reported in Table 5.2.

Table 5.2: Regression analysis results for gold price, exchange rate and inflation rate.

Predictor	Coefficient	St. Error	t-statistic	p-value
Constant	-19732	22630	-0.87	$\mathbf{0.40}$
Inflation Rate	-414	272	-1.52	$\mathbf{0.14}$
Exchange Rate	505	176	2.86	0.00

The values in Table 5.2 give an idea about the regression analysis conducted on the data. The coefficient of exchange rate is significant at 5% level of significance and with reference to Table 5.2. coefficients of inflation rate and the constant value in the regression model are not significant since p-values are greater than 0.05.

The identified linear regression equation with estimated parameters is given as follows;

$$Gold\ Price = -19732 - 414\ Inflation\ Rate + 505\ \ Exchange\ Rate \ldots\ldots\ldots (5.1)$$

The relationship of the gold price with inflation rate and exchange rate is explained by the regression equation (5.1). The coefficient of exchange rate suggests that gold price will increase 505 times when exchange rate goes up by one and decrease 414 times when inflation rate goes up by one.

Table 5.3: Model Summary

$\boldsymbol{S = 8718}$	$R - sq = \mathbf{42.3}\%$
$R - sq(adj) = 38.8\%$	
$Sum\ of\ squares\ of\ residual\ \ 250805$	
$Durbin - Waston\ statistic\ \ \mathbf{0.14}$	
$F - statistic\ 37.533 Prob(F - statistic)\ 0.0000$	

The value of R^2(42.3%) describes the percentage of variation in the response that is explained by the model which indicates how good the model fit is although it does not indicate the accuracy of the model. Standard error of regression S assess how well the model describes the response. The lower value of standard error of regression(S), the better the description of the response that is provided by the model. But here $S = 8718$. Durbin-Watson statistic 0.14 which is less than $d_L = 1.28$, $k = 2$, $n = 36$ (refer APPENDIX C) also implies that fitted model is affected by serial correlation which can be a **spurious** regression (R^2.>DW value).

Therefore, it can be concluded that the regression analysis is not sufficiently accurate to explain the relationship between the factors. In order to that it is decided to check the existence of co-integration among three factors.

5.3 Co-integration Equation

The existence of co-integration equation among three factors was examined for optimum lag based on minimum AIC and SIC values.

5.3.1 Lag Selection Criteria

The optimum lag length to develop models was checked by conducting lag length criteria based on restricted VAR model.

VAR Lag Order Selection Criteria
Endogenous variables: GOLD EX IR
Exogenous variables: C
Date: 02/26/17 Time: 08:34
Sample: 1 36
Included observations: 33

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-569.5051	NA	2.35e+11	34.69728	34.83333	34.74306
1	-459.7061	192.9801	5.25e+08	28.58825	29.13243	28.77135
2	-436.1026	37.19333	2.20e+08	27.70319	28.65551*	28.02362
3	-423.7304	17.24610*	1.87e+08*	27.49881*	28.85928	27.95657*

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion

Figure 5.1: Lag selection criteria for VAR model

Lag selection has been done to develop co-integration models for quarter value gold price, inflation rate and exchange rate based on minimum AIC and SIC values. In order to that lag 3 and lag 2 are selected as optimum lag lengths respectively to develop VEC models.

5.3.2 Pre Condition for Johansen Co-integration Test

The pre- condition of Johansen co-integration test was checked by employing ADF test with intercept and results are shown in Table 5.4 and Table 5.5.

Table 5.4: ADF test results for three variables on level

Time Series	t-statistic	p −value
Gold Price	-1.793313	0.3774
Inflation rate	-2.895111	0.0564
Exchange rate	0.369260	0.9785

As $p-$ value of each original data is not less than 5% of significance indicate the non-stationarity of data at level$[I(0)]$. The first difference of each has been obtained to make them stationary. Thus, ADF test has been carried out to prove them statistically.

Table 5.5: ADF test results for first difference of three variables

First difference series	t-statistic	p −value
Gold Price	-4.030739	0.0037
Inflation rate	-4.252959	0.0021
Exchange rate	-4.25139	0.0021

ADF test for the first difference series of each variable is statistically recommended that all three series are stationary as p- value of each variable is less than 5% of significance level. This implies that all the time series are stationary$[I(1)]$. It means that the variables meet pre condition. In order to that it was decided that a Johansen co-integration test on VAR based upon two different lag lengths would be carried out.

5.4 The Development of VEC Model Based on Minimum AIC Value

5.4.1 Johansen Co-integration Test

Recall the hypotheses to be checked by Johansen co-integrating test.

H_0: There is no long run co-integration between GOLD, EX and IR

H_1: There is a long run co-integration between GOLD, EX and IR

The trace statistic and maxeigen value statistic in Figure5.2 indicated co-integration existence among the three variables. The p −value 0.02 of null hypothesis H_0 in trace statistic does not exceed the 5% of significance level. Therefore, H_0 " there is no long run co-integration among the three variables" is rejected at 5% level of significance, and the alternative hypothesis H_1 which says that " there is a long run co-integration among the three variables" is accepted. This is justified by the max eigen value test as its $p -$ value 0.006 of H_0does not exceed the significance level of 5%. This means that the H_0 can be rejected and the H_1 which says that there is a long run co-integration relation among three variables can be accepted.

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.571276	32.33144	29.79707	0.0250
At most 1	0.150181	5.229290	15.49471	0.7840
At most 2	0.000684	0.021886	3.841466	0.8823

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.571276	27.10215	21.13162	0.0064
At most 1	0.150181	5.207404	14.26460	0.7155
At most 2	0.000684	0.021886	3.841466	0.8823

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level

Figure 5.2: Johansen co-integration test for minimum AIC value

The following Figure 5.3 indicates there is one co-integration equation among three variables. It explains variation of gold price ($GOLD$) when EX and IR varies. Both indicate positive correlation to the $GOLD$; it implies that both factors have a positive impact on gold price. When the exchange rate is increased by one unit gold price is affected 2927.45 times and when the inflation rate is increased by one unit the gold price is affected 3569.44 times.

Cointegrating Eq:	CointEq1
GOLD(-1)	1.000000
EX(-1)	2927.456 (703.169) [4.16323]
IR(-1)	3569.446 (693.789) [5.14486]
C	-418705.3

Figure 5.3: Co-integration equation output by *E-views 7*

5.4.2 Vector Error Correction Model (VECM)

There are enough statistical evidences to fit a Vector Error Correction Model (VECM) for three variables. The following hypotheses are developed for the VECM.

H_0^a: There is no long run causality running from IR and EX to GOLD

H_1^a: There is a long run causality running from and IR and EX to GOLD

H_0^b: There is no short run causality between GOLD and IR

H_1^b: There is a short run causality between GOLD and IR

H_0^c: There is no short run causality between GOLD and EX

H_1^c: There is a short run causality between GOLD and EX

The tentative VECM equation for the dependent variable $GOLD$ can be written as follows;

$$D(GOLD) = C(1) * (GOLD(-1) + 2927.45 * EX(-1) + 3569.44 * IR(-1) - 418705.30) + C(2) * D\big(GOLD(-1)\big) + C(3) * D\big(GOLD(-2)\big) + C(4) * D\big(GOLD(-3)\big) + C(5) * D\big(EX(-1)\big) + C(6) * D\big(EX(-2)\big) + C(7) * D\big(EX(-3)\big) + C(8) * D\big(IR(-1)\big) + C(9) * D\big(IR(-2)\big) + C(10) * D\big(IR(-3)\big) + C(11)$$

In VECM equation, $C(1)$ is the coefficient of co-integrating equation $(GOLD(-1) + 2927.45 * EX(-1) + 3569.44 * IR(-1) - 418705.30)$. Coefficient $C(1)$ is the error correction term or speed of adjustment towards equilibrium of gold price variable. To determine the significance of the coefficients EC model and the p- values of each term is produced by *E-views 7;* the results are given in Figure 5.4.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.121595	0.025596	-4.750597	0.0001
C(2)	-0.367627	0.191439	-1.920334	0.0685
C(3)	-0.187495	0.163805	-1.144626	0.2652
C(4)	-0.124662	0.154800	-0.805312	0.4297
C(5)	272.1600	192.5900	1.413158	0.1723
C(6)	185.4331	213.7318	0.867597	0.3954
C(7)	327.0341	194.8907	1.678039	0.1082
C(8)	346.0934	176.9396	1.955998	0.0639
C(9)	568.3110	183.1002	3.103825	0.0054
C(10)	153.5847	193.2883	0.794589	0.4357
C(11)	588.3339	368.5940	1.596157	0.1254

Figure 5.4: The probability values of coefficients of the model

Coefficients recorded in Figure 5.4 have both positive and negative values. It means that this VEC model consists of both long run and short run causalities. Here C_1 has a negative sign and it is significant at level of 5% with the probability value of 0.0001. The H_0^a is rejected and that there is a long run causality running from inflation rate and exchange rate to gold price for one period of lag(one quarter) is accepted. The speed of adjustment towards to the equilibrium is 12.1%. $C(2)$, $C(3)$, $C(4)$, $C(5)$, $C(6)$, $C(7)$, $C(8)$, $C(9)$ and $C(10)$ are coefficients of short run causalities and $C(11)$ is the constant for whole system.

$P -$ value of the coefficient of C_{11} is 0.12 which is not significant at 5% since it is greater than the probability value 0.05. Wald test has been employed to examine the short run causality from *EX* and *IR* to *GOLD*.

5.4.3 Wald Test Results for Short Run Causalities

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	1.514771	(3, 21)	0.2398
Chi-square	4.544314	3	0.2084

Null Hypothesis: C(2)=C(4)=C(3)=0
Null Hypothesis Summary:

Figure 5.5: Wald test results for coefficients of $C(2), C(3)$ and $C(4)$

Figure 5.5 reported that p-value of χ^2 is 0.2084 which is greater than 5%of significance and infers that the null hypothesis, $C(2) = C(3) = C(4) = 0$ is accepted and there is no short run causality effect from lag one, lag two and lag three from $GOLD$ to current $GOLD$.

From the VECM the short run causality from EX to $GOLD$ is examined by using the coefficients $C(5)$, $C(6\ )$, and $C(7)$ as they are the coefficients of EX variable to $GOLD$ which has a positive sign. To check short run causality from EX to $GOLD$; Wald test is implemented. Figure 5.6 shows the Wald test results for coefficients for exchange rate(EX). The null hypothesis $\mathrm{H}_0^{\mathrm{b}}$ is $C(5) = C(6) = C(7)$=0 checked by p-value of χ^2

is 0.029, less than the significance level of 5%. Thus, H_0^b can be rejected and H_1^b can be accepted. It is inferred that these coefficients are jointly influence on $GOLD$ then it is accepted that a short run causality from EX to $GOLD$ exists. In other words, all EX having one, two and three lags (here lag represents quarters) causes $GOLD$.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	3.005842	(3, 21)	0.0533
Chi-square	9.017527	3	0.0291

Null Hypothesis: C(5)=C(6)=C(7)=0
Null Hypothesis Summary:

Figure 5.6: Wald test results for coefficients of $C(5), C(6)$ and $C(7)$.

Wald test has been carried out to check the significance of joint influence of IR to $GOLD$ in Figure 5.7. The test results inferred that χ^2 is 0.000 which is < 5% of significance level. Hence H_0^c which says "there is no short run causality between $GOLD$ and IR" is rejected, and accepted H_1^c which implies that there is a joint influence between IR and $GOLD$. In other words, it is concluded that there is a short run causality from IR to $GOLD$.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	7.619081	(3, 21)	0.0012
Chi-square	22.85724	3	0.0000

Null Hypothesis: C(8)=C(9)=C(10)=0

Figure 5.7: Wald test results for coefficients of $C(8)$, C(9) and $C(10)$

The p −value of coefficient of $C(8), C(9)$ and $C(10)$ is 0.0000 which is significant at 5% level. It implies that IR of two quarters before current quarter affects $GOLD$ price.

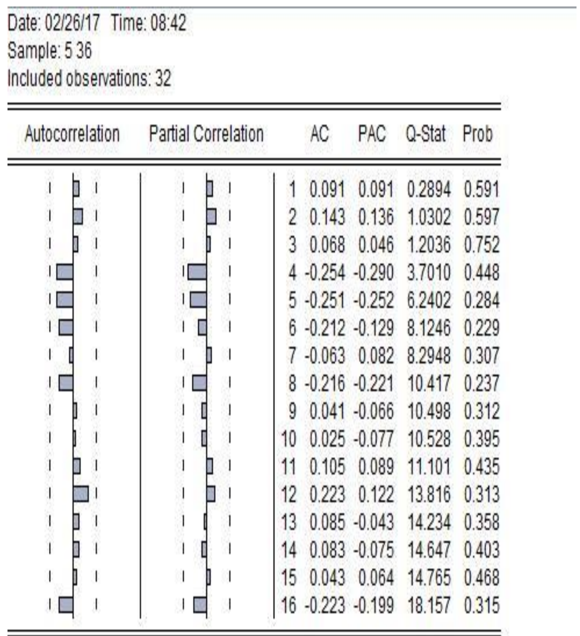
Therefore, the final VECM for gold price, inflation rate and exchange rate can be written as follows;

$$D(GOLD) = -0.12 * (GOLD(-1) + 2927.45 * EX(-1) + 3569.44 * IR(-1) - 418705.30) + 272.16 * D\big(EX(-1)\big) + 185.43 * D\big(EX(-2)\big) + 327.03 * D\big(EX(-3)\big) + 346.09 * D\big(IR(-1)\big) + 568.3 * D\big(IR(-2)\big) + 153.58 * D\big(IR(-3)\big)$$

R^2 in this model is 69.3% inferred the current gold price is explained by 69.3% of previous gold price, exchange rate and inflation rate. P-value of F-statistic of the model is 0.001 which is significant at 5% meaning that the data fits well with the model.

5.4.4 Residual Analysis for the Model

Under a model diagnostic check, a residual analysis criterion has been carried out to assess whether the model is correctly specified. Firstly, to check the correlogram of the model so that autocorrelation is available among residuals, secondly to check that the residuals of the model are normally distributed, thirdly, to verify that the model is able to capture the heteroscedasticity affect and finally to check any serial correlation effect on the model. In Figure 5.8, the correlogram of residuals indicates that residuals between the 95% confidence interval are such that none of the lags exceeds the limits. This inferred that there is no reason to reject the null hypothesis H_0 which said that no autocorrelation exists among the residuals of the model.



Date: 02/26/17 Time: 08:42
Sample: 5 36
Included observations: 32

	AC	PAC	Q-Stat	Prob
1	0.091	0.091	0.2894	0.591
2	0.143	0.136	1.0302	0.597
3	0.068	0.046	1.2036	0.752
4	-0.254	-0.290	3.7010	0.448
5	-0.251	-0.252	6.2402	0.284
6	-0.212	-0.129	8.1246	0.229
7	-0.063	0.082	8.2948	0.307
8	-0.216	-0.221	10.417	0.237
9	0.041	-0.066	10.498	0.312
10	0.025	-0.077	10.528	0.395
11	0.105	0.089	11.101	0.435
12	0.223	0.122	13.816	0.313
13	0.085	-0.043	14.234	0.358
14	0.083	-0.075	14.647	0.403
15	0.043	0.064	14.765	0.468
16	-0.223	-0.199	18.157	0.315

Figure 5.8: The correlogram of the residuals of the model

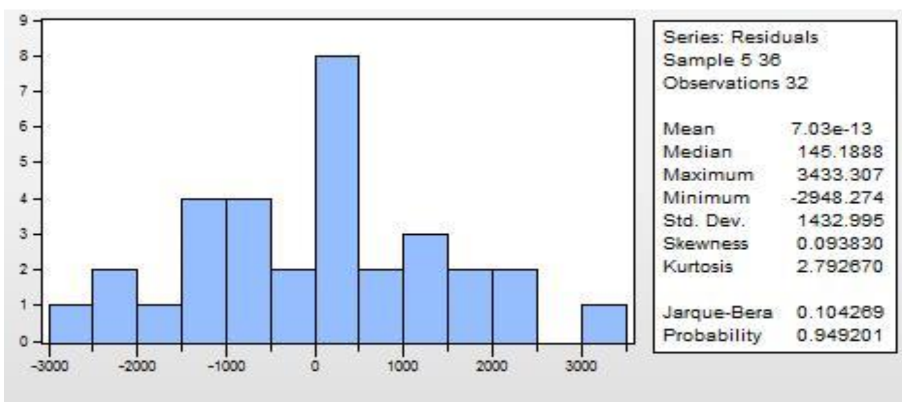


Figure 5.9: The histogram of the residuals of the model.

Figure 5.9 shows the histogram of the residuals of the developed co-integration model. Since the probability of Jarque-Bera statistics is greater than 5%, the H_0 that residuals are normally distributed can be accepted. The kurtosis of the histogram is less than 3 and its left tail is heavy while skewness is closer to 0.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.876341	Prob. F(2,19)	0.4325
Obs*R-squared	2.702581	Prob. Chi-Square(2)	0.2589

Figure 5.10: The serial correlation of the residuals of the model

The serial correlation effect is tested for the residuals of the model and given in Figure 5.10. The $p-$ value of $\chi^2 >$ 5% and hence the H_0 is accepted, indicating that no serial correlation effect can be seen among residuals of the model.

The homoscedasticity of the residuals of the model is tested for non- constant variance of the model. The results are given in Figure 5.11.

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.412691	Prob. F(12,19)	0.9400
Obs*R-squared	6.616204	Prob. Chi-Square(12)	0.8819
Scaled explained SS	2.553982	Prob. Chi-Square(12)	0.9980

Figure 5.11: The heteroscedasticity of the residuals of the model

The H^0 that there is a homoscedasticity, is tested and the results shown in Figure 5.11. The $p-$value of $F-$test statistics of 0.9400 which is greater than 5% significance level means that one cannot reject the null hypothesis and accept the alternative hypothesis. Thus the non- constant variance of residuals exists. Therefore the model is significant for all residual tests.

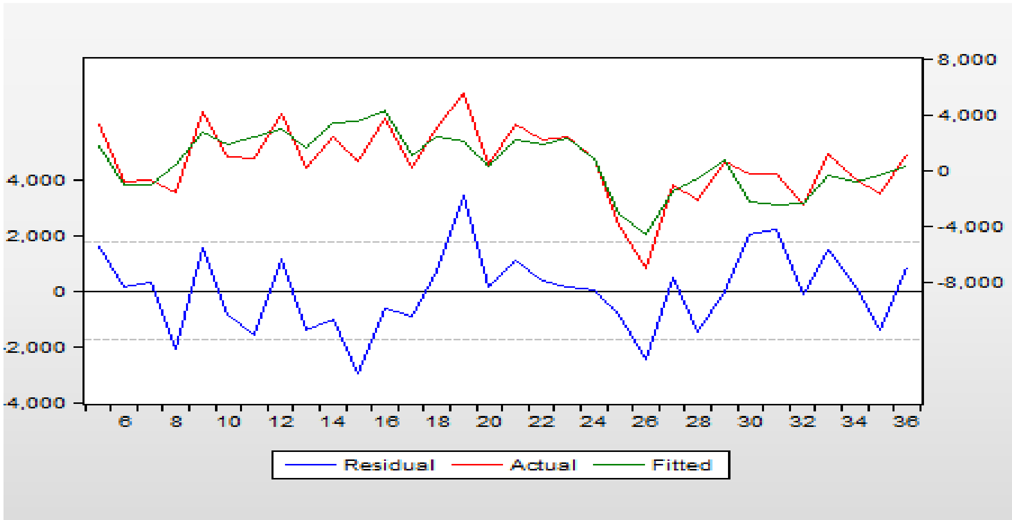


Figure 5.12: The comparison of actual and fitted values for VECM with gold price as the dependent variable.

Figure 5.12 shows the fitted co-integration model for data series which is better than ordinary regression model and residuals that also follows the property of $iid \sim (0, \sigma^2)$.

5.5 The Development of VEC Model Based on Minimum SIC Value

5.5.1 Johansen Co-integration Test

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.584539	36.61910	29.79707	0.0070
At most 1	0.159048	7.632978	15.49471	0.5054
At most 2	0.056427	1.916699	3.841466	0.1662

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.584539	28.98612	21.13162	0.0032
At most 1	0.159048	5.716278	14.26460	0.6499
At most 2	0.056427	1.916699	3.841466	0.1662

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Figure 5.13: Johansen co-integration test for minimum SIC value

Johansen trace statistics and maxeigen values are given in Figure 5.13. Trace statistic and maxeigen value test indicate that one co-integration equation exists. It is inferred that both statistic values are greater than the critical values (None), that establishes a long run co-integration relationship in the model. The H_0 of trace statistic that " there is no long run co-integration among variables" is rejected since the p-value, 0.007 is less than 5% significance level. Thus we accept H_1 that "there is a long run co-integration among variables".

As p-value is 0.003 in maxeigen value test is justified that the co-integration is exists among three variables. H_0 of maxeigen value test that "there is no long run co-integration among variables" is rejected and H_1 that "there is a long run co-integration among variables" is accepted.

Cointegrating Eq:	CointEq1
GOLD(-1)	1.000000
EX(-1)	3015.910 (664.210) [4.54060]
IR(-1)	3316.334 (777.458) [4.26561]
C	-426924.5

Figure 5.14: The co-integration equation output by *E-views 7*

Figure 5.14 indicates coefficients of variables of EX and IR in co-integration equation. It explains the variation of gold price when EX and IR values are vary. Both factors make a positive impact on the gold price.

5.5.2 Vector Error Correction Model (VECM)

There is enough statistical evidence to fit a Vector Error Correction Model (VECM) for three variables. The following hypotheses are developed for the VECM.

H_0^{ae}: There is no long run causality running from IN_RATE and EX_RATE to GOLD

H_1^{ae}: There is a long run causality running from and IN_RATE and EX_RATE to GOLD

H_0^{b}: There is no short run causality between GOLD and IN_RATE

H_1^{b}: There is a short run causality between GOLD and IN_RATE

H_0^{c}: There is no short run causality between GOLD and EX_RATE

H_1^{c}: There is short run causality between GOLD and EX_RATE

The tentative VEC model equation for the dependent variable $(GOLD)$ can be written as follows;

$$\begin{aligned} D(GOLD) = \ & C(1) * (GOLD(-1) + \ 3015.91 * EX(-1) + \ 3316.33 * IR(-1) \\ & - 426924.51)) + C(2) * D\big(GOLD(-1)\big) + C(3) * D\big(GOLD(-2)\big) \\ & + C(4) * D\big(EX(-1)\big) + C(5) * D\big(EX(-2)\big) + C(6) * D(IR(-1)) \\ & + C(7) * D(IR(-2)) + C(8) \end{aligned}$$

The equation $GOLD(-1) + \ 3015.91 * EX(-1) + \ 3316.3 * IR(-1) - 426924.51)$ in the VEC model indicates the co-integration equation with error correction term $C(1)$. The rest of the model indicates short run causalities among three variables and need to justify the significance of coefficients $C(2)$, $C(3)$, $C(4)$, $C(5)$, $C(6)$ and $C(7)$. Thus their p -values are examined for 5% of significance level and recorded in Figure 5.15 as follows;

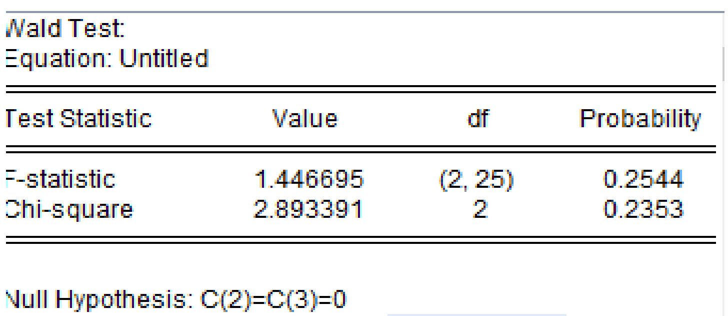
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.091353	0.015956	-5.725360	0.0000
C(2)	-0.224150	0.153492	-1.460338	0.1567
C(3)	-0.126355	0.147618	-0.855958	0.4002
C(4)	278.7885	130.5580	2.135361	0.0427
C(5)	474.6591	146.2819	3.244823	0.0033
C(6)	62.83319	140.2590	0.447980	0.6580
C(7)	304.7598	148.6022	2.050843	0.0509
C(8)	447.4871	337.3095	1.326637	0.1966

Figure 5.15: The probability values of coefficients of the model

Figure 5.15 depicts the coefficient values and their probabilities in each coefficient. Except $C(1)$ the other coefficients indicate short run causalities in the VEC model. The negative sign of the coefficient of $C(1)$ indicates that there is a long run causality among the three variables. The $C(1)$ coefficient value lies between 0 and 1 and the p-value is less than 0.05 which implies that the coefficeint value is statistically significant at 5% level. Although the p-value of other coefficients are more than 5% of significance level they cannot be rejected as insignificant as it is necessary to confirm whether they have a joint influence on $GOLD$. The Wald test has been carried out to justify this argument statistically.

5.5.3 Wald Test Results for Short Run Causalities

The Figure 5.16 indicates the results of the Wald test for the coefficients $C(2)$ and $C(3)$ as they are the coefficients for short run causality of $GOLD$. The p-value of 0.23 for χ^2 is greater than 5% significance level which indicates that null hypothesis $C(2) = C(3) = 0$ is accepted. It infers that there is no short run causality from lag 1 and lag 2 of gold price to current gold price.


Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	1.446695	(2, 25)	0.2544
Chi-square	2.893391	2	0.2353

Null Hypothesis: C(2)=C(3)=0

Figure 5.16: Wald test results for coefficietns $C(2)$ and $C(3)$

Figure 5.17 shows the p-value of $C(4)$ and $C(5)$ is 0.0000 and it is significance at 5% of significance level. Therfore, null hypothesies can be rejected and an alternative hypothesies that says there is a short run causality from EX to $GOLD$ is accepted.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	10.69786	(2, 25)	0.0004
Chi-square	21.39571	2	0.0000

Null Hypothesis: C(4)=C(5)=0

Figure 5.17: Wald test results for coefficients $C(4)$ and $C(5)$

Figure 5.18 indicates the p-value of χ^2 is 0.037 which is less than 5% of significance level inferring that null hypothesis can be rejected and an alternative hypothesis can be accepted. It implies the joint influence from lag 1 and lag 2 of IR to current $GOLD$.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	3.294104	(2, 25)	0.0537
Chi-square	6.588209	2	0.0371

Null Hypothesis: C(6)=C(7)=0
Null Hypothesis Summary:

Figure 5.18: Wald test results for coefficietns $C(6)$ and $C(7)$

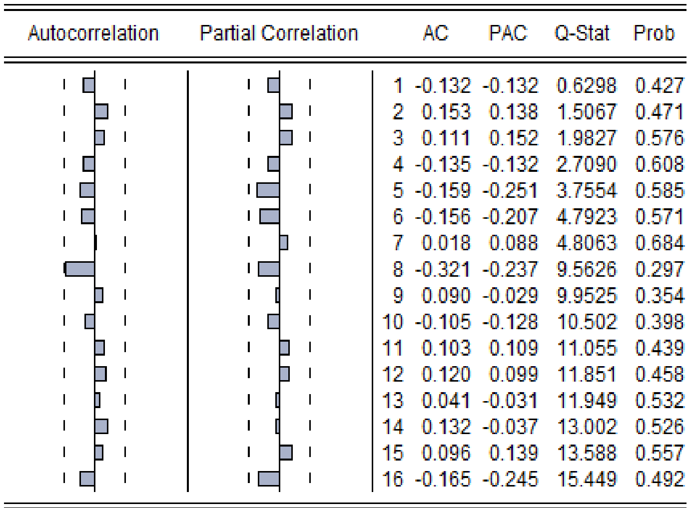
Therefore, the final VECM for gold price, inflation rate and exchange rate can be written as follows;

$$\begin{aligned} D(GOLD) = & -0.09 * (GOLD(-1) + 3015.91 * EX(-1) + 3316.33 * IR(-1) \\ & - 426924.51)) + 278.78 * D\big(EX(-1)\big) + 474.65 * D\big(EX(-2)\big) \\ & + 62.83 * D(IR(-1)) + 304.75 * D(IR(-2)) + 447.48 \end{aligned}$$

The R^2 in the model is 65.1%, infers that the current gold price is explained by 65.1% of previous gold price, exchange rate and inflation rate. The developed model is significant at 5% of significance level since p −value of the F −statistic is 0.000.

5.5.4 Residual Analysis for the Model

Residual analysis has been carried out under four criteria for a model diagnostic.



	AC	PAC	Q-Stat	Prob
1	-0.132	-0.132	0.6298	0.427
2	0.153	0.138	1.5067	0.471
3	0.111	0.152	1.9827	0.576
4	-0.135	-0.132	2.7090	0.608
5	-0.159	-0.251	3.7554	0.585
6	-0.156	-0.207	4.7923	0.571
7	0.018	0.088	4.8063	0.684
8	-0.321	-0.237	9.5626	0.297
9	0.090	-0.029	9.9525	0.354
10	-0.105	-0.128	10.502	0.398
11	0.103	0.109	11.055	0.439
12	0.120	0.099	11.851	0.458
13	0.041	-0.031	11.949	0.532
14	0.132	-0.037	13.002	0.526
15	0.096	0.139	13.588	0.557
16	-0.165	-0.245	15.449	0.492

Figure 5.19: The correlogram of the residuals of the model

All $p-$values of the correlogram of residuals in Figure 5.18 lie between 95% confidence interval inferring that all lags do not exceed the limits. It implies there is no reason to reject the null hypothesis H_0which says that "no autocorrelation exists among the residuals of the model".

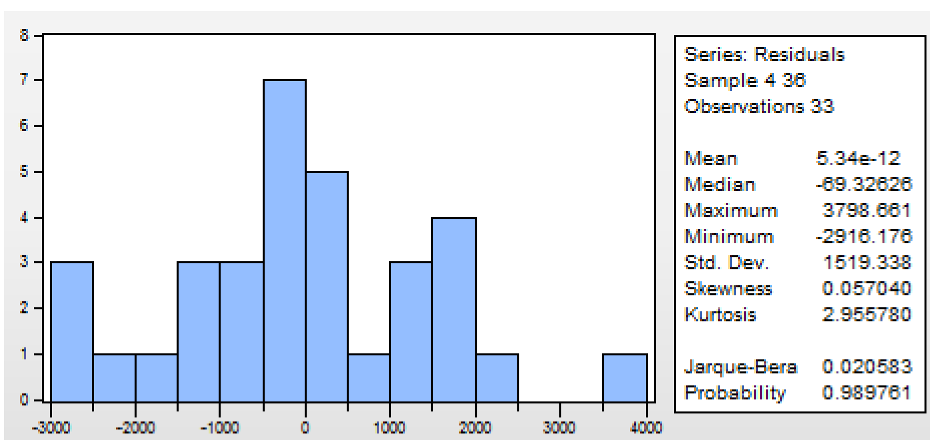


Figure 5.20: The histogram of the residuals of the model

Figure 5.20 shows the histogram of the residuals of the developed model. Since the p −value of Jarque-Bera statistics is greater than 5%, the H_0 that residuals are normally distributed can be accepted. That skewness is closer to zero implies that the distribution of the residual series is symmetric. Since kurtosis of the residual series is 2.95 (closer to 3) this implies the normality of residual distribution.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3.575285	Prob. F(2,23)	0.0445
Obs*R-squared	7.826347	Prob. Chi-Square(2)	0.0200

Figure 5.21: The serial correlation of the residuals of the model.

The p −value of χ^2 is 0.02 which is less than 0.05. Hence the H_0 is rejected and it is accepted that there is serial correlation affect among residuals of the developed model.

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.400266	Prob. F(9,23)	0.9221
Obs*R-squared	4.468735	Prob. Chi-Square(9)	0.8779
Scaled explained SS	2.507995	Prob. Chi-Square(9)	0.9807

Figure 5.22: The heteroscedasticity of the residuals of the model

The p-value of F-statistic is 0.9221 which is greater than 5% significance level means that one cannot reject the null hypothesis. Thus the non constant variance of residuals exists in the model.

5.6 Appropriate Final VEC Model

The developed VEC model based on minimum SIC value inferred by the lag selection criteria is rejected as an affect of serial correlation. Thus, the developed VEC model based on minimum AIC value is accepted as the appropriate final fitted model.

The appropriate final fitted VEC model can be written as follows;

$$D(GOLD) = -0.12 * (GOLD(-1) + 2927.45 * EX(-1) + 3569.44 * IR(-1) - 418705.30) + 272.16 * D\big(EX(-1)\big) + 185.43 * D\big(EX(-2)\big) + 327.03 * D\big(EX(-3)\big) + 346.09 * D\big(IR(-1)\big) + 568.31 * D\big(IR(-2)\big) + 153.58 * D\big(IR(-3)\big) + 588.33$$

The speed of adjustment towards to equilibrium of $GOLD$ is 1.2%, AIC value is 18. 02, R^2 is 69.2% and standard error of regression is 1741.06.

5.7 Forecasting and Accuracy of the Model

Gold price forecasted for three quarters ahead is calculated by the fitted VEC model and pieces are given in the Table 5.6.

Table 5.6: Forecasting gold price for VEC model

Index	**Quarter**	**Predicted**	**Actual**	**Error**	**Error %**
37	1st 2016	42,110.96	43736. 69	1625.73	3.71
38	2nd 2016	43,856.02	46 956.76	3100.74	6.60
39	3rd 2016	45,601.08	49992.39	4391.31	8.78

From the equation (3.11.1) the calculated Mean Absolute Percentage Error (MAPE) is equal to 6.36 %.

5.8 Summary of Findings

- Ordinary regression cannot be employed for multivariate time series data because of the existence of serial correlation among data. Thus, it is necessary to move to a co-integration model to identify the relation between the factors under considered.

- Lag selection has done based on minimum AIC value when develop a VEC model provide better results.
- Thus, the appropriate VEC model to forecast gold price is

$$D(GOLD) = -0.12 * (GOLD(-1) + 2927.45 * EX(-1) + 3569.44 * IR(-1) - 418705.30) + 272.16 * D\big(EX(-1)\big) + 185.43 * D\big(EX(-2)\big) + 327.03 * D\big(EX(-3)\big) + 346.09 * D\big(IR(-1)\big) + 568.31 * D\big(IR(-2)\big) + 153.58 * D\big(IR(-3)\big) + 588.33$$

- The previous gold price has an impact on the current gold price. The previous exchange rates and inflation rates also have a joint impact on the current gold price.

- The MAPE value for the appropriate final fitted VEC model is 6.36% **.**

Chapter 6

Discussion

6.1 Summary in Chapter 4

Although throughout the ages Sri Lanka has neither been a leading producer nor a purchaser of gold; the country in unable to free itself from the effects of the rise and fall of the gold prices. Therefore forecasting gold prices is an important aspect of economic decisions making. In this research work the primary focal area of examination is the inclusion of a model for the determination of monthly gold price using Box-Jenkins and ARIMA models.

Basically, two models have been developed to forecast gold prices depending on statistical analysis that has carried out. Two models were named $D[GOLDPRICE]$ and $D\big[Ln[GOLDPRICE]\big]$. The $D[GOLDPRICE]$ models with terms AR (1) and MA (1) and AR (2) and MA (2) were rejected because of the coefficients are not significant for 5% of significance level. Even though the coefficients are significant for 5% of significance level in the model $D[GOLDPRICE]$ with AR (3) and MA (3) terms rejected due to the high value of standard error of regression. As per the results of Box-Cox transformation, two models were developed for $D[Ln[GOLD\ PRICE]]$. Model 1 has AR(2) and MA(2) terms and model 2 has AR(3) and MA(3) terms. Correlograms of residuals of the developed models for $D[Ln[GOLD\ PRICE]]$ were suggested that the appropriate model to forecast the monthly gold price is the $D[Ln[GOLD\ PRICE]]$ with terms AR(3) and MA(3).

Therefore, the appropriate model can be written as follows;

$$D(1-\alpha_3 B^3)\,Ln\,X_t = \beta_1 e_{t-3}$$

$$\boldsymbol{D\big[Ln[GOLD\ PRICE]\big]_t} = 0.9086\,D(LnX)_{t-3} - 0.9443\,e_{t-3}$$

Where AR(3) term written as $D\,(Ln\,X)_{t-3}$, MA(3) term written as e_{t-3} and X is the $GOLD\ PRICE$.

6.2 Summary in Chapter 5

Gold price in the market is always influenced by multiple factors such as inflation rates, oil prices, exchange rates etc. In this study the two factors considered were inflation rate and exchange rate. Three different analyses have been carried out to determine the relationship between gold price and these factors. The Granger causality affect has been tested. Since the test was proved to fail the affect of Granger causality among tested data range statistically, as a result of that there is no statistical facilities to fit VAR model.

Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
D1EX does not Granger Cause D1GOLD	32	2.22265	0.1104
D1GOLD does not Granger Cause D1EX		1.62278	0.2093
D1IR does not Granger Cause D1GOLD	32	0.57645	0.6359
D1GOLD does not Granger Cause D1IR		0.18704	0.9042
D1IR does not Granger Cause D1EX	32	0.90940	0.4505
D1EX does not Granger Cause D1IR		1.41591	0.2615

Figure 6.1: Granger causality test for first difference of three variables. Here EX; exchange rate, IR; inflation rate and GOLD means gold price

The hypothesis is to be checked by implementing Granger Causality test for three time series data.

H_0: X does not Granger Cause Y

H_1: X does Granger Cause Y

Where X and Y are three variables. In order to test the null hypothesis, F statistic is appointed.

The corresponding probability of F-statistic of null hypothesis in each variable does not Granger cause as it is greater than 5% of significance level recommends the acceptance of the null hypothesis H_0. Thus, there is no Granger causality affect among three variables in the tested time period can be determined. The ordinary regression analysis has been carried out for the quarterly value of factors and the low DW statistic(0.14) is justified by the existence of spurious regression. Thus, to understand the relationship among three variables co-integration has been carried out. The lag values have been selected to carry out Johansen co-integration test is based on minimum AIC and SIC values that indicate the existence of co-integration vectors. A VEC model was developed for a minimum AIC value, which has a lag length of 3, and the inferred gold price of the current quarter is explained by 69.3% of the gold price of the previous quarter and the exchange and inflation rates. The speed of adjustment towards the equilibrium is 12.1%. Residual analysis is carried out on the developed VEC model for a minimum SIC value with a lag length 2, failed the serial correlation LM test implying that a serial correlation exists among residuals of the model. Thus, it is decided that VEC model developed with a minimum AIC value is the appropriate model to explain the relationship between gold price, exchange rates and inflation rates. Therefore, the appropriate VEC model to forecast the gold price can be written as follows;

$$\begin{aligned} D(GOLD) = -0.12 & \\ & * (GOLD(-1) + 2927.45 * EX(-1) + 3569.44 * IR(-1) \\ & - 418705.30) + 272.16 * D\big(EX(-1)\big) + 185.43 * D\big(EX(-2)\big) \\ & + 327.03 * D\big(EX(-3)\big) + 346.09 * D\big(IR(-1)\big) + 568.31 * D\big(IR(-2)\big) \\ & + 153.58 * D\big(IR(-3)\big) + \ 588.33 \end{aligned}$$

6.3 Model Comparison.

When comparing two models it can be seen that increasing of percentage errors of VEC model is higher than increasing of percentage error of ARIMA model.

Table 6.1: Forecast gold prices for appropriate ARIMA model

Month	**ARIMA model**			
	Predicted	**Actual**	**Error %**	
January 2016	39576.86	40631.69	2.59	**MAPE 9.4 %**
February 2016	39509.64	44424.71	11.06	
March 2016	39434.64	46153.65	14.55	

The R^2 is 0.043, standard error of regression is 0.0394 and AIC value is -3.606 for the appropriate ARIMA model with terms AR (3) and MA (3).

Table 6.2: Forecast gold prices for appropriate VEC model

Quarter	**VEC model**			
	Predicted	**Actual**	**Error %**	
First quarter 2016	42110.96	43736. 69	3.71	**MAPE 6.36 %**
Second quarter 2016	43856.02	46 956.76	6.60	
Third quarter 2016	45601.08	49992.39	8.78	

The speed of adjustment towards to equilibrium of $GOLD$ is 1.2%, AIC value is 18. 02, R^2 is 69.2% and standard error of regression is 1741.06.

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

At the end of this study depending on the statistical results the following conclusions were made.

- For the first phase of the study a time series model was fitted to forecast the monthly gold price in Sri Lanka. A systematic and iterative methodology of Box-Jenkin ARIMA was developed forecasting the gold price. Terms with AR(3) and MA(3) in model $\boldsymbol{D}[\boldsymbol{Ln}[\boldsymbol{GOLD\ PRICE}]]$ was selected a suitable fitted model to forecast the gold price in Sri Lanka depending on forecasting accuracy, that is MAPE 9.4 %.

The following conclusions were made on the basis of the study carried out on the regression analysis and non stationary analysis.

- Gold price and exchange rate show positive correlation which can be understood from the ordinary regression analysis. However, regression analysis indicates the existence of spurious **regression** among the three variables.

- The gold price is considered as the dependent variable in the VEC model developed based on minimum AIC value inferring long run equilibrium and the existence of short run causalities among three variables. The speed of adjustment towards the equilibrium is 12.1%. Gold price of the current quarter is explained by 69.3% of the gold price of the previous quarter, exchange rate and inflation rate in the VEC model.

- LM test which is done for serial correlation for residuals indicate further existence of serial correlation in the VEC model developed based on minimum SIC value. Thus it is concluded, the model is not acceptable to forecast the gold price in Sri Lanka.

- Therefore the VEC model developed based on minimum AIC value that is

$$D(GOLD) = -0.12 * (GOLD(-1) + 2927.45 * EX(-1) + 3569.44 * IR(-1) - 418705.30) + 272.16 * D(EX(-1)) + 185.43 * D(EX(-2)) + 327.03 * D(EX(-3)) + 346.09 * D(IR(-1)) + 568.31 * D(IR(-2)) + 153.58 * D(IR(-3)) + 588.33$$

is a suitable model to forecast the gold price in Sri Lanka. MAPE value is 6.36% .

- When comparing two models it can be identified that increasing the percentage error of ARIMA model is higher than the VEC model.

- According to the percentage errors it can be concluded these two models are more suitable to forecast Sri Lankan gold price for short term of periods. Since the mean absolute percentage error (MAPE) as the forecasting accuracy measure, the most appropriate model is to forecast the gold price in Sri Lanka is VEC model than the ARIMA model.

- Granger causality test based on VAR estimation indicates that no Granger causality exists in considered data range. Hence, VAR models were not developed.

7.2 Recommendations

- The study could have been improved by adding more variables such as oil prices, stock market index, interest rate and consumer price index with gold price.
- The expanded data range would give more generalized models and such models should be checked for different frequencies of data like annual and spot prices.
- Co-integration models can be developed by changing the dependent variables which would provide a more rational view on how the relationships would be affected.

LIST OF REFERENCES

Adrangi, B(2003), Inflation Economic activity, hedging: The case of gold, and silver investments, Gold price modeling using system dynamics, *The journal of wealth management 6.*

Baguley, T.,(2012), Serious Stats: A guide to advanced statistics for the behavioral Sciences, *Palgrace Macmillan, pp 402.*

Banumathy, K., Azhagaiah, R. (2015), Long- run and short-run causality between stock price and gold price: Evidence of VECM analysis from India, *Management studies and Economics Systems (MSES).*-1(4), 247-256.

Bilal, A.R., Talib, N.B.A., Haq I.U., Khan, M.N. and Naveed, A.A., (2013), How gold prices correspondence to stock index: A comparative analysis of Karachi stock exchange and Bombay stock exchange, *World Applied Sciences 21(4) pp 1818- 4952.*

Bisgaard, S. and Kulahci, M. (2011), Time series analysis and forecasting by example, *Wiley and Sons, Inc.*

Bishnoi, R. and Lan, J. (2014), An empirical analysis of factors affecting gold prices, *IJHPD 3, No.2.*

Capie, F., Terence, C.M. and Wood, G. (2015), Gold as a hedge against the dollar, *International Financial Markets, Institution and Money 15, 343.*

Cochrane, J.H.(1997), Time series for macroeconomics and finance, *Graduate School of Business, University of Chicago.*

Deshmukh, R.B. and Pawar, D.D.(2017), Prediction of gold prices using time series analysis, *Bulletin of Mathematics and Statistics research, A peer reviewed international research journal Vol.5, Issue 1, 40-44.*

Dhanalakshmi, P.M. and Reddy, R.R.S.(2016), An analytical study on forecasting model with special attention to gold price, *International Journal of Advanced Research, Ideas and Innovations in Technology, Volume 2, Issue 3.*

Sepanek, E., How these 10 factors regularly influence on gold prices, *www.sbcgold.cm/blog/10-factors-regularly-influence-gold-prices/ (2014).*

Gangopadhyay, K. Jangir, A. and Sensarma, R. (2015), Forecasting the price of gold; an error correction approach, *IIMB Management Review.*

Guha, B. and Bandyopadhyay, G.(2016), Gold price forecasting using arima model, *Advanced Management Science 4, No.2, 117.*

Hipel, K.W. and Mcleod, A.I. (1994), Time series modeling of water resources and environmental systems, *Amsterdam, Elsevier.*

Ibrahim, S.N., Kamaruddin, N.I. and Hasan, R. (2014), The determinants of gold prices in Malaysia, *Journal of Advanced Managment Science, Vol. 2, No.1.*

Ismail, Z., Yahya, A. and Shabri,(2009), Forecasting gold prices using multiple linear regression method, *American Journal of Applied Sciences 6, No.8, 1509.*

Khan, M.M.A.(2013), Forecasting of gold prices (box jenkins approach), *International Journal of Emerging Technology and Advanced Engineering 3,* pp 662- 670.

Kirkemo, H., Newman, W.L. and Ashley, R.P., A brief history of gold uses, prospecting, mining and production, *Republished from a USGS general interest publication(2005-2017)* .

Kumari, K.W.S.N. and Tan, A.(2014), Modeling volatity series: with reference to gold price, *Proceedings of the Peradeniya Univ, International Research Sessions, Sri Lanka 18,* pp 391.

Mombeini, H. and Chamzini, A.Y. (2015), Modeling gold price via artificial neural network, *Journal of Economics, Business and Management, Vol.3, No.7*.

Montgomery, D.C., Jennings, C.L. and Kulahci, M. (2015), Introduction to time series analysis and forecasting, *Wiley and Sons, Inc.*

Moore, G. H. (1990), Gold *prices and a leading index of inflation*, Challenge 33 3, No.7

Nair, G. K., Choudhary, N. and Purohit, H. (2015), The relationship between gold prices and exchange value of us dollar in India, *Emerging Markets Journal 5, No.1*.

Navin and Vadivu, G.(2015), Big data analytics for gold price forecasting based on decision tree algorithm and support vector regression(SVR*), International Journal of Science and Research(IJSR), Volume 4, Issue 3.*

Pitigalaarachchi, P.A.A.C., Jayasundara, D.D.M. and Chandarasekara, N.V.(2015), *Forecasting Sri Lankan gold prices using time series analysis*, Proceedings of the Peradeniya University International Research sessions, Sri Lanka, Volume 19.

Poonam, B. and Sharma, B.(2014), Impact of crude oil and us dollar on gold prices in India, *Asia Pacific journal of Research, 60-66, Vol:1, Issue XV.*

Prakash, P. and Sundararajan, S. (2014), An empirical analysis on the relationship between gold and silver with special reference to the national level commodity exchanges India.

Shafiee, S. and Topal, E. (2010), An overview of global gold market and gold price forecasting, *Resources Policy 35, 178-189.*

Sherman, E. J. (1983), A gold pricing model, *Journal of portfolio management 9.*

Simakova, J. (2011), Analysis of the relationship between oil and gold prices, *pp 651-662.*

Sindhu (2013), A study on impact of select factors on the price of gold, *IOSR Journal of Business and Management (IOSR-JBM) 8, PP 84-93.*

Jadhav, S., Kakade, S, Utpat, K. and Deshpande, H.(2015), Indian share market forecasting with arima model, *International Journal of Advanced Research in Computer and Communication Engineering, Vol.4, Issue 11.*

Sparks, J.J. and Yurova, Y.V. *Comparative performance of arima and ARCH/GARCH models on time series of daily equity prices for large companies.*

Tharmmaphornphilas, W., Lohasiriwat, H. and Vannasetta, P.(2012), Gold price modeling using system dynamics, *Engineering Journal 16, Issue.5.*

Tripathi, L.K., Parashar, A. and Singh, R. (2014), Global factors and gold price in India causal study, *International Journal of Advanced Research in Management and Social Sciences 3, No.7.*

Woodward, W.A., Gray, H.L. and Elliott, A.C. (2003), Applied time series analysis, *Pearson Education Pte, Ltd. Indian Branch.*

Perera, W.T.K, Impact of Petroleum Prices on Inflation in Sri Lanka, Staff Study, Central Bank of Sri Lanka.

Yaffee, R.A. and Mc Gee, M.(2003), Introduction to time series analysis and forecasting, with applications of SAS and SPSS, *Academic Press, INC.*

Appendix A

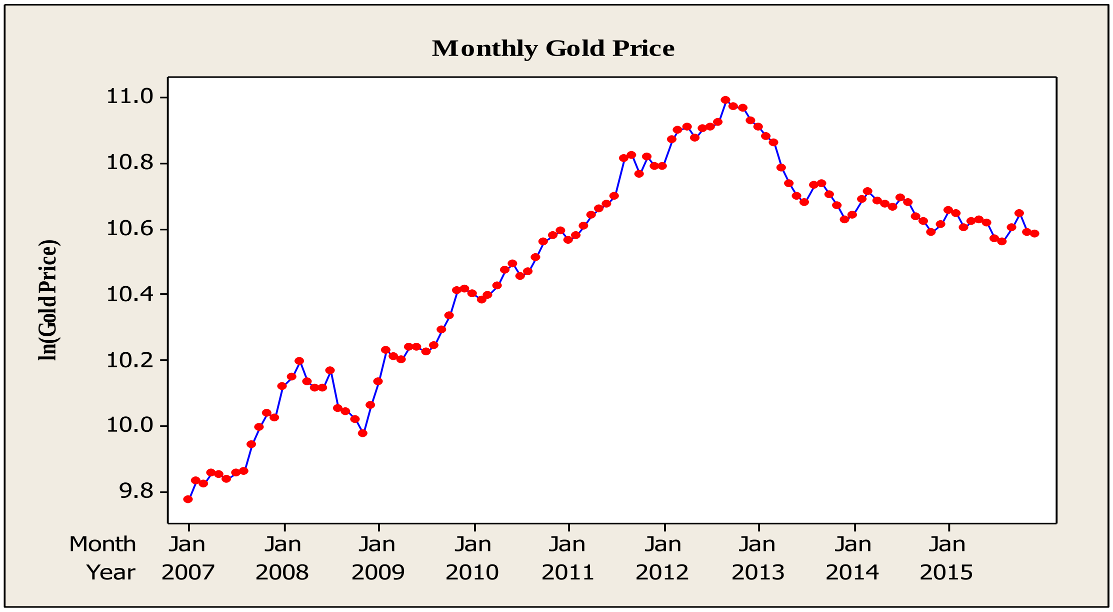
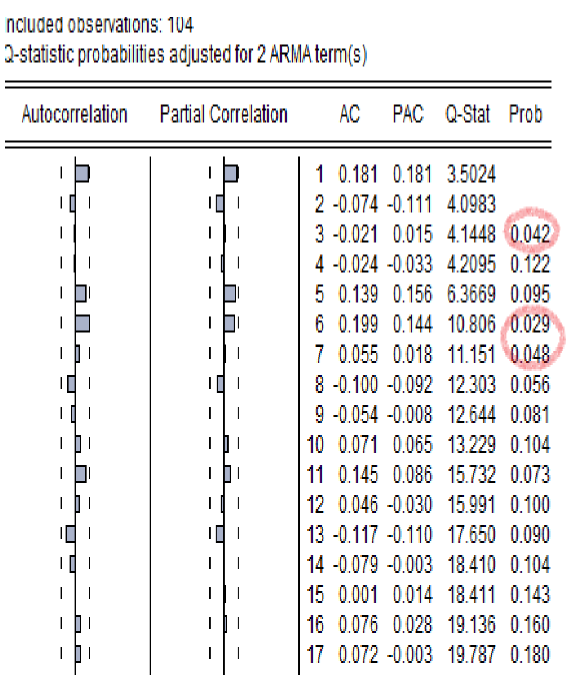


Figure A1: Log value of monthly gold price



ncluded observations: 104
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.181	0.181	3.5024	
		2	-0.074	-0.111	4.0983	
		3	-0.021	0.015	4.1448	0.042
		4	-0.024	-0.033	4.2095	0.122
		5	0.139	0.156	6.3669	0.095
		6	0.199	0.144	10.806	0.029
		7	0.055	0.018	11.151	0.048
		8	-0.100	-0.092	12.303	0.056
		9	-0.054	-0.008	12.644	0.081
		10	0.071	0.065	13.229	0.104
		11	0.145	0.086	15.732	0.073
		12	0.046	-0.030	15.991	0.100
		13	-0.117	-0.110	17.650	0.090
		14	-0.079	-0.003	18.410	0.104
		15	0.001	0.014	18.411	0.143
		16	0.076	0.028	19.136	0.160
		17	0.072	-0.003	19.787	0.180

Figure A2: Correlogram of $D[GOLD\ PRICE]$

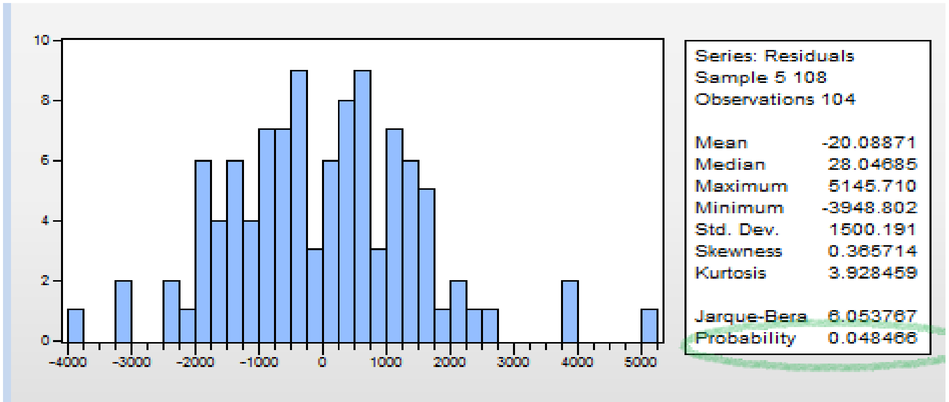


Figure A3: The histogram of residuals of $D[GOLD\ PRICE]$

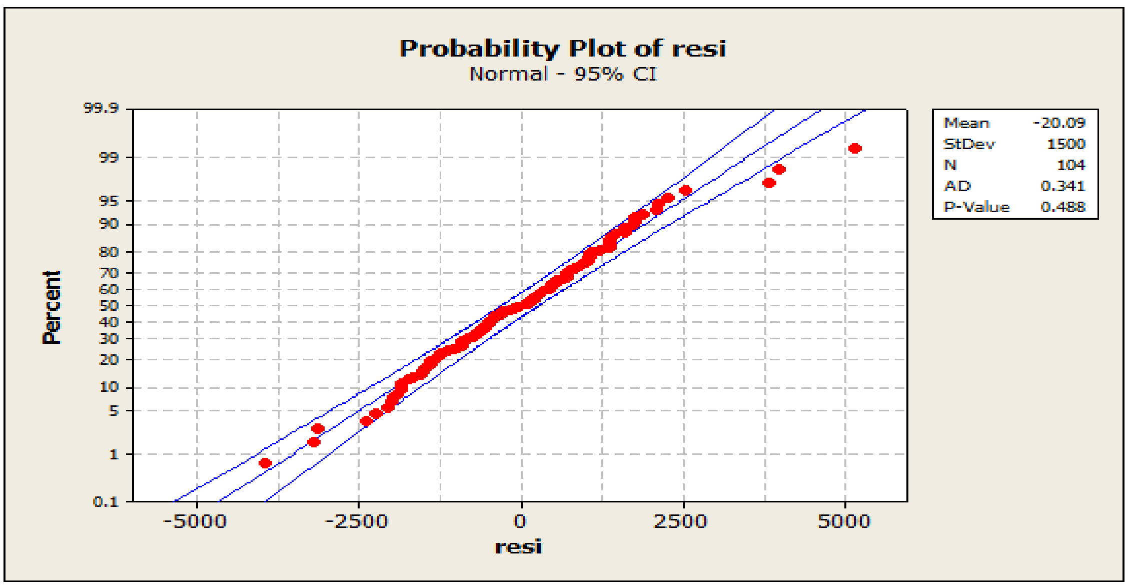
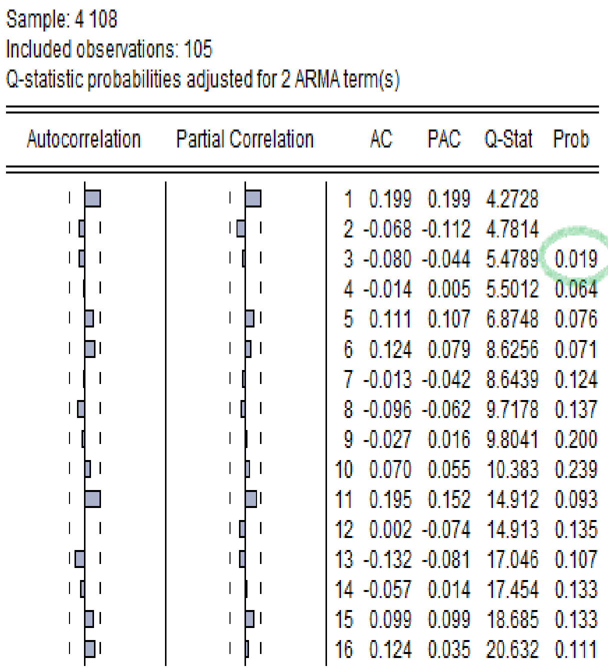


Figure A4: Normal probability plot of residuals of $D[GOLD\ PRICE]$



Sample: 4 108
Included observations: 105
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.199	0.199	4.2728	
		2	-0.068	-0.112	4.7814	
		3	-0.080	-0.044	5.4789	0.019
		4	-0.014	0.005	5.5012	0.064
		5	0.111	0.107	6.8748	0.076
		6	0.124	0.079	8.6256	0.071
		7	-0.013	-0.042	8.6439	0.124
		8	-0.096	-0.062	9.7178	0.137
		9	-0.027	0.016	9.8041	0.200
		10	0.070	0.055	10.383	0.239
		11	0.195	0.152	14.912	0.093
		12	0.002	-0.074	14.913	0.135
		13	-0.132	-0.081	17.046	0.107
		14	-0.057	0.014	17.454	0.133
		15	0.099	0.099	18.685	0.133
		16	0.124	0.035	20.632	0.111

Figure A5: Correlogram of $D[Ln[GOLD\ PRICE]]$- Model 1

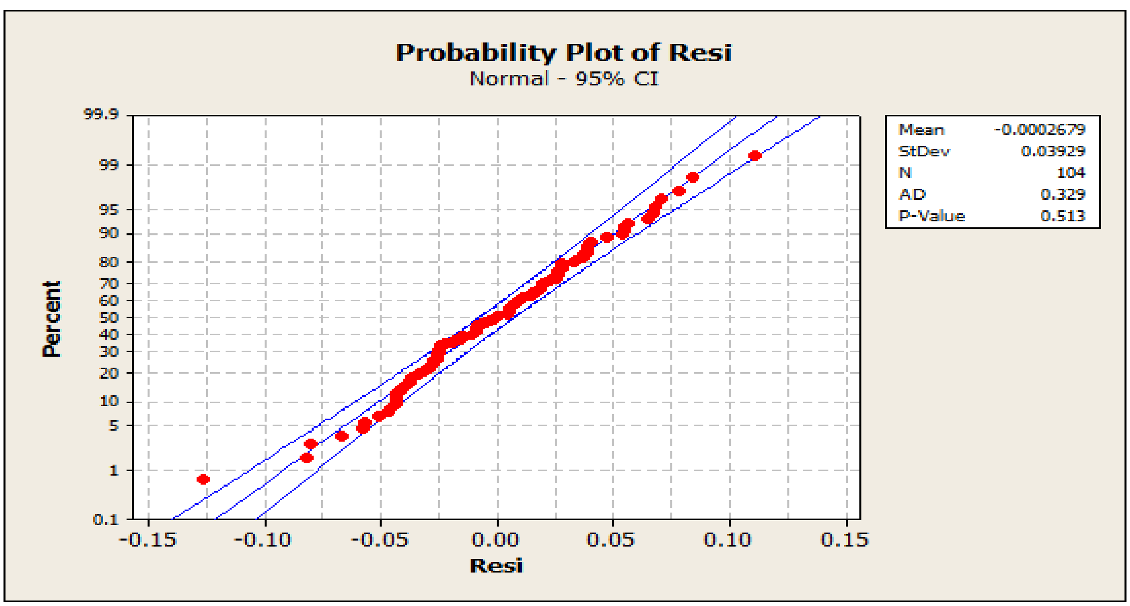


Figure A6: Normal probability plot of residuals of $D[Ln[GOLD\ PRICE]]$-Model 1

Appendix B

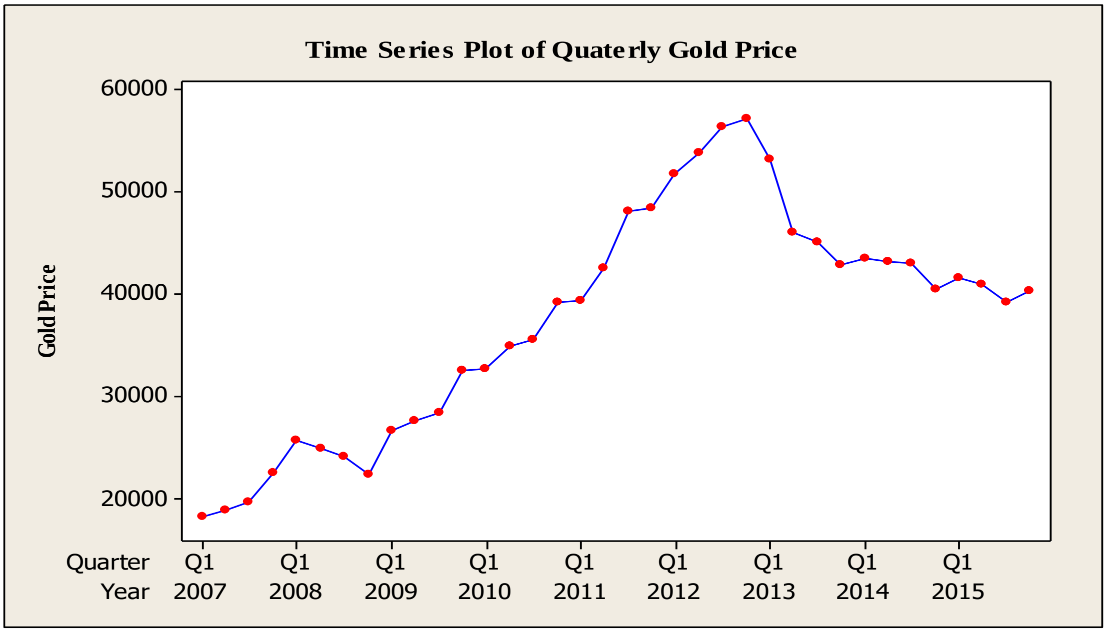


Figure:B1 Time Series plot of quarterly gold price from 2007, January to 2015, December

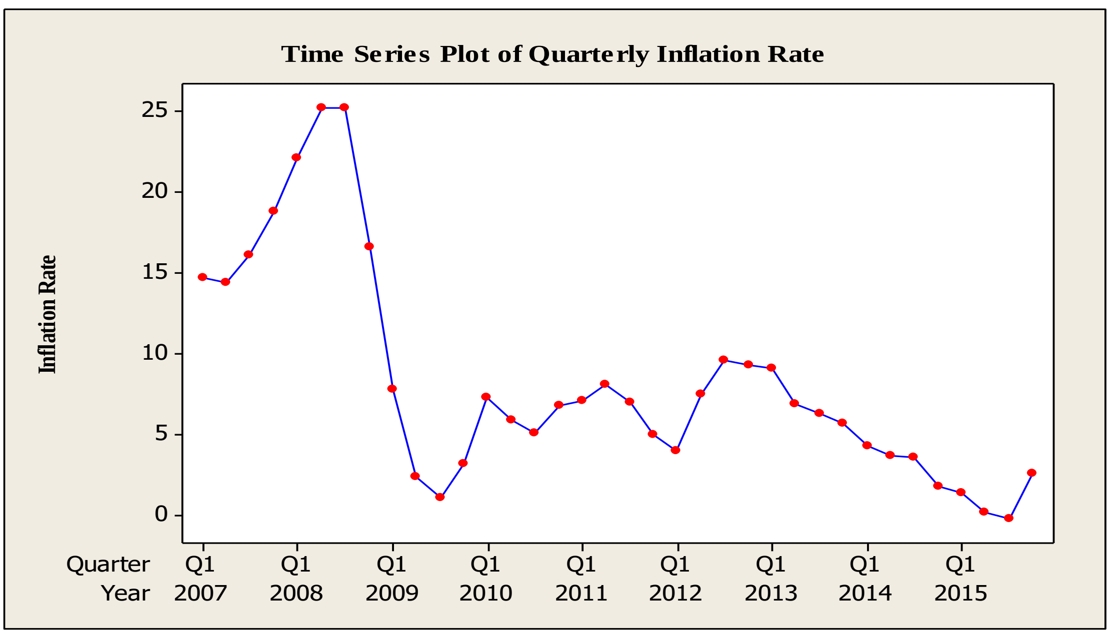


Figure:B2 Time Series plot of quarterly of inflation rate from 2007, January to 2015, December

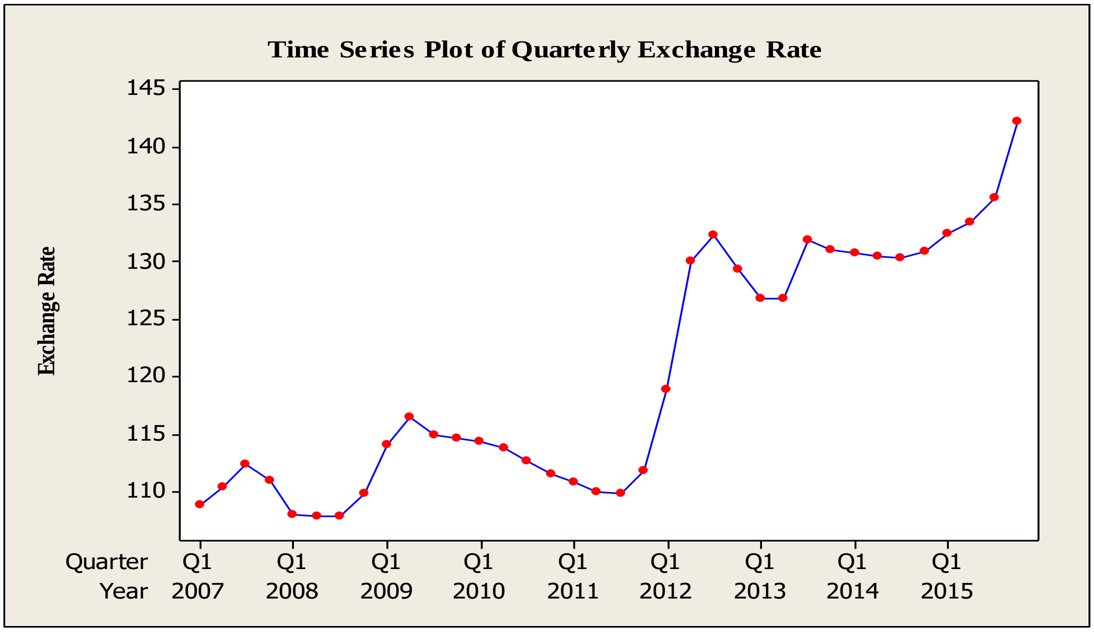


Figure:B3 Time Series plot of quarter values of exchange rate from 2007, January to 2015, December

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-19731.58	22630.05	-0.871920	0.3896
EX	504.9510	176.5174	2.860629	0.0073
IR	-414.7010	272.3866	-1.522472	0.1374

R-squared	0.422910	Mean dependent var	37515.66
Adjusted R-squared	0.387935	S.D. dependent var	11143.27
S.E. of regression	8717.890	Akaike info criterion	21.06380
Sum squared resid	2.51E+09	Schwarz criterion	21.19576
Log likelihood	-376.1484	Hannan-Quinn criter.	21.10985
F-statistic	12.09173	Durbin-Watson stat	0.136598
Prob(F-statistic)	0.000115		

Figure:B4 Regression model results for quarterly values of three variables

Appendix C

Critical Values of the Durbin–Watson Statistic

Sample Size	Probability in Lower Tail (Significance Level = α)	k = Number of Regressors (Excluding the Intercept)									
		1		2		3		4		5	
		d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
	0.01	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
15	0.025	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
	0.05	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
	0.01	0.95	1.15	0.86	1.27	0.77	1.41	0.63	1.57	0.60	1.74
20	0.025	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
	0.05	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
	0.01	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
25	0.025	1.13	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
	0.05	1.20	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
	0.01	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
30	0.025	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
	0.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
	0.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
40	0.025	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
	0.05	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
	0.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
50	0.025	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
	0.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
	0.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
60	0.025	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
	0.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
	0.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
80	0.025	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
	0.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
	0.01	1.52	1.56	1.50	1.58	1.48	1.60	1.45	1.63	1.44	1.65
100	0.025	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72
	0.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78