MODIFIED BOX-JENKINS AND GARCH FOR FORECASTING HIGHLY VOLATILE TIME SERIES DATA

SITI ROSLINDAR BINTI YAZIZ

DOCTOR OF PHILOSOPHY

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : SITI ROSLINDAR BINTI YAZIZ

Date of Birth : 02 JUNE 1979

Title : MODIFIED BOX-JENKINS AND GARCH FOR FORECASTING HIGHLY VOLATILE TIME SERIES DATA

Academic Session : SEMESTER II 2018/2019

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organisation where research was done)*
☑	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Universiti Malaysia Pahang reserves the following rights:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies of the thesis for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

_______________________ (Student's Signature)

_______________________ (Supervisor's Signature)

790602-03-5630
Date: May 2019

Assoc. Prof. Dr. Roslinazairimah Binti Zakaria
Date: May 2019

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Doctor of Philosophy.

____

(Supervisor's Signature)

Full Name : DR. ROSLINAZAIRIMAH BINTI ZAKARIA

Position : ASSOCIATE PROFESSOR

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

____

(Student's Signature)

Full Name : SITI ROSLINDAR BINTI YAZIZ

ID Number : PSS12001

Date :

MODIFIED BOX-JENKINS AND GARCH FOR FORECASTING HIGHLY VOLATILE TIME SERIES DATA

SITI ROSLINDAR BINTI YAZIZ

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy

Faculty of Industrial Sciences & Technology

UNIVERSITI MALAYSIA PAHANG

MAY 2019

ACKNOWLEDGEMENTS

Alhamdulillah, I am grateful to Allah The Almighty for conferring me the passion, strength, health and the patience to finally complete this thesis.

My deepest appreciation and gratitude to my supervisors, Dr. Roslinazairimah Zakaria (main supervisor) and Assoc. Prof. Dr. Maizah Hura Ahmad (field supervisor-UTM) for enlightening me through the art of writing critically, motivating me, their guidance and continuous support, valuable suggestions and perhaps critically, their patience. Also thanks to Prof. Dr. Noor Azlinna Azizan (co-supervisor) for her marginal contribution.

Special thanks go to Prof. John Boland and Dr. Manju Agrawal from the University of South Australia (UniSA) who gave me their guidance, specifically on the theoretical part, during my 2-month attachment study. I am greatly indebted to them both for the knowledge and ideas as well as countless support during my presence there.

I also would like to express my deepest gratitude and appreciation to Dr. Suhartono from Institut Teknologi Sepuluh Nopember (ITS), Indonesia for his contribution of ideas, significant assistance and continuous support.

I thank faculty members, Prof. Dr. Mashitah Mohd Yusoff, Assoc. Prof. Dr Hasbi Ab. Rahim, Assoc. Prof. Dr. Mohd Zuki Salleh and all Mathematics staff who granted me the study leave and tolerated my teaching workload. I also wish to express my appreciation to all my friends who helped me directly and indirectly.

I would also like to extend my thanks to the Ministry of Higher Education (Skim Latihan Akademik Bumiputera, SLAB) and UMP research grants (RDU130369 and RDU1703198) for the financial support.

Last but not least, my gratitude goes to my loving family especially to my beloved husband, Dr. Izan Izwan Misnon and our children, Siti Sarah and Luqman Hakim for their extreme understanding, support and encouragement.

ABSTRAK

Model Box-Jenkins digunakan secara meluas sama ada sebagai model peramalan, piawaian atau bersepadu dalam kajian terkini siri masa. Pemodelan Box-Jenkins adalah salah satu teknik peramalan paling berkuasa yang digunakan dalam praktis kajian analisis siri masa. Kebanyakan data siri masa contohnya data ekonomi dan sains persekitaran adalah bervarians tidak malar secara semulajadi. Walau bagaimanapun, untuk data siri masa yang bervarians tidak malar yang tinggi, model Box-Jenkins adalah tidak sesuai untuk diaplikasikan kerana andaian ralat varians malar tidak dipenuhi dan ia juga tidak dapat mengendalikan sifat heteroskedastisiti. Menggabungkan model Box-Jenkins dengan model stokastik heteroskedastisiti seperti model *generalised autoregressive conditional heteroscedastic* (GARCH) merupakan satu kaedah yang berkesan untuk mengatasi kekangan model Box-Jenkins bagi data varians tidak malar. Kajian ini menilai prestasi model kombinasi antara Box-Jenkins dan variasi GARCH dalam pemodelan dan peramalan data univariat siri masa yang bervarians tidak malar yang tinggi dengan pemodelan Box-Jenkins sebagai asas prosedur. Empat prosedur dicadangkan dalam kajian ini dalam menilai prestasi model kombinasi tersebut di mana tiga cadangan prosedur awal adalah menggunakan model Box-Jenkins dengan standard GARCH (or BJ-G). Prosedur cadangan pertama adalah berdasarkan prosedur asas Box-Jenkins dan ia digunakan sebagai kajian tinjauan awal. Prosedur cadangan kedua adalah berdasarkan prosedur cadangan pertama yang difokuskan untuk mengendalikan data siri masa bervarians tidak malar yang tinggi secara spesifik, menggunakan model BJ-G dengan penekanan kepada pengecaman ciri data bervarians tidak malar yang tinggi pada peringkat awal. Manakala, prosedur cadangan ketiga adalah lanjutan daripada prosedur cadangan kedua, yang digunakan untuk menilai keupayaan model BJ-G untuk peramalan jangka panjang. Prosedur cadangan keempat adalah kombinasi prosedur cadangan kedua dan ketiga yang mana ia merupakan prosedur komprehensif untuk pemodelan dan peramalan data siri masa yang bervarians tidak malar yang tinggi menggunakan model Box-Jenkins – variasi GARCH. Kesemua prosedur cadangan diilustrasikan dengan data harian harga emas dunia kerana data ini adalah data siri masa yang bervarians tidak malar yang tinggi. Berdasarkan kajian awal ke atas 5000 data harian data harian harga emas mengunakan prosedur cadangan pertama BJ-G, nilai ralat yang kecil membuktikan model BJ-G adalah model yang diyakini untuk pemodelan dan peramalan data bervarians tidak malar yang tinggi. Keputusan empirik daripada data harian harga emas dunia menggunakan prosedur cadangan kedua menyatakan prosedur ini adalah lebih praktikal berbanding prosedur cadangan pertama dalam pemodelan data bervarians tidak malar yang tinggi menggunakan model BJ-G dan secara langsung dapat menentukan bilangan data yang optimal. Keputusan empirik mencadangkan 25% daripada data yang terkini atau 1250 data adalah mencukupi untuk model BJ-G dengan prestasi peramalan yang sama seperti menggunakan kesemua data. Manakala, berdasarkan kajian empirik ke atas 1250 data harian harga emas itu menggunakan prosedur cadangan ketiga, didapati model BJ-G berkeupayaan untuk mengikuti pola data sebenar sehingga tujuh hari ke hadapan, khasnya dalam selang peramalan 95%. Prosedur cadangan keempat diuji ke atas model Box-Jenkins dengan variasi GARCH menggunakan data siri yang sama digunakan untuk prosedur cadangan ketiga. Sebagai kesimpulan, model kombinasi Box-Jenkins dan variasi GARCH mempunyai potensi yang besar, oleh itu prosedur cadangan keempat BJ-G memberikan satu prosedur peramalan siri masa yang komprehensif, sistematik dan praktikal bagi data siri masa yang bervarians tidak malar yang tinggi.

ABSTRACT

The Box-Jenkins model has widely been used either as the forecasting, benchmarking or as the integrated model in the current research of time series. The Box-Jenkins modelling is one of the most powerful forecasting techniques available in research practice of the time series analysis. Most of the time series data such as in economics and in environmental sciences are volatile in nature. However, for a highly volatile time series data, the Box-Jenkins model is inappropriate to be applied since it violates the errors assumption of constant variance and it is not able to handle the heteroscedasticity property. Combining the model with a heteroscedastic stochastic model such as the generalised autoregressive conditional heteroscedastic model (GARCH) can be an effective way to overcome the limitation of the Box-Jenkins model in handling the non-constant variance. This study evaluates the performance of the combination model of Box-Jenkins and GARCH-type in modelling and forecasting univariate highly volatile time series data with Box-Jenkins modelling as the base procedure. In evaluating the performance of the model, four procedures are proposed in this study where the first three procedures are using the model of Box-Jenkins and standard GARCH (or BJ-G). The first proposed procedure is developed based on the theoretical Box-Jenkins's procedure and it is used for the preliminary study. The second proposed procedure is developed based on the first proposed procedure to focus on handling the highly volatile time series data specifically, using BJ-G model by emphasizing on the identification of highly volatile characteristic in the data at the early stage. While the third proposed procedure is an extension from the second procedure, which evaluates the multistep ahead forecasting performance of the BJ-G model. The fourth procedure of BJ-G is developed from the second and third procedures and it is a comprehensive procedure for modelling and forecasting highly volatile time series data using Box-Jenkins – GARCH-type model. The proposed procedures are illustrated using the daily world gold price data since it is a highly volatile type of time series. Based on the preliminary study on 5000 world daily gold price data set using the first procedure of BJ-G, the small magnitude of error proves that BJ-G is a reliable model in modelling and forecasting highly volatile data. The empirical results of the world daily gold price using the second proposed procedure indicate that the procedure is more practical than the first propose procedure to be used in modelling a univariate highly volatile data using BJ-G model which simultaneously ensures an optimal number of data in dealing with the model. The empirical results suggested that the latest 25% of data or 1250 data is sufficient to be employed using BJ-G model with similar forecasting performance as by using all data. Meanwhile, based on the empirical results on the 1250 world daily gold prices and by employing the third procedure, it is found that the BJ-G model is able to follow the trend of the actual data up to seven days ahead, specifically within 95% prediction interval. The fourth proposed procedure is also tested on the Box-Jenkins with various GARCH-type models using the same data series as in the third proposed procedure. In conclusion, the combination model of Box-Jenkins and GARCH-type has great potential, thus the fourth proposed procedure of BJ-G provides a comprehensive, systematic and practical procedure of time series forecasting for univariate highly volatile time series data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

y_t	The original series or the observed data at time *t* period
y_t^*	The transformed data at time *t*
s_t	The stationary series at time *t*
h	The forecasting horizon
$\hat{y}_{T+h}$	The forecast data for *h*-step ahead
$\hat{s}_{T+h}$	The simulated stationary series for forecasting horizon h
T	The number of data in-sample series; the origin
n	The number of data in out-of-sample series
k	The number of lag
$k_{\max}$	The maximum number for lag *k*
w	The number of series
m	The number of auxiliary regressors
μ	The mean data (or model)
μ_t	The conditional mean of s_t
σ^2	The variance data (or model)
σ_X^2	The variance of X
σ_t^2	The conditional variance of y_t
σ_t	The volatility of a_t
a_t	The random errors at time *t* period
$\{\hat{a}_t\}$	The residuals of data
$\hat{a}_t$	The residual at time *t*
v_t	The error in the squared returns, that is $a_t^2 - \sigma_t^2$
ε_t	The standardised error (innovations) of model
$e_T(h)$	The *h*-step ahead forecast error at origin T
γ_k	Autocovariance coefficient at lag *k*
$\{\gamma_k\}$	The plot of γ_k versus lag *k* or the autocovariance function
ρ_k	Autocorrelation coefficient at lag *k*

$\{\rho_k\}$	The plot of ρ_k versus lag k or the autocorrelation function
r_k	The sample ρ_k
ϕ_{kk}	Partial autocorrelation coefficient at lag k
r_{kk}	The sample ϕ_{kk}
p	The order of the autoregressive model
q	The order of the moving average model
φ_p	The autoregressive parameters with order p
θ_q	The moving average parameters with order q
d	The order of differencing
P	The order of seasonal autoregressive
Q	The order of the seasonal moving average
D	The order of seasonal differencing
S	The seasonal period
B	The backward shift operator
r	The order of the generalised autoregressive conditional heteroskedastic model
s	The order of the autoregressive conditional heteroskedastic model
α_i	The coefficient of the parameters ARCH
β_i	The coefficient of the parameters GARCH
c	The constant
C_i	The coefficient for the Δy_{t-i}
X	The continuous random variable
$f(x)$	The probability density function of X
$K(x)$	The kurtosis of X
$S(x)$	The skewness of X
m'_ℓ	The ℓth moment of a continuous random variable X about the origin
m_ℓ	The ℓth central moment of X about the mean
λ	The minimum residual mean square error value
$\hat{\lambda}$	The estimated λ
$J(\lambda; y)$	The Jacobian of the transformation

$S(\lambda)$	The residual sum of squares in the analysis of variance of y_t^*
υ_λ	The number of independent components in λ
v	The degrees of freedom
ξ	The skewness parameter
κ	The shape parameter
$\Gamma(\cdot)$	The gamma function
a	The $(T \times T)$ matrix
θ	The $(T \times 1)$ vector of unknown parameters associated with the transformed data
Γ_T	The covariance matrix of symmetric form
P_T	The autocorrelation matrix
L	The likelihood function
$\ln L$	The log likelihood function
$\ln L_{\max}$	The maximised log likelihood function
Z	The integers
x_t'	The deterministic time trend
ϕ_a	The parameter that defines the relationship between successive values of a_t and a_{t-1}
ϕ_{df}	The parameter to be estimated in DF-test and ADF-test
π	The parameter to be estimated in DF-test and ADF-test where $\pi = \phi - 1$
$\hat{b}$	The estimated coefficient for ARCH-LM
TD_t	The deterministic terms
M	The risk premium parameter
N_{t-i}	The indicator for negative a_{t-i}
g_i	The leverage effect term of a_{t-i}
δ	The positive real number

LIST OF ABBREVIATIONS

AR	Autoregressive model
MA	Moving average model
ARMA	Autoregressive moving average model
ARIMA	Autoregressive integrated moving average model
SARIMA	Seasonal autoregressive integrated moving average model
ACF	Autocorrelation function
PACF	Partial autocorrelation function
MLE	Maximum likelihood estimation
OLS	Ordinary least squares
ARCH	Autoregressive conditional heteroskedastic
GARCH	Generalised autoregressive conditional heteroskedastic
ARCH LM	ARCH Lagrange Multiplier
EACF	Extended autocorrelation function
IID	Independent identically distributed
NID	Normal independently distributed
pdf	Probability density function
AIC	Akaike Information Criteria
SIC	Schwarz Information Criterion
ADF	Augmented Dickey-Fuller
SSR	Residual sum of squares
dof	Degrees of freedom
MAE	Mean absolute error
MSE	Mean square error
RMSE	Root mean square error
MAPE	Mean absolute percentage error
PIs	Prediction intervals
CV	Cross-validation
LBQ-test	Ljung-Box *Q*-test
DW-test	Durbin-Watson test
JB-test	Jarque-Bera test

CHAPTER 1

INTRODUCTION

1.1 Background of Study

A time series is a set of sequential observations with respect to time. Hence, time series analysis is the study about the data collected through time. Many data sets appear as time series, such as commodity price, exchange rate, electricity load demand and weekly rainfall data. Therefore, time series analysis covers various fields of study that attracts high interest among the researchers since the 1960s. The development of four-stage iterative procedure of time series i.e. identification, estimation, diagnostic checking and forecasting developed by Box and Jenkins (1968), known as the Box-Jenkins modelling, is considered as the catalyst for the time series research (De Gooijer & Hyndman, 2006). The Box-Jenkins modelling is one of the most powerful forecasting methods available in research practice of the time series analysis.

A forecasting method is a procedure for computing forecasts from present and past values. Forecasting methods can be broadly classified into three, which are judgemental forecasts, univariate methods and multivariate methods, and can be a combination of more than one of the three methods. The judgemental forecasts are based on subjective judgement, intuition, commercial knowledge and any other relevant information. The Delphi technique is one of the famous judgemental methods. The univariate and the multivariate methods are the statistical-based method. According to Chatfield (2001), the statistical methods tend to be superior than judgemental methods in general.

The univariate methods deal with forecasts, which depend only on present and past values of the single series, while the multivariate methods deal with forecasts of a given variable depend, at least partly, on values of one or more additional time series variables. The multivariate models include multiple regression, transfer function and distributed lag models, econometric models and multivariate versions of autoregressive (AR) and autoregressive moving average (ARMA) models, including vector autoregressive (VAR) models. This study only focuses on the statistical method, specifically univariate methods, and does not attempt to cover judgemental forecasting and multivariate forecasting.

The study is limited to the univariate method and model as it is very useful for many purposes, including forecasting large number of series and providing a benchmark in comparative forecasting studies (Chatfield, 2001; De Gooijer & Hyndman, 2006). Figure 1.1 presents the overview of all related univariate time series methodology and the corresponding models, which are graphically drawn using freemind software from the input of the studies of Bisson and Gurpinar (2017), Hyndman and Athanasopoulos (2014), Adhikari and Agrawal (2013), Tsay (2013), Box, Jenkins and Reinsel (2008), De Gooijer and Hyndman (2006) and Chatfield (2001). The Box-Jenkins modelling and its models are highlighted in the figure.

The practicality of the Box-Jenkins modelling and its good performance in analysing time series data makes the Box-Jenkins model continuously considered as the forecasting, the benchmark or as the integrated model in current research. Furthermore, the popularity of this model is also due to its capability to analyse almost any set of time series data either for profit or non-profit applications (Christodoulos, Michalakelis & Varoutas, 2010). Some of its profit applications, for example in business and economics, the model is extensively used in exchange rate forecasting (Allen & Taylor, 1990; Giddy & Dufey, 1975; Khashei & Bijari, 2010; Singh & Jain, 2018; Zhang, 2003) and commodity prices (Darekar & Reddy, 2017; Ho, Xie & Goh, 2002; Yaziz, Ahmad, Nian & Muhammad, 2011). For non-profit application, such as in environmental sciences area, the Box-Jenkins model has been applied in ozone concentrations (Awang, Kar Yong & Yin Hoeng, 2017; G. Liu, Tarasick, Fioletov, Sioris & Rochon, 2009; Robeson & Steyn, 1990), air quality (Polydoras, Anagnostopoulos & Bergeles, 1998; Taneja, Ahmad, Ahmad & Attri, 2016) and hydrological study (Castellano-Méndez, González-Manteiga, Febrero-Bande, Manuel Prada-Sánchez & Lozano-Calderón, 2004; Fouli, Fouli, Bashir & Loni, 2017).

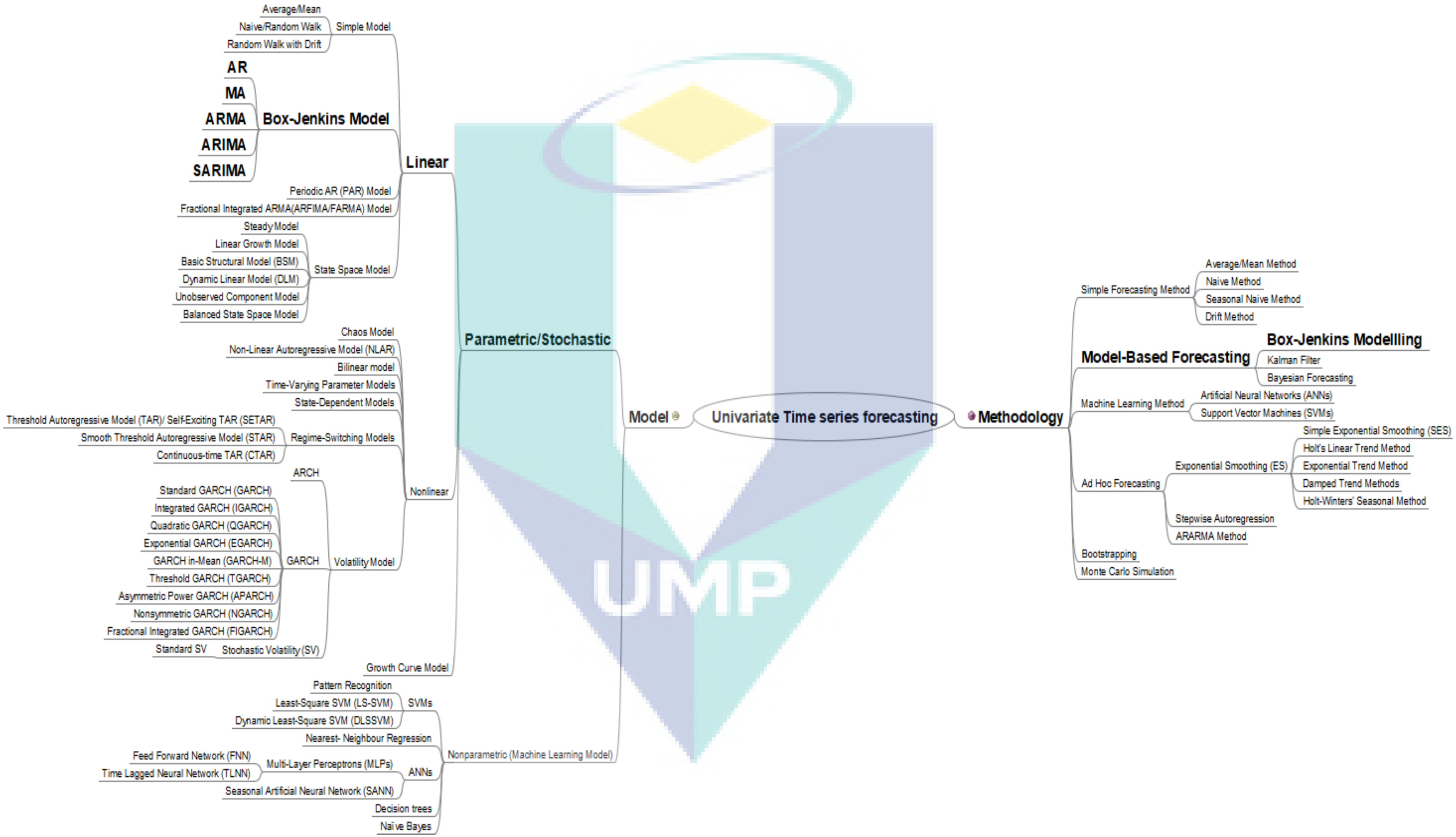


Figure 1.1 Overview of methodologies and models in univariate time series forecasting
Source: Bisson and Gurpinar (2017), Hyndman and Athanasopoulos (2014), Adhikari and Agrawal (2013), Tsay (2013), Box et al. (2008), De Gooijer and Hyndman (2006) and Chatfield (2001).

Most of the time series data are volatile in nature where the data varies over time. If the volatility in a data series is low, the Box-Jenkins model is appropriate as it assumes that the variance of the errors is constant, known as homoscedasticity property. However, for highly volatile data, the variance for errors is non-constant and the Box-Jenkins model is found inappropriate since it violates the errors assumption of constant variance. The characteristic of non-constant errors in variance is known as heteroscedasticity or autoregressive conditional heteroscedastic (ARCH) effects.

The ARCH effects in a highly volatile data are commonly seen in economics and financial data. The highly volatile characteristic initially detected using time series plot where it shows large variation and the volatility cluster (i.e. data is high for certain time period and low for certain time period) in the plot. However, it is hard to detect accurately the ARCH effects in a data series using graphical presentation. Hence, a statistical test namely the heteroscedasticity test, is needed to confirm the highly volatile characteristic in a time series data. Therefore, if the decision of the test rejects the null hypothesis of no ARCH effect in the residuals of the model, then the series is classified as a highly volatile data.

Hence, the Box-Jenkins model should not be applied to highly volatile data since it fails to handle the heteroscedasticity property that is present in the data series. Therefore, if a study wants to use the Box-Jenkins model to analyse a highly volatile data because of its good reputation in research practice, then a modification on the model needs to be done. Combining the model with a heteroscedastic stochastic model or a well-known volatility model can be an effective way to overcome the limitations of the Box-Jenkins model in handling ARCH effects in the data series.

Previous studies have shown that generalised autoregressive conditional heteroscedastic (GARCH)-type model is widely applied to handle volatility in a data series (S. Hammoudeh & Yuan, 2008; Qadan & Yagil, 2012; Trück & Liang, 2012). In recent years, many studies proposed the incorporation of GARCH-type model into the Box-Jenkins model due to its good performance in dealing with highly volatile data. Some of the studies that incorporate the Box-Jenkins model with GARCH-type are ARIMA-GARCH (Babu & Reddy, 2015; Chen, Hu, Meng & Zhang, 2011; Girish, 2016; Liu & Shi, 2013; Loi & Ng, 2018; Tan, Zhang, Wang & Xu, 2010; Zhou, He & Sun, 2006), AR-EGARCH (Ahmed, 2017; Ferenstein & Gasowski, 2004; Girish, 2016; Walid, Chaker, Masood & Fry, 2011), AR-GARCH (Ferenstein & Gasowski, 2004; Gaglianone & Marins, 2017; Harrison & Paton, 2004; Sohn & Lim, 2007), ARIMA-PARCH (Girish, 2016), ARIMA-TGARCH (Ahmad, Ping, Yaziz & Miswan, 2015; Freedi, Shamiri & Isa, 2012), ARMA-GARCH (Liu & Shi, 2013; Pham & Yang, 2010; Wang, Gelder, Vrijling & Ma, 2005), ARMA-EGARCH (Ord, Koehler, Snyder & Hyndman, 2009) and ARIMA-GARCH-M (Liu, Erdem & Shi, 2011; Liu & Shi, 2013; Liu, Shi & Qu, 2013). An extensive discussion on the model of Box-Jenkins with GARCH-type is provided in Chapter 2.

Although these studies obtain promising results by applying the Box-Jenkins model with GARCH-type, there is no study that focuses on the development of procedure or procedure on the combination model of Box-Jenkins and GARCH-type to deal with highly volatile data. Hence, this motivates a study to develop a procedure in modelling and forecasting a highly volatile time series data with the Box-Jenkins as the base model by incorporating GARCH-type model to capture the heteroscedasticity in the data series.

However, this study focuses on the standard GARCH at first, or simply called as GARCH, due to its popularity and parcimonious characteristics, in developing a basic procedure of the combination model of Box-Jenkins and GARCH-type. The procedure is then applied to other GARCH-type models that work practically for highly volatile data. The combination model of Box-Jenkins and GARCH (or BJ-G) has great potential, thus the proposed procedure in this study would give significant procedural contribution to the basis for research that deals with highly volatile data. Not only that, the development of procedure in this study will also provide useful guidelines for using the combination model to address highly volatile data which is in line with big data analytics and support the 4th industrial revolution (IR 4.0).

1.2 Problem Statement

Improving forecasting method is one of the main issues in time series research. Therefore, the research continues to improve the effectiveness of the forecasting models. According to Chatfield (2001), the forecasting literatures concentrate on how to implement particular forecasting method, whereas most forecasters probably need much more help with the strategy of forecasting. There is plenty of software available to make it easy to fit the model to the data series, however it is still hard to decide when to use the model and how to choose the appropriate model for the data series. The recent publications in time series such as Loi and Ng (2018), Ahmed (2017), Hyndman and Athanasopoulos (2017) and Tsay (2013) are still lacking in demonstrating a clear strategy of forecasting. Therefore, this study would present a comprehensive strategy of forecasting in the form of a procedure, specifically for forecasting highly volatile time series data, which will assist the researcher in forecasting. The proposed procedure will complement the existing procedure in time series modelling and forecasting, specifically the Box-Jenkins modelling.

One of the established time series methods in many research practices is the Box-Jenkins modelling (De Gooijer & Hyndman, 2006; Christodoulos, Michalakelis & Varoutas, 2010). In this study, the focus is modelling and forecasting the univariate highly volatile time series data by applying the Box-Jenkins as the base model. It is vital for a model to be able to analyse and predict data which reflects the data series pattern. Therefore, in this study, the GARCH model is considered to be incorporated with the Box-Jenkins model due to its capability in handling heteroscedasticity in the data series.

Hence, the main issue in this study is how to develop an appropriate procedure in modelling and forecasting a univariate highly volatile time series data using Box-Jenkins – GARCH (BJ-G) model, or simply called as procedure of BJ-G. The proposed procedure is based on the standard Box-Jenkins's procedure, which consists of four stages. This proposed procedure will be used to justify and evaluate the performance of the BJ-G model in analysing and forecasting (at one-step ahead) the data series.

If the first proposed procedure of BJ-G has shown promising results, then the next issue that needs to be considered is how to develop a procedure of BJ-G model in handling univariate highly volatile time series data specifically. This proposed procedure of BJ-G emphasizes on the identification of highly volatile characteristics in the data at the early stage before further analysis is conducted. Therefore, the first proposed procedure should be improved by introducing the steps of heteroscedasticity test and the BJ-G model identification in Stage I of the second proposed procedure of BJ-G instead of in Stage III in the first proposed procedure. The second proposed procedure would simultaneously ensure the optimal number of data required for practical application in handling univariate highly volatile time series data using BJ-G model. Determination of the optimal number of data using a statistical model for practical application is one of the main issues in time series forecasting (Chatfield,2001; Hyndman & Athanasopoulos, 2014; Hyndman & Konstenko, 2017).

However, the second proposed procedure of BJ-G is only applied for one-step ahead forecasting performance, which is not practical for real data due to its limitation of the prediction period (Babu & Reddy, 2015; Pham & Yang, 2010; Byström, 2005). Therefore, the third issue that needs to be considered in this study is how to develop a procedure of BJ-G that can be used for multistep forecasting. It is observed that many statistical software only provides the analysis and forecasting results for one-step ahead forecast. Hence, in evaluating the forecasting performance for multistep ahead using BJ-G model, a set of codes using programming language needs to be constructed to analyse the data up to *n*-step ahead forecasting. The codes will be associated with the third proposed procedure of BJ-G.

Since the combination model of Box-Jenkins and GARCH has great potential for research that deals with univariate highly volatile time series data, the comprehensive procedure of BJ-G is considered in the study. Therefore, the fourth procedure is developed from the second and third procedures. The fourth procedure is then applied to all GARCH-type models related to highly volatile data. Hence, this study proposes a comprehensive procedure using Box-Jenkins with GARCH-type model in improving forecasting method specifically for forecasting univariate highly volatile time series data.

1.3 Objectives of the Study

The objectives of this study are:

1. To propose a procedure using Box-Jenkins as the base model in modelling and forecasting univariate highly volatile time series data.
2. To propose a procedure of BJ-G in specifically handling univariate highly volatile time series data.
3. To propose a procedure of BJ-G in evaluating the multistep forecasting performance for the univariate highly volatile time series data.
4. To propose a comprehensive procedure of BJ-G for all GARCH-type models in modelling and forecasting univariate highly volatile time series data.

1.4 Scope of the Study

This study focuses on parametric-based models, namely Box-Jenkins and GARCH in modelling and forecasting highly volatile time series data. These two models are combined to form a BJ-G model, which is a combination of both linear and nonlinear models. The data used in this study is the world daily gold price obtained from www.kitco.com which is reliable in the market. The data in this website is used by many related companies and in research articles. Scope of this research is to undertake a comprehensive investigation on the proposed procedure of the BJ-G model in forecasting univariate highly volatile time series data only. Itemised research activities of each objective are as follows:

1. To propose a procedure as Box-Jenkins as the base model in modelling and forecasting univariate highly volatile time series data.
    a. The proposed procedure of BJ-G is based on the standard Box-Jenkins's procedure, which consists of four stages, that are Stage I (Model identification), Stage II (Parameter estimation), Stage III (Diagnostic checking) and Stage IV (Forecasting).
    b. In the proposed procedure of BJ-G, the best Box-Jenkins model to analyse the data series is identified first. The heteroscedasticity in a highly volatile data series is detected in the diagnostic checking stage. The proposed procedure is meant for a combination of Box-Jenkins model with standard GARCH (GARCH) model.
    c. New steps and methods are suggested for the proposed procedure of BJ-G including data descriptive, the improvement of autocorrelation function (ACF) and partial autocorrelation function (PACF) methods, the use of Box-Cox transformation method, linearity test, portmanteau test, extended autocorrelation function (EACF) method, the Ljung-Box *Q*-test and the ARCH test.
    d. The proposed procedure of BJ-G investigates the appropriateness of distribution of innovations for the BJ-G model by considering Normal, *t*, skewed-*t*, generalised error distribution (GED) and Skewed-GED.
    e. A 5000 daily world gold price starting from 24th November 1993 to 17th December 2013 is used in the proposed procedure of BJ-G as a case study.

2. To propose a procedure of BJ-G in specifically handling univariatehighly volatile time series data.
   a. The proposed procedure is a modification from the first proposed procedure of BJ-G.
   b. In the proposed procedure, the heteroscedasticity in a data series is detected in the model identification stage since the study focuses on highly volatile data.
   c. The data used in the first proposed procedure of BJ-G is divided into six different sets of data based on previous literatures (Babu & Reddy, 2015; Ferenstein & Gasowski, 2004; Gaglianone & Marins, 2017; García-Ferrer et al., 2012; Harrison & Paton, 2004; Koopman et al., 2007; Sohn & Lim, 2007) and each sample is tested using the new proposed procedure in determining the optimal number of data for BJ-G model.

3. To propose a procedure of BJ-G model in evaluating the multistep forecasting performance for the univariate highly volatile time series data.
   a. The proposed procedure of BJ-G is an extension from the proposed procedure of BJ-G in the second objective, specifically in Stage IV.
   b. In this proposed procedure, sets of codes in R language are constructed.
   c. The multistep forecasting performance for the combination model using the new proposed procedure is considered with 80% and 95% prediction intervals as suggested by Hyndman and Athanasopoulos (2013).
   d. The data used in this procedure is the data series of the optimal sample based on the procedure of BJ-G in objective 2.

4. To propose a comprehensive procedure of BJ-G for all GARCH-type models in modelling and forecasting univariate highly volatile time series data.
   a. The proposed procedure of BJ-G is limited to apply for univariate time series data.
   b. The proposed procedure is applied to all GARCH-type models that are used previously for highly volatile data including the standard GARCH (GARCH), the GARCH in the mean (GARCH-M), the exponential GARCH (EGARCH), the threshold GARCH (TGARCH) and the asymmetric power ARCH (APARCH)).

1.5 Significance of the Study

This study will evaluate the performance of the combination model of BJ-G in forecasting univariate highly volatile time series data. As one of the established method in time series analysis, the Box-Jenkins model is used as the base model in the combination model. The good reputation of Box-Jenkins in handling univariate time series data and the practicality of GARCH in handling heteroscedasticity in a data series will contribute to a new potential approach in forecasting highly volatile time series data.

This study will propose four procedures of BJ-G to cater the four objectives in this doctorate research. The first proposed procedure of BJ-G is developed based on the standard Box-Jenkins's procedure since it will be used in evaluating the performance of the combination model to forecasting univariate highly volatile data for the preliminary study. The second proposed procedure of BJ-G model is developed specifically dealing for univariate highly volatile data at the early stage which simultaneously ensure the optimal number of data required for BJ-G model. While the third proposed procedure of BJ-G is an extension from the second one, which is applied in evaluating the multistep ahead forecasting performance of the BJ-G model. The fourth proposed procedure is developed from the second and third procedures and it is a comprehensive BJ-G procedure for modelling and forecasting univariate highly volatile time series data using the Box-Jenkins with GARCH-type model.

The proposed procedures are illustrated using the daily world gold price data since it is expected to be a highly volatile type of time series data. The original data series of the gold price is applied to the first and second proposed procedures of BJ-G, while the data series from an optimal number of data (based on the empirical results from the second procedure of BJ-G) would be used in evaluating the third and fourth proposed procedures. It is expected that the empirical results would demonstrate a good result in forecasting evaluations using the proposed procedures of BJ-G.

At the final stage, this study would contribute to a comprehensive procedure in modelling and forecasting up to *n*-step ahead for highly volatile time series data using Box-Jenkins with all GARCH-type models, as proposed by the fourth procedure of BJ-G. This study indirectly enhances the capability of the Box-Jenkins model in forecasting data series for further improvement of procedures and results. The guidelines given by the proposed procedures of BJ-G package with R codes developed would provide a good tool to demonstrate an element of data science which support IR 4.0.

1.6 Thesis Organisation

This thesis comprises of five (5) chapters with summary drawn for each study as the last section in the respective chapters. Chapter 1 (this chapter) presents a general introduction and motivation behind the research. Chapter 2 provides the literature reviews of highly volatile time series with a focus on univariate modelling, specifically the Box-Jenkins models, the heteroscedastic stochastic model especially GARCH-type and the combination of BJ-G model. Chapter 3 outlines the research methodology and concepts used in the study; theory of the Box-Jenkins modelling, the procedure for the combination of BJ-G model and the proposed procedures of BJ-G as in the objectives in this study with theoretical explanations for methods and tests used in each stage of the procedures. Chapter 4 presents the analysis of the data series, by taking daily world gold prices as a case study, using the four procedures of BJ-G as proposed in Chapter 3. Chapter 5 provides an overall summary of this thesis and highlights future research prospects. It is then followed by reference and appendices to facilitate a better understanding of this thesis. In general, the framework of this study is outlined in Figure 1.2.

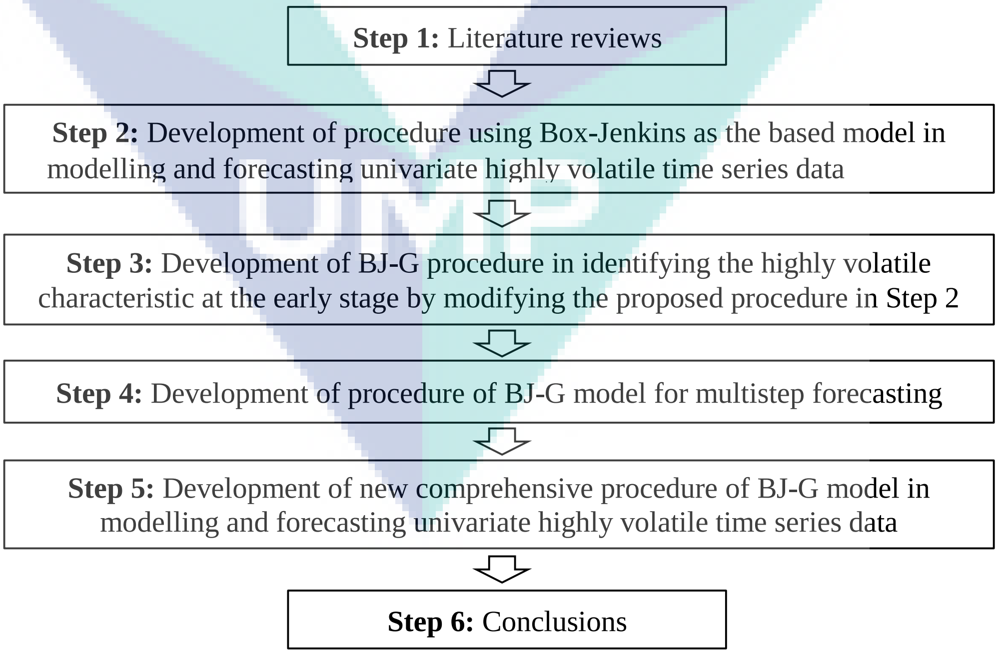


Figure 1.2 Framework of study

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter provides a brief review of Box-Jenkins modelling and the research scenario for the highly volatile time series data. The statistics for the number of publications in the related study from ISI web of science are provided. The modelling of Box-Jenkins and Box-Jenkins – GARCH (BJ-G) model in terms of procedure in the present study are critically reviewed to determine the gap for further development and contributions.

2.2 Box-Jenkins Modelling and Highly Volatile Time Series Data

Box-Jenkins modelling is proposed by Box and Jenkins in 1968 by introducing a procedure of four iterative stages namely model identification, parameter estimation, diagnostic checking and forecasting (Box & Jenkins, 1968). The modelling provides a systematic methodology for identifying and estimating models that incorporate autoregressive (AR) and moving average (MA) models. The AR, MA, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) are the models under consideration of the Box-Jenkins modelling, or so-called as the Box-Jenkins model.

The Box-Jenkins model is essentially a model for transforming the original series y_t into a series that consists of random errors component a_t; or $y_t \rightarrow a_t$, where y_t is often highly correlated series and a_t is an uncorrelated series. In general, the random errors component will exist in all construction of Box-Jenkins models. In the Box and Jenkins modelling, the principle of parsimony is an important principle in the choice of models, meaning that a simpler (having fewer parameters) model whilst adequately representing the data should be selected (Box & Jenkins, 1968; Box, et. al, 2008).

In the modelling, Box and Jenkins developed the class of stochastic models that are capable to represent stationary and nonstationary behaviour in obtaining the appropriate stochastic models since they believed that the optimal forecasts of future values of time series are determined by the stochastic model that describes the series. Their main effort then goes to statistical analysis in the stochastic model for the series that is directed to forecasting. In their initial study, Box and Jenkins employed significant relationships of two main univariate time series models, namely autoregressive and moving average which are given by Equation 2.1 and 2.2, respectively (Box & Jenkins, 1968). These models originally developed by Yule (Yule, 1927).

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \ldots + \phi_p(y_{t-p} - \mu) + a_t \quad 2.1$$

$$y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \quad 2.2$$

Equation 2.1 is the mathematical expression for the autoregressive model which shows that the deviation $y_t - \mu$ is linearly dependent on previous deviations and on a_t, where μ is the mean of data. Meanwhile, in the moving average model, the $y_t - \mu$ is made linearly dependent on a_t and on one or more previous random errors, as expressed in Equation 2.2. The forms in Equation 2.1 and 2.2 have linear relationships.

By definition, a univariate time series is a single random variable at time *t*, y_t while a univariate model describes the distribution of y_t in terms of its relationship with past values of y_t with a series of random errors, a_t. Hence, the Box-Jenkins models are able to handle linear stochastic series and is considered as one of the linear stochastic models. In general, the linear stochastic models often work well and able to provide an adequate approximation for modelling and forecasting data series (De Gooijer & Hyndman, 2006). Due to its capability, linear stochastic model as well as the corresponding linear method provides a useful benchmark for comparison of the results with alternative models. According to Tsay (2013), the univariate linear time series model covers simple econometric models that are useful in business, finance and economics.

The practicality of the Box-Jenkins model in handling linear stochastic series either as forecasting, the benchmark or as the integrated model makes the Box-Jenkins modelling is one of the established method in time series analysis. A book of Time Series Analysis: Forecasting and Control by Box and Jenkins in 1970 sparks a significant contribution in time series forecasting studies using the Box-Jenkins models. The book has had an enormous impact on the theory and practice of modern time series forecasting, specifically on univariate time series (De Gooijer & Hyndman, 2006).

Since the Box-Jenkins models contribute a significant effect to univariate time series data, hence the study on the Box-Jenkins modelling and its models still receives great interest till today. Figure 2.1 shows a trend in citation indexed journal papers for Box-Jenkins model during the years 2000 to early 2018 in various research areas. The top ten research areas related to time series data using Box-Jenkins model between year 2000 and 2018 is given in Table 2.1. The data was generated from the ISI web of science (Thompson Reuters, http:apps.webofknowledge.com/) using the keywords 'time series', 'Box-Jenkins model' and all related Box-Jenkins models on 23 January 2018. It can be observed that the number of publications has been continuously increasing year after year, thereby opening new ideas and opportunities for researchers to develop better understanding on the methods and procedures of the Box-Jenkins modelling.

Figure 2.1 Number of papers published in the time series research using Box-Jenkins model between 2000 and 2018 generated from ISI web of science on 23 January 2018

Despite the fact that the Box-Jenkins model is popular and practical to be used, it is not able to handle the non-constant variance that are present in a data series. In general, the Box-Jenkins models assume that the variance of the error, $\sigma_{a_t}^2$ is constant. Note that, the error term a_t in the Box-Jenkins model is usually assumed normal and having mean zero and constant variance $\sigma_{a_t}^2$, or can be written as $a_t \sim N\left(0, \sigma_{a_t}^2\right)$ (Box, Jenkins & Reinsel, 2008). However, for a highly volatile time series, the error term does not satisfy the homoscedastic assumption of constant variance.

Table 2.1 Top ten research areas related to time series data using Box-Jenkins model between 2000 and 2018 generated from ISI web of science on 23 January 2018

Research area	Number of papers	Percentage (%)
Mathematics & Statistics	1598	22.0
Engineering	1519	20.9
Business & Economics	1157	15.9
Computer science & Artificial intelligence	1115	15.3
Geology	616	8.5
Environmental sciences ecology	564	7.8
Mathematical methods in social sciences	399	5.5
Water resources	294	4.0
Geochemistry geophysics	285	3.9
Energy fuels	266	3.6

The time series studies for highly volatile is reportedly rapidly increasing in recent years, which indicates that the nature of current data is highly volatile especially in the commodity markets. This statement is supported by the statistics data generated from ISI web of science using the keywords "time series" and "volatility" or "highly volatile" on 23 January 2018, as shown graphically in Figure 2.2 and tabulated in Table 2.2. Therefore, a good forecasting method that is able to handle well a highly volatile time series data is vital to study in providing a useful forecasting tool especially for applications in economics and other related areas.

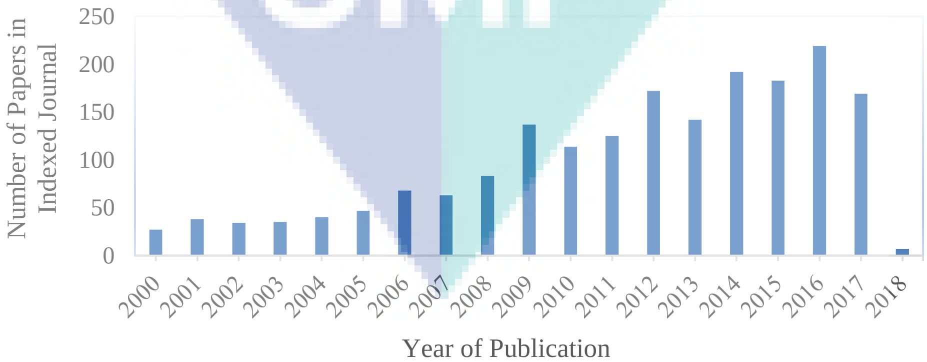


Figure 2.2 Number of papers published related to highly volatile time series between 2000 and 2018 generated from ISI web of science on 23 January 2018

Table 2.2 Top ten research areas related to highly volatile time series data between 2000 and 2018 generated from ISI web of science on 23 January 2018

Research area	Number of papers	Percentage (%)
Business & Economics	898	47.4
Mathematics & Statistics	588	31.0
Computer science & Artificial intelligence	312	16.5
Mathematical methods in social sciences	293	15.5
Engineering	258	13.6
Environmental sciences ecology	126	6.6
Operations research management science	101	5.3
Meteorology atmospheric sciences	66	3.5
Energy fuels	65	3.4
Science technology other topics	48	2.5

2.3 ARCH/GARCH Model and Volatility

In a highly volatile series, the time-varying variance (volatility or heteroscedasticity) which depends on the observations of the immediate past, is called conditional variance. The autoregressive conditional heteroscedastic (ARCH) model as introduced by Engle is used to model the conditional variance of the innovations, σ_t^2 (Engle, 1982). An ARCH model is the first model that provides a systematic framework for univariate volatility modelling to detect heteroscedasticity in financial time series in understanding volatility in time series data (Engle, 1982; Tsay, 2005). However, the ARCH models are difficult to estimate since they often produce negative estimates of the coefficient of ARCH parameters. Note that, all ARCH parameters must be positive. To solve this problem, Bollerslev proposed an extended form of heteroscedastic model known as the generalised autoregressive conditional heteroscedastic (GARCH) (Bollerslev, 1986).

In general, GARCH-type models can easily accommodate volatility clustering in a data series and being used in research practice of time series modelling and forecasting (refer Figure 1.1). Previous studies showed that GARCH-type models such as standard GARCH, GARCH-M and asymmetric GARCH (EGARCH, APARCH) are widely applied in time series forecasting to handle volatility in a data series (Ahmed, 2017; Chen et al., 2011; Girish, 2016; Loi & Ng, 2018; Liu & Shi, 2013; Pham & Yang, 2010). Table 2.3 presents the top ten research areas using ARCH/GARCH-type model between 2000 and 2018 generated from ISI web of science on 23 January 2018. From the statistics, it is shown that GARCH-type models have widely been used in forecasting economic or financial related data, where estimation of variance is important in the assessment of risk.

Table 2.3 Top ten research areas related to time series data using ARCH/GARCH - type model between 2000 and 2018 generated from ISI web of science on 23 January 2018

Research area	Number of papers	Percentage (%)
Business & Economics	388	40.7
Mathematics & Statistics	376	39.4
Computer science & Artificial intelligence	150	15.7
Mathematical methods in social sciences	139	14.6
Engineering	117	12.3
Operations research management science	47	4.9
Energy fuels	37	3.9
Environmental sciences ecology	35	3.7
Science technology other topics	21	2.2
Energy fuels	266	3.6

Due to its homoscedastic assumption of constant variance, the Box-Jenkins model is found inappropriate for modelling and forecasting highly volatile time series data. However, because of the good reputation of Box-Jenkins model in research practice, the model is worth considering in forming a forecasting model for the highly volatile time series. Since the forecasting model must reflect its structure and pattern, the conditional variance in a highly volatile data series should be considered in forming a reliable forecasting model. Hence, the incorporation of GARCH model to Box-Jenkins model can be an effective way to overcome the limitation of the Box-Jenkins model in forecasting highly volatile time series.

There are various forecast models in dealing with highly volatile time series data such as support vector machines (SVMs) (Villegas, Pedregal & Trapero, 2018), artificial neural networks (ANNs) (Alasali, Haben, Becerra & Holderbaum, 2018; Chakravarty, Mohapatra & Dash, 2016; Chen et al., 2012; Chitsaz, Shaker, Zareipour, Wood & Amjady, 2015; Jun, Lingyu, Yuyan & Peng, 2017; Panapakidis, 2016; Pandey, Jagadev, Dehuri & Cho, 2019; Yu, Choi & Hui, 2012), hybrid ANNs and Box-Jenkins (Weerathunga & Silva, 2018), artificial intelligence (AI) (Yang, Zhang & Wang, 2019), dynamic window size algorithm (DyWiSA) (Dalmazo, Vilela & Curado, 2017), hybrid fuzzy system and ANNs (Barros & de Medeiros, 2017), dynamic model averaging (DMA) (Naser, 2016), bootstrapped regime switching (BRS) (Gel, Lyubchich & Ahmed,

UMP

2016), wavelet smoothing method (Michis, 2015), sparse modelling (Tzagkarakis, Caicedo-Llano & Dionysopoulos, 2015), value-at-risk (VaR) estimation method (Liu, Chung & Wen, 2014), hybrid of ANNs and SVMs (Hu, Wang & Zeng, 2013) and Taguchi method (Wang & Huang, 2007). Although these models achieve a certain effect in forecasting highly volatile data, many studies in recent years applied the combination models of Box-Jenkins with GARCH-type to time series data in various fields for their good performance.

Data generated from ISI web of science on 23 January 2018, as presented in Figure 2.3, shows the number of papers published using the combination model between 2000 and 2017. It shows that the number of publications is increasing significantly especially since year 2009, which indicates that the model of Box-Jenkins – GARCH-type is a promising one in forecasting highly volatile time series data. However, the information about the procedure in applying the combination model is not documented clearly. Hence, the development of a comprehensive procedure of BJ-G would provide useful basic guidelines for using the combination model of Box-Jenkins – GARCH-type to address univariate highly volatile data.

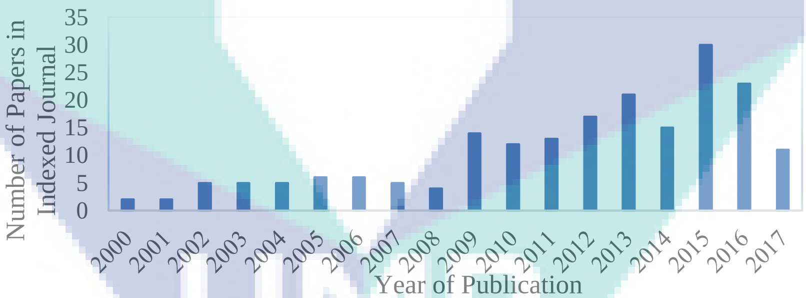


Figure 2.3 Number of papers published using BJ-G model between 2000 and 2017 generated from ISI web of science on 23 January 2018

2.4 A Review of Some Studies on Box-Jenkins – GARCH Model

The study on highly volatile time series for modelling and forecasting purposes using Box-Jenkins with GARCH-type model is supported by many researchers. Some of the recent studies, specifically for univariate data, are summarised in Table 2.4. To the best of our knowledge, although these studies achieve a certain effect in modelling and forecasting highly volatile time series data, very limited studies or literatures focus on the development of BJ-G model procedure.

Table 2.4 Review on the selected studies on Box-Jenkins – GARCH-type model

Researcher	Data	Model	Methods/Procedure
Loi and Ng, 2018	- electricity prices - 524 weekly wholesale prices (6/5/06-20/4/16) - ratio 503:21 (or 96:4)	ARIMA-GARCH	i) Descriptive statistics and volatility checking (use standard deviation at level and log price) ii) Stationarity test: differenced, Perron/Voselgang breakpoint unit root tests iii) Identify model: ACF, PACF iv) Parameter estimate: MLE, AIC v) Diagnostic test: ARCH test, ACF on $\hat{a}_t^2$ vi) Forecast evaluation: 1-step, MAPE, MAE, RMSE, TIC *Handling outlier *Apply structural breaks to ARIMA-GARCH *Provides brief procedure
Gaglianone and Marins, 2017	- exchange rate - 183 monthly data (Jan 00-Mac 2015) with ratio 72:111 (or 40:60) - 3977 daily data (3/1/00-31/3/2015) with ratio 1565: 2412 (or 40:60)	AR-GARCH	i) Analysis on exchange rate ii) Diagnostic test: *t* innovations iii) Forecasting: Point forecast (RMSE), multiple step ahead up to h = 20 (70% PIs, LPDS, Knüppel test, Berkowitz test) iv) Model ranking: local analysis, risk analysis *AR-GARCH is 1 of 14 model *more on multiple step forecasting *Brief procedure for AR-GARCH
Ahmed, 2017	- stock market index prices - energy commodity prices: oil and natural gas 541 weekly closing data (1/2-8/6/2016)	AR-EGARCH - for univariate data VAR-EGARCH -for bivariate data	i) Stationarity: ADF, PP, KPSS ii) Properties for volatility: Mean, standard deviation, skewness, kurtosis, JB-test, LBQ-test, ARCH test iii) lag order: SIC iv) Parameter estimate: AIC, SIC, QMLE v) Diagnostic Test: JB, LBQ, LBQ2, ARCH *No discussion on forecasting part *No framework is proposed
Girish, 2016	- Spot electricity prices - 27384 hourly data (1/10/10-15/11/13) - ratio 26304:1080 (or 96:4)	ARIMA-GARCH, ARIMA-PARCH, ARIMA-EGARCH	i) Data stationarity: first differenced price ii) Heteroscedasticity test: White test (for data) iii) Identify model: ACF and PACF iv) Diagnostic test: SIC, ACF, PACF, LBQ, LBQ^2, ARCH test v) Forecasting evaluation: 1-step ahead forecast (RMSE, MAE, MAPE and TIC) *Not discuss on parameter estimate

Table 2.4 Continued

Researcher	Data	Model	Methods/Procedure
Liu and Shi, 2013	- short term real time electricity prices - 18960 hourly data (1/1/08-28/2/2010) - ratio 17544:1416 (or 92:8)	ARMA-GARCH, ARMA-GARCH-M	i) Descriptive statistics on price ii) ARMA model: use EACF method iii) Parameter estimation: MLE iv) Diagnostic test: adjusted R^2, *F*-test, AIC, SIC, ACF and PACF on residuals v) Forecasting evaluation: 1-step ahead forecast (RMSE, MAE, MAPE, TIC) *no discussion on data stationarity *no test on heteroscedasticity *no graphical framework on the procedure
Babu and Reddy, 2015	- Stock price - daily closing data (Jan 2010-Dec 2011)	Hybrid ARIMA-GARCH	i) Pre-processing step: decompose series to highly and low volatile using MA filter. ii) The highly volatile data is partitioned and interpolated to apply ARIMA-GARCH model iii) Check volatility (use standard deviation), volatility clustering, fat tail distribution. iv) Forecasting evaluation: MAPE, MaxAPE, MAE, RMSE, 20-step ahead *no discussion on how to choose the model order, diagnostic test. *discussion more on *k*-fold cross validation in choosing model and forecasting.
García-Ferrer, González-Prieto and Peña, 2012	- stock price - 1250 daily closing data - (1/1/00 – 31/12/04) - ratio 1000:250 (or 80:20)	AR-GARCH, ARMA-GARCH	i) Analysis on log return, descriptive statistics, JB-test and LBQ-test, LBQ^2-test ii) ARMA model: SIC, GARCH: QMLE iii) Innovations: normal, *t*, GED iv) Forecast error: MSE, 1-step ahead *Propose procedure using GICA-GARCH to fit a ARMA-GARCH model

Table 2.4 Continued

Researcher	Data	Model	Methods/Procedure
Liu, Erdem and Shi, 2011	- wind speed data - hourly data (1/1/02-31/12/08)	ARMA-GARCH, ARMA-GARCH-M	i) ARMA model: use EACF method ii) Parameter estimation: MLE iii) Diagnostic test: adjusted R^2, *F*-test, AIC, SIC, LBQ-test, BG-test *no test on heteroscedasticity, show data volatility using time plot * no discussion on stationarity data *no discussion on forecasting part *no graphical framework/procedures
Chen et al., 2011	- Short-time traffic flow (transportation) - 3-min, 5-min, 10-min and 15-min traffic flow - (1/10 – 30/11/09) - Ratio 50:50	Hybrid ARIMA-GARCH	i) Stationary data: differenced ii) Model identification: ACF, PACF iii) Diagnostic test: LBQ, ARCH-test, LBQ^2 iv) Forecast evaluation: MAE, MSE, MRE, 1-step ahead *no discussion on data transformation, parameter estimate *provides graphical framework on procedure
Walid, Chaker, Masood and Fry, 2011	- stock market index prices - exchange rate - weekly closing data - (Dec 1994-March 2009)	AR-EGARCH	i) Return price and rate, descriptive statistics, JB-test, LBQ, LBQ^2-test ii) Stationarity test: ADF, PP, KPSS iii) Lag order: AIC, Hannan and Quinn iv) Diagnostic test: LBQ, LBQ^2 * no discussion on forecasting part *no framework/procedure is proposed
Harrison and Paton, 2004	- Stock market price - 1384 daily data (7/5/97 -16/9/00)	AR-GARCH	i) Analysis on stock market return ii) Descriptive statistics, Shapiro-Wilk for normality test iii) Parameter estimation: MLE iv) Model selection: AIC v) Diagnostic test: ARCH-test, LBQ-test *no discussion on forecasting

Table 2.4 Continued

Researcher	Data	Model	Methods/Procedure
Pham and Yang, 2010	- Vibration signal (highly volatile time series) for fault prediction - 470 of 6-hour data - (Sept – Nov 2005) - ratio 250:220 (or 53:47)	ARMA(1,1)-GARCH(1,1)	i) Stationarity: ACF plot ii) Identify model: ACF and PACF iii) Parameter estimate: MLE iv) Forecasting: MSE, 1-step to 10-step ahead *not emphasis on diagnostic checking *no checking on the existence of heteroscedasticity in the data. *no transformation data step. *shows framework, but not enough details on method used.
Ord, Koehler, Snyder and Hyndman, 2009	- gasoline price - monthly data - (Jan 1991 – Nov 2006) - ratio 132:59 (or 70:30)	ARMA(1,1)-EGARCH(1,1)	i) Analyse log gas price ii) Parameter estimation: MLE iii) Forecasting: up to *h*-step ahead *use the model in monitoring heteroscedastic processes *no procedure is given on identification stage and diagnostic test
Koopman, Ooms and Carnero, 2007	- daily electricity spot prices - 4480 daily data (4/1/93 – 10/4/05)	periodic seasonal Reg-ARFIMA-GARCH	i) Analyse log data, first difference of log prices, descriptive statistics ii) Parameter estimation: MLE, AIC iii) Diagnostic test: residual plot, LBQ, LBQ^2-test, histogram and ACF for residuals, ACF of squared residuals, *t* innovations *no discussion on forecasting * provides basic modelling framework in sentence form, no graphical framework
Byström, 2005	- electricity spot prices (highly volatile) - 41665 hourly data (1/1/96 - 1/10/00) - Ratio 17472:24193 (or 40:60)	AR-GARCH(1,1)	i) Analysis on return price, descriptive statistics, LBQ-test, LBQ^2-test ii) Parameter estimation: MLE iii) Diagnostic test: descriptive statistics on residuals, LBQ-test, LBQ^2-test, normal and *t* iv) Forecasting: 1-step to 24-step, use tail quantiles at different probability *no explanation on how to choose order of GARCH model. *procedure/procedure used is not clear presented

Table 2.4 Continued

Researcher	Data	Model	Methods/Procedure
Sohn and Lim, 2007	- rates of Dow Jones Industrial Average - data stock price - (24/9/01-10/11/03)	AR-GARCH	i) Analysis on log return rates ii) Forecasting evaluation: *h*-step ahead, MRE *no procedure to choose the order of the model, assume data follows AR(2)-GARCH(1,1)
Ferenstein and Gasowski, 2004	- stock price - Daily closing price - (26/3/92 – 9/12/02)	AR-GARCH, AR-EGARCH	i) Analysis on log returns of prices ii) Identification: use PACF of log return for AR model, LBQ-test iii) Parameter estimation: QMLE, AIC, SIC iv) Diagnostic test: LBQ-test, LM ARCH-test, QQ-plot, innovations distribution (normal, *t*, GED, hyperbolic) v) Forecasting evaluation: MSE, MAE, MAPE, MMEO, MMEU, 1-step and 2-step *the ratio of estimate: forecast is not stated

2.5 Gap of Knowledge in the Study

In this study, the Box-Jenkins modelling is used as the basic approach in forecasting highly volatile time series data. There are four iterative stages in the modelling, Stage I (Model identification), Stage II (Parameter estimation), Stage III (Diagnostic checking) and Stage IV (Forecasting). In finding the gap of knowledge in the study, specifically in the development of the procedure in the Box-Jenkins modelling in handling highly volatile time series data, published books that have been highly influential which were written by experts on the topic of time series forecasting that is related to the Box-Jenkins modelling for univariate data have been referred. The four main references used are for theoretical and application of Box-Jenkins modelling. Table 2.5 presents the review of the procedure of Box-Jenkins modelling from the publications.

Table 2.5 Review on procedure of Box-Jenkins modelling for univariate data

Publication	Contribution on Procedure	Limitation/Remarks
Time Series Analysis: Forecasting and Control 4th and 5th ed. (Box, et al., 2008; Box, Jenkins, Reinsel & Ljung, 2015)	**Stage I**: (i) Stationarity test: ACF and PACF, unit root test, use differenced method; (iii) Identify model: ACF and PACF (k_{max} = 20); (iv) Model selection: AIC, SIC. The Box-Cox transformation is suggested for seasonal series. **Stage II**: MLE, OLS **Stage III**: autocorrelation check on residual, LBQ-test at k_{max} = 20) **Stage IV**: MSE for 1-step ahead, PIs 50% and 95% for multiple step. • Provide thorough explanation of the theoretical parts on the methods used. • Provide a brief graphical presentation on the stages in the procedure. • Provide some basic concepts of ARCH/GARCH in errors part of BJ. • To detect ARCH/GARCH effect: ACF and PACF of squared errors, LBQ-test on squared errors, ARCH LM test. • Illustrate the procedures using R software in the latest publication.	• The procedure to combine the Box-Jenkins model with ARCH/GARCH is not well explained. • No detailed graphical presentation (include methods and tests) on the procedure of BJ modelling either for BJ models or BJ-G. • The steps in each stage is not clearly explained. • No discussion on the procedure to determine the optimal sample size either for BJ models or BJ-G. • Lack of discussion on distribution of errors which is important in Stage III.
Forecasting: Principles and Practice 1st - 2nd online edition (OText) (Hyndman & Athanasopoulos, 2014; Hyndman & Athanasopoulos, 2017)	**Stage I**: (i) Stationarity: time plot, data transformation, differencing (ACF, ADF-test ($k=(T-1)^{1/3}$), LBQ-test at $k=10$; (ii) Identify model: ACF, PACF **Stage II**: MLE, AIC and SIC **Stage IV**: 80% and 95% PIs, 1 to *n*-step • Steps in forecasting: define problem, collect info, preliminary analysis, choose and fit models, forecasting evaluation. • Forecasting tools:(i) Graphics: time plot, lag plot, scatter plot, ACF (ii) data transformation: Box-Cox, calendar adjustments; (iii) residual diagnostics: plot, histogram, ACF, LBQ-test; (iv) forecast accuracy: in-sample to out-of-sample (80:20), forecast error (MAE, RMSE, MAPE, scaled error, MASE), cross-validation, 80% and 95% PIs. • Provides a framework of general process of forecasting using ARIMA. • Emphasise graphical methods to explore, analyse and forecast the data. • Provides data analysis using R language specifically forecast package. • Provides a general idea to determine the optimal sample size for time series.	• The procedures (steps, methods) using BJ model is not structured well in the stages of BJ modelling. • Unclear stages of the Box-Jenkins modelling in the framework of general process of forecasting using ARIMA. • Did not provide thorough discussion on the theoretical details behind each method. • No discussion on GARCH, suggest to use data transformation (Box-Cox) to address heteroscedasticity. • No discussion on Box-Jenkins model and GARCH in handling volatility in data series

Table 2.5 Continued

Publication	Contribution on Procedure	Limitation/Remarks
An Introduction to Analysis of Financial Data with R (Tsay, 2013)	**Stage I**: (i) time plots, stationarity (scatter plot, ACF, PACF, differencing, ADF-test), LBQ-test at $k = \ln T$, (ii) Identify model: AR (use ACF, AIC, SIC), MA (use PACF), ARMA and ARIMA (use EACF, AIC, SIC)) **Stage II**: OLS, MLE **Stage III**: ACF, LBQ-test on residuals, DW-test, LBQ-test on squared residuals, LMARCH-test, error distribution. **Stage IV**: Backtesting (MSE, RMSE, MAE, Bias), 1 to *n*-step (95% PIs). • Properties of financial data: asset return, simple return, compounded return. • Statistical tool: moments, graphics (time plot, histogram, scatter plot), JB-test. • linear models in financial time series: AR, MA, ARMA, ARIMA, SARIMA, random walk, random walk with drift. • Provides steps in ARCH/GARCH model • Discuss ARCH and GARCH-type model • Demonstrate analysis using R language.	• The procedures (steps, methods) using BJ model is not structured well. • No detailed graphical presentation (include methods and tests) on the procedure of BJ modelling either for BJ models or BJ-G. • No discussion on the combination of Box-Jenkins model with GARCH-type in handling volatility in data series • No discussion on the procedure to determine the optimal sample size either for BJ models or BJ-G.
Time Series Forecasting (Chatfield, 2001)	**Stage I**: (i) Model formulation: preliminary analysis; (ii) Model selection: ACF and PACF, AIC and SIC. **Stage III**: residual plot, autocorrelation check, LB*Q*-test (*k*= 20). **Stage IV**: forecast evaluation (RMSE, MAE, MSE, PMSE, ME, MAPE), 80% or 90% PIs • objectives in time series analysis: data description, modelling, forecasting, control. • Preliminary part: (i) Data description: time plot, descriptive statistics, histogram; (ii) Box-Cox transformation when series has severe changes in variance; (iii) data cleaning: outlier, errors, missing; (iv) emphasise on ACF • Stationarity: differencing • in-sample (Stage I – Stage III) and out-of-sample (Stage IV), ratio 90:10 • Nonlinearity checking: plot y_t vs y_{t-1} • Brief discussion on sample size • Discuss briefly on ARCH/GARCH. • Provides details on prediction intervals.	• No detailed graphical presentation (include methods and tests) on the procedure of BJ modelling either for BJ models or BJ-G. • No discussion on the procedure to determine the optimal sample size. • A brief discussion on **Stage II** (MLE, OLS). • No discussion on the distribution of errors which is important in model forecasting. • No detailed discussion on models in dealing with univariate highly volatile data.

*BJ is simplified for Box-Jenkins and BJ-G is simplified for Box-Jenkins with GARCH

The books of Time Series Analysis: Forecasting and Control 4th ed. by Box et al. (2008) have been cited 44 241 times, as generated in Google Scholar up to 23 January 2018. Hence, it shows that the Box-Jenkins modelling has been very influential in the development of time series modelling. Since the study focuses on the development of the Box-Jenkins modelling, the books have been used as the main reference for the basic theories and procedures in this study. Other than that, the books by Hyndman and Athanasopoulos (2014,2017) and Chatfield (2001) and have also been used as key references in the development of the procedures for practical application of univariate highly volatile. Whereas, the book by Tsay has been used as the main reference for financial tools since the study focuses on highly volatile data that is closely related to financial data (Tsay, 2013). Therefore, based on the critical reviews on books presented in Table 2.5 and the existing publications, the knowledge gaps are identified, as the details are given in Table 2.6.

Table 2.6 The knowledge gaps in the study

New contribution	Limitation/Remarks
1. Proposed a modified procedure in modelling and forecasting univariate highly volatile data series using BJ-G model based on the Box-Jenkins modelling, by considering the methods and tests used in **Stage I** to **Stage IV** as suggested in the key publications used. 2. Proposed a modified BJ-G procedure in dealing specifically with univariate highly volatile time series data which simultaneously ensure the number of data required for practical application using BJ-G model. 3. Proposed a modified procedure in evaluating the multistep forecasting performance using BJ-G model. 4. Proposed a comprehensive procedure of Box-Jenkins with all GARCH-type model in forecasting highly volatile time series up to *n*-step ahead. 5. Provides a well-structured graphical presentation of each BJ-G stage (**Stage I – Stage IV**) for each proposed procedure. 6. Evaluate the performance of the proposed procedures and the corresponding graphical presentations for world gold price. 7. The procedures and the corresponding graphical presentation are also applicable when applying the Box-Jenkins model to data series, by ignoring the parts of heteroscedasticity test.	• For univariate data time series only. • Consider the model of Box-Jenkins for the mean model. • Consider the model of GARCH-type for the variance model.

The knowledge gaps identified are focused on the development of time series forecasting model in handling univariate highly volatile time series data using

Box-Jenkins modelling by incorporating the Box-Jenkins with GARCH-type model. The positioning of Box-Jenkins and the GARCH models in the univariate time series forecasting can be referred to in Figure 1.1 in Chapter 1.

2.6 Concluding Remarks

The current study aims to develop a new procedure for modelling and forecasting highly volatile time series data with the Box-Jenkins as the base model. Since the focus in this study is to develop the procedure of Box-Jenkins that deals with univariate highly volatile time series data, then the GARCH-type model is considered in the proposed procedures, namely the Box-Jenkins – GARCH's procedure or procedure of BJ-G. The proposed procedure of BJ-G is demonstrated using the world gold prices as the case study.

CHAPTER 3

METHODOLOGY AND STATISTICAL TOOLS

3.1 Introduction

This chapter theoretically describes the concepts and methodologies used in the study. Based on these concepts and methodologies, the current study aims to develop a conceptual procedure for modelling and forecasting highly volatile time series data with the Box-Jenkins as the base model. The world gold prices will be the case study. A general Box-Jenkins framework in modelling and forecasting is illustrated in Figure 3.1.

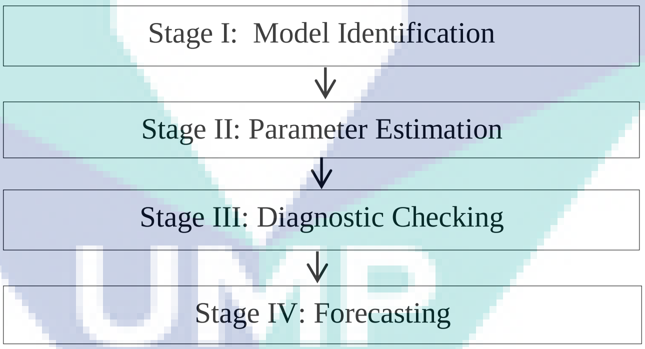


Figure 3.1 General Box-Jenkins's framework

However, the framework can be enhanced for further improvement of procedures and results to develop a new procedure. In our work, while a new Box-Jenkins's procedure is developed, its two important principles which are stationarity and parsimony will not be violated. This chapter starts with the discussion on Box-Jenkins modelling and forecasting procedures. Since the focus in this study is to develop a procedure of Box-Jenkins that deals with highly volatile time series data, then the theory of the GARCH models is considered in developing the combination model of Box-Jenkins –

GARCH, simply called as BJ-G. In the final section, a procedure of BJ-G developed by the current study is presented.

3.2 Box-Jenkins Modelling

Box-Jenkins modelling involves five types of models (Box et al., 2008). The models are autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA). The models which are associated with stationary behaviours are AR, MA and ARMA. Stationary models assume that the process remains in equilibrium statistically with probabilistic properties that do not change over time, in particular varying about a fixed constant mean level and with constant variance. ARIMA is the only model that handles nonstationary time series with nonseasonal characteristics, meanwhile, SARIMA is the only model of Box-Jenkins that is dedicated to nonstationary with seasonal time series. The procedure in choosing the appropriate Box-Jenkins model is shown in Figure 3.2.

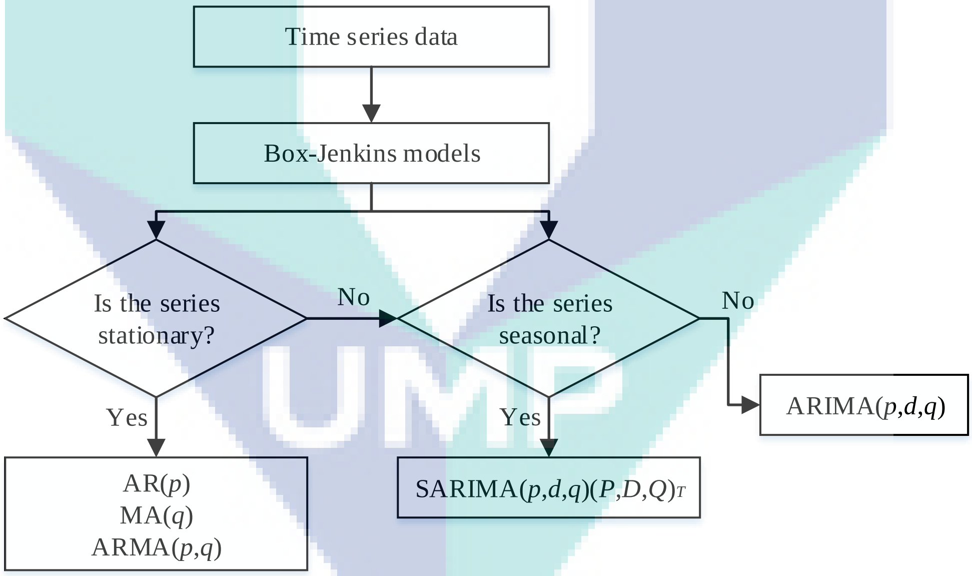


Figure 3.2 Schematic diagram for the procedure in choosing the appropriate Box-Jenkins Model
Source: Box, et al. (2008).

Figure 3.2 shows how to identify different types of time series data as either stationary or nonstationary. The general forms of the Box-Jenkins's stationary and nonstationary models are given as follows.

i) Stationary Models

Let y_t and a_t be the observed value and random error at time period *t*, respectively; with *c* is a constant, μ is the mean of the model, $\varphi_1, \varphi_2, ..., \varphi_p$ are the autoregressive parameters with order *p*, $\theta_1, \theta_2, ..., \theta_q$ are the moving average parameters with order *q*, *d* is the order of differencing and B is the backward shift operator. The operator of $\varphi_p(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i$ and $\theta_q(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ are polynomials in terms of *B* of degree *p* and *q*. Note that *p*, *q* and *d* are integers. The stationary Box-Jenkins models have the form given in Equation 3.1 to 3.3. In these models, the random errors a_t are assumed to be independently and identically distributed (IID) with mean zero and constant variance of σ^2.

The autoregressive model of order *p*, or AR(*p*) is expressed as in Equation 3.1,

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_p y_{t-p} + a_t \quad \text{for } |\varphi_1|, |\varphi_2|, ..., |\varphi_p| < 1 \qquad 3.1$$

or equivalently to

$$(1 - \varphi_1 B - \varphi_2 B^2 - ... - \varphi_p B^p)(y_t - \mu) = a_t$$

where $c = (1 - \varphi_1 - \varphi_2 - \ldots - \varphi_p)\mu$. Meanwhile, the moving average model of order *q*, or abbreviated as MA(*q*), is expressed as in Equation 3.2.

$$y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} \quad \text{for } |\theta_1|, |\theta_2|, ..., |\theta_q| < 1 \qquad 3.2$$

or can be written as $y_t = \mu + (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q) a_t = \mu + \theta(B) a_t$.

The autoregressive moving average model of order *p* and *q*, or particularly an ARMA(*p*,*q*), is a model that mixes the AR(*p*) and MA(*q*) models as in Equation 3.3.

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} \qquad 3.3$$

or can be written as $\varphi(B) y_t = \theta_0 + \theta(B) a_t$, where $c = (1 - \varphi_1 - \varphi_2 - ... - \varphi_p)\mu$.

ii) Nonstationary Models

The autoregressive integrated moving average model of order *p* and *q*, ARIMA(*p*,*d*,*q*) is the extended model of ARMA(*p*,*q*) with order of differencing, *d*. This model suggests that in a nonstationary case, the series need to be differenced in order to form a stationary series. The general form of ARIMA(*p*,*d*,*q*) has the form as in Equation 3.4,

$$\varphi_p(B)(1-B)^d(y_t-\mu)=\theta_q(B)a_t \quad \text{for } |\varphi_1|,|\varphi_2|,\ldots,|\varphi_p|<1,\ \ |\theta_1|,|\theta_2|,\ldots,|\theta_p|<1. \qquad 3.4$$

The seasonal autoregressive integrated moving average model of Box-Jenkins model denoted by SARIMA $(p,d,q)(P,D,Q)_S$, is designed for the nonstationary and seasonal series. This seasonal model is extended from ARIMA model and represented by Equation 3.5,

$$\Phi_P(B^S)\varphi_p(B)(1-B)^d(1-B^S)^D\dot{y}_t=\Theta_Q(B^S)\theta_q(B)a_t \qquad 3.5$$

where

$$\dot{y}_t=\begin{cases} y_t-\mu, & \text{if } d=D=0 \\ y_t, & \text{otherwise} \end{cases}$$

and $\Phi_P(B^S)=1-\sum_{I=1}^{P}\Phi_I(B^S)^I$, $\Theta_Q(B^S)=1-\sum_{J=1}^{Q}\Theta_J(B^S)^J$ which are polynomials in terms of B^s of order *P* and *Q*, $\nabla_D^S=(1-B^S)^D$, *S* is the seasonal period, *P* is the order of seasonal autoregressive, *Q* is the order of the seasonal moving average, and *D* is the order of seasonal differencing.

The Box-Jenkins approach is different from most methods in a time series because it uses an iterative approach of identifying a possible model from a general class of models. As depicted in Figure 3.1, the general Box-Jenkins framework includes four iterative stages namely Stage I: Model identification, Stage II: Parameter estimation, Stage III: Diagnostic checking and Stage IV: Forecasting. The general step by step procedure in the Box-Jenkins framework is briefly reviewed as follows [see (Box et al., 2008)].

Stage I (Identification): The stationarity in the data is tested since the Box-Jenkins models are applicable for stationary series data. The use of Box-Cox transformation is suggested for seasonal series. The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary series then are used to identify the possible model of the Box-Jenkins for the data. For the Box-Jenkins models, the ACF and PACF properties of a stationary data series are employed in determining the order of the models. Basically, the sample ACF is used to obtain the possible values of order q for MA(q), ARIMA(p,d,q) and SARIMA(p,d,q)(P,D,Q)$_S$ models, and the sample PACF is used to obtain the possible values of order p for AR(p), ARIMA(p,d,q) and SARIMA(p,d,q)(P,D,Q)$_S$ models.

Stage II (Parameter estimation): The parameters for the possible model are estimated using widely known parameter estimation approach such as maximum likelihood estimation (MLE) or ordinary least square (OLS). The model with significant parameter(s) will be considered for the next stage. Essentially, the method of MLE works by finding the most likely values of the parameters given the actual data. More specifically, a log-likelihood function is formed and the values of the parameters that maximise it are required. The MLE can be employed to find parameter values for both linear and nonlinear models. Meanwhile, the method of OLS is simply a procedure that finds the minimum of the sum of squared error function.

Stage III (Diagnostic checking): In diagnostic checking or data cleaning stage, the chosen model will be statistically verified against the original data to see whether it adequately describes the series. The model fits the data well if its estimated errors or its residuals $\{\hat{a}_t\}$ are generally small, randomly distributed, homoscedasticity (constant and finite variance process) and contain no useful information, for the closeness to the white noise criteria. If the specified model is not satisfactory, the process of stage I to stage III is repeated by using a new identified model in order to improve the previous model. This iterative procedure continues until a satisfactory model is obtained.

Stage IV (Forecasting): When the satisfactory model is obtained, the model then can be used for forecasting. The forecasting evaluation for one-step ahead is the minimum mean square error (MSE) while for multiple step ahead, the forecasting evaluation used is the prediction intervals (PIs) of 95% and 50%.

The graphical visualisation of the theoretical procedure of Box-Jenkins modelling is illustrated by Figure 3.3, where Box-Jenkins is abbreviated as BJ. The current practices in Box-Jenkins modelling are quite general and not thorough enough to describe the nature of time series data studied as well as the stages. By not being thorough, researchers might overlook the fact that some time series, specifically the highly volatile time series, which cannot be analysed using the current procedure of Box-Jenkins since the modelling is not able to handle the non-constant variance that exist in the time series data.

Previous studies have shown that the GARCH models are widely applied to handle volatility in a data series (Ahmed, 2017; Baur & Lucey, 2010; Chen et al., 2011; Girish, 2016; Loi & Ng, 2018; Liu & Shi, 2013; Pham & Yang, 2010; Trück & Liang, 2012). Therefore, the volatility model is being considered in constructing the proposed procedure of Box-Jenkins model that deals with highly volatile time series data. In addition, by considering the GARCH model in the errors, it would be useful in providing a more accurate prediction interval in the future forecast (Ruppert & Matteson, 2015).

3.3 The Combination of Box-Jenkins and GARCH Model

In dealing with univariate highly volatile time series data using Box-Jenkins approach, the consideration then is to study the combination of Box-Jenkins and the GARCH (BJ-G) model. Based on previous studies on Box-Jenkins with GARCH-type model as summarised in Table 2.4 (in Chapter 2), the use of the combination model in handling heteroscedasticity in a data series is supported by many researchers. However, to the best knowledge of the researcher, the previous studies are lack in the development of procedure of Box-Jenkins – GARCH-type model. Therefore, this study initiates to develop a comprehensive procedure of the model by focusing on the combination of Box-Jenkins with GARCH-type model.

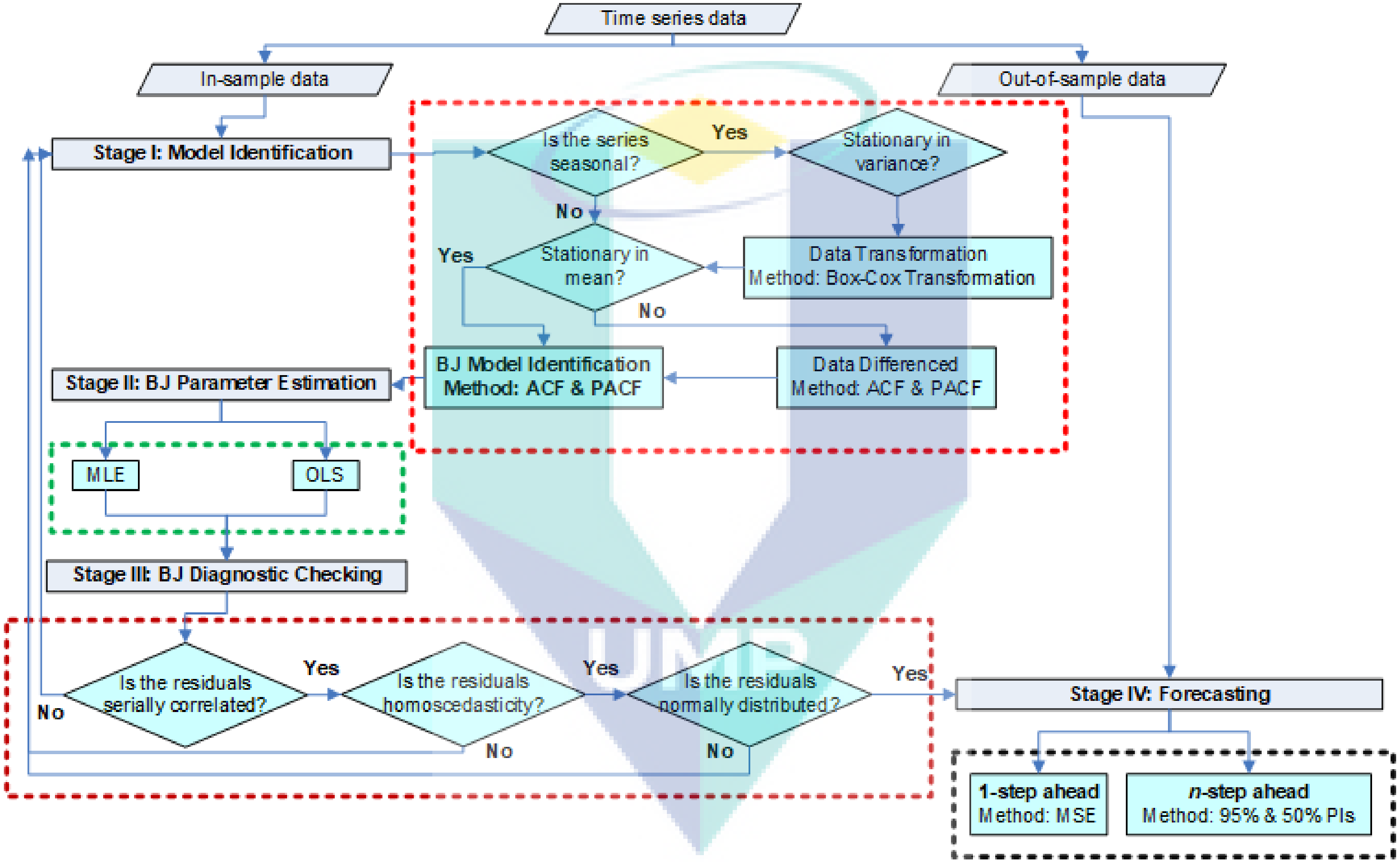


Figure 3.3 Theoretical procedure of the Box-Jenkins modelling

The standard GARCH is considered in the preliminary stage of this study due to its parsimonious characteristic as well as its popularity in handling heteroscedasticity in a data series (Babu & Reddy, 2015; Chen et al., 2011; Gaglianone & Marins, 2017; Girish, 2016; Liu & Shi, 2013; Loi & Ng, 2018; Pham & Yang, 2010; Tan et al., 2010; Tsay, 2013; Zhou et al., 2006). By applying standard GARCH model, or simply called as GARCH, as a base model to handle the volatility in the data series, the basic theory related to the volatility model is discussed. In this volatility model, the key concept is the conditional variance, that is, the variance conditional on the past. Suppose that the mean model at time *t* for a univariate series is given as in Equation 3.6,

$$s_t = \mu_t + a_t \qquad 3.6$$

where s_t and a_t be the stationary data and random error at time period *t*, respectively; with μ_t is conditional mean of s_t and $a_t = \sigma_t \varepsilon_t$ where ε_t is the innovations of the model and has zero-mean independent and identically distributed sequences with continuous distributions. The term a_t follows a GARCH(*r*,*s*) model if the conditional variance of s_t, denoted by σ_t^2, is given as in Equation 3.7 (Bollerslev, 1986; Francq & Zakoïan, 2010).

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i a_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 \qquad 3.7$$

where α_i and β_i are the coefficient of the parameters ARCH and GARCH, respectively. The random variable σ_t is called the volatility of a_t. According to Francq and Zakoian, there is no general agreement concerning the definition of volatility; volatility sometimes refer to a conditional standard deviation, and sometimes to a conditional variance (Francq & Zakoïan, 2010).

There are two definitions regarding the GARCH process (Francq & Zakoïan, 2010). The first one is, the a_t is called a GARCH(*r*,*s*) process (or sometimes called semi-strong) if its first two moments exist and satisfy:

(i) $E(a_t | a_u, u < t), \quad t \in Z$.

(ii) There exist constant α_0, $\alpha_i, i = 1,2,\ldots,r$ and β_j, $j = 1,2,\ldots,s$ in Equation 3.7.

However, the first definition of the GARCH process does not directly provide a solution process satisfying those conditions. While, the second definition is, the a_t is called a strong GARCH(r,s) process if

$$\begin{cases} a_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i a_{t-1}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-1}^2 \end{cases}$$

where the α_i and β_j are nonnegative constants $(\alpha_i \geq 0, \beta_i \geq 0)$ and α_0 is a (strictly) positive constant $(\alpha_0 > 0)$. The second definition is more restrictive but allows explicit solutions to be obtained, which is introduced by Bollerslev. If r = 0, the GARCH(r, s) process reduces to the ARCH(s) process, and for r = s = 0, a_t is simply white noise. In the ARCH(s) process, the conditional variance is specified as a linear function of past sample variances only, whereas the GARCH(r, s) process allows lagged conditional variances to enter as well (Bollerslev, 1986).

In the combination model of BJ-G, a two-phase procedure is proposed. In the first phase, the best model identified from the Box-Jenkins models is first used to model the mean data of time series and the residuals of this model will then be investigated for heteroscedasticity to detect the existence of volatility in the data series. In the second phase, the GARCH is used to model the variance equation of the residuals. In this combination model, the Box-Jenkins model with GARCH error components is applied to analyse the univariate series and to predict the values of approximation series (Chen et al., 2011; Liu & Shi, 2013; Tan et al., 2010; Zhou et al., 2006). In this procedure, the error term a_t of the Box-Jenkins model is said to follow a GARCH process of orders r and s. The flowchart of this combination procedure of Box-Jenkins with standard GARCH can be summarised as shown in Figure 3.4 and this procedure is applicable for other GARCH-type models. Note that, the error distribution for GARCH model denoted by ε_t is an independent and identically distributed (IID) as normal, t and GED distribution (Tsay, 2013).

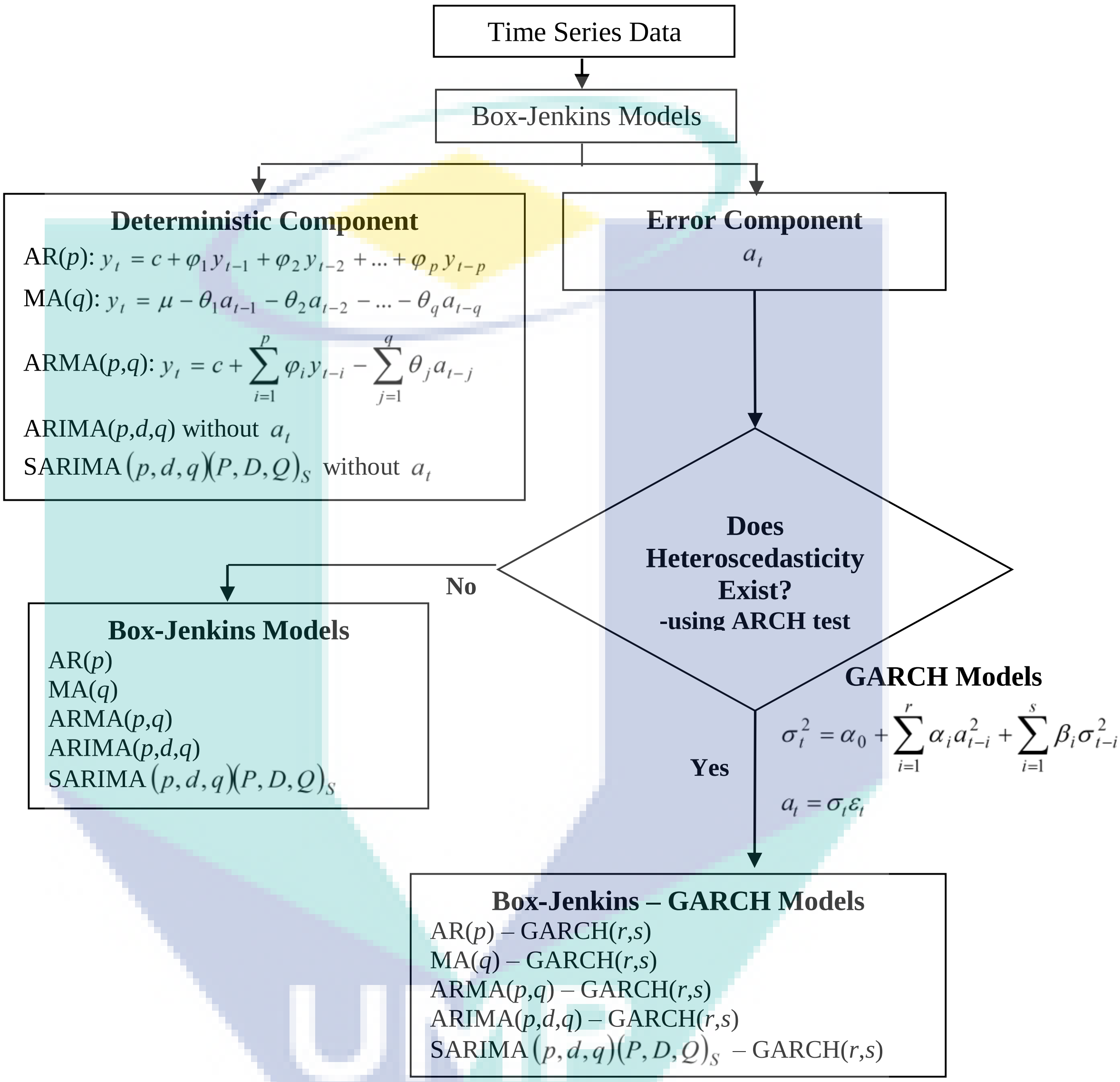


Figure 3.4 Procedure of combination of BJ-G, specifically with standard GARCH

The ACF and PACF of the squared residuals help to specify the GARCH orders, r and s, respectively (Pham & Yang, 2010). In practice, the modelling procedure of GARCH model consists of three steps as follows:

Step 1: Build an econometric model for the observed data and remove any serial correlation in the data. Use the residual series of the econometric model to check GARCH effects. The ARCH test and Lagrange Multiplier test are used to check conditional heteroscedasticity.

Step 2: ACF and PACF of the squared residuals help to specify the GARCH orders, r and s, respectively. However, normally these orders are not very accurate.

Step 3: To check the adequacy of the fitted model, Ljung - Box Q-test (LBQ-test) of estimated residuals is used for mean model and squares of estimated residuals for the variance model.

3.4 The Proposed Modified Procedure of Box-Jenkins – GARCH Model for Modelling and Forecasting Highly Volatile Time Series Data

Figure 3.5 illustrates the new procedure of Box-Jenkins – GARCH (or BJ-G) model for modelling and forecasting highly volatile time series data as proposed by the current study. The proposed procedure takes into account the steps that were usually omitted by researchers but yet should be considered to ensure the validity of the Box-Jenkins results specifically to highly volatile time series data. Before the four stages as shown in Figure 3.1 are conducted, the time series need to be partitioned into in-sample and out-of-sample series, in a typical ratio of 90:10 (Chatfield, 2001). The in-sample series is used to estimate model which involves the identification, parameter estimation and diagnostic checking stages. Meanwhile, the out-of-sample series will be used to validate the model developed in the previous stage. This cross-validation procedure has been proposed and practiced by Box and Jenkins (Box & Jenkins, 1968), although it is not clearly emphasised. It is stated that cross-validation procedure is a more sophisticated version of in-sample/out-of-sample in evaluating forecast accuracy (Hyndman & Athanasopoulos, 2014).

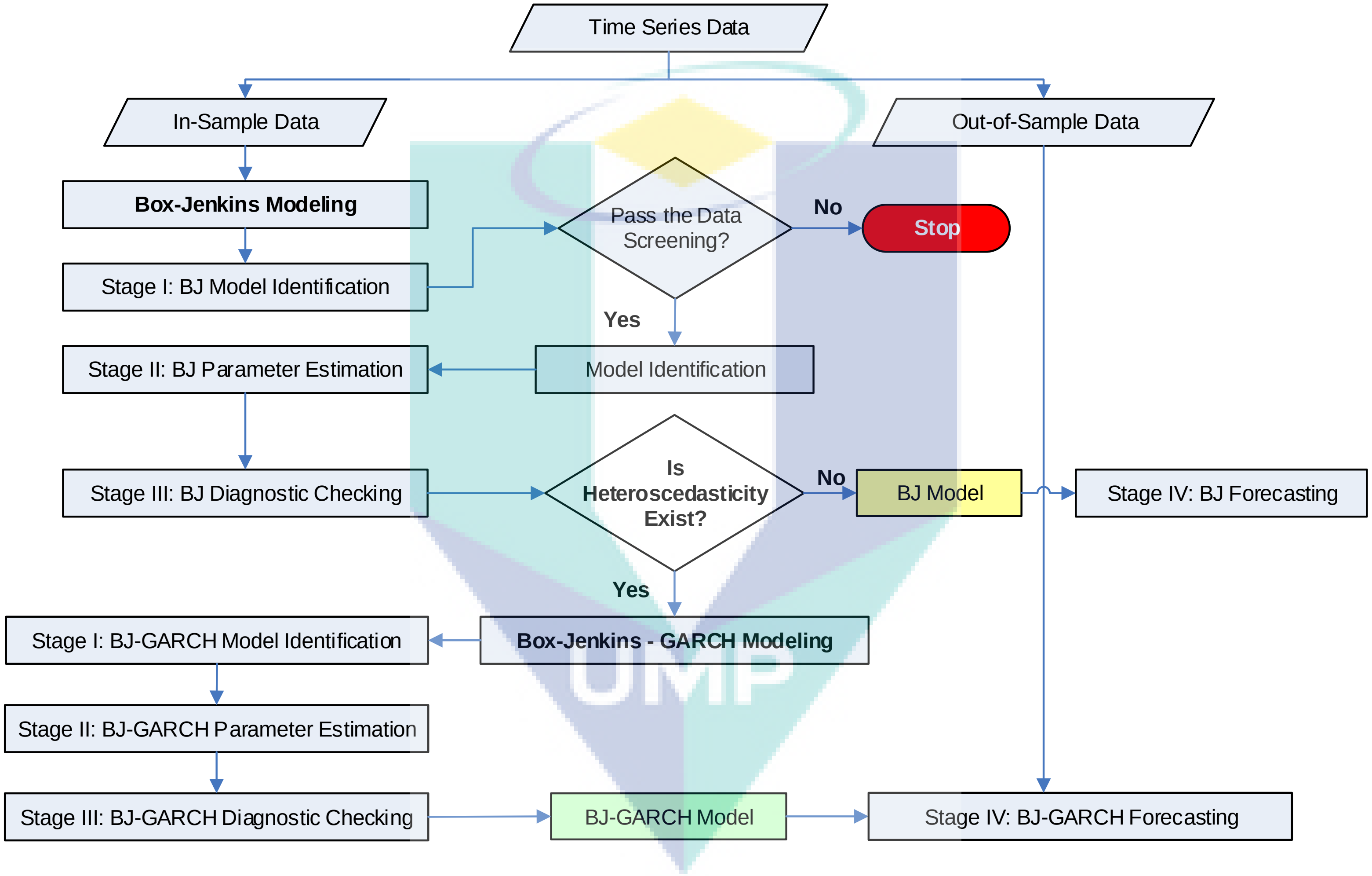


Figure 3.5 Proposed procedure of BJ-G for highly volatile time series data (Note: Box-Jenkins is abbreviated as BJ)

In general, time series data go through the same identification and parameter estimation stages as the standard Box-Jenkins procedures. However, for a highly volatile time series data, the existence of heteroscedasticity in the diagnostic checking stage will violate the assumption of constant variance in the Box-Jenkins models. Therefore, a combination of BJ-G is proposed in the standard Box-Jenkins procedures in handling volatility in the data series, namely the procedure of BJ-G. The details of each stage for the proposed procedure are described in Section 3.4.1 to Section 3.4.4.

3.4.1 Stage I: Model Identification

In the model identification stage, the procedures can be divided into two parts, data screening and model identification, as can be illustrated in Figure 3.6. In the data screening part, there are five procedural steps that should be considered in this proposed procedure, which are (i) data plotting, (ii) descriptive statistics, (iii) data stationarity, (iv) preliminary linearity test, and (v) Portmanteau test.

In practice, the first step of data screening is by plotting a time series graph to detect three important characteristics of the data series: (i) variation that increases or decreases as the series increases (nonstationary in-variance), (ii) occurrence of an upward or downward trend (nonstationary in-mean), and (iii) seasonality; in tracing the data series movement. Meanwhile, the part of descriptive statistics is important in developing the basic statistical measures such as mean, variance, skewness and kurtosis in studying the characteristics of a time series data. These basic statistical measures can be explained using the concept of moments of a random variable, as can be referred to Section 3.4.1.1.

In dealing with a time series data via the Box-Jenkins modelling, stationarity is one of the important aspects that need to be considered, as thoroughly explained in Section 3.4.1.2. Stationarity in data can be classified into two: (i) stationarity in-variance and (ii) stationarity in-mean. Whenever the data exhibits a large variation within a given period, the stationarity in-variance can be achieved using suitable transformation.

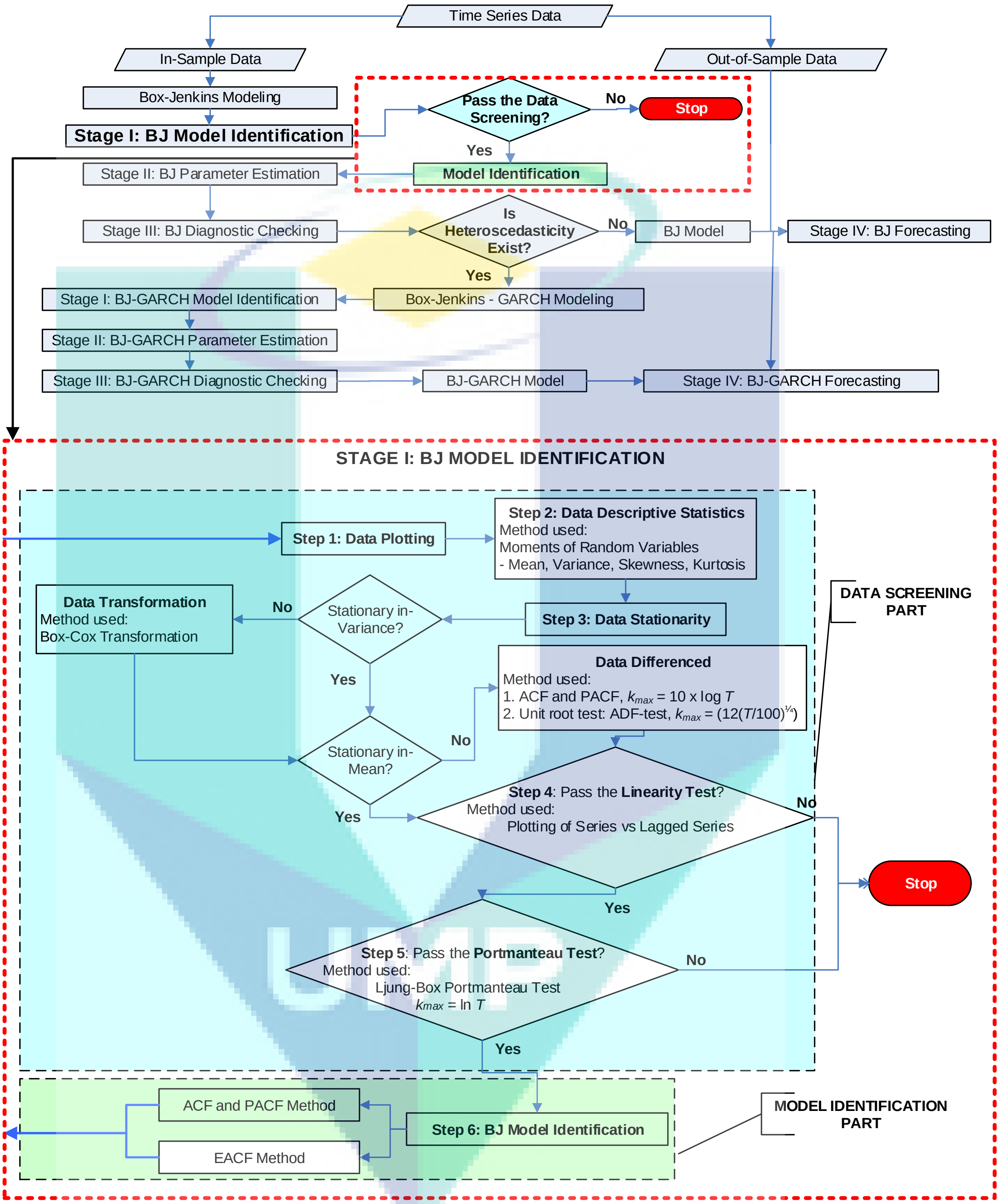


Figure 3.6 Detail procedures in Stage I of the procedure of BJ-G for highly volatile data

If the series is nonstationary in-variance, the data series need to be transformed first in order to stabilise the variance. The objective of the transformation is to simplify the pattern in the historical data by removing known sources of variations or by making the pattern more consistent across the whole data set since simpler patterns usually lead to more accurate forecasts (Hyndman & Athanasopoulos, 2014). When dealing with historical data which exhibit increasing or decreasing pattern as the series increases, the data transformation is required for better results. In this study, the Box-Cox transformation is used as the data transformation method due to its potential best practice in normalising data, stabilising variance and reducing heteroscedasticity (Box & Cox, 1964; Osborne, 2010). The details of Box-Cox transformation method can be referred to Section 3.4.1.2 (a).

When the variance in the data series is stabilised, the trend in-mean is investigated. If the series shows a trend either upward or downward within a given period, which demonstrates nonstationary in-mean, then the series need to be differenced to achieve the stationarity. The stationarity in-mean can be graphically identified and tested using the concept of autocorrelation functions (ACF) and partial autocorrelation functions (PACF) as the theory for the concepts can be referred to Section 3.4.1.2 (b(i)). On the other hand, the stationarity in-mean can be checked statistically using the unit root test such as the Augmented Dickey-Fuller test (ADF-test), as presented in Section 3.4.1.2 (b(ii)).

Once the time series data has achieved the stationarity, the preliminary linearity test is suggested in the proposed procedure as shown in Figure 3.5 to verify that a linear model, specifically the Box-Jenkins model, is appropriate to model the stationary data, as can be referred to Section 3.4.1.3. Then, the Portmanteau test using LBQ-test, as discussed in Section 3.4.1.4, is recommended to check the existence of serial correlations in the linear series since the Box-Jenkins model is only valid for correlated series. However, these two procedures have always been neglected in many studies, even though these two procedures are important as an initial procedure either to continue or to stop in applying the Box-Jenkins modelling in analysing as well as forecasting data series.

After the data series pass the screening part, the data is well prepared for the model identification part. In practice, the ACF and PACF of the stationary data will be used to identify the possible models of Box-Jenkins. However, the autocorrelations function method is uncertain and not quite informative in identifying the appropriate Box-Jenkins model. Hence, in this proposed procedure, the extended autocorrelation function (EACF) method is strongly recommended to be used as an alternative to identify the most appropriate order of the Box-Jenkins model. The details of the EACF approach can be referred to Section 3.4.1.5.

3.4.1.1 Moments of a Random Variable

Volatile time series always arises in financial data. In finance, the first fourth moment of a random variable are used to describe the behaviour of asset prices and returns (Tsay, 2013). The prices and returns are considered as continuous data since it takes any value in the interval on a given number line. By definition, the ℓ th moment of a continuous random variable *X* about the origin is defined as

$$m'_\ell = E\left(X^\ell\right) = \int_{-\infty}^{\infty} x^\ell f(x)\,dx, \tag{3.8}$$

where $E\left(X^\ell\right)$ denotes the expectation of X^ℓ and $f(x)$ is the probability density function (pdf) of *X*. The first moment is defined by

$$m'_1 = E(X) = \int_{-\infty}^{\infty} x\ f\ (x)\,dx. \tag{3.9}$$

The first moment is known as the mean, denoted by μ_X , $m'_1 = \mu_X = E(X)$. It measures the central location of the data series.

The ℓ th central moment of *X* about the origin is defined as in Equation 3.10, provided that the integral exists. Therefore, the second central moment m_2 is defined by Equation 3.11, is known as the variance of *X*, denoted by σ_X^2 which measures the variability of the data series. The positive square root of variance represents the standard deviation of *X*, denoted by σ_X . For asset prices and returns, variance (or standard deviation) is a measure of uncertainty and often used as a risk measure.

$$m_\ell = E\left[(X - \mu_X)^\ell\right] = \int_{-\infty}^{\infty} (x - \mu_X)^\ell f(x)\,dx \tag{3.10}$$

$$m_2 = E\left[(X - \mu_X)^2\right] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\,dx \tag{3.11}$$

The third central moment, m_3 relates to the symmetry of the data series with respect to its mean. The normalisation of the third central moment is called the skewness of *X*, denoted by *S*(*x*). The skewness indicates the degree of asymmetry of the distribution about the mean, which is defined as Equation 3.12.

$$S(x) = E\left[\frac{(X-\mu_X)^3}{\sigma_X^3}\right] \qquad 3.12$$

Meanwhile, the normalisation of the fourth central moment is called the kurtosis of *X*, denoted by *K*(*x*). It measures the tail thickness of the distribution of *X*. The kurtosis indicates the peakedness of a distribution about its mean which is defined as in Equation 3.13.

$$K(x) = E\left[\frac{(X-\mu_X)^4}{\sigma_X^4}\right] \qquad 3.13$$

The quantity of $K(x)-3$ is called the excess kurtosis since $K(x)=3$ is for normal distribution. A distribution with positive excess kurtosis is said to have heavy tails, implying that the distribution has more mass on the tails as compared to normal distribution. This means that a random variable from such distribution contains more extreme values, and the distribution is said to be leptokurtic. On the other hand, a distribution with negative excess kurtosis is said to have short tails, and the distribution is said to be platykurtic.

In finite sample, the moments of a random variable can be estimated. Let $\{x_1, x_2, \ldots, x_T\}$ be a random sample of *X* with *T* observations. The estimated sample mean, sample variance, sample skewness and sample kurtosis are given by Equation 3.14 to Equation 3.17, respectively. The hypothesis and its testing regarding the mean, skewness and kurtosis of the series can be summarised as in Table 3.1.

$$\hat{\mu}_X = \frac{1}{T}\sum_{t=1}^{T} x_t \qquad 3.14$$

$$\hat{\sigma}_X^2 = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \hat{\mu}_X)^2 \qquad 3.15$$

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_X^3} \sum_{t=1}^{T} (x_t - \hat{\mu}_X)^3 \qquad 3.16$$

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_X^4} \sum_{t=1}^{T} (x_t - \hat{\mu}_X)^4 \qquad 3.17$$

Table 3.1 Test of hypothesis for mean, skewness and kurtosis for a data series

Hypothesis	**Test statistic**	**Critical point**	**Decision rule**
$H_0 : \mu_X = 0$ $H_1 : \mu_X \neq 0$	$t_{\text{test}} = \frac{\sqrt{T}\hat{\mu}_X}{\hat{\sigma}_X}$	$t_{\frac{\alpha}{2},T-1}$ or $Z_{\frac{\alpha}{2}}$	reject H_0 if $\lvert t_{\text{test}} \rvert > t_{\frac{\alpha}{2},T-1}$ *or $\lvert t_{\text{test}} \rvert > Z_{\frac{\alpha}{2}}$ or p-value $\leq \alpha$
$H_0 : S(x) = 0$ $H_1 : S(x) \neq 0$	$t_{\text{test}} = \frac{\hat{S}(x)}{\sqrt{\frac{6}{T}}}$	$t_{\frac{\alpha}{2},T-1}$ or $Z_{\frac{\alpha}{2}}$	reject H_0 if $\lvert t_{\text{test}} \rvert > t_{\frac{\alpha}{2},T-1}$ *or $\lvert t_{\text{test}} \rvert > Z_{\frac{\alpha}{2}}$ or p-value $\leq \alpha$
$H_0 : K(x) - 3 = 0$ $H_1 : K(x) - 3 \neq 0$	$t_{\text{test}} = \frac{\hat{K}(x) - 3}{\sqrt{\frac{24}{T}}}$	$t_{\frac{\alpha}{2},T-1}$ or $Z_{\frac{\alpha}{2}}$	reject H_0 if $\lvert t_{\text{test}} \rvert > t_{\frac{\alpha}{2},T-1}$ *or $\lvert t_{\text{test}} \rvert > Z_{\frac{\alpha}{2}}$ or p-value $\leq \alpha$

*For a sufficiently large *T*, the test statistic approaches a standard normal distribution.

In the hypothesis testing on μ_X, the property of the consistent estimate of μ_X under rather weak conditions, specifically $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{T}\right)$, makes the test statistic approaches a standard normal distribution for a sufficiently large *T*. While, according to Snedecor and Cochran (1980), if *X* is a normal random variable, then $\hat{S}(x)$ and $\hat{K}(x) - 3$ are distributed asymptotically as normal with zero mean and variances $\frac{6}{T}$ and $\frac{24}{T}$, respectively. Jarque and Bera proposed a normality test based on these asymptotic properties, which can be used to test the normality of the asset prices and returns (Jarque & Bera, 1987).

3.4.1.2 Stationarity in Time Series

A time series is a set of observations sequentially in time, hence a time series analysis is about the study of data collected through time. A stationary process is a special class of stochastic processes, which is based on the assumption that the process is in a particular state of statistical equilibrium. A stochastic process is a statistical phenomenon that evolves in time according to probabilistic laws or can be said as a model that describes the probability structures of a sequence of observations. A stochastic process is said to be strictly stationary if its properties are unaffected by a change of time origin, that is, if the joint probability distribution associated with *T* observations $y_{t_1}, y_{t_2}, ..., y_{t_T}$, made at any set of times $t_1, t_2, ..., t_T$, is the same as the *T* observations $y_{t_1+k}, y_{t_2+k}, ..., y_{t_T+k}$, made at times $t_1 + k, t_2 + k, ..., t_T + k$. Thus, for a time series to be strictly stationary, the joint distribution of any set of observations must be unaffected by shifting all the times of observation forward or backward by any integer of *k*.

The stationarity assumption implies that the probability distribution $f(y_t)$ is the same for all times *t* and may be written as $f(y)$. Hence, the stochastic process has a constant mean as given in Equation 3.18,

$$\mu = E(y_t) = \int_{-\infty}^{\infty} y \ f(y)\,dy \tag{3.18}$$

which defines the level about which it fluctuates, and a constant variance as given by Equation 3.19,

$$\sigma_y^2 = E\left[(y_t - \mu)^2\right] = \int_{-\infty}^{\infty} (y - \mu)^2 \ f(y)\,dy \tag{3.19}$$

which measures its spread on this level. In practice, the mean μ and the variance σ_y^2 of the stochastic process can be estimated by the sample mean and the sample variance of the time series, given by Equation 3.20 and Equation 3.21, respectively (Tsay, 2013).

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \tag{3.20}$$

$$s_y^2 = \frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y})^2 \tag{3.21}$$

The stationarity assumption also implies that the joint probability distribution $f(y_{t_1}, y_{t_2})$ is of constant interval for all times t_1, t_2. Under the stationarity assumption, the covariance between y_t and y_{t+k}, separated by *k* intervals of time or by lag *k*, must be the same for all times *t*. In time series, covariance is used to investigate how observations are related to each other in time and it measures the degree of second order variation between two data at two different times. Therefore, the covariance of y_t and y_{t+k} is known as the autocovariance coefficient at lag *k*, denoted by γ_k, is given as Equation 3.22. On the other hand, a stochastic process is weakly stationary if the mean μ is a fixed constant for all times *t* and the covariance γ_k depends only on the time difference or time lag *k* for all times *t*.

$$\gamma_k = \text{cov}[y_t, y_{t+k}] = E[(y_t - \mu)(y_{t+k} - \mu)] \qquad 3.22$$

Similarly, the autocorrelation coefficient at lag *k* or the correlation between y_t and y_{t+k}, denoted by ρ_k, is defined by Equation 3.23,

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E\left[(y_t - \mu)^2\right]E\left[(y_{t+k} - \mu)^2\right]}} = \frac{Cov[y_t, y_{t+k}]}{\sigma_y^2} = \frac{\gamma_k}{\gamma_0} \qquad 3.23$$

since for a stationary process, the variance $\sigma_{y_t}^2 = \sigma_{y_{t+k}}^2$ and $\sigma_y^2 = \gamma_0$. Particularly, if $k = 0,$ Equation 3.23 implies that $\rho_0 = 1$.

The covariance matrix of symmetric form, Γ_T associated with a stationary process for observations $y_1, y_2, ..., y_T$ made at *T* successive time is given by

$$\Gamma_T = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \cdots & \gamma_0 \end{bmatrix} = \sigma_y^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{T-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix} = \sigma_y^2 \mathrm{P}_T$$

where $\gamma_T = \sigma_y^2 \rho_T$, refer to Equation 3.23. The covariance matrix is formed from the autocovariance coefficients associated with constant elements on any diagonal, is called an autocovariance matrix. The corresponding correlation matrix P_T is called an autocorrelation matrix. Both Γ_T and P_T are positive definite for any stationary process (Box et al., 2008). This positive definiteness of the P_T indicates that its determinant and all principal minors are greater than zero, which then implies that $-1 < \rho_1, \rho_2, ..., \rho_q < 1$ for q orders.

The plot of γ_k versus lag k is called the autocovariance function of the stochastic process, similarly the plot of ρ_k versus lag k is called the autocorrelation function (ACF) of the process. The autocovariance function and the ACF are denoted by $\{\gamma_k\}$ and $\{\rho_k\}$, respectively. The $\{\rho_k\}$ is a plot of the diagonals of P_T and is necessarily symmetric about zero, implies $\rho_k = \rho_{-k}$. Examples of ACF plot is given in Figure 3.7.

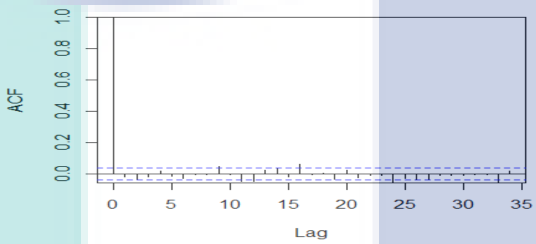


Figure 3.7 ACF plot

Therefore, it can be concluded that a stationarity process for y_t is completely characterised by its mean μ and its autocovariance function $\{\gamma_k\}$. The stationarity process can also be characterised by its mean μ, variance σ_y^2 and autocorrelation function $\{\rho_k\}$ or ACF. However, the later approach is mostly used in practical applications. In time series modelling, the model cannot be directly applied if the series is nonstationary. A stationary time series is one whose properties do not depend on the time at which the series is observed or more precisely, if y_t is a stationary time series, then for all n, the distribution of $(y_t, y_{t+1,...,} y_{t+N})$ does not depend on t. Figure 3.8 illustrates the case of stationary series which shows that the series appear to vary about a fixed level, that is zero.

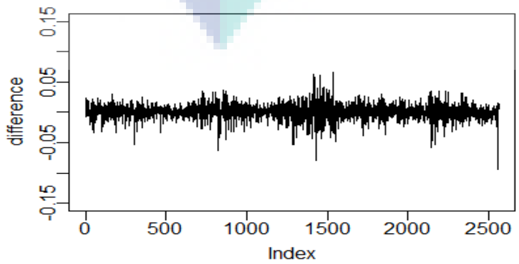


Figure 3.8 Stationary series

White noise process is an example of a stationary process. It is a sequence of IID random variables, denoted as $a_1, a_2, ..., a_T$, which is assumed to have mean zero and variance σ_a^2. The independence implies that a_t are uncorrelated, hence its autocovariance is given by Equation 3.24.

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 , & k = 0 \\ 0 \quad , & k \neq 0 \end{cases} \tag{3.24}$$

Since stationarity is the initial aspect that needs to be considered when dealing with time series data, therefore it is important to know whether the data contains any trend or seasonal characteristics. The time series with trends or with seasonal characteristics are not stationary since the trend and seasonality will affect the value of the time series at different times. The graphical representation for cases of nonstationary series is illustrated in Figure 3.9. Figure 3.9(a) shows the series does not vary about a fixed level, exhibits an overall upward trend and the variances increases as the series increases. Time series that exhibit these phenomena are said to be nonstationary in-mean and in-variance. Figure 3.9(b) presents the case of nonstationary in-mean since it exhibits a trend in the series. The case of seasonality series is shown in Figure 3.9(c), since the series presents a characteristic of repetitive pattern in nature. Meanwhile, Figure 3.9(d) illustrates the case of seasonality and nonstationary in-mean and in-variance.

The stationarity process for the case illustrated in Figure 3.9(a) is presented in Figure 3.10. Figure 3.10(a) shows the observed series of daily gold prices in the Malaysia market from year 2003 to 2014, which indicates the case of nonstationary in-mean and in-variance, as well as exhibiting overall upward and nonseasonal trends. In order to make such data series stationary, data transformation and data differencing are needed. The log transformation is suggested to handle nonstationary in-variance based on certain criteria and the differencing method is used to cater nonstationary in-mean.

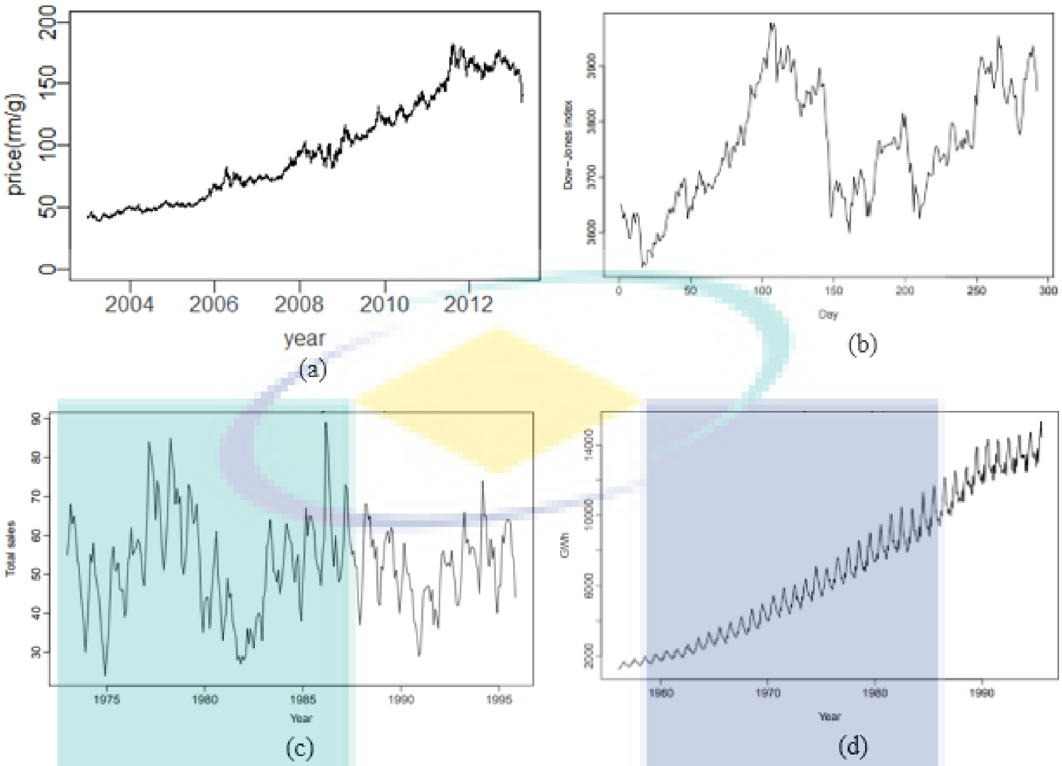
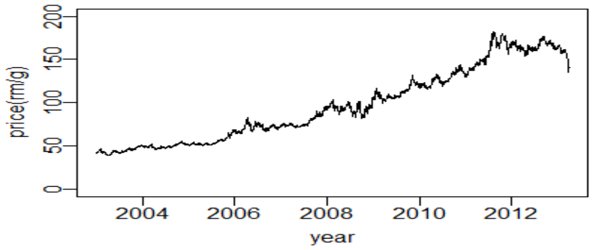
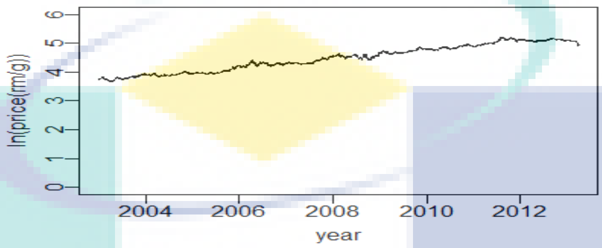


Figure 3.9 Graphical representations for several cases of nonstationary series

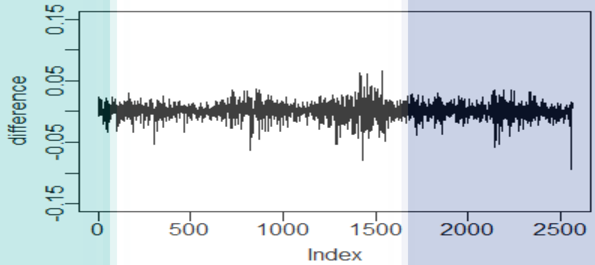
The plot of transformed series is shown in Figure 3.10(b). After the transformation, the series is found to be less volatile, however, the upward trend still exists in the series, therefore the transformed series need to be differenced. The differenced process is given by $y_t^* = \log y_t - \log y_{t-1}$, where y_t^* is the transformed data at time t and graph for the differenced data is shown in Figure 3.10(c). Figure 3.10(c) shows the series is stationary after the first differenced on the transformed series, which presents that the transformation and differencing have made the nonstationary series into stationary series. The stationary series presents variances that change over time as shown in Figure 3.10(c) and it is usually observed in financial time series data. This volatility behaviour (variances change over time) can be captured by conditional heteroscedasticity models such as ARCH or GARCH models. If the first differenced is insufficient to make the series stationary, Box and Jenkins suggested that the second differenced is always sufficient for most series to achieve stationarity (Box & Jenkins, 1968).



(a)



(b)



(c)

Figure 3.10 The plots of nonstationary series to obtain stationary series after transforming and differencing

3.4.1.2(a) Stationarity in-Variance: Box-Cox Transformation Method

The Box-Cox transformation has found more practical data transformation in a variety of fields, especially in econometrics (Sakia, 1992). The Box-Cox transformation is popular in financial time series analysis and has been considered in forecasting volatility (Gonçalves & Meddahi, 2011; Higgins & Bera, 1992). The use of Box-Cox power transformation in improving forecasting accuracy is also supported by many researchers (Lee, Sadaei, & Suhartono, 2013; Luetkepohl & Xu, 2011).

The use of Box-Cox transformation which is a preliminary step in the identification stage of fitting Box-Jenkins model was recommended by Box and Jenkins for seasonal series (Box & Jenkins, 1976). Box and Jenkins suggest power or log transformations to achieve stationarity in the variance of the time series data. The power and log transformation is known as the Box-Cox transformation named after Box and

Cox (Box & Cox, 1964). The Box-Cox transformation is a modification from a family of power transformations introduced by Tukey (Tukey, 1957). The formula of the Box-Cox transformation for positive series, $y_t > 0$ is given in Equation 3.25,

$$y_t^* = \begin{cases} \dfrac{y_t^{\lambda} - 1}{\lambda}, & \text{for } \lambda \neq 0 \\ \log_e(y_t), & \text{for } \lambda = 0 \end{cases} \tag{3.25}$$

where y_t is the actual data at time *t*, y_t^* is the transformed data at time *t*, and λ is the minimum residual mean square error value.

Box and Cox proposed the Box-Cox transformation as a solution to satisfy these assumptions: (i) the variables (or their error terms) are normally distributed, (ii) the additivity of the errors structures (or mean model is linear), and (iii) the variance of error terms is homoscedastic. However, the data transformation approach is also well applied to the case of heteroscedastic in the error terms as well as able to reduce the noise and volatility effect in the data (Gonçalves & Meddahi, 2011; Lee et al., 2013). According to Osborne (2010), Box-Cox transformation represents a potential best practice whenever normalising data or equalising variance is desired. Frequently, the transformation not only stabilises the variance, but also improves the approximation of the distribution by normal distribution. Therefore, the Box-Cox transformation can be a solution for simultaneously correcting normality, linearity and reducing volatility in the variance (Box & Cox, 1964; Osborne, 2010; Sakia, 1992).

The Box-Cox transformation represents a family of power transformations that incorporates and extends the conventional options to help researchers easily find the optimal transformation of data. The approach suggests a λ value that corresponds to an understandable transformation to make it easier to transform the data back (back-transform) to obtain forecasts on the original scale. Table 3.2 shows some commonly used values of λ and its associated transformation.

Table 3.2 Some commonly used values of λ and its associated transformation

Values of λ	**-1.0**	**-0.5**	**0**	**0.5**	**1.0**
Transformation	$\dfrac{1}{y_t}$	$\dfrac{1}{\sqrt{y_t}}$	$\log_e y_t$	$\sqrt{y_t}$	y_t

As reported by the previous literatures, it is recommended to use log function as the power transformation especially for financial data (Luetkepohl & Xu, 2011; Proietti & Lütkepohl, 2013). Note that the log function is a subset of the class of Box-Cox transformation whenever λ equals to zero. According to Nelson and Granger (1979), the log transformation is frequently used by econometricians, either because the change in logarithm of variables approximates percentage changes, or rate of return, or because it is observed that the variability of a series appears to be related to the level, so that using logarithms may produce relationships with more homogeneous residual. The use of log function in the financial data also seems related to the distribution of data itself since most of the financial data tends to be positively skewed and the logarithmic transformation is recommended for positively skewed data (Olivier & Norberg, 2010).

According to Box and Cox (1964), in estimating the λ, an assumption is made that for some unknown λ, the transformed data y_i^*, for $i = 1,2,\ldots,T$, satisfies the full normal theory assumptions (i.e. independently normally distributed with constant variance σ^2) and with expectation $E\left(y^{(\lambda)}\right) = a\theta$, where a is known $(T \times T)$ matrix and θ is a $(T \times 1)$ vector of unknown parameters associated with the transformed data, or can be simplified as $y_t^{(\lambda)} \sim NID\left(a\theta, \sigma^2\right)$.

Consider a time series $x_1, x_2, \ldots, x_T$ with normal pdf, as given by

$$f\left(x_t; \mu, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_t - \mu\right)^2\right).$$

According to Box and Cox (1964), the probability density for the untransformed (original) data y_t is obtained by multiplying the normal density by the Jacobian of the transformation. By considering the assumption of $y_t^* \sim NID\left(a\theta, \sigma^2\right)$, the pdf of y_t is given by Equation 3.26.

$$f\left(y_t; a\theta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(y_t - a\theta\right)^2\right)\frac{dy_t^*}{dy_t} \qquad 3.26$$

The likelihood function (L) is the product of the pdf for each y_t, given by

$$L(\lambda; y_i^*) = \prod_{i=1}^{T} f_{Y_i}(y_i^*, a\theta, \sigma^2)\frac{dy_i^*}{dy_i}$$

$$= \left(f_{Y_1}(y_1^*, a\theta, \sigma^2) f_{Y_2}(y_2^*, a\theta, \sigma^2) \ldots f_{Y_T}(y_T^*, a\theta, \sigma^2)\right)\left(\prod_{i=1}^{T}\left|\frac{dy_i^*}{dy_i}\right|\right)$$

$$= \left(\left(2\pi\sigma^2\right)^{-\frac{T}{2}}\left(e^{\left(-\frac{1}{2\sigma^2}\left(y_1^*-a\theta\right)^2\right)+\left(-\frac{1}{2\sigma^2}\left(y_2^*-a\theta\right)^2\right)+\ldots+\left(-\frac{1}{2\sigma^2}\left(y_T^*-a\theta\right)^2\right)}\right)\right)\left(\prod_{i=1}^{T}\left|\frac{dy_i^*}{dy_i}\right|\right)$$

$$= \left(2\pi\sigma^2\right)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{T}\left(y_i^* - a\theta\right)^2\right) J(\lambda; y)$$

where $J(\lambda; y) = \prod_{i=1}^{T}\left|\frac{dy_i^*}{dy_i}\right|$. By considering y_t^*, a and θ are $(T \times 1)$, $(T \times T)$ and $(T \times 1)$ matrices, respectively, therefore

$$y_t^* - a\theta = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_T^* \end{bmatrix} - \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1T} \\ a_{21} & a_{22} & \ldots & a_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ a_{T1} & a_{T2} & \ldots & a_{TT} \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_T \end{bmatrix} = \begin{bmatrix} y_1^* - (a_{11}\theta_1 + a_{12}\theta_2 + \ldots + a_{1T}\theta_T) \\ y_2^* - (a_{21}\theta_1 + a_{22}\theta_2 + \ldots + a_{2T}\theta_T) \\ \vdots \\ y_T^* - (a_{T1}\theta_1 + a_{T2}\theta_2 + \ldots + a_{TT}\theta_T) \end{bmatrix}$$

Then, by multiplying the transpose matrix $(y_t^* - a\theta)'$ to the matrix $(y_t^* - a\theta)$, the product can be expressed as

$$\left(y_t^* - a\theta\right)'\left(y_t^* - a\theta\right) = \left[y_1^* - (a_{11}\theta_1 + a_{12}\theta_2 + \ldots + a_{1T}\theta_T)\right]^2 + \left[y_2^* - (a_{21}\theta_1 + a_{22}\theta_2 + \ldots + a_{2T}\theta_T)\right]^2$$

$$+ \ldots + \left[y_T^* - (a_{T1}\theta_1 + a_{T2}\theta_2 + \ldots + a_{TT}\theta_T)\right]^2$$

$$= \left(y_1^* - a\theta\right)^2 + \left(y_2^* - a\theta\right)^2 + \ldots + \left(y_T^* - a\theta\right)^2$$

$$= \sum_{i=1}^{T}\left(y_i^* - a\theta\right)^2$$

Therefore, the likelihood function can be simplified as Equation 3.27.

$$L\left(\lambda; y_i^*\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{T}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left(\left(y^* - a\theta\right)'\left(y^* - a\theta\right)\right)\right) J(\lambda; y) \qquad 3.27$$

Box and Cox applied the MLE approach in estimating parameter of λ. The MLE method was developed by Fisher in the 1920s and since then it has been widely used due to sufficiency, consistency and efficiency properties it has (Myung, 2003). Therefore, by taking the natural logarithm of the likelihood function,

$$\begin{aligned}
\ln L\left(\lambda; y_i^*\right) &= \log_e\left(L\left(\lambda; y_i^*\right)\right) \\
&= \log_e\left(\frac{1}{\left(2\pi\sigma^2\right)^{\frac{T}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left(\left(y^* - a\theta\right)'\left(y^* - a\theta\right)\right)\right) J(\lambda; y)\right) \\
&= \log_e\left(2\pi\sigma^2\right)^{-\frac{T}{2}} + \log_e\left(\exp\left(-\frac{1}{2\sigma^2}\left(\left(y^* - a\theta\right)'\left(y^* - a\theta\right)\right)\right)\right) + \log_e J(\lambda; y) \\
&= -\frac{T}{2}\log_e(2\pi) - \frac{T}{2}\log_e \sigma^2 - \frac{1}{2\sigma^2}\left(\left(y^* - a\theta\right)'\left(y^* - a\theta\right)\right) + \log_e J(\lambda; y)
\end{aligned}$$

Thus, for fixed λ, the maximised log likelihood is given by Equation 3.28,

$$\ln L_{\max}(\lambda) = -\frac{T}{2}\log_e \hat{\sigma}^2(\lambda) + \log_e J(\lambda; y) \qquad 3.28$$

where $\hat{\sigma}^2(\lambda) = \frac{S(\lambda)}{T}$ and $S(\lambda)$ is the residual sum of squares in the analysis of variance of y^*.

Gold price consists of positive value, therefore the data considered in this study is the case for $y_t > 0$. According to Box and Cox, the second term in Equation (3.28) is replaced by $(\lambda - 1)\sum_{t=1}^{T}\log_e y_t$ in the case of $y_t > 0$ (Box & Cox, 1964). By substituting $S(\lambda)$ and the updated second term to Equation 3.28, therefore the maximised log likelihood becomes an expression as Equation 3.29.

$$\ln L_{\max}(\lambda) = -\frac{T}{2}\log_e\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_t^* - \bar{y}_t^*\right)^2\right] + (\lambda - 1)\sum_{t=1}^{T}\log_e y_t \qquad 3.29$$

Then, by plotting $\ln L_{\max}(\lambda)$ versus λ for a trial series of values, the maximizing value of $\hat{\lambda}$ may be read-off, with $100(1-\alpha)\%$ confidence interval as given by $\ln L_{\max}(\hat{\lambda}) - \ln L_{\max}(\lambda) < \frac{1}{2}\chi^2_{\upsilon_\lambda}(\alpha)$, where υ_λ is the number of independent components in λ.

Since λ is employed significantly in the Box-Cox transformation and the transformation is widely applied in analysing and forecasting data, many software offer the plotting with estimated value of λ with its 95% confidence interval based on the Box-Cox power transformation approach, such as Minitab and R language. There are many functions in the forecast packages in R that is specially built to compute and plot λ for the Box-Cox transformation such as AID package (Asar, Ilk, & Dag, 2017), `BoxCox` function in `car` package (Fox, Weisberg, Adler, & Bates, 2015), `BoxCox` function in `MASS` package (Ripley, Venables, Bates, & Hornik, 2017) and the `BoxCox` function in `forecast` package (Hyndman et al., 2015). In this study, the estimation of λ is obtained using `BoxCox` function in `forecast` package or using `AID` package. However, there are limitations of the R packages such as the `forecast` package give the value of λ without plotting and the `AID` package is only valid for $T \leq 5000$. Therefore, for the case of $T > 5000$, the estimated λ can be obtained using Minitab.

3.4.1.2(b) Stationary in-Mean

Most volatile time series specifically in finance and economics exhibit trending behaviour or nonstationary in the mean. Since the data should be in a stationary form to be analysed, then if the data exhibits trending characteristics, some form of trend removal is required. There are two common procedures to remove the trend that are the autocorrelation functions (ACF and PACF) and the unit root test methods.

(i) Sample Autocorrelation Function and Sample Partial Autocorrelation Function

Box and Jenkins proposed to use the ACF and the PACF of the sample data as the basic tools in checking the stationarity in the mean as well as to identify the order of the time series model (Box & Jenkins, 1968). They provided both a theoretical framework and practical rules for determining appropriate values for *p* and *q* as well as their seasonal counterparts of *P* and *Q* by using the ACF and the PACF. The use of autocorrelation functions in a linear time series model is able to capture the linear dynamic of the data (Tsay, 2005). The ACF and the PACF provide a useful measure of the degree of dependence between values of a time series at specific intervals of separation and play an important role in the prediction of future values (Boland, 2008). On the other hand, the ACF and PACF of the squared residuals from a stationary series are used to get the possible values of r and s, respectively, for the $\text{GARCH}(r,s)$ models.

The ACF is a measure of the linear relationship between time series observations separated by some time period, denoted the lag *k*. Note that the correlation coefficient between y_t and y_{t-k} is called the lag-*k* autocorrelation of y_t and is commonly denoted by ρ_k, which is specifically defined as Equation 3.30,

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\text{Var}(y_t)\text{Var}(y_{t-k})}} = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)} = \frac{\gamma_k}{\gamma_0} \qquad 3.30$$

where $\text{Var}(y_{t-k}) = \text{Var}(y_t)$, for the case of weakly stationary. A weakly stationary time series y_t is not serially correlated if and only if $\rho_k = 0$ for all $k > 0$. The collection of autocorrelations, $\{\rho_k\}$, is called the ACF of y_t.

Box concluded that most satisfactory estimate of the *k*th lag autocorrelation ρ_k from data of $y_1, y_2, ..., y_T$ is denoted by r_k as given in Equation 3.31 (Box et al., 2008).

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0} = \frac{\frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2} \qquad 3.31$$

The value r_k in Equation 3.31 is called the sample autocorrelation function at lag *k*. The plot of r_k versus lag *k* is called the sample autocorrelation function of the process, refer to Figure 3.7. Since the ACF is a plot of the diagonals of the P_T and symmetric about zero, therefore in practice, it is necessary to plot the positive half of the sample ACF.

Sometimes, the sample ACF is called as the correlogram. The characteristic of positive definiteness of the P_T which indicates that $-1 < \rho_1, \rho_2, ..., \rho_q < 1$, implies the values of r_k is $-1 < r_k < 1$.

In identifying a model for a time series, it is necessary to have a rough check on whether ρ_k (or r_k), is effectively zero beyond a certain lag. Thus, Bartlett's approximation is employed to approximate large-lag standard error of the estimated autocorrelation r_k at lags k greater than some value q beyond which the theoretical autocorrelation function ρ_k may be deemed to have "died out", which is expressed by Equation 3.32.

$$\mathrm{se}[r_k] \approx \sqrt{\frac{1}{T}\left(1 + 2\sum_{i=1}^{q} \rho_i^2\right)} \quad , \qquad k > q \qquad 3.32$$

In practice, the r_k $(k = 1,2,...,q)$ are substituted to replace the ρ_k. The large-lag standard error approximates the standard deviation of r_k for appropriately large lags $(k > q)$, with the assumption that the ρ_k are all essentially zero beyond some hypothesised lag $k = q$.

Similarly, the large-lag Bartlett's approximation for the covariance between the estimated autocorrelations r_k and r_{k+s} at two different lags k and $k+s$ is given in Equation 3.33. This approximation result is required in the interpretation of individuals since large covariance can exist between neighbouring values.

$$\mathrm{cov}[r_k, r_{k+s}] \approx \frac{1}{T} \sum_{\upsilon=-q}^{q} p_\upsilon \, p_{\upsilon+s} \, , \qquad k > q \qquad 3.33$$

A special case of the large-lag standard errors occur when $p = 0$. In this case, the ρ_k are taken to be zero for all lags (other than lag 0), hence the series is completely random or white noise. Thus, the standard errors for r_k can be expressed in the simple form as in Equation 3.34.

$$se[r_k] \approx \frac{1}{\sqrt{T}} \, , \qquad k > 0 \qquad 3.34$$

For the white noise series, the result in Equation 3.33 indicates that the estimated autocorrelations between r_k and r_{k+s} are uncorrelated, therefore a collection of estimated autocorrelations for different lags will tend to be independently and normally distributed with mean 0 and variance $1/T$. Noted that, the r_k is also known to be approximately normally distributed for large *T*.

While, the partial autocorrelation function is a device that exploits the fact that an AR(*p*) process has an autocorrelation function that is infinite in extent, it can by its nature be described in terms of *p* nonzero functions of the autocorrelations. The PACF shows the relation between two observations y_t and y_{t+k} after they are separated with other observations between y_t and y_{t+k}. Theoretically, PACF is a function of two observations y_t and y_{t+k} that can be separated by a lag of *k* time units, denoted by ϕ_{kk}. The partial autocorrelation function of a *p*th-order autoregressive process has a cut off after lag *p*. In practice, the estimate of the *k*th lag partial autocorrelation from data of $y_1, y_2, ..., y_T$, is denoted by r_{kk}, as given in Equation 3.35.

$$r_{kk} = \hat{\phi}_{kk} = \begin{cases} r_1 & , \quad \text{if } k = 1 \\ \dfrac{r_k - \sum_{i=1}^{k-1} r_{k-1,i}\, r_{k-i}}{1 - \sum_{i=1}^{k-1} r_{k-1,i}\, r_i} & , \quad \text{if } k = 2, 3, ... \end{cases} \tag{3.35}$$

where $r_{k,j} = r_{k-1} - r_{kk}\, r_{k-1,k-j}$ for $j = 1, 2, ..., k-1$. The r_{kk} in Equation 3.35 is called the sample partial autocorrelation function at lag *k*. The plot of r_{kk} versus lag *k* is called the sample partial autocorrelation function of the process. The characteristics of symmetric and positive definiteness for r_{kk} are similar to the sample ACF, which implies $-1 < r_{kk} < 1$.

As for the sample ACF, it also needs to have a rough check on whether r_{kk} is effectively zero beyond a certain lag *p* in identifying a model related to autoregressive. On the hypothesis that the process is autoregressive of order *p*, it was shown that the estimated partial autocorrelations of order $p+1$ and higher, are approximately independently and normally distributed with zero mean (Daniels, 1956; Jenkins, 1961;

Quenouille, 1949). Also, if T is the number of observations used in estimating, the variance of the estimated partial autocorrelations of order $p+1$ and higher is given by Equation 3.36.

$$\mathrm{var}[r_{kk}] \cong \frac{1}{T} \quad , \qquad k > p \tag{3.36}$$

Thus, for the partial autocorrelations, the standard error of the estimated partial autocorrelations of order $p+1$ and higher is expressed by Equation 3.37.

$$se[r_{kk}] \approx \frac{1}{\sqrt{T}} \quad , \qquad k > p \tag{3.37}$$

A time series is said to be stationary when the sample ACF and PACF dies down or cuts off drastically in the correlograms, as given in Equation 3.38,

$$\begin{aligned} r_k &= 0, \qquad \text{for } k > q \\ r_{kk} &= 0, \qquad \text{for } k > p \end{aligned} \tag{3.38}$$

where p and q are the number of lag that the sample ACF and PACF cuts off, respectively. For the Box-Jenkins models, the sample ACF and the sample PACF are used to get the possible order q of models that consist of moving average components and to get the possible values of order p for models that consist of autoregressive components, respectively, for a stationary data series. Box et al. provides a convenient reference table of the properties of the theoretical autocorrelation and partial autocorrelation functions for autoregressive and moving average processes of first and second order, as summarised in Table 3.3 (Box et al., 2008).

In obtaining a useful estimate of the ACF and PACF, Box and Jenkins recommend that the number of data be at least 50 and the value of k not larger than $\frac{T}{4}$ (Box & Jenkins, 1968). Alternatively, Hyndman and Athanasopoulos (Hyndman & Athanasopoulos, 2014) proposed a particular formula of the maximum number of lag k of the ACF and PACF, denoted by $k_{\max}$, given by Equation 3.39,

$$k_{\max} = 10\log_{10}\left(\frac{T}{w}\right) \tag{3.39}$$

where T is number of observations and w is number of series. Since the scope in this study is univariate data, therefore the maximum lag can be rewritten as Equation 3.40. Note that, the suggested $k_{\max}$ value is still consistent with the idea from Box and Jenkins.

$$k_{\max} = 10 \times \log_{10} T \qquad 3.40$$

Table 3.3 Behaviour of the ACF and PACF for the d^{th} difference of an ARIMA process of first and second order

Source: Box et al. (2008)

Order	**Behaviour of ρ_k**	**Behaviour of ϕ_{kk}**	**Preliminary estimates**	**Admissible region**
$(1,d,0)$	Decays exponentially	Only ϕ_{11} nonzero	$\varphi_1 = \rho_1$	$-1 < \varphi_1 < 1$
$(0,d,1)$	Only ρ_1 nonzero	Exponential dominates decay	$\rho_1 = \dfrac{-\theta_1}{1+\theta_1^2}$	$-1 < \theta_1 < 1$
$(2,d,0)$	Mixture of exponentials or damped sine wave	Only ϕ_{11} and ϕ_{22} nonzero	$\varphi_1 = \dfrac{\rho_1(1-\rho_2)}{1-\rho_1^2}$ $\varphi_2 = \dfrac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$-1 < \varphi_2 < 1$ $\varphi_2 + \varphi_1 < 1$ $\varphi_2 - \varphi_1 < 1$
$(0,d,2)$	Only ρ_1 and ρ_2 nonzero	Dominated by mixture of exponential or damped sine wave	$\rho_1 = \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$ $\rho_2 = \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$-1 < \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$
$(1,d,1)$	Decays exponentially from first lag	Dominated by exponential decay from first lag	$\rho_1 = \dfrac{(1-\theta_1\varphi_1)(\varphi_1-\theta_1)}{1+\theta_1^2-2\varphi_1\theta_1}$ $\rho_2 = \varphi_1\rho_1$	$-1 < \varphi_1 < 1$ $-1 < \theta_1 < 1$

ii. Unit Root Test: Augmented Dickey-Fuller Test

Despite the visual inspection of the autocorrelation functions, the formal method to test the trend stationarity of a series is the unit root test. Unit root test is a method for detecting unit roots in time series since the presence of a unit root indicates that the time series is not stationary in-mean so that the series should be differenced in order to make it stationary. One of the most widely used unit root test is the Augmented Dickey-Fuller test (ADF-test).

Dickey and Fuller proposed a useful tool for testing a series for the presence of a unit root, known as standard Dickey-Fuller test (DF-test) (Dickey & Fuller, 1979). This test is used to determine whether the series is stationary or should it undergo differencing to achieve stationarity, and they proved that the test is more powerful compared to Box-Pierce *Q*-statistic in testing the Box-Jenkins model. In the standard DF-test, a simple AR(1) process is considered which is given by

$$y_t = \phi_{df}\, y_{t-1} + TD_t + a_t$$
$$TD_t = c + x_t^{'}$$

where TD_t is a deterministic term consisting of *c* as a constant and $x_t^{'}$ as the deterministic time trend, $y_0 = 0$, ϕ_{df} is a parameter to be estimated that consist of real number and $a_t \sim NID(0, \sigma^2)$. If $|\phi_{df}| > 1$, the y_t is a nonstationary series with the variance increases exponentially as *t* increases. The time series with $|\phi_{df}| = 1$ is called a random walk series and might be in the cases of nonstationary series. While if $|\phi_{df}| < 1$, y_t converges to a stationary series. Thus, the hypothesis of unit root test is to test the stationarity in a series which can be evaluated by using whether $|\phi_{df}| < 1$ that indicates that no unit root in the data series. Hence, the hypotheses of the DF-test may be written as $H_0 : \phi_{df} = 1$ against $H_1 : |\phi_{df}| < 1$.

Alternatively, the AR(1) model may be rewritten as Equation 3.41,

$$\begin{aligned} \Delta y_t &= \pi y_{t-1} + TD_t + a_t \\ TD_t &= c + x_t^{'} \end{aligned} \qquad 3.41$$

where $\pi = \phi_{df} - 1$. The unit root tests are often computed using this alternative model (Zivot & Wang, 2003). Therefore, the null and alternative hypothesis for the model in Equation 3.41 may be written as $H_0 : \pi = 0$ versus $H_1 : \pi < 0$ with the test statistic given by $t_{\text{test}} = \dfrac{\hat{\pi}}{se(\hat{\pi})}$, where $\hat{\pi}$ is the estimate of π and $se(\hat{\pi})$ is the standard error estimate. Dickey and Fuller showed that this test statistic does not follow the *t* distribution and derived the asymptotic results and simulated critical values for various sample sizes

(Dickey & Fuller, 1979). Mackinnon then implemented a much larger size of simulations, allowing the calculation of Dickey-Fuller critical values and *p*-values for any sample sizes (MacKinnon, 1996). The Mackinnon critical value calculations are used by many softwares including EViews and SPlus in constructing test output.

The standard DF unit root test is valid if the time series is well characterised by an AR(1) with white noise errors. However, many financial time series have a more complicated structure than a simple AR(1) model. Furthermore, if the series is correlated at higher order lags, the assumption of the a_t is violated. To overcome these constraints, Said and Dickey constructed the augmented Dickey-Fuller (ADF) unit root test for higher-order correlation to accommodate general ARMA(p,q) and ARIMA(p,d,q) models, where the number d equals the number of unit roots in the characteristic equation for the time series (Said & Dickey, 1984). For the ADF-test, the following regression model is estimated, as shown in Equation 3.42,

$$\begin{aligned} \Delta y_t &= \pi\, y_{t-1} + TD_t + C_1 \Delta y_{t-1} + C_2 \Delta y_{t-2} + \ldots + C_k \Delta y_{k-1} + a_t \\ TD_t &= c + x_t' \end{aligned} \qquad 3.42$$

where $\Delta y_t = y_t - y_{t-1}$ denotes the first differenced series, $\pi = \phi_{df} - 1$, k is the number of lags to include in the regression, C_i is the coefficient for the Δy_{t-i}, $a_t \sim NID(0, \sigma^2)$ and TD_t is the deterministic term which may consist of constant, or a constant and trend. The augmented model as in Equation 3.42 is then used to test the hypothesis as the model in Equation 3.41 by using the same test statistic. If the y_t needs differencing, then the coefficient $\hat{\pi}$ should be approximately zero, while if y_t is already stationary, then $\hat{\pi} < 0$. In practice, *p*-value approach is easier to apply than *t*-test statistic in testing the hypothesis. Based on the hypothesis given, the null hypothesis for an ADF-test is that the data series is nonstationary, therefore large *p*-values are indicative of nonstationarity, while small *p*-values suggest stationarity. At 5% significance level, differencing is required if the *p*-value is greater than 0.05.

There are two practical issues in performing an ADF-test. First, the choice of TD_t either as constant or constant and trend since the chosen TD_t will reflect the hypothesis appropriately and characterise the trend properties of the data. According to Zivot and

Wang, the constant TD_t is appropriate for non-trending series such as interest rates, exchange rates, and spreads, meanwhile the constant and trend case is appropriate for trending time series like asset prices (Zivot & Wang, 2003). The second practical issue in ADF-test is the specification of the lag length, $k_{\max}$. If the lag length is too small then the remaining serial correlation in the errors will be biased to the test, on the other hand, if it is too large then the power of the test will suffer. Schwert suggested the useful rule of thumb formula for determining the $k_{\max}$, as given by Equation 3.43,

$$k_{\max} = \left(12 \left(\frac{T}{100} \right)^{\frac{1}{4}} \right) \qquad 3.43$$

where T is the number of observations (Schwert, 1989).

In this study, the ADF-test is performed using EViews since the software provides a user friendly tool for testing a series for the presence of a unit root. The ADF unit root test that EViews provides tests the null hypothesis of $H_0 : \phi_{df} = 1$ against the left-tailed alternative $H_1 : \phi_{df} < 1$ with Mackinnon critical value approach. In the ADF-test, if null hypothesis is not rejected means the data needs to be differenced to make it stationary.

3.4.1.3 Preliminary Linearity Test in Time Series

Let y_t be an original time series with mean μ and a_t be a white noise series as $a_t \sim IID\left(0, \sigma_a^2\right)$. In proposing autoregressive model, Box and Jenkins stated that the deviation of a stationary time series, $y_t - \mu$ can be made linearly dependent on previous deviations and on a_t, as can be referred to Equation 2.1 (Box & Jenkins, 1968). Meanwhile, for moving average model, the deviation can be made linearly dependent on a_t and on one or more previous $a's$, refer to Equation 2.2. Hence, before considering the Box-Jenkins model to the data series, it is wise to check the linearity of the stationary data first. This preliminary linearity test is done in the identification stage to clarify whether the linear model is appropriate to model the stationary series. One other way to diagnostically do that is to plot the deviation series from stationary data versus the lagged series. If the stationary series is linear, then it indicates that a linear fit is appropriate. Thus, the Box-Jenkins model is applicable in analysing the data series.

3.4.1.4 Portmanteau Test

The Box-Jenkins model works well for time series data or serially correlated data. The ACF plays an important role in linear time series analysis in testing the serial correlation in data. In testing the autocorrelations of y_t in finite samples, Ljung and Box modified the Portmanteau test statistic proposed by Box and Pierce (Box & Pierce, 1970; Ljung & Box, 1978). The modified Portmanteau test, known as Ljung-Box *Q*-statistic, or simply called as LBQ-test, is given by Equation 3.44,

$$Q(k_{\max}) = T(T+2)\sum_{k=1}^{k_{\max}} \frac{r_k^2}{T-k} \tag{3.44}$$

is a test statistic for the $H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$ versus $H_1 : \rho_k \neq 0,\ \ k \in \{1,2,\ldots,k_{\max}\}$; where T is the total observations, r_k is the sample ACF of y_t, k is the number of lag and $k_{\max}$ is the maximum lag being considered. The decision rule is to reject H_0 if $Q(k_{\max}) > \chi_\alpha^2$ denotes the $100(1-\alpha)$th percentile of a chi-squared distribution with $k_{\max}$ degrees of freedom (dof), or if *p*-value is used, the rule then is to reject H_0 when $p\text{-value} < \alpha$. According to Tsay, simulation studies suggest that the choice of $k_{\max} \approx \ln T$ provides better power performance for nonseasonal time series (Tsay, 2013). Meanwhile, for seasonal time series, the rule needs modification in which autocorrelations with lags at the multiples of the seasonality are more important.

3.4.1.5 Extended Autocorrelation Function

Tsay and Tiao proposed a new approach that uses the extended autocorrelation function (EACF) to specify the order of Box-Jenkins model specifically to an ARMA process (Tsay & Tiao, 1984). The output of EACF is a two-way table, where the rows correspond to AR order *p* and the columns to MA order *q*. The theoretical version of EACF for ARMA(1,1) is given in Table 3.4.

Table 3.4 Theoretical EACF table for an ARMA(1,1) model

				MA				
AR	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	*	X	O	O	O	O	O	O
3	*	*	X	O	O	O	O	O
4	*	*	*	X	O	O	O	O
5	*	*	*	*	X	O	O	O

The key feature of the EACF table is that, for an ARMA(p,q) model, the triangle of "O" will have it upper left vertex at the (p,q) position. The EACF table consist of "X" and "O", where "X" denotes that the absolute value of the corresponding EACF is greater than or equal to twice of its asymptotic standard error, while "O" denotes that the corresponding EACF is less than twice of its standard error in modulus. The standard error of EACF can be computed using $2/\sqrt{T}$ where T is the sample size.

3.4.2 Stage II: Parameter Estimation

As current practice in the Box-Jenkins modelling, the order of the Box-Jenkins model is confirmed first, and then the parameters of the model chosen are estimated. In this study, two estimation methods that are commonly used in previous literatures are applied, maximum likelihood estimation (MLE) and ordinary least squares (OLS). These two methods are built-in in many statistical software including EViews and R language. The method of MLE and OLS for Box-Jenkins models can be referred to Box et al. (2008) and Wei (2006) for details.

However, it is possible that two or more significant models are considered, in which the order of the models are identified by the ACF and PACF approach, might come from the estimation methods. In identifying the best Box-Jenkins model to the series, model selection criteria method is implemented. In general, if the numbers of parameters of the models are the same, then the model with the smallest mean squared error is selected. While if the number of parameters of the models are different, then the parsimony principle is applied by selecting the simpler model yet is adequately significant. In this proposed procedure, the well-known Akaike Information Criteria

(AIC) and another commonly used criterion function which is the Schwarz Information Criterion (SIC) are applied in identifying the best significant Box-Jenkins model, as illustrated in Figure 3.11. These criteria determine the appropriateness of the Box-Jenkins model that identified by EACF method.

As for the time series model, Akaike (Akaike, 1974) proposed the AIC and Schwarz (Schwarz, 1978) proposed the SIC which are defined as in Equation 3.45 and Equation 3.46, respectively.

$$\text{AIC}(p,q) = T\ln\left(\tilde{\sigma}_\ell^2\right) + 2(p+q) \tag{3.45}$$

$$\text{SIC}(p,q) = T\ln\left(\tilde{\sigma}_\ell^2\right) + (p+q)\ln(T) \tag{3.46}$$

The $\text{AIC}(p,q)$ is the value of AIC for the model with AR order *p* and MA order *q*, $\text{SIC}(p,q)$ is the value of SIC for the model with AR order *p* and MA order *q*, $\tilde{\sigma}_\ell^2$ is the maximum likelihood estimate of σ_a^2 and *T* is the number of observations. Note that, σ_a^2 is the variance of a_t. SIC is also known as the Bayesian information criterion (BIC).

In the model selection criteria, the smallest value of the AIC or SIC are preferred. The difference between AIC and SIC is the value of penalty used. For AIC, the penalty value is fixed to two while $\ln(T)$ for SIC. The AIC and SIC are extensively adopted to guide the choice of alternative models for both linear and nonlinear modelling (Verbeek, 2004).

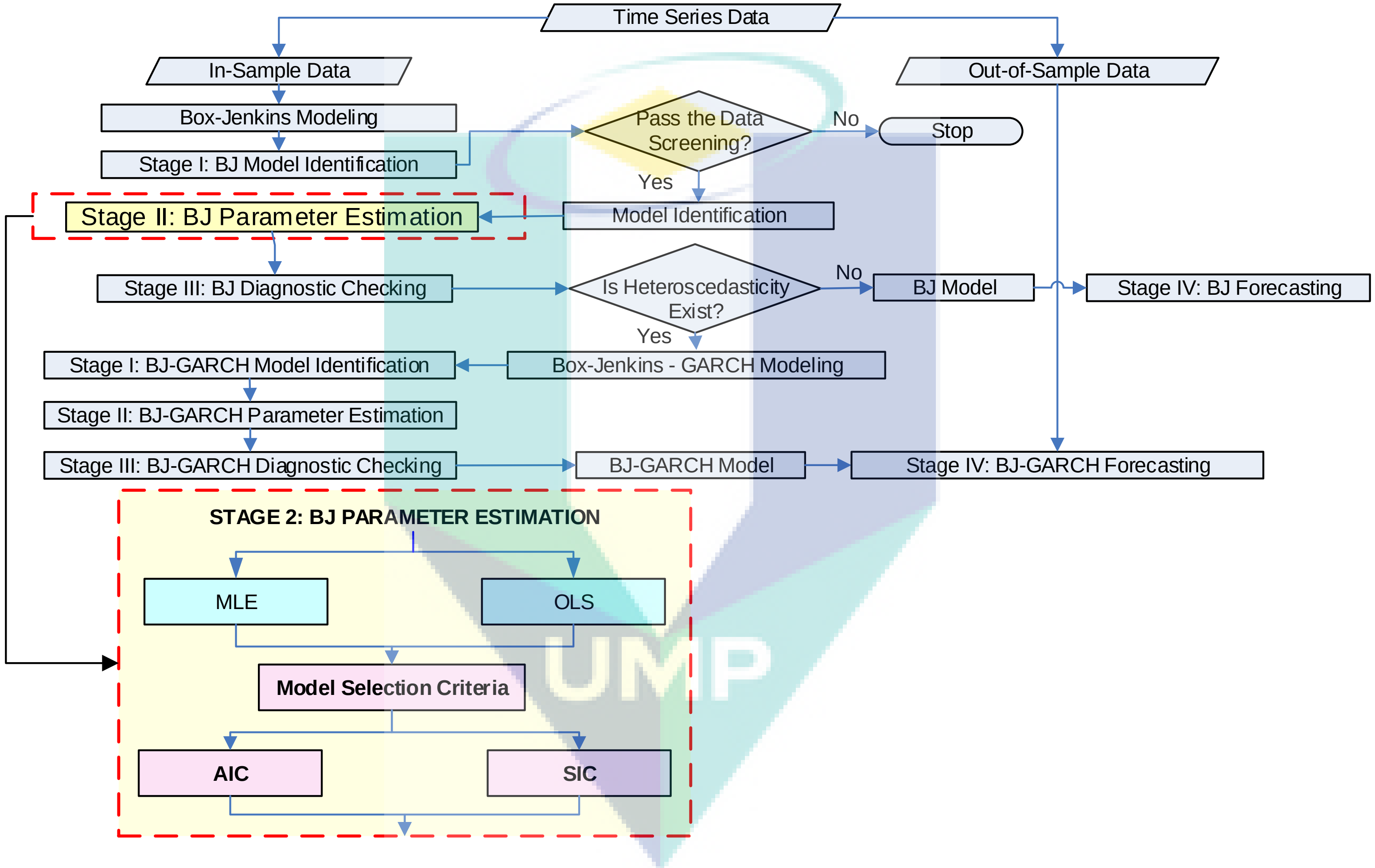


Figure 3.11 Procedures in Stage II of the procedure of BJ-G for highly volatile data

3.4.3 Stage III: Diagnostic Checking

Once the most appropriate model has been chosen, the model must be examined carefully in the diagnostic checking stage. In this stage, the residual series $\{\hat{a}_t\}$ of the chosen model is investigated in checking the model adequacy. If the model is adequate, then the residual series should behave as a white noise (Tsay, 2013). Otherwise, if the chosen model is found to be inadequate, the Box-Jenkins model identification procedure will be repeated in order to obtain a new appropriate model. In the diagnostic checking stage, the tests considered are serial correlation test, heteroscedasticity test and normality test to assure that the errors behave like white noise. Figure 3.12 shows the detailed procedures of the diagnostic checking stage in the proposed procedure of BJ-G in handling highly volatile time series data.

By definition, a time series y_t is a white noise process if the series has zero mean $(E(y_t)=0)$, has constant and finite variance process $(\text{Var}(y_t)=\sigma^2)$ for all t where $\sigma^2<\infty$, and a serially uncorrelated $(\text{Cov}(y_t, y_s)=0, \text{if } t \neq s)$. Particularly, the white noise errors in time series consists of a sequence of independent and identically distributed (IID) random variables, denoted by $a_1, a_2, \ldots, a_T$, that is assumed to have zero mean, serially uncorrelated and homoscedastic variance. The plot of residuals versus time or residual plots can be used to examine graphically the IID assumption as well as to spot possible outliers in the series.

To assure white noise process on the errors, the residual series $\{\hat{a}_t\}$ of the model is investigated in terms of independence, homoscedasticity (constant and finite variance process) and its distribution. The Durbin-Watson test, the autocorrelation functions and the LBQ-test on the residuals are commonly used to check correlations in the residuals of a series. The tests of serial correlations are further discussed in Section 3.4.3.1.

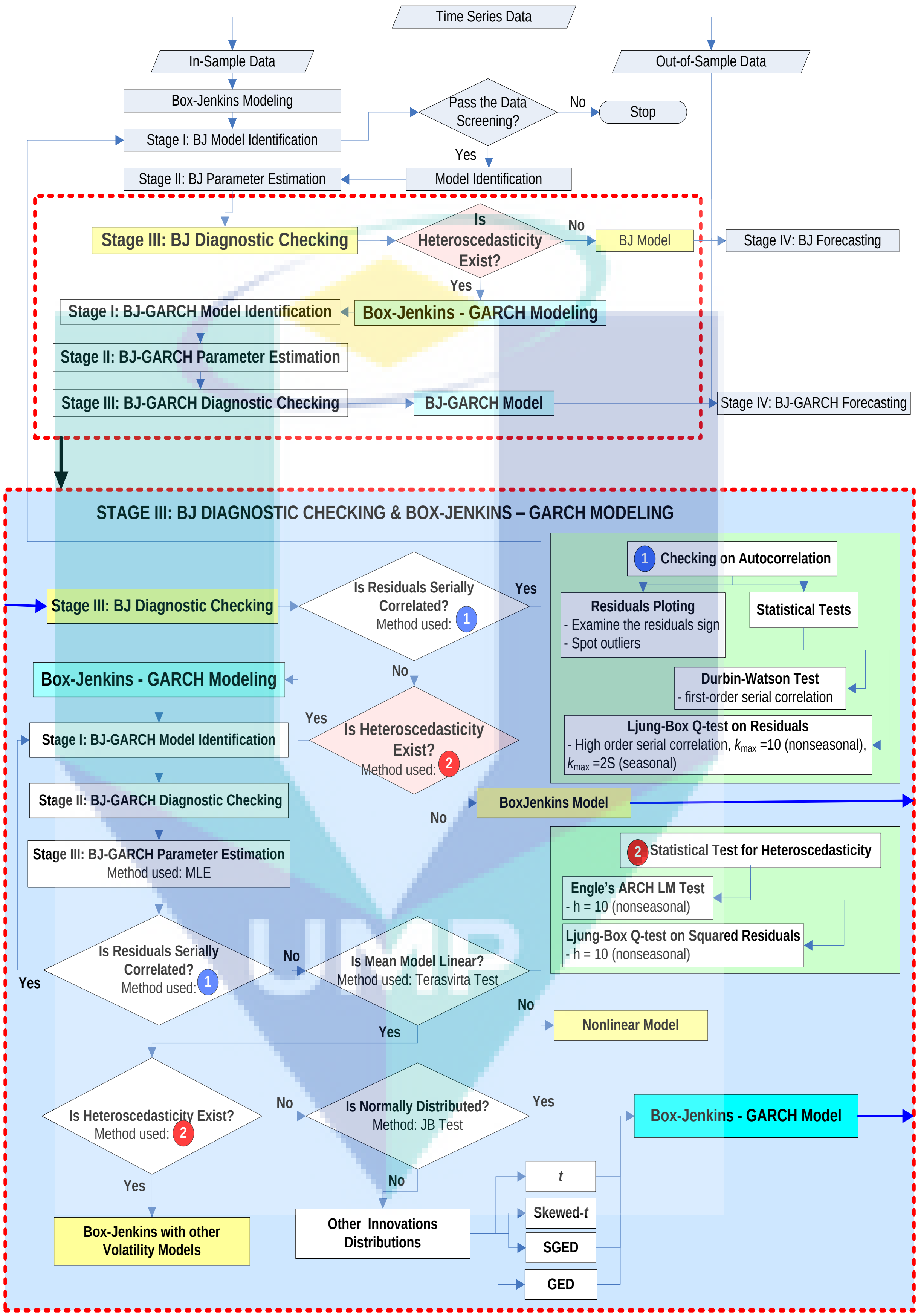


Figure 3.12 Detail procedures in Stage III of the procedure of BJ-G for highly volatile data

Meanwhile, to detect the existence of heteroscedasticity in the residuals, the ARCH test, the autocorrelation functions on squared residuals and the Ljung-Box *Q*-statistics on the squared residuals are used, as presented in Section 3.4.3.2. For a highly volatile time series data, the tests of heteroscedasticity will show that the variance of the series is not correctly specified by the model since the variance is not constant. In handling the existence of volatility clustering in the series, this study proposed the BJ-G model since GARCH model is highly recommended in the previous literatures as volatility model.

In the combination model of BJ-G, the chosen model of the Box-Jenkins is used to model the mean data of time series while the GARCH is used to model the variance equation of the residuals. The identification method of order r and s for the GARCH model can be referred to in Section 3.2 for details. In the diagnostic checking stage, once the residuals of the chosen combination model are not serially correlated, then it is strongly suggested that the linearity checking of the mean model is conducted by applying the Terasvirta test, as will be discussed further in Section 3.4.3.3. The consideration of the linearity test is made for validating the appropriateness of the use of Box-Jenkins in the combination model to model the mean data series. Meanwhile, the appropriateness of GARCH to model the variance equation for a volatile time series is shown by the tests of heteroscedasticity.

On the other hand, the distribution of errors (or innovations) in white noise is initially assumed to be normal, therefore normality test on the residual series as described in Section 3.4.3.4 is done to check the assumption. However, the non-normal characteristic that typically exists in the residuals of volatile time series also leads to failure of the normality test. Therefore, distributions such as t, generalized error distribution (GED) and their skewed version are considered to model the errors. The detail of the distributions for the innovations can be referred to in Section 3.4.3.5. Therefore, in the procedure, a BJ-G model is proposed to analyse the univariate volatile series as well as to forecast highly volatile series.

3.4.3.1 Serial Correlation Tests

The errors in time series data usually exhibit some type of autocorrelated structure such as the errors are correlated with themselves at different time periods. Note that, the errors of a time series data should be independent but the data itself is serially correlated. The residual plots can be useful for the detection of autocorrelation. If the sign of the residuals is randomly distributed or correlation is close to zero, then the model errors are uncorrelated. Meanwhile, if there are not enough changes of sign in the pattern of residuals or the sign of residuals occur in cluster, there is positive correlation in the errors. On the other hand, if the residuals alternate signs too rapidly, there exists negative correlation in the errors (Tsay, 2013).

The autocorrelation between two errors that are one period apart, or the lag one autocorrelation is defined as Equation 3.47,

$$\rho_1 = \frac{Cov(a_t, a_{t+1})}{\sqrt{\sigma^2_{a_t}}\sqrt{\sigma^2_{a_t}}} = \frac{\phi_a \sigma^2_{\varepsilon_t}\left(\frac{1}{1-\phi_a{}^2}\right)}{\sqrt{\sigma^2_{\varepsilon_t}\left(\frac{1}{1-\phi_a^2}\right)}\sqrt{\sigma^2_{\varepsilon_t}\left(\frac{1}{1-\phi_a^2}\right)}} = \phi_a \qquad 3.47$$

where a_t is the error term in the model at time period *t*, ε_t is an NID$\left(0, \sigma_e^2\right)$ random variable, ϕ_a is a parameter that defines the relationship between successive values of the model errors a_t and a_{t-1} with $|\phi| < 1$, and the time index is $t = 1, 2, ..., T$. Consequently, the autocorrelation between two errors that are *k* periods apart is $\rho_k = \phi_a^k$, which is called the autocorrelation function. Two common statistical tests in detecting the presence of autocorrelation are the Durbin-Watson test and the LBQ-test on residuals. Meanwhile, the residual plot shows graphical evidence to support the result of these serial correlation tests.

The Durbin-Watson test (DW-test) is the test developed by Durbin and Watson (Durbin & Watson, 1971). The test is one of widely used procedure to detect the positive presence of autocorrelation in time series model errors. The DW-test measures the first-order serial correlation in the residuals of the estimated equation. The considered serial correlation in the DW-test is given by Equation 3.48,

$$a_t = \phi_a a_{t-1} + \varepsilon_t \qquad 3.48$$

where a_t is the error term in the model at time period *t*, ε_t is an $\mathrm{NID}\left(0, \sigma_e^2\right)$ random variable, ϕ_a is a parameter that defines the relationship between successive values of the model errors a_t and a_{t-1} with $|\phi| < 1$, and the time index is $t = 1,2,\ldots,T$. The null hypothesis for the DW test is no serial correlation in the residuals, while presence of serial correlation is the alternative hypothesis, or can be simplified as $H_0 : \phi_a = 0$ versus $H_1 : \phi_a > 0$.

Let *T* is the number of observations and $\hat{a}_t$ is the residual at time *t*, the DW-test statistic is given by Equation 3.49.

$$DW = \frac{\sum_{t=2}^{T} \left(\hat{a}_t - \hat{a}_{t-1}\right)^2}{\sum_{t=1}^{T} \hat{a}_t^{\,2}} = \frac{\sum_{t=2}^{T} \hat{a}_t^2 + \sum_{t=2}^{T} \hat{a}_{t-1}^2 - 2\sum_{t=2}^{T} \hat{a}_t \hat{a}_{t-1}}{\sum_{t=1}^{T} \hat{a}_t^{\,2}} \approx 2\left(1 - r_1\right) \qquad 3.49$$

where r_1 is the lag one sample autocorrelation coefficient defined as

$$r_1 = \frac{\sum_{t=1}^{T-1} \hat{a}_t \hat{a}_{t+1}}{\sum_{t=1}^{T} \hat{a}_t^2}.$$

The value of DW-test should be approximately 2 for uncorrelated errors $\left(r_1 = 0\right)$. According to Johnston and DiNardo, if $\mathrm{DW} \approx 2$ means there is no serial correlation, if $DW < 2$ means there is positive serial correlation and if $2 < \mathrm{DW} < 4$ means there is negative serial correlation in the residuals (Johnston & DiNardo, 1997).

There are limitations of the DW-test as a test for serial correlation. One of the main limitations is, if there are lagged dependent variables on the right-hand side of the model, the DW-test is no longer valid. Another main limitation is the test is only valid for testing the possibility of serial correlation in a first-order time series model. To overcome these limitations, the LBQ-test is preferred in most applications.

The LBQ-test on residuals is used in the diagnostic checking to recognise whether there is autocorrelation in the residuals of fitted time series models. The null hypothesis for this test is no remaining serial correlation in the residuals. Failure to reject the null hypothesis means that the mean equation is correctly specified up to lag $k_{\max}$. The test statistic for the test can be referred to Equation 3.44. In the test, statistics $Q(k_{\max})$ follows $\chi^2(k_{\max})$ distribution with $k_{\max}$ degrees of freedom if there is no autocorrelation among residuals.

In practice, the choice of $k_{\max}$ may affect the performance of the LBQ-test statistic. As for the selection of $k_{\max}$ value, the general principlel is that a larger value is better and the literature normally uses of $k_{\max}$ value of no more than 20 (Živkov, Njegić, Momčilović, & Milenković, 2016). Often $k_{\max} = 15$ is used as it is able to detect model failures (Engle, 2001). Simulation studies suggest that $k_{\max} \approx \ln T$ provides better power performance (Tsay, 2013). Alternatively, Hyndman and Athanasopoulos suggest using $k_{\max} = 10$ for nonseasonal data and $k_{\max} = 2S$ for seasonal data, where S is the period of seasonality. They believed that, the suggestion value of $k_{\max}$ is adequate to ensure that the number of lag is large enough to capture any meaningful and troublesome correlations. The LBQ-test is not good when $k_{\max}$ is large, therefore if the $k_{\max}$ value is larger than $T/5$, then use $k_{\max} = T/5$ (Hyndman & Athanasopoulos, 2014).

In EViews, the result of LBQ-test is displayed together with ACF and PACF of the residuals for high order serial correlation. If there is no serial correlation in the residuals, the ACF and PACF at all lags will be closed to zero, and all *Q*-statistics will be insignificant with large *p*-values.

3.4.3.2 Heteroscedasticity Test

Let $a_t = y_t - \mu_t$ be the errors of the mean equation of model. The squared series of errors $\{a_t^2\}$ is used in checking the conditional heteroscedasticity of the model for the data series, which is also known as the ARCH effects. This particular heteroscedasticity specification is motivated by the observation that in many financial time series, the magnitude of the residuals appeared to be related to the magnitude of the recent residuals. Hence, ignoring ARCH effects may result in the loss of efficiency.

There are two ARCH tests available and commonly used in testing heteroscedasticity in data series. The first test is to apply the Lagrange multiplier test, which is known as the ARCH LM test (Engle, 1982). The Engle's ARCH LM test is equivalent to the F statistic for testing $\alpha_i = 0$ for $i = 1,2,...,k_{\max}$ in the linear regression as given by the equation below,

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_{k_{\max}} a_{t-k_{\max}}^2 + e_t \qquad \text{for } t = k_{\max}+1, k_{\max}+2,...,T$$

where e_t denotes the error term, $k_{\max}$ is a positive integer and T is the number of observation. This is a regression of the squared residuals on a constant and lagged squared up to order $k_{\max}$. The selection of $k_{\max}$ value for ARCH test is the same as explained in Section 3.4.3.1.

The null hypothesis for the ARCH LM test is $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_{k_{\max}} = 0$ and the alternative hypothesis is $H_1 : \alpha_i \neq 0$ for some i between 1 and $k_{\max}$. The ARCH LM test, as described by Tsay (2013), is given by Equation 3.50,

$$f_{\text{test}} = \frac{(\text{SSR}_\text{A} - \text{SSR}_\text{B})/k_{\max}}{\text{SSR}_\text{B}/(T - 2k_{\max} - 1)} \qquad 3.50$$

where $\text{SSR}_\text{A} = \sum_{t=k_{\max}+1}^{T} (\hat{a}_t^2 - \bar{a})$, $\bar{a} = \frac{1}{T}\sum_{t=1}^{T} \hat{a}_t^2$ is the sample mean of $\hat{a}_t^2$, $\hat{a}_t^2$ is the estimated errors or residuals of the model, $\text{SSR}_\text{B} = \sum_{t=k_{\max}+1}^{T} \hat{e}_t^2$ and $\hat{e}_t^2$ is the least squares residual of the prior linear regression. Under the H_0, the ARCH LM test follows an F distribution with degrees of freedom of $k_{\max}$ and $T - 2k_{\max} - 1$, $f_{k_{\max}, T-2k_{\max}-1}$. For sufficiently large T, one can use $k_{\max} f_{test}$ as the test statistic, which is asymptotically a chi-squared distribution with $k_{\max}$ degrees of freedom, $\chi_{k_{\max}}^2$. The rule is to reject the H_0 if $k_{\max} f_{test} > \chi_{\alpha,k_{\max}}^2$ or the $p-$value $\leq \alpha$. If the decision is to not reject H_0, it can be concluded that there is no ARCH effect in the residuals of the model.

The second test for conditional heteroscedasticity is the LBQ-test on a squared residual series $\{\hat{a}_t^2\}$. The test on the squared residuals of the best fitting Box-Jenkins model could be useful in improving forecasts of the series since numerous time series data in which the squared residuals appear to be autocorrelated even though the residuals do not (Granger & Andersen, 1978; McLeod & Li, 1983). The details of the LBQ-test on $\{\hat{a}_t^2\}$ can be referred to McLeod and Li (McLeod & Li, 1983).

The null hypothesis of the test is no ARCH in the residuals or the variance equation is correctly specified up to lag $k_{\max}$. In EViews, the result of LBQ-test is displayed together with ACF and PACF of the squared residuals for high order of lag. If there is no ARCH in the residuals, the ACF and PACF at all lags will approach zero, and all *Q*-statistics value should be insignificant with $p-\text{value} > \alpha$. On the other hand, if there is presence of ARCH in the residuals, the PACF for the squared residuals can be used to determine the suitability of ARCH or GARCH model in handling the heteroscedasticity in the data series.

3.4.3.3 Linearity Test for Mean Model

To validate the linearity assumption of the mean model to data series, a widely used linearity test in neural networks known as Terasvirta test is used (Teräsvirta, 1994; Teräsvirta, Lin, & Granger, 1993). The Terasvirta test for time series proposed by Teräsvirta et al. (1993) is based on the concepts of neural networks theory. In this linearity test, the null hypothesis is that the mean model is linear and the test is designed for the autoregressive model of order *p*. There are three stages in implementing this test, given as follows:

Step 1: Regress y_t on $1, y_{t-1}, y_{t-2}, \ldots, y_{t-p}$. Compute the residuals $\hat{u}_t$ and the sum of the squared residuals SSR_0 where $\text{SSR}_0 = \sum_{t=1}^{T} \hat{u}_t^2$.

Step 2: Regress $\hat{u}_t$ on $1, y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ and *m* auxiliary regressors. Compute the residuals $\hat{v}_t$ and the residual sum of squares SSR_1 where $\text{SSR}_1 = \sum_{t=1}^{T} \hat{v}_t^2$.

Step 3: Compute the test statistic, $f_{\text{test}} = \dfrac{(\text{SSR}_0 - \text{SSR}_1)/m}{\text{SSR}_1/(T-p-1-m)}$.

Note that, p is the order of autoregressive model, T is the number of data and m is the number of auxiliary regressors. Under the linearity test, f_{test} is approximately f distributed with m and $T-p-1-m$ degrees of freedom, $f_{\alpha,m,T-p-1-m}$. If $f_{\text{test}} > f_{\alpha,m,T-p-1-m}$ or $p\text{-value} \le \alpha$, then the null hypothesis is rejected. In R language, the linearity test for time series data is implemented in the tseries package by the function `terasvirta.test( )`.

3.4.3.4 Normality Test

Jarque-Bera test (JB-test) is one of the common test statistics for testing whether the series or the errors of the series is normally distributed. The test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is computed as in Table 3.5.

Table 3.5 Jarque-Bera test statistic

Hypothesis	Test statistic	Critical point	Decision rule
$H_0 : X \sim N(\mu, \sigma^2)$ $H_1 : X \nsim N(\mu, \sigma^2)$	$\text{JB}_{\text{test}} = T\left[\dfrac{\hat{S}^2(x)}{6} + \dfrac{(\hat{K}(x)-3)^2}{24}\right]$	$\chi^2_{\frac{\alpha}{2},2}$	reject H_0 if $\text{JB}_{\text{test}} > \chi^2_{\frac{\alpha}{2},2}$ or $p\text{-value} \le \alpha$

Under the null hypothesis of a normal distribution, the JB-test statistic is distributed as $\chi^2_{\frac{\alpha}{2},2}$. A small *p*-value leads to the rejection of the null hypothesis of a normal distribution at 5% significance level. It is important to test the validity of normality assumption since violation of the assumption may lead to the use of wrong estimators, invalid inferential statements and inaccurate conclusions (Jarque & Bera, 1987).

3.4.3.5 Distribution of Errors

The distribution of the standardised error or innovations ε_t in the part of diagnostic checking is investigated in order to find the appropriate innovations to make the model fit the data well. The considered distributions for ε_t in this study are Normal, *t*, the skewed-*t*, the generalised error distribution (GED) and the skewed generalised error distribution (SGED).

Under the normality assumption on the errors, the pdf of ε_t is given by Equation 3.51,

$$f(\varepsilon_t \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right). \qquad 3.51$$

If the data exhibit heavy tails characteristics, it is more appropriate to assume that ε_t follows a *t* distribution. The *t* distribution also known as Student's *t* distribution, is introduced by William Sealy Gosset (known as "Student") in 1908 (Boland, 1984). Let ε_t be a *t* distribution with υ degrees of freedom, the pdf of ε_t is given by Equation 3.52 where $\Gamma(.)$ is the gamma function. For a *t* distribution with υ, the mean is 0, its variance is $\upsilon/(\upsilon-2)$ if $\upsilon > 2$, the skewness is 0 if $\upsilon > 3$ and the excess kurtosis is $6/(\upsilon-4)$ if $\upsilon > 4$.

$$f(\varepsilon_t \mid \upsilon) = \frac{\Gamma((\upsilon+1)/2)}{\Gamma(\upsilon/2)\sqrt{\upsilon\pi}}\left(1+\frac{\varepsilon_t^2}{\upsilon}\right)^{-\frac{(\upsilon+1)}{2}}, \quad \upsilon > 2, \qquad 3.52$$

To handle the data that exhibit heavy tails with skewness characteristics, the *t* distribution has been modified to become a skewed-*t* distribution. For the innovation ε_t of an ARCH process, the pdf of a standardised skewed- *t* distribution is given by Equation 3.53,

$$f(\varepsilon_t \mid \xi, \upsilon) = \begin{cases} \dfrac{2}{\xi + \dfrac{1}{\xi}} \vartheta f\left[\xi(\vartheta\varepsilon_t + \varpi) \mid \upsilon\right] & \text{if } \varepsilon_t < -\varpi/\vartheta \\ \dfrac{2}{\xi + \dfrac{1}{\xi}} \vartheta f\left[(\vartheta\varepsilon_t + \varpi)/\xi \mid \upsilon\right] & \text{if } \varepsilon_t \geq -\varpi/\vartheta \end{cases} \qquad 3.53$$

where $f(\cdot)$ is the pdf of the standardised t distribution in Equation 3.51, ξ is the skewness parameter, υ is the degrees of freedom and $\upsilon > 2$, and the parameters ϖ and ϑ are given as,

$$\varpi = \frac{\Gamma((\upsilon-1)/2)\sqrt{\upsilon-2}}{\sqrt{\pi}\Gamma(\nu/2)}\left(\xi - \frac{1}{\xi}\right), \qquad \vartheta^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - \varpi^2$$

where ξ^2 is a measure of the skewness.

The generalised error distribution (GED) is a possible candidate for the description of financial market price changes (Giller, 2005). The GED is a symmetrical unimodal and a member of the exponential family. Box and Tiao (1992) call the GED distribution as an exponential power distribution. The GED distribution is proposed by Subbotin (1923) and is defined by three parameters: μ locates the mean of the distribution; σ defines the dispersion of the distribution; and κ controls the shape of the distribution. If ε_t may assume a GED, therefore the pdf of the errors is given by Equation 3.54,

$$f(\varepsilon_t \mid \mu, \sigma, \kappa) = \frac{\exp\left(-\frac{1}{2}\left(|\varepsilon_t|/\sigma\right)^{\frac{1}{\kappa}}\right)}{\sigma\left(2^{(1+\kappa)}\right)\Gamma(1+\kappa)}, \quad -\infty < \varepsilon_t < \infty, \quad \kappa < \infty \qquad 3.54$$

where $\Gamma(\cdot)$ is the gamma function. This distribution reduces to a Normal distribution if $\kappa = \frac{1}{2}$, i.e. $G\left(\mu, \sigma^2, \frac{1}{2}\right) = N(\mu, \sigma^2)$; if $\kappa = 1$ then the distribution is recognized as Double Exponential or Laplace distribution, i.e. $G(\mu, \sigma^2, 1) = L(\mu, 4\sigma^2)$; and if $\kappa \to 0$, the distribution tends to uniform distribution $U(\mu-\sigma,\mu+\sigma)$. For $\kappa<\frac{1}{2}$ the distribution is platykurtic, while for $\kappa>\frac{1}{2}$, it is leptokurtic.

The skewed generalised error distribution (SGED) introduced by Theodossiou (2008) is used to accommodate the skewness and leptokurtosis in the ε_t. If ε_t follows the SGED, then the pdf for the errors is given by Equation 3.55.

$$f(\varepsilon_t \mid \mu,\sigma,\kappa,\xi)=\frac{C}{\sigma}\exp\left(-\frac{1}{[1-sign(\varepsilon_t+\delta\sigma)\xi]^{\kappa}\theta^{\kappa}\sigma^{\kappa}}\left|\varepsilon_t+\delta\sigma\right|^{\kappa}\right) \qquad 3.55$$

where

$$C=\frac{\kappa}{2\theta}\Gamma\left(\frac{1}{\kappa}\right)^{-1},$$

$$\theta=\Gamma\left(\frac{1}{\kappa}\right)^{\frac{1}{2}}\Gamma\left(\frac{3}{\kappa}\right)^{\frac{1}{2}}S(\xi)^{-1},$$

$$\delta=2\xi AS(\xi)^{-1},$$

$$S(\xi)=\sqrt{1+3\xi^2-4A^2\xi^2},$$

$$A=\Gamma\left(\frac{2}{\kappa}\right)\Gamma\left(\frac{1}{\kappa}\right)^{-\frac{1}{2}}\Gamma\left(\frac{3}{\kappa}\right)^{-\frac{1}{2}}.$$

Note that μ , σ and ξ are the expected value, the standard deviation and the skewness parameter for the distribution, respectively, *sign* is the sign function and $\Gamma(.)$ is the gamma function. The scaling parameters κ and ξ obey the following constraints $\kappa>0$ and $-1<\xi<1$. The parameter κ controls the height and tails of the density function and the ξ controls the rate of descent of the density around the mode of the distribution which is defined as $\mu-\delta\sigma$.

3.4.4 Stage IV: Forecasting

The best BJ-G model is identified from Stage I to Stage III. Then, this model will be used in the forecasting and the detailed procedures in the forecasting stage of the proposed procedure as shown in Figure 3.13. Note that, if the stationary series is a transformed series, then the out-of-sample series must be transformed as well before the chosen model is applied to the series since the selection of the model is based on the analysis of the stationary data. By applying the chosen model to the out-of-sample data series in stationary form, a series of forecast data is obtained. The forecast series in the stationary form is then compared to the out-of-sample series in obtaining the forecast error.

However, the forecast data obtained is in the stationary form which is not in the original scale. To obtain the forecast data series as well as forecast evaluations in its original scale, there are two approaches. The first approach is by modifying the selected model based on the transformation chosen in the identification stage. The model with the retransformed scale is then used to get the out-of-sample series. The second approach is by retransforming the forecast transformed data. The forecast data series in its original scale then is used in evaluating the forecasting performance.

In the forecasting stage, the series of out-of-sample data are used to obtain the forecast results since the accuracy of forecasts can only be determined by considering how well a model performs on new data that were not used when fitting the model (Hyndman & Athanasopoulos, 2017). The validity and accuracy of a forecasting model is assessed by the cross-validation (CV) method. The valid forecasting model will demonstrate good predictive accuracy. Note that, the out-of-sample one-step ahead forecast is also known as one-step time series cross-validation.

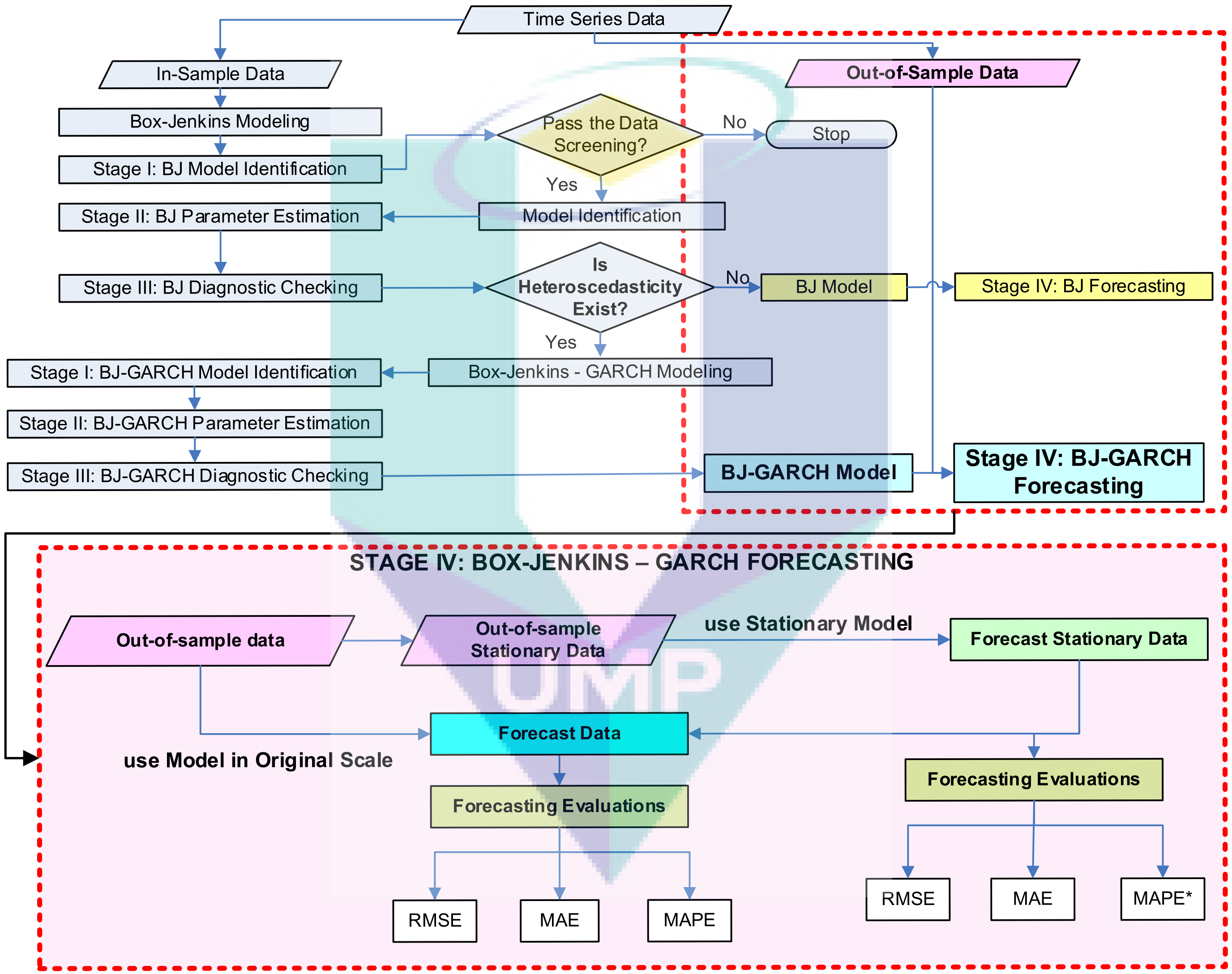


Figure 3.13 Detail procedures in Stage IV of the procedure of BJ-G for highly volatile data

3.4.4.1 Time Series Cross-Validation

Time series cross-validation is a version of CV that searches for a good one-step ahead forecast of the data (Hart, 1994). Let n be the number of forecasts and $\hat{y}_T(h)$ be the forecast made at origin T of the actual value y_{T+1} at future time $T+1$, that is, at lead time or forecasting horizon h. Here y_{T+1} refers alternatively to the out-of-sample series. For the one-step ahead forecast, the procedure of CV of a statistical model is given in the following steps (Hyndman & Athanasopoulos, 2017; Tsay, 2013):

Step 1: Divide the data set into two parts; in-sample and out-of-sample. There is no specific rule to guide the division, but each subsample should contain sufficient data points. Typical analyst used a ratio of 90:10 (Chatfield, 2001).

Step 2: Perform model estimation using in-sample data and use the fitted model to obtain 1-step ahead forecast and its forecast error. Suppose the in-sample data is $\{y_t \mid t = 1,2,\ldots,T\}$, then the fitted model using the first T data points is used to compute the one-step ahead forecast, $\hat{y}_T(1)$ and its forecast error, $e_T(1) = y_{T+1} - \hat{y}_T(1)$.

Step 3: Reestimate the model using $T+1$ data points and compute the 1-step ahead forecast and its forecast error. That is, compute $e_{T+1}(1) = y_{T+2} - \hat{y}_{T+1}(1)$, where $\hat{y}_{T+1}(1)$ is the 1-step ahead forecast of the newly fitted model at the forecast origin $T+1$.

Step 4: Repeat step 3 until $e_{T+(n-1)}(1) = y_{T+n} - \hat{y}_{T+(n-1)}(1)$.

Note that the CV procedure is also applicable for multistep ahead forecasting, i.e. $h = 2,3,\ldots,n$. The procedure of CV is known as backtesting in the finance literature.

3.4.4.2 Forecasting Evaluations

By applying the procedure of CV, the forecasting performance is compared across models based on three evaluation criteria which are commonly used in the previous literatures. The evaluation criteria are the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). In the forecasting evaluations, if the actual values and forecast values are closer to each other, a small forecast error will be obtained. Thus, smaller RMSE, MAE and MAPE values are preferred. The best forecasting model is the one that generates the lowest prediction error. However, if the results are not consistent among these criterions, it is suggested to choose MAPE since it is relatively more stable than others (Wang, Huang, & Wang, 2012). The evaluation criteria for one-step ahead forecast are given in Equation 3.56 to Equation 3.58. The evaluation criteria in Equation 3.56 to Equation 3.58 are also applicable for multistep ahead forecasting or when $h = 2, 3, ..., n$.

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}\left| e_j(1)\right| \tag{3.56}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{n}\left(e_j(1)\right)^2}{n}} \tag{3.57}$$

$$\text{MAPE} = \frac{100}{n}\sum_{j=1}^{n}\left|\frac{e_j(1)}{y_{T+j}}\right| \tag{3.58}$$

3.4.4.3 Prediction Intervals

In addition to assessing the best forecast data $\hat{y}_t$, it is necessary to specify their accuracy to measure the risks associated with decisions based upon the forecasts may be assessed. The accuracy of the forecasts can be expressed by calculating probability limits on either side of each forecast or also known as prediction intervals (PIs) (Box et al., 2008). Computing prediction intervals (PIs) is an important part of the forecasting process as it is useful to quantify the accuracy of the forecast data (Chatfield, 2001). These limits may be calculated for any convenient set of probabilities, such as 95%, 80% or 50%. However, suitable percentages may be used. These limits are such that the given information available at origin *T*, there is a probability of $1-\alpha$ that the actual value is in the out-of-sample series, y_{T+1}, when it occurs, will be within them, which is expressed as Equation 3.59.

$$P\left(y_{T+1}(-) < y_{T+1} < y_{T+1}(+)\right) = 1-\alpha \tag{3.59}$$

Then, a 95% prediction interval for one-step ahead forecast of y_{T+1} for a_t normally distributed, is given by Equation 3.60, where $\text{Var}\left[e_T(1)\right]$ is the variance of the

1-step ahead forecast error that is defined as Equation 3.61. The a_{T+1} is referred to as the shock to the series at $T+1$, which is also known as the one-step ahead forecast error at the forecast origin *T*. In practice, the estimated value of $\text{Var}[e_T(1)]$ can be obtained from the variance of the one-step ahead forecast residuals of the model considered. The limit in Equation 3.59 and the PIs in Equation 3.60 are also applicable to multistep ahead forecasting, i.e. $h = 2,3,...,n$.

$$\hat{y}_T(1) \pm Z_{0.025}\sqrt{\text{Var}[e_T(1)]} \qquad 3.60$$

$$\text{Var}[e_T(1)] = \text{Var}(a_{T+1}) = \sigma_a^2 \qquad 3.61$$

Consequently, a $(1-\alpha)100\%$ PI for *h*-step ahead forecasting and a_t follows normal distribution is given by Equation 3.62. If a_t has heavy-tailed characteristics, then Equation 3.62 is modified by changing $Z_{\alpha/2}$ to the appropriate error distribution with heavier tails than the normal.

$$\hat{y}_T(h) \pm Z_{\alpha/2}\sqrt{\text{Var}[e_T(h)]} \qquad 3.62$$

There are several discussions on how to determine the value of $\text{Var}[e_T(h)]$. According to Chatfield (2001), for a series that shows no obvious trend, autocorrelation or seasonality, the $\text{Var}[e_T(h)]$ can be determined using Equation 3.63. Box et al. (2008) recommend that the use of Equation 3.63 is optimal for an ARIMA(0,1,1) model. For a random walk model, the $\text{Var}[e_T(h)]$ can be estimated using Equation 3.64. The Equation 3.64 is true for a random walk model and it can be in error when it is applied to other types of model (Chatfield, 2001).

$$\text{Var}[e_T(h)] = [1+(h-1)\alpha^2]\text{Var}[e_T(1)] \qquad 3.63$$

$$\text{Var}[e_T(h)] = h\,\text{Var}[e_T(1)] \qquad 3.64$$

3.5 The Modified Procedure for Univariate Highly Volatile Data using BJ-G

The earlier proposed procedure of BJ-G in Section 3.4 is further modified specifically for univariate highly volatile time series data. Therefore, a significant modification is done on Stage I of the procedure of BJ-G in Figure 3.5 to ensure the data series is a highly volatile data at the early stage, which is the data is prepared appropriately to use the BJ-G model. Hence, in Stage I of the newly proposed procedure of BJ-G, two steps are introduced that are the preliminary heteroscedasticity test and the BJ-G model identification. These two steps are used to justify the significance of adding volatility model to the Box-Jenkins and to identify the appropriate BJ-G model, respectively.

The modified procedure will simultaneously able to determine the optimal number of data required for BJ-G model. Determination of the optimal number of data using a statistical model for practical application is one of the main issues in time series forecasting (Chatfield,2001; Hyndman & Athanasopoulos, 2014; Hyndman & Konstenko, 2017). According to Hyndman and Kostenko (2017), the number of data required for any statistical model depends on at least two items: the number of model parameters to be estimated and the amount of random variation existing in the data. It means, a reasonable approach to determine the appropriate number of data for forecasting is to ensure that there is enough data to estimate the model and the model performs well for out-of-sample evaluation. By referring to Figure 3.5, these two items are incorporated in Stage II and Stage IV of the proposed procedure by considering the selection criteria of AIC and SIC in the model estimation stage and by applying the out-of-sample one-step ahead forecasting evaluations using MSE, RMSE, MAE and MAPE in the forecasting stage. Konishi and Kitagawa (2008) reported that as the number of data increases, minimising the AIC is equivalent to minimising the MSE.

Figure 3.14 illustrates the new proposed procedure of BJ-G in modelling and forecasting highly volatile time series data that will simultaneously ensure the optimal number of data required for BJ-G model. In Stage I of the new proposed procedure, there are eight steps that are data plotting, data descriptive statistics, data stationarity, preliminary linearity test, Portmanteau test, BJ model identification, preliminary heteroscedasticity test and BJ-G model identification. For Stage II to Stage IV in the newly proposed procedure, the procedure and the method used are the same as in the proposed procedure in Section 3.4. In this study, the new proposed framework is employed to different data series of daily world gold price in determining the optimal number of data using BJ-G model. Note that, this new proposed procedure of BJ-G is the improvised version of procedure for BJ-G as proposed in Section 3.4.

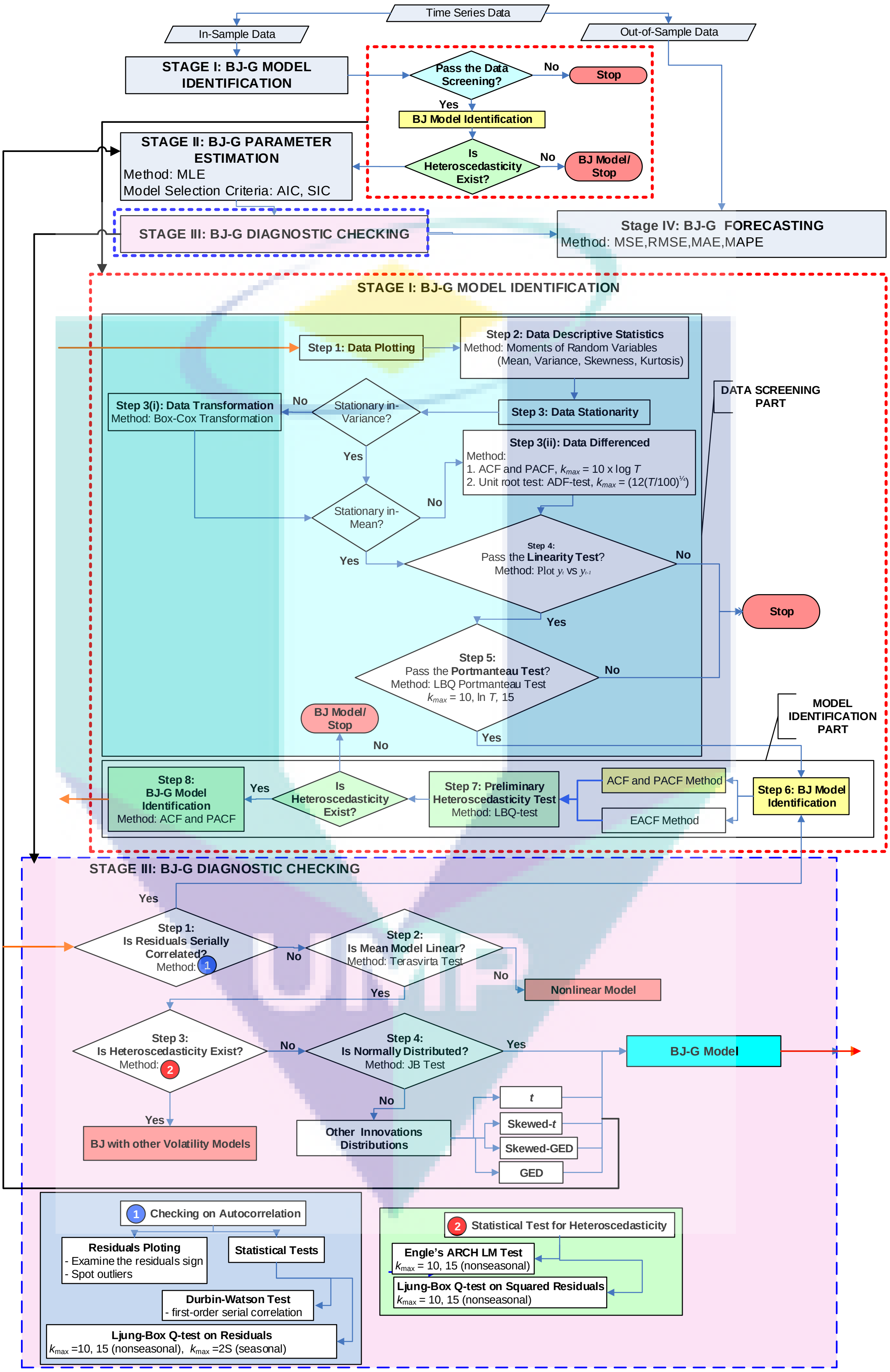


Figure 3.14 New proposed procedure of BJ-G in forecasting highly volatile data

3.6 Multistep Forecasting for Highly Volatile Data using Modified BJ-G Procedure

The second proposed procedure of BJ-G is only applied for one-step ahead forecasting performance, which is not practical for real data due to its limitation of the prediction period (Babu & Reddy, 2015; Pham & Yang, 2010; Byström, 2005). Hence, the study on the multistep ahead forecast is important since it is significant for practical application purposes using the BJ-G model. Therefore, the following study is aimed at proposing a modified procedure of BJ-G in evaluating the multistep forecasting performance of Box-Jenkins – GARCH (or BJ-G) model for highly volatile time series data. In investigating the performance of multistep ahead forecasting for the BJ-G model, Stage IV in the procedure of Figure 3.14 is extended to *n*-step ahead forecasting by proposing a new procedure of BJ-G as presented in Figure 3.15.

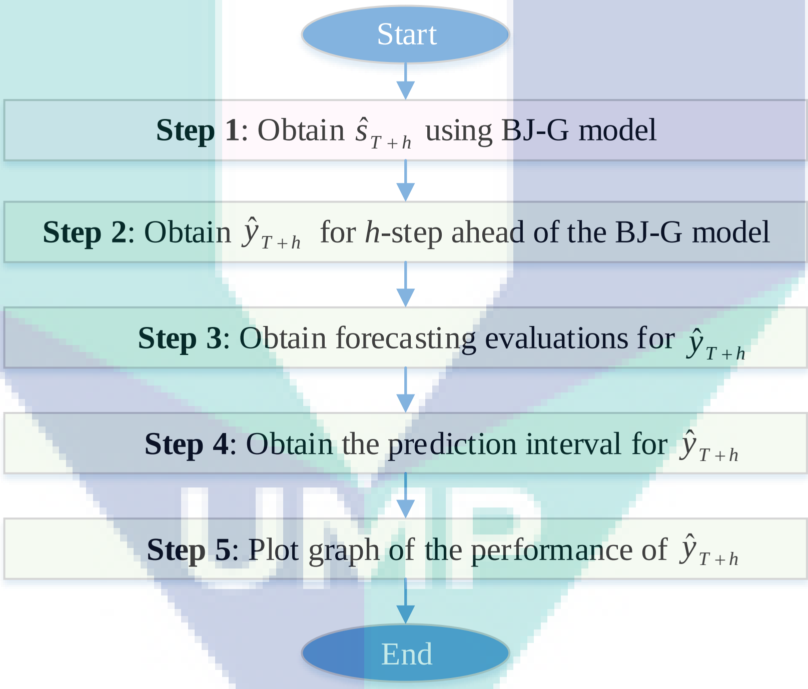


Figure 3.15 Proposed procedure of BJ-G for multistep ahead forecasting

Note that, the procedure for Stage I to III of the procedure of BJ-G are the same as discussed in Section 3.5. In order to achieve the objective, the procedure and programming codes are constructed for multistep ahead forecast using BJ-G model. It has been observed that available software is only able to provide the results for one-step ahead forecast. In the proposed procedure, sets of codes are constructed in R for evaluating the forecasting performance up to *n*-step ahead, which is based on the proposed model of BJ-G. The proposed procedure in Figure 3.15 is explained explicitly in the following steps:

Step 1**:** Obtain the simulated stationary series, $\hat{s}_{T+h}$ for forecasting horizon $h = 1,2,3,\ldots,n$ using the proposed BJ-G model. There are two approaches to obtain $\hat{s}_{T+h}$; the first approach is using the results of one-step ahead forecast of BJ-G model from available software (EViews). In the second approach, the $\hat{s}_{T+h}$ series is obtained through simulation on BJ-G model using programming codes.

Step 2**:** Obtain the forecast data for *h*-step ahead, $\hat{y}_{T+h}$ of the BJ-G model. The corresponding R codes of $\hat{y}_{T+h}$ for one-step ahead are written.

Step 3**:** Obtain forecasting evaluations of MAE, RMSE and MAPE for *h*-step ahead forecast by comparing $\hat{y}_{T+h}$ and the out-of-sample data y_{T+h}. The corresponding R codes for the forecasting evaluations of $\hat{y}_{T+h}$ for one-step ahead are written.

Step 4**:** Obtain the prediction intervals (PIs) for $\hat{y}_{T+h}$. The PIs gives an interval within which the actual data, y_t is expected to lie with a specified probability by using the forecast, $\hat{y}_{T+h}$. In this study, the PIs used are 80% and 95%, which is commonly used in forecasting method as suggested by Hyndman and Athanasopoulos (2013). The R codes for PIs of 80% and 95% of $\hat{y}_{T+h}$ using one-step ahead forecast are written.

Step 5**:** Graphical presentation for the performance of the forecast data is shown by plotting the graph of actual data in the out-of-sample series, y_{T+h} and the *h*-step ahead forecast, $\hat{y}_{T+h}$ with its prediction intervals. The R codes for plotting the performance with PIs of 80% and 95% for one-step ahead forecast are written.

The procedure from Step 1 to 5 for $h = 2,3,\ldots,n$ is repeated in order to obtain the multistep ahead forecast evaluations for BJ-G model.

3.7 Modified BJ-G Procedure for GARCH-type Models

In recent years, many studies proposed the incorporation of Box-Jenkins with GARCH-type model due to its good performance in dealing with highly volatile data. Based on literatures, some of the studies on univariate highly volatile data that incorporate the Box-Jenkins model with GARCH-type are ARIMA-GARCH (C. Chen et al., 2011; Tan et al., 2010; Zhou et al., 2006), AR-EGARCH (Ahmed, 2017; Ferenstein & Gasowski, 2004; Girish, 2016; Walid, Chaker, Masood & Fry, 2011), AR-GARCH (Gaglianone & Marins, 2017), ARIMA-APARCH (Girish, 2016), ARIMA-TGARCH (Ahmad et al., 2015; Freedi et al.,2012), ARMA-GARCH (Liu & Shi, 2013; Pham & Yang, 2010; Wang, Gelder, Vrijling, & Ma, 2005), ARMA-EGARCH (Ord, Koehler, Snyder & Hyndman, 2009) and ARIMA-GARCH-M (Liu, Erdem & Shi, 2011; Liu & Shi, 2013; Liu, Shi, & Qu, 2013).

Since the combination model of Box-Jenkins and GARCH-type has great potential for research that deals with univariate highly volatile time series data, the comprehensive procedure of BJ-G should be considered in the study. Therefore, the comprehensive procedure of BJ-G which is the fourth proposed procedure in this study is developed from the second and third procedures (refer to Figure 3.14 and Figure 3.15 for the second and the third proposed procedures, respectively). The fourth proposed procedure of BJ-G is illustrated by Figure 3.16.

The fourth proposed procedure of BJ-G is not only applicable for Box-Jenkins with standard GARCH model but it is also can be applied to Box-Jenkins with all GARCH-type models under consideration in previous studies on highly volatile data that are GARCH-M EGARCH, TGARCH and APARCH. Note that, the procedure of combination Box-Jenkins with standard GARCH model as in Figure 3.4 and the modelling procedure in estimating GARCH parameters as discussed by Pham and Yang (2010) are also applicable to other GARCH-type models under study.

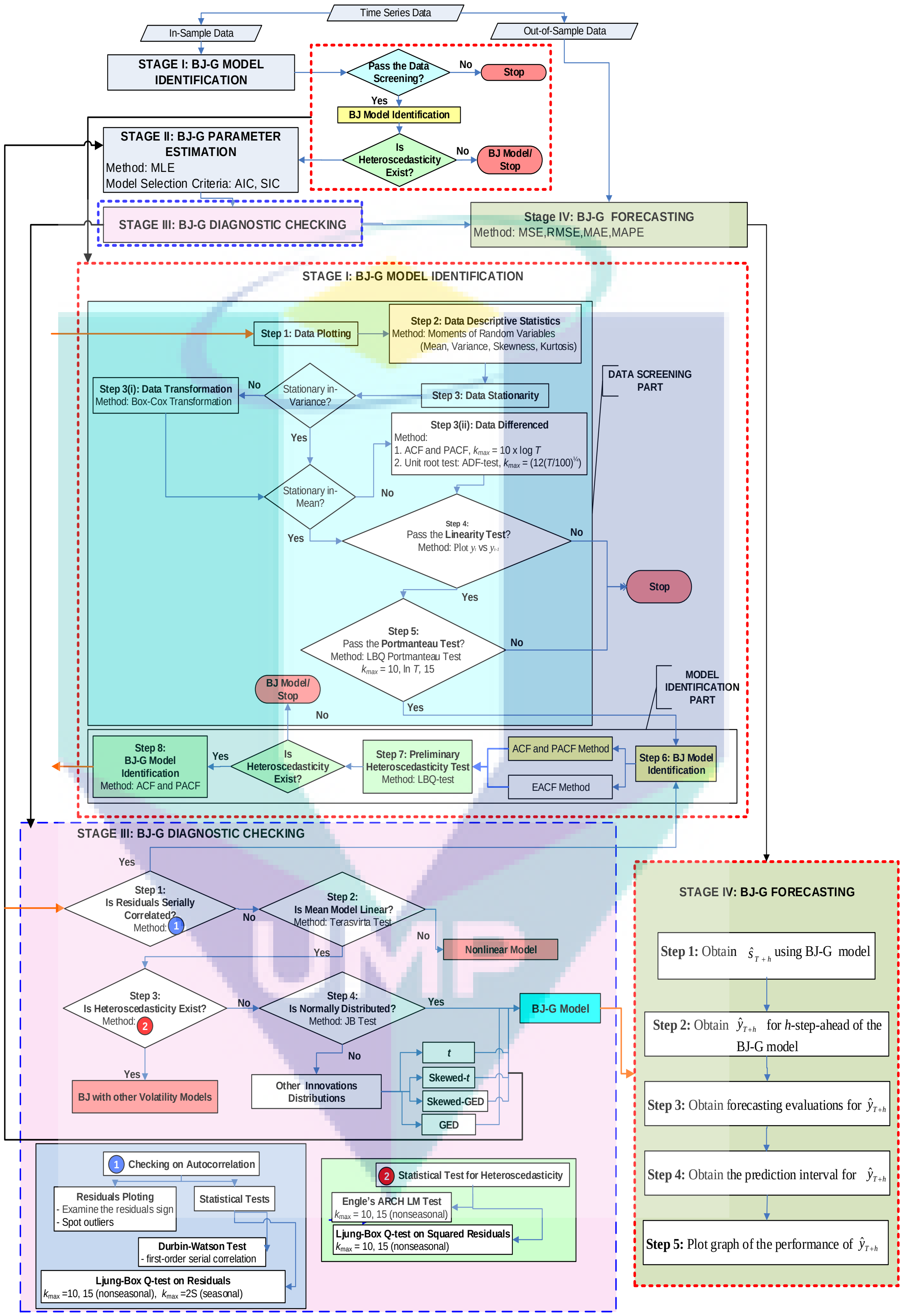


Figure 3.16 A modified comprehensive proposed procedure of BJ-G for modelling and forecasting univariate highly volatile time series data

The GARCH-M model, where "M" stands for GARCH in the mean, is one of the symmetric GARCH-type models, as it is symmetric in response to the past volatility. The $\text{GARCH}(r,s)-\text{M}$ model can be written as Equation 3.65 (Tsay, 2013).

$$\begin{gathered} s_t = \mu + M\sigma_t^2 + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim IID(0,1) \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i a_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 \end{gathered} \tag{3.65}$$

where s_t is the stationary series, σ_t^2 is the conditional variance of s_t, M is the risk premium parameter and α_i and β_i are parameters satisfying conditions similar to those of GARCH model. A positive M indicates that the series is positively related to its past volatility. The GARCH-M model is used to model the phenomenon of a series that may depend on its volatility. The existence of risk premium in the GARCH-M model implies that there are serial correlations in the series.

The exponential GARCH or EGARCH is proposed by Nelson (1991) to overcome some weaknesses of the GARCH model in terms of the leverage effect and parameter restrictions (Freedi et al., 2012). EGARCH is one of the asymmetric GARCH-type models. An $\text{EGARCH}(r,s)$ model can be written as Equation 3.66, where s_t is the stationary series, σ_t^2 is the conditional variance of s_t, μ_t is conditional mean of s_t, α_i and β_i are parameters satisfying conditions similar to those of GARCH model and g_i signifies the leverage effect of a_{t-i} (Tsay, 2013; Freedi et al., 2012). It is expected that g_i to be negative in real applications, if it is exists. Note that, a positive a_{t-i} or there is "good news" contributes $\alpha_i(1+g_i)|\varepsilon_{t-i}|$ to the log volatility, in contrast a negative a_{t-i} or there is "bad news" gives $\alpha_i(1-g_i)|\varepsilon_{t-i}|$ where $\varepsilon_{t-i} = \dfrac{a_{t-i}}{\sigma_{t-i}}$.

$$\begin{gathered} s_t = \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim IID(0,1) \\ \ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{r} \alpha_i \frac{|a_{t-i}| + g_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{s} \beta_j \ln \sigma_{t-j}^2 \end{gathered} \tag{3.66}$$

The threshold generalised autoregressive conditional heteroscedastic or TGARCH is one of the commonly used volatility models in handling leverage effects in a data series. For a univariate series of s_t, a TGARCH (r,s) model is given by Equation 3.67 (Tsay, 2013).

$$\begin{aligned} s_t &= \mu_t + a_t, \qquad a_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim IID(0,1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{r} (\alpha_i + g_i N_{t-i}) a_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 \end{aligned} \qquad 3.67$$

where N_{t-i} is an indicator for negative a_{t-i}, that is

$$N_{t-i} = \begin{cases} 1 & \text{if } a_{t-i} < 0, \qquad \text{for } i = 1,2,\ldots,r \\ 0 & \text{if } a_{t-i} \geq 0, \end{cases}$$

and s_t is the stationary series, μ_t is conditional mean of s_t, σ_t^2 is the conditional variance of s_t, α_i and β_i are parameters satisfying conditions similar to those of GARCH model. Noted that, g_i signifies the leverage effect of a_{t-i}. It is expected that g_i to be negative in real applications (Tsay, 2013). This TGARCH model is also known as the GJR model since Glosten et al. (1993) essentially proposed the same model.

The asymmetric power autoregressive conditional heteroscedastic or APARCH model is proposed by Ding, Granger, & Engle, 1993. The APARCH(r,s) model can be written as Equation 3.68 where s_t is the stationary series, σ_t^2 is the conditional variance of s_t, μ_t is conditional mean of s_t, α_i and β_i are parameters satisfying conditions similar to those of GARCH model, g_i signifies the leverage effect of a_{t-i} and δ is a positive real number (Tsay, 2013). If $\delta = 0$ in Equation 3.68, then the APARCH model becomes the EGARCH model of Nelson (1991), while if $\delta = 2$, the APARCH model becomes to TGARCH. Similar to GARCH model, the APARCH (1,1) model is often used in practice (Tsay, 2013).

$$\begin{aligned} s_t &= \mu_t + a_t, \qquad a_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim IID(0,1) \\ \sigma_t^\delta &= \alpha_0 + \sum_{i=1}^{r} \alpha_i \left(|a_{t-i}| + g_i a_{t-i} \right)^\delta + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^\delta \end{aligned} \qquad 3.68$$

3.8 Concluding Remarks

The Box-Jenkins – GARCH's procedure based on Box-Jenkins modelling was proposed to handle highly volatile data. There are four procedures of BJ-G proposed in this study, as given by Figure 3.5, Figure 3.14, Figure 3.15 and Figure 3.16. The first proposed procedure of BJ-G as shown by Figure 3.5 is developed based on the standard Box-Jenkins's procedure since it will be used in evaluating the performance of the combination model to forecasting univariate highly volatile data for the preliminary study. The second proposed procedure of BJ-G model as shown in Figure 3.14 is developed specifically dealing for univariate highly volatile data at the early stage which simultaneously ensure the optimal number of data required for BJ-G model. The third proposed procedure of BJ-G as given by Figure 3.15 is applied in evaluating the multistep ahead forecasting performance of the BJ-G model. While, the fourth proposed procedure as given by Figure 3.16 is a comprehensive BJ-G procedure for modelling and forecasting univariate highly volatile time series data using the Box-Jenkins with GARCH-type model.

The performance of the four proposed procedures of BJ-G will be illustrated using real life data, specifically the world gold prices as will be discussed in details in Chapter 4. The daily gold price data are selected since it is expected to be a highly volatile type of time series data.

CHAPTER 4

GOLD PRICE FORECASTING USING MODIFIED PROCEDURE OF BOX - JENKINS – GARCH FOR HIGHLY VOLATILE TIME SERIES: A CASE STUDY

4.1 Introduction

Gold is noted as a volatile monetary asset commodity (Batten, Ciner & Lucey, 2010; Lucey, Larkin & O'Connor, 2013; Yaya, Tumala & Udombodo, 2016). The world gold prices are used to illustrate the proposed procedures of Box-Jenkins – GARCH (BJ-G) in Chapter 3. This chapter presents the empirical results of the highly volatile gold prices series. Section 4.2 presents the preliminary analysis on gold price using the modified BJ-G as illustrated in Figure 3.5. The analysis describes a step-by-step for four stages in the procedure using BJ-G model to determine its suitability in modelling and forecasting gold prices, specifically at one-step ahead.

In section 4.3, the second proposed procedure of BJ-G is applied to the gold price series in evaluating the performance of the procedure in specifically handling highly volatile data for practical application. The promising results from one-step ahead out-of-sample forecast series using the second procedure BJ-G has motivated the extension to multiple-step ahead forecast as will be discussed in details in Section 4.4. In Section 4.5, the fourth proposed procedure of BJ-G is applied to Box-Jenkins with all GARCH-type models as in the previous studies on highly volatile data including standard GARCH, GARCH-M, EGARCH, TGARCH and APARCH in determining the best GARCH-type model in handling volatility, specifically in the gold price series. The data is analysed using EViews and R programming language. The empirical results are thoroughly explained and the detail analyses can be referred to Appendix 2 to 5 for Section 4.2 to 4.5, respectively.

4.2 Preliminary Analysis on Gold Price Forecasting using Modified BJ-G

Based on the previous studies of Box-Jenkins – GARCH-type model on highly volatile data, there are various number of data used starting 180 to almost 42000, depending on frequency of data either monthly, weekly, daily or hourly (Ahmed, 2017; Babu & Reddy, 2015; Byström, 2005; Ferenstein and Gasowski, 2004; Gaglianone & Marins, 2017; García-Ferrer et al., 2012; Girish, 2016; Harrison & Paton, 2004; Koopman et al., 2007; Liu & Shi, 2013; Loi & Ng, 2018; Ord et al., 2009; Pham & Yang, 2010; Sohn & Lim, 2007; Walid et al., 2011). In the preliminary analysis, the daily basis price series is chosen due to it is the shortest frequency of available data. Noted that, in general, the capability to forecast in shorter time periods means faster response to fluctuation of the data. Since the data in this study is daily basis, the number of data from 500 to 5000 are usually considered in the related literatures (Babu & Reddy, 2015; Ferenstein & Gasowski, 2004; Gaglianone & Marins, 2017; García-Ferrer et al., 2012; Harrison & Paton, 2004; Koopman et al., 2007; Sohn & Lim, 2007). Therefore, in the preliminary study, 5000 daily gold price series is considered to ensure all significant characteristics related to volatility of data are captured.

The daily world gold prices price data used in the study starts from 24th November 1993 to 17th December 2013 of 5-day-per-week frequencies. Values are quoted in US dollars per ounce and the data is obtained from a reliable source of www.kitco.com. However, there are some missing price values in the original series due to holiday and stock market closing day. The data is divided into two parts: (i) in-sample data of period from 24th November 1993 to 20th December 2011 with 4500 observations, and (ii) out-of-sample period from 21st December 2011 to 17th December 2013 with 500 observations. The in-sample data is used to estimate model, whereas the out-of-sample data is used in model forecasting, with the ratio of estimate to forecast 90:10. The ratio of 90:10 is used in this preliminary study since it is a typical ratio used by analyst (Chatfield, 2001).

4.2.1 Stage I: Gold Price Data Identification

In employing Box-Jenkins modelling, the model cannot be directly applied if the series is nonstationary. It is important to know whether the data contains any trend or seasonal components. By referring to Figure 3.5, the first step of identification is to check the occurrence of an upward or downward trend as well as seasonality in gold price movement by plotting in-sample series as shown in Figure 4.1. From the figure, we can observe that the gold price series does not vary around a fixed level which indicates that the series is nonstationary in both mean and variance, exhibiting an overall upward and nonseasonal trend.

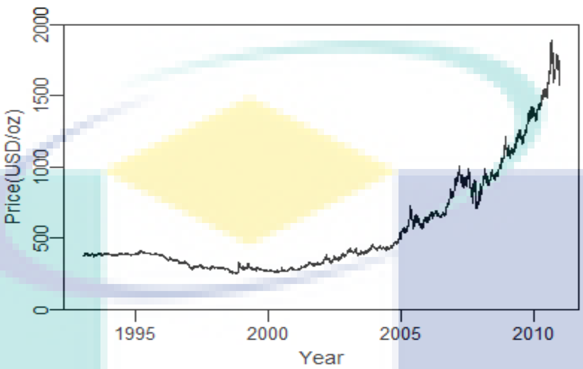


Figure 4.1 In-sample series of daily gold price

Based on the descriptive statistics of the series as given in Table 4.1, it can be seen that most of the data is around USD 560/oz with a standard deviation of USD 363/oz. The sample skewness and kurtosis for the original series are 1.5925 and 4.7121, respectively. In testing the skewness of the series, the hypothesis is $H_0 : S(y_t) = 0$ versus $H_1 : S(y_t) \neq 0$. Since the *t*-test statistic of 43.6124 and its *p*-value close to zero, hence, the null hypothesis of zero skewness is rejected at the 5% significance level. For excess kurtosis, the hypothesis is $H_0 : K(y_t) - 3 = 0$ versus $H_1 : K(y_t) - 3 \neq 0$. The *t*-test statistic of excess kurtosis for the in-sample series is 23.4439 with *p*-value close to zero, thus the null hypothesis is rejected. The values and test of hypotheses for skewness and kurtosis imply that the series is asymmetric, positively skewed and leptokurtic, as graphically shown by histogram in Figure 4.2. These characteristics imply that the normality assumption for the daily gold price is rejected at any level of significance by the Jarque-Bera test (JB-test), with test statistic of 2454.4068, which is very large as compared to a chi-square distribution with 2 degrees of freedom.

Table 4.1 Descriptive statistics for in-sample series

Min	**Max**	**Mean**	**Median**	**Std. dev.**	**Skewness**	**Kurtosis**	**JB-test**	**NoO**
252.8	1895	558.7877	387.1250	363.4643	1.5925	4.7121	2454.4068 (0.0000)	4500

* Std. dev is abbreviated for standard deviation, the values in parenthesis denotes *p*-value and NoO is abbreviated for number of observation.

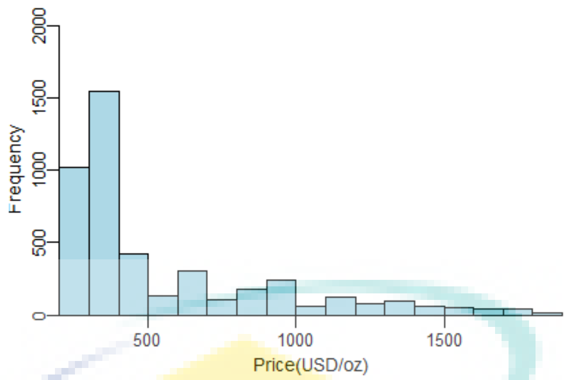


Figure 4.2 Histogram for in-sample series

4.2.1.1 Data Stationarity

The original gold price series depicts nonstationary behaviour, therefore the data needs to be handled first by stabilising the variance which will suggest the appropriate transformation series (Hyndman & Athanasopoulos, 2014). The transformation series, if needed, will be checked on the stationarity in-mean. In this study, the Box-Cox transformation is used as the variance stabilising method. Meanwhile, the ACF and PACF as proposed by Box and Jenkins (1968) are used in analysing stationarity in-mean of the transformation series, which is supported by ADF-test results.

Based on the Box-Cox transformation analysis, the best estimated power value of $\lambda = -0.2147$ which is close to 0, implies that the transformation of $y_t^* = \ln y_t$, where y_t^* is the transformed data and y_t is the observed data, is appropriate to stabilise the variance in the data series. The log transformation is chosen as compared to the transformation with exact value of λ, since it is easier to back-transform to the original price data for the forecasting purpose. Note that the in-sample gold price series is positively skewed, and this agrees with well-established guidelines of transformation that a log transformation is recommended for positively skewed data (Olivier & Norberg, 2010).

The plot of the log transformed series is shown in Figure 4.3. Based on descriptive statistics of the transformed data given in Table 4.2, it can clearly be seen that the series is less volatile up to 99.85%, suggesting that the log transformation indeed helps in stabilising the amplitude of the gold price. On the other hand, the log transformation not only stabilises the variance, but also improves the approximation of data normality. Figure 4.4 shows the histogram of the transformed data which is symmetrically distributed as compared to the histogram of the original price data shown in Figure 4.2, which indicates that the log transformation did improve the normality of the data.

Table 4.2 Descriptive statistics for the transformed data series

Min	Max	Mean	Median	Std. dev.	Skewness	Kurtosis	JB-test	NoO
5.5326	7.5470	6.1652	5.9587	0.5315	0.8959	-0.4107	633.8854 (0.0000)	4500

* Std. dev is abbreviated for standard deviation, the values in parenthesis denotes *p*-value and NoO is abbreviated for number of observation.

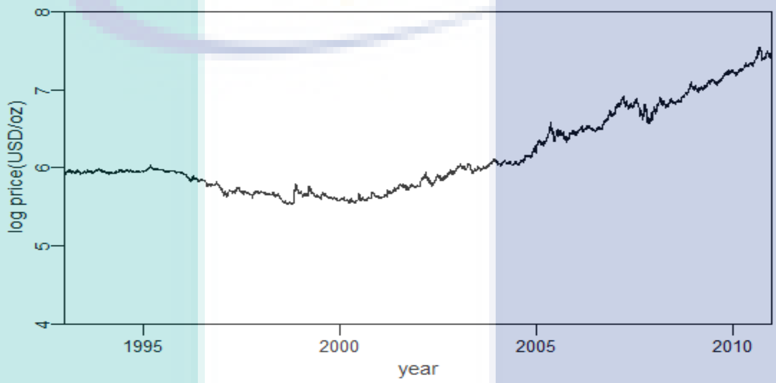


Figure 4.3 The transformed data of daily gold price for in-sample period

However, the $p\text{-value} = 0.0000$ of the JB-test rejects normality in the log data. Even though the transformation method is able to stabilise effectively the variance in the data, Figure 4.3 graphically shows that the trend still exists in the log series. This indicates that the transformed series is still not stationary, specifically in-mean. Therefore, the ACF and PACF of the log series are investigated as recommended by Box and Jenkins (1968) in handling the in-mean nonstationary case. The in-mean nonstationary behaviour is then statistically investigated using unit root test of ADF-test as proposed by Said and Dickey (1984).

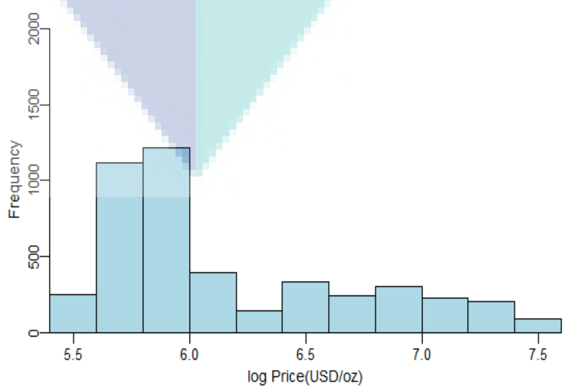
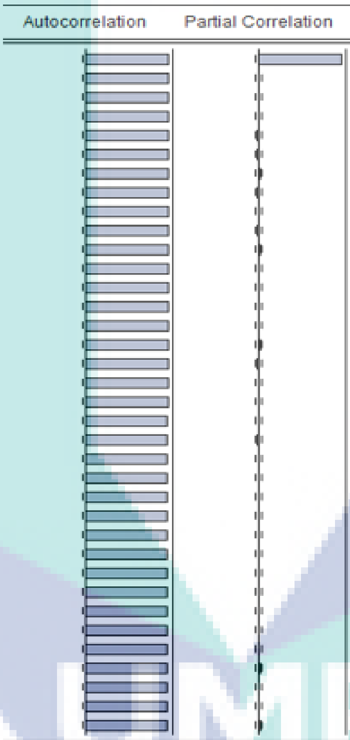


Figure 4.4 Histogram for in-sample transformed series

The correlogram or sample ACF, r_k of the log series, as shown in Figure 4.5, suggested that the transformed data is nonstationary, specifically in-mean, due to the spikes which is slowly decaying to zero. The pattern of the ACF that shows spikes in the sample on one side indicates that the log series does not have seasonal behaviour (Box & Jenkins, 1968; Hanke, Reitsch, & Wichern, 2001), which confirms the nonseasonal trend as shown in Figure 4.1. The number of data used for this univariate in-sample series is $T = 4500$ implies that the maximum lag for the ACF and PACF is $k_{\max} = 10\log(4500) \approx 36$, by applying Equation 3.40. The sample ACF with slowly decaying to zero up to $k_{\max} \approx 36$ suggested that this log series needs differencing.



	AC	PAC	Q-Stat	Prob
1	0.999	0.999	4495.7	0.000
2	0.998	0.007	8985.4	0.000
3	0.998	-0.001	13469.	0.000
4	0.997	0.005	17946.	0.000
5	0.996	-0.023	22417.	0.000
6	0.995	-0.023	26882.	0.000
7	0.994	0.013	31339.	0.000
8	0.993	-0.013	35791.	0.000
9	0.993	-0.004	40235.	0.000
10	0.992	-0.011	44672.	0.000
11	0.991	0.012	49103.	0.000
12	0.990	-0.003	53527.	0.000
13	0.989	0.007	57945.	0.000
14	0.988	-0.003	62356.	0.000
15	0.987	-0.001	66760.	0.000
16	0.987	0.015	71157.	0.000
17	0.986	-0.011	75548.	0.000
18	0.985	0.008	79933.	0.000
19	0.984	-0.002	84311.	0.000
20	0.983	0.003	88683.	0.000
21	0.982	-0.015	93048.	0.000
22	0.982	0.003	97406.	0.000
23	0.981	-0.006	101758	0.000
24	0.980	-0.002	106103	0.000
25	0.979	0.007	110442	0.000
26	0.978	-0.003	114774	0.000
27	0.977	0.006	119099	0.000
28	0.976	0.006	123419	0.000
29	0.976	0.004	127731	0.000
30	0.975	-0.004	132038	0.000
31	0.974	-0.002	136338	0.000
32	0.973	0.005	140631	0.000
33	0.972	0.010	144919	0.000
34	0.972	0.004	149200	0.000
35	0.971	0.003	153476	0.000
36	0.970	0.011	157745	0.000

Figure 4.5 The ACF and PACF of the log series

To confirm statistically that the transformed data is nonstationary in-mean, the ADF-test is employed to the log series as given in Table 4.3. The ADF-test using EViews is based on the left-tailed t distribution as proposed by MacKinnon (1996). Based on the ADF-test with the maximum lag $k_{\max} = 12\left((4500/100)^{\frac{1}{4}}\right) \approx 31$, the t-test statistics is -1.0016, which is greater than the t-critical of -3.4108 at 5% significance level. Hence, the null hypothesis of a presence of unit root in the data series is not rejected. If we consider the analysis based on p-value, the p-value is 0.9422 that is greater than 0.05, therefore we arrive at the same conclusion. The presence of unit root suggests that the data series is nonstationary. Hence, data differencing is needed for the transformed data (log data) to make the series stationary.

Table 4.3 Augmented Dickey-Fuller unit root test on transformed data

t_{test}	$-t_{0.05,37}$	***p*-value**	**Number of lag**
-1.0016	-3.4108	0.9422	31

Thus, the log gold price series need to be differenced, $y_t - y_{t-1}$ in order to remove the trend and obtain a stationary series. The ADF-test for first order difference of the log series is -13.9454 which is much smaller than the 5% significance level of test critical value, as given by Table 4.4. The p-value = 0.0000 indicates that the ADF t-statistic is significant and there is no unit root in the data series, which suggests that the first differenced of log price series is stationary. The stationarity of the first differenced log price series is then supported by the sample ACF and PACF patterns for the series as shown in Figure 4.6, where the values are reduced drastically to zero. This is agreed with the previous studies that gold price has nonstationary characteristics (Dunis & Nathani, 2007; Shafiee & Topal, 2010; Smith, 2002). Consequently, the gold price series is stationary after one lagged difference from the daily log price series or simply it is stationary in the form of the daily log return price series.

Table 4.4 Augmented Dickey-Fuller unit root test on first differenced log data

t_{test}	$-t_{0.05,37}$	***p*-value**	**Number of lag**
-13.9454	-3.4108	0.0000	31

Figure 4.7 graphically illustrates the stationarity of the first order differenced log gold price series since most of the data are located around the mean of zero, $\mu = 0$. However, there are some spikes in the figure which represents volatility clustering specifically starting year 2001 (around data of 1700-day in Figure 4.7) due to relatively weak supply of gold, geopolitical tensions since the 11 September terrorist attacks, the emergence of new markets in developing economies, growing speculation about the large US current-account imbalances and the required correction through a significant depreciation of the dollar which have contributed to the upward trend and higher volatility (Alcidi, De Grauwe, Gros, & Oh, 2010). As shown in the figure, there is clear evidence

of volatility clustering that is large or small asset price changes tend to be followed by other large or small price changes of either positive or negative sign. This implies that gold price return volatility changes over time.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.2247	0.635
		2	0.006	0.006	0.3854	0.825
		3	0.002	0.002	0.4010	0.940
		4	0.046	0.046	9.7870	0.044
		5	-0.009	-0.009	10.179	0.070
		6	-0.037	-0.038	16.498	0.011
		7	-0.017	-0.018	17.838	0.013
		8	0.005	0.003	17.958	0.022
		9	0.016	0.017	19.070	0.025
		10	-0.006	-0.003	19.245	0.037
		11	-0.042	-0.041	27.100	0.004
		12	-0.032	-0.035	31.745	0.002
		13	0.008	0.005	32.020	0.002
		14	0.010	0.012	32.494	0.003
		15	-0.028	-0.022	35.963	0.002
		16	0.049	0.051	46.834	0.000
		17	-0.001	-0.005	46.843	0.000
		18	0.008	0.003	47.163	0.000
		19	-0.007	-0.005	47.400	0.000
		20	0.029	0.027	51.287	0.000
		21	-0.021	-0.020	53.282	0.000
		22	-0.004	-0.004	53.344	0.000
		23	-0.022	-0.023	55.541	0.000
		24	-0.046	-0.049	64.928	0.000
		25	-0.025	-0.024	67.831	0.000
		26	-0.029	-0.028	71.667	0.000
		27	-0.013	-0.010	72.435	0.000
		28	-0.006	0.001	72.585	0.000
		29	-0.022	-0.023	74.712	0.000
		30	-0.007	-0.011	74.954	0.000
		31	-0.006	-0.004	75.108	0.000
		32	0.003	-0.001	75.147	0.000
		33	-0.031	-0.031	79.460	0.000
		34	0.009	0.006	79.829	0.000
		35	0.011	0.008	80.361	0.000
		36	-0.011	-0.022	80.917	0.000

Figure 4.6 ACF and PACF for first differenced log series

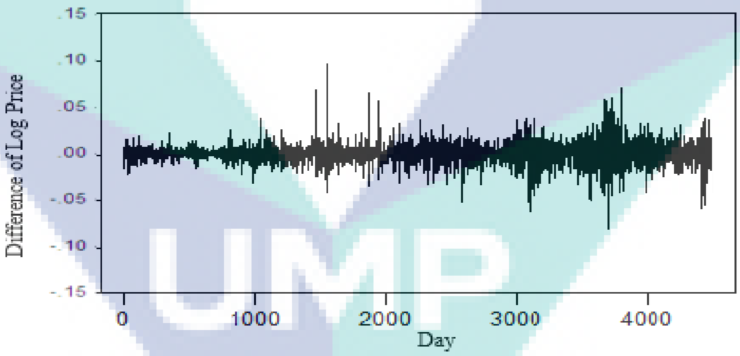


Figure 4.7 The first order difference of daily log gold price series

The mean of stationary series as shown in Table 4.5 is 0.0003. It is numerically supports that the average return is positive but very close to zero, however, the value is statistically significant at 5% significance level. The standard deviation for stationary series is 0.0106 which shows that the log return series is less volatile up to 99.99% as compared to the original series. The distribution of stationary data is symmetric as the hypothesis of zero skewness is accepted since the *t*-test statistic is 1.1008 with *p*-value more than 5% significance level. The symmetric characteristic of the stationary series is supported graphically by Figure 4.8.

Table 4.5 Descriptive statistics for stationary series

Min	Max	Mean	Median	Std. dev.	Skewness	Kurtosis	JB-test	NoO
-0.0797	0.0964	0.0003 (0.0398)	0.0002	0.0106	0.0402 (0.2710)	6.9142 (0.0000)	8974.38 (0.0000)	4499

* Std. dev is abbreviated for standard deviation, the values in parenthesis denotes *p*-value and NoO is abbreviated for number of observation.

On the other hand, the kurtosis for the log return has increased more than four times compared to the price data. The *t*-test of the excess kurtosis for the series is 94.6661 with *p*-value close to zero, imply that the log return series is leptokurtic with a higher peak and fatter tails as compared to the price data. The leptokurtic or heavy-tailed characteristic implies that the log return of gold price puts more mass on the tails and contains more extreme values. The test of hypotheses for mean, skewness and kurtosis indicate that, in general, the stationary series or log return series has more gains than losses, but the profit obtained is close to zero for most of the time (refer to Figure 4.7).

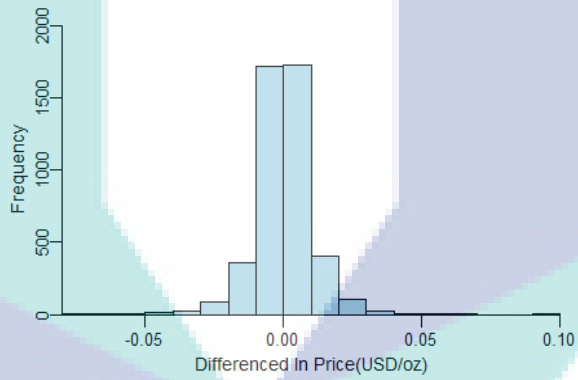


Figure 4.8 Histogram for stationary series

4.2.1.2 Preliminary of Linearity Test

Box-Jenkins approach is one of the widely used forecasting methods for linear data. Therefore, it is wise to conduct linearity test for the gold price data, specifically to stationary data, before the Box-Jenkins model is applied. The preliminary linearity step is necessary in identifying whether the data series fits the linear model or not at the early stage. Therefore, in the proposed Box-Jenkins framework, this step is applied in the identification stage.

Figure 4.9 shows the plot of differenced log price series versus its lagged 1 series. It can be seen that the plot is nearly a straight line, implying that the Box-Jenkins's linear model is appropriate for the data series. This is following the concept driven by Box and Jenkins (1968) for their models, that the deviation is linearly dependent on previous deviations. The deviation in this study is the stationary data itself, y_t^* since the mean of y_t^* is very close to 0.

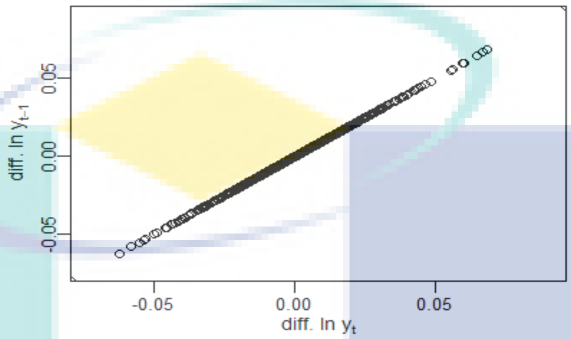


Figure 4.9 The plot of the first differenced log price and its previous deviations

4.2.1.3 Portmanteau Test

Since the linearity is proven for the stationary series of gold price, then the serial correlation of the series is checked either it is serially correlated data or not. It can be seen that most values of the sample ACF and PACF for the stationary series, as shown in Figure 4.6, are close to zero which indicates that the stationary series has tendency to be uncorrelated series. On the assumption that the stationary series completely random, therefore the standard error limits for sample ACF and PACF are the same, that is $\mathrm{se}[r_k] = \mathrm{se}[r_{kk}] \approx \frac{1}{\sqrt{4499}} = 0.0149$. Referring to the figure, most of the values of sample ACF and PACF are within two standard error limits of 0.0298, suggesting that the serial correlations in daily gold price returns are small, if any (Box et al., 2008; Box & Jenkins, 1968; Tsay, 2013). Based on these results, it can be concluded that the Box-Jenkins model is not appropriate to analyse the stationary series of gold price. However, it is not wise to make the conclusion just based on graphical related results.

Therefore, in the screening part of the proposed procedure of BJ-G, a Portmanteau test of LBQ-test is applied to verify whether the stationary series has no serial correlations as well as to verify the justification of application of the Box-Jenkins model to the data series. In the Portmanteau test, the hypothesis null of $H_0 : \rho_1 = \rho_2 = \ldots = \rho_k$ with $k = 9$ is tested for the stationary data. The choice of $k = \ln 4499 \approx 9$ is based on Tsay (2013) suggestion, since the in-sample log return price series is nonseasonal. For $k = 9$, the LBQ-test is 19.0700 with p-value 0.0246. It is clearly that the LBQ-test rejects the null hypothesis of no serial correlations in the log return price series at 5% significance level. Consequently, there exists serial correlations in the stationary series of the daily gold price series. Therefore, the serial correlations test leads to the justification of using Box-Jenkins models for the data series. The stationary data is now prepared well for the Box-Jenkins model identification part.

4.2.1.4 Box-Jenkins Model Identification

Based on the previous analysis, the daily gold price series has shown nonstationary and nonseasonal patterns. The series is stationary in-mean and variance after differencing of one lagged to the log price series which is agreed with the ADF unit root test and spikes patterns of the ACF and PACF for the differenced log series. The preliminary linearity test supports the use of the linear model, while the Portmanteau test for the stationary series leads to the application of Box-Jenkins models in handling the serial correlation in the series. Therefore, these results then reflect to the $\text{ARIMA}(p,1,q)$ as the appropriate Box-Jenkins model in analysing the daily gold price data.

The Box-Jenkins modelling makes use of the sample ACF and PACF to specify a model that can capture the dynamic dependence of the data. Thus, the r_k and r_{kk} for the stationary series as shown in Figure 4.7 is investigated to identify the order of the ARIMA model. Referring to the figure, most of the values of r_k and r_{kk} are close to zero and it is hard to identify graphically the appropriate order for the ARIMA model. Due to parsimony approach as practiced in the Box-Jenkins modelling, the values of both parameters p and q are suggested for 0, 1, and 2. These values of p and q are always appropriate for the stationary series in most application (Box & Jenkins, 1968).

On the assumption that ρ_k to be zero for all lags since most of the values of r_k graphically are close to zero, that is $q = 0$, then the estimated large lag standard error is $\text{se}[r_k] \approx \frac{1}{\sqrt{4499}} = 0.0149$. Referring to Figure 4.6, there are only seven lags that are greater than the two standard error limit of 0.0298, therefore the model with $q = 0$ can be accepted. Since the value of $r_1 = -0.0070$ is less than the two standard error of

0.0298, it can be concluded that ρ_1 is zero. Therefore, it might be reasonable to investigate the model of $q = 1$, that is to test the hypothesis of $H_0 : \rho_k = 0\,(k \geq 2)$ versus $H_1 : \rho_1 \neq 0$. Using Equation 3.32, the estimated large-lag standard error under this hypothesis is $\mathrm{se}[r_k] \approx \sqrt{\frac{1}{4499}\left[1 + 2(-0.0070)^2\right]} = 0.0149$ where $k > 1$. Based on the values of r_k in Figure 4.6, there are seven of the estimated autocorrelations for lags greater than 1 are greater than two standard error limit, therefore there is no reason to doubt the adequacy of the model that consists of $q = 1$.

The same method of model adequacy is applied to other considered parameters *p* and *q*. The estimated standard errors of autocorrelations and partial autocorrelations for all considered models are the same that is 0.0149. It is observed that the number of lags of sample ACF and PACF with greater than two standard error limit (or 0.0298) for the ARIMA models of (0,1,0), (0,1,1), (0,1,2), (1,1,0), (1,1,1), (1,1,2), (2,1,0), (2,1,1) and (2,1,2) are not more than seven, indicates that all considered models are adequate.

Alternatively, instead of using autocorrelation method in identifying the ARIMA model, we recommend one to apply the EACF method as introduced by Tsay and Tiao (1984) in the proposed framework. As shown in Table 4.6, the EACF result suggests the order of (0,1,0) since the triangle of "O" have its upper left vertex at the (0,0) position with the standard error of $2/\sqrt{4499} = 0.0298$. Referring to Figure 4.6, the behaviour of sample ACF and PACF of the stationary series likely agreed with the ARIMA(0,1,0) as suggested by the EACF result, but all other considered models are also worth entertaining.

Table 4.6 The simplified EACF table for the differenced log series

	MA Order:*q*												
p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	O	O	O	X	O	X	O	O	O	O	X	X	O
1	X	O	O	X	O	O	O	O	O	O	O	X	O
2	X	X	O	O	O	O	O	O	O	O	O	O	O
3	X	X	X	O	O	X	O	O	O	O	O	O	O
4	X	X	X	X	O	O	O	O	O	O	O	O	O
5	X	X	X	X	X	O	O	O	O	O	O	O	O
6	X	X	X	X	X	O	O	O	O	O	O	O	O

4.2.2 Stage II: Parameter Estimation of the Box-Jenkins Model

In the estimation stage, as can be referred to Figure 3.10, all the possible ARIMA models are estimated using MLE and OLS methods. Table 4.7 shows the results from the estimation stage using EViews at 5% significance levels with normalised AIC and SIC values for the nine possible ARIMA models. In general, the model with smaller AIC and SIC values are concluded to be the better estimation model.

Table 4.7 The results of estimation stage of the possible ARIMA models

Models	OLS		MLE	
	AIC	SIC	AIC	SIC
ARIMA(0,1,0)	- 6.2646	- 6.2632	- 6.2642	- 6.2613
ARIMA(0,1,1)*	- 6.2642	- 6.2614	- 6.2638	- 6.2595
ARIMA(0,1,2)*	- 6.2638	- 6.2595	- 6.2634	- 6.2577
ARIMA(1,1,0)*	- 6.2640	- 6.2612	- 6.2638	- 6.2595
ARIMA(1,1,1)	- 6.2643	- 6.2600	- 6.2633	- 6.2576
ARIMA(1,1,2)*	- 6.2638	- 6.2581	- 6.2629	- 6.2558
ARIMA(2,1,0)*	- 6.2634	- 6.2591	- 6.2634	- 6.2577
ARIMA(2,1,1)*	- 6.2632	- 6.2575	- 6.2629	- 6.2558
ARIMA(2,1,2)*	- 6.2632	- 6.2561	- 6.2625	- 6.2539

*The insignificant model

Based on the values of AIC and SIC for all significant models in Table 4.7, it can be concluded that the estimation using OLS is similar than MLE for ARIMA models, with very slight difference. Referring to the table, there are two possible models that are found to be significant at 5% significance level, ARIMA(0,1,0) and ARIMA(1,1,1). Since the AIC and SIC values for both models are equivalent, hence the models need to be chosen wisely. These results are in line with the models suggested by Box and Jenkins (1968) since there is similar characteristics in the sample ACF and the stationary series.

In the proposed procedure of BJ-G, the EACF approach introduced by Tsay and Tiao (1984) is recommended to be used to overcome the uncertainty of model chosen. The result of the EACF approach is supported by the smallest values of AIC and SIC using OLS estimation method as well as agreed with the model suggestion based on behaviour of autocorrelation functions by Box et al. (2008). Furthermore, according to the principle of parsimony that simple models are preferred to complex models when all things being equal, hence the model of ARIMA(0,1,0) is chosen as the best estimate model to model daily gold price.

Consequently, the estimation results of ARIMA(0,1,0) for stationary data of in-sample daily gold price data using OLS method is given by Table 4.8.

Table 4.8 Estimation result for ARIMA (0,1,0) model

Variable	**Coefficient**	**Standard error**	t_{test}	***p*-value**
c	0.0003	0.0002	2.0565	0.0398

In Table 4.8, the constant c are statistically significant at 5% level, hence the model of ARIMA(0,1,0) for the daily log return of the gold price data is given by Equation 4.1, where $a_t \sim NID(0,0.0001)$ with $\hat{\sigma}_a^2 = 0.0001$ and ∇y_t^* is the differenced log price data at time *t*. As given by the equation, it implies that the series has no significant serial correlation.

$$\nabla y_t^* = 0.0003 + a_t \qquad 4.1$$

The significance of c in the estimated model implies that the expected daily log return for gold market is about 0.03%, which is positive and remarkable. In fact, it is small, but has an important long term implication, supporting the common belief that the return of gold investment performs well in the long term, as can be explained as follows.

By using the *n*-period simple gross return as defined in Equation 4.2,

$$1 + R_t(n) = \prod_{j=1}^{n-1}\left(1 + R_{t-j}\right) \qquad 4.2$$

where R_t is simple return, therefore, the average annual simple gross return for daily log return gold price is

$$r_t = \left[\prod_{t=1}^{4499}\left(1 + y_t^*\right)\right]^{\frac{250}{4499}} - 1 = 0.0692$$

where $r_t = \ln(1 + R_t)$ which is called the log return and y_t^* is the daily log return (or stationary data). Hence, the average annual simple return is given by

$$R_t = \exp(r_t) - 1 = \exp(0.0692) - 1 = 0.0717$$

This shows that the daily simple return of gold investment grew about 7.17% per annum from 1993 to 2011. The return value for the investment is given by the compound return for *n*-period as defined by Equation 4.3,

$$\mathrm{FV} = \mathrm{PV}(R_t + 1)^n \qquad 4.3$$

where FV is the future value and PV is the present value. By substituting $\mathrm{PV} = 1$, and $R_t = 0.0717$ in Equation 4.3, therefore, a one-dollar investment in gold at the end of 1993 would be worth about $1(0.0717 + 1)^{18} \approx \mathrm{USD}\ 3.48$ at the end of 2011.

4.2.3 Stage III: Diagnostic Checking of the Box-Jenkins Model

The chosen model, ARIMA(0,1,0) is then examined carefully in the diagnostic checking stage in detecting the adequacy of the model to the data series. In the diagnostic checking stage, referred to Figure 3.11, the residual series $\{\hat{a}_t\}$ of the model is investigated in terms of independence, homoscedasticity and normality for the closeness to the white noise criteria. The residuals plot for ARIMA(0,1,0) of the in-sample stationary series, shown as in Figure 4.10, illustrates randomness in the residuals with some spikes representing volatility clustering in certain periods as reflected by the differenced log data.

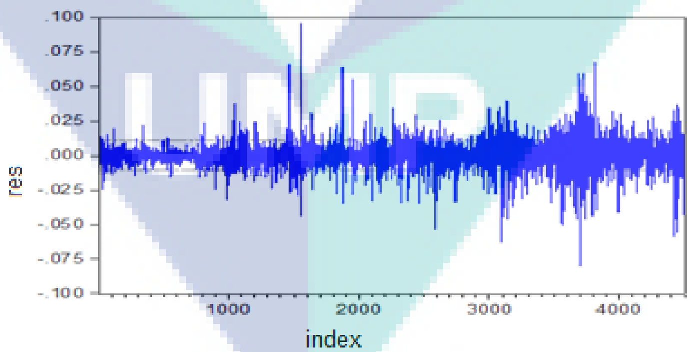


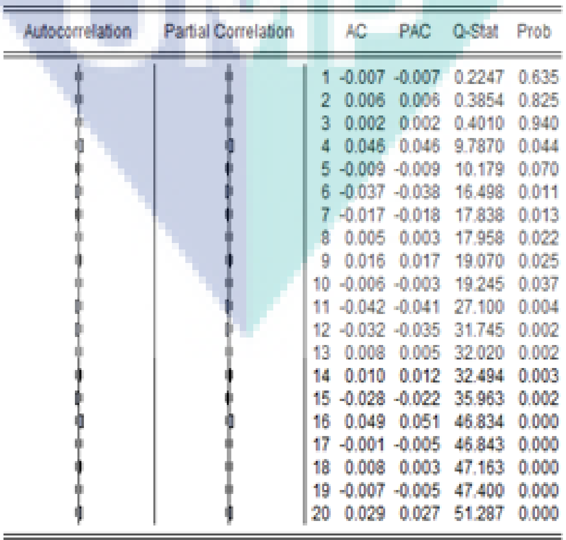
Figure 4.10 Residuals plot for ARIMA(0,1,0)

As in the serial correlation test, the result of DW-test is approximated to two $(2.0139 \approx 2)$ which shows that there is no first-order serial correlation in the residuals. The LBQ-test is then tested on the residuals to test the null hypothesis that there is no remaining serial correlation in the residuals for higher lags. From the test, as shown in

Figure 4.11, the *p*-value of the LBQ-test is not significant up to lag 5 shows that the mean equation of the ARIMA(0,1,0) to the data is correctly specified up to lag 5, at 5% significance level. The ACF and PACF of the residuals are both relatively small and approximately equal to zero, support the independency in the residuals, as shown graphically in Figure 4.10.

In testing of heteroscedasticity in the residuals, the ARCH LM test, or simply called as ARCH test is applied to the residuals of the ARIMA(0,1,0) model. The number of lag 10 and lag 15 are used for the ARCH test (Engle, 2001; Hyndman & Athanasopoulos, 2014). The ARCH test shows significant *p*-value which indicates the presence of ARCH in the residuals up to lag 15 as shown in Table 4.9.

It is also very clear that various spikes of ACF and PACF of squared residuals of ARIMA(0,1,0), as shown in Figure 4.12, are beyond the two standard error limits of 0.0298, showing that the residuals under consideration have ARCH effect. The LBQ-test which is tested on the squared residuals also agreed that there is the ARCH effect in the residuals. This means that the variance equation for the ARIMA model is not correctly specified due to the existence of volatility clustering in the data series. Furthermore, the PACF of the squared residuals of ARIMA(0,1,0) shows insignificant results up to lag 12, which indicate that at least 12 variables should be considered in ARCH model at 5% significance level. Hence, by applying parsimonious method, it is suggested to use GARCH model as compared to ARCH in handling the existence of heteroscedasticity in the residuals.

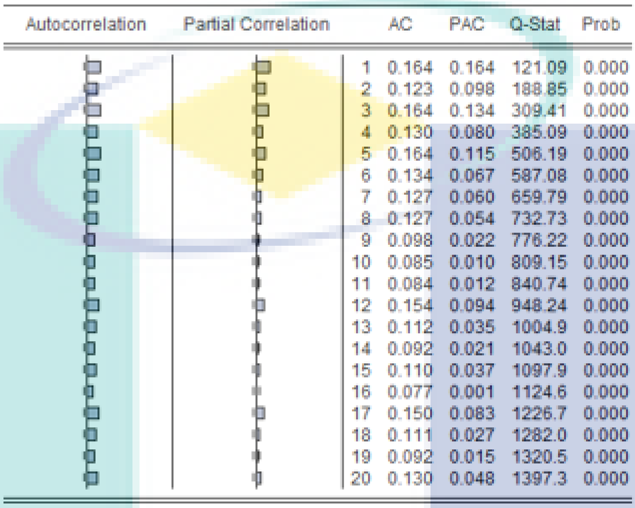


Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.2247	0.635
		2	0.006	0.006	0.3854	0.825
		3	0.002	0.002	0.4010	0.940
		4	0.046	0.046	9.7870	0.044
		5	-0.009	-0.009	10.179	0.070
		6	-0.037	-0.038	16.498	0.011
		7	-0.017	-0.018	17.838	0.013
		8	0.005	0.003	17.958	0.022
		9	0.016	0.017	19.070	0.025
		10	-0.006	-0.003	19.245	0.037
		11	-0.042	-0.041	27.100	0.004
		12	-0.032	-0.035	31.745	0.002
		13	0.008	0.005	32.020	0.002
		14	0.010	0.012	32.494	0.003
		15	-0.028	-0.022	35.963	0.002
		16	0.049	0.051	46.834	0.000
		17	-0.001	-0.005	46.843	0.000
		18	0.008	0.003	47.163	0.000
		19	-0.007	-0.005	47.400	0.000
		20	0.029	0.027	51.287	0.000

Figure 4.11 LBQ-test on residuals for ARIMA(0,1,0)

Table 4.9 Heteroscedasticity test using ARCH test for ARIMA(0,1,1)

χ_{test}	***p*-value**	$k_{\max}$
372.0444	0.0000	10
421.2442	0.0000	15



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.164	0.164	121.09	0.000
		2	0.123	0.098	188.85	0.000
		3	0.164	0.134	309.41	0.000
		4	0.130	0.080	385.09	0.000
		5	0.164	0.115	506.19	0.000
		6	0.134	0.067	587.08	0.000
		7	0.127	0.060	659.79	0.000
		8	0.127	0.054	732.73	0.000
		9	0.098	0.022	776.22	0.000
		10	0.085	0.010	809.15	0.000
		11	0.084	0.012	840.74	0.000
		12	0.154	0.094	948.24	0.000
		13	0.112	0.035	1004.9	0.000
		14	0.092	0.021	1043.0	0.000
		15	0.110	0.037	1097.9	0.000
		16	0.077	0.001	1124.6	0.000
		17	0.150	0.083	1226.7	0.000
		18	0.111	0.027	1282.0	0.000
		19	0.092	0.015	1320.5	0.000
		20	0.130	0.048	1397.3	0.000

Figure 4.12 Ljung-Box *Q*-test on squared residuals for ARIMA(0,1,0)

On the other hand, the JB-test of normality as given by Table 4.10 strongly rejects the null hypothesis that the white noise innovation ε_t is a normal distribution. The rejection is supported by the existence of many outliers on the left and right tails of the normal QQ-plot as shown in Figure 4.13. Based on the descriptive statistics for the residuals of ARIMA(0,1,0), it can be seen that both mean and variance of the residuals are approximately zero, with value of 1.58×10^{-17} and 0.0001, respectively.

Table 4.10 Descriptive statistics for the residuals of ARIMA(0,1,0)

Min	**Max**	**Mean**	**Median**	**Std. dev.**	**Skewness**	**Kurtosis**	**JB-test**	**NoO**
-0.0800	0.0961	1.58×10^{-17}	-0.0001	0.0106	0.0402	9.9186	8974.38 (0.0000)	4199

*Std. dev is abbreviated for standard deviation, the values in parenthesis denotes *p*-value and NoO is abbreviated for number of observation.

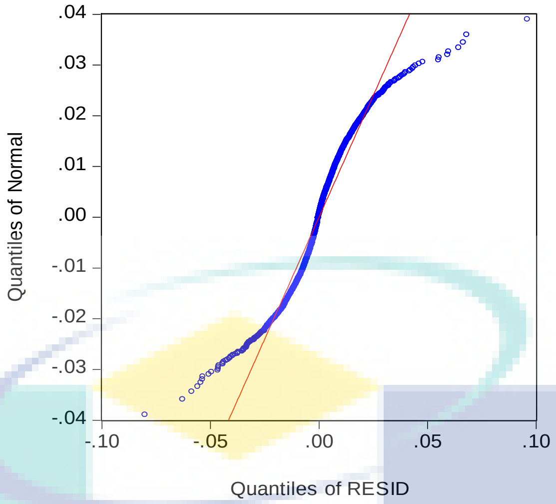


Figure 4.13 Normal QQ-plot for ARIMA(0,1,0)

Based on the results of diagnostic checking, ARIMA(0,1,0) model is good in handling serial correlation test in the residuals, however it fails to satisfy other tests of white noise criteria. This is due to the presence of volatility clustering in the data series and non-Gaussian characteristics in the residuals of the series as shown in Figure 4.10 and Figure 4.13, respectively. Therefore, in handling the existence of heteroscedasticity in the residuals, the stationary series is reanalysed using ARIMA-GARCH model. Then, the innovations of residuals of the proposed ARIMA-GARCH model will be tested including normal, *t*, skewed-*t*, GED and skewed-GED.

4.2.4 Modelling Gold Price using Box-Jenkins – GARCH

The previous analysis on the Box-Jenkins specifically ARIMA models determines that the ARCH effect occurred in the data series where conditional variance is not constant throughout the time, due to the presence of volatility in daily gold price data series. The considered ARIMA model did not handle the heteroscedasticity that exist in the data series. By referring to Figure 3.11, the standard GARCH models as the recommended volatility model is used to satisfy the non-constant behaviour in the residuals of the ARIMA models, by applying ARIMA-GARCH model to the gold price.

4.2.4.1 Stage I: Model Identification of ARIMA-GARCH

In the combination model of ARIMA-GARCH, the best model of the Box-Jenkins that is ARIMA(0,1,0) is used as the mean model while GARCH(r, s) is used as the variance model. In identifying the appropriate value of r and s for GARCH model, the ACF and PACF for the squared residuals of the considered ARIMA is used, where the ACF is used to specify the r value and the PACF is used to specify the s value. Based on Figure 4.12, the ACF and PACF for squared residuals for ARIMA(0,1,0) suggested the values for $r = 1, 2, 3, 4, 5$ and $s = 1, 2, 3, 4$, respectively. Therefore, there are 20 possible model combinations between ARIMA(0,1,0)-GARCH(r,s), as the details in Appendix 2.

4.2.4.2 Stage II: Parameter Estimation of ARIMA-GARCH

Even though OLS shows the best parameter estimate for the ARIMA models, but the method has disadvantages when volatility or ARCH effect is present in the series (Chand, Kamal, & Ali, 2012). Therefore, the parameter of the ARIMA-GARCH model is estimated using the MLE since it is widely applied in the combination of Box-Jenkins - ARCH/GARCH model in various data series (Chand et al., 2012; Chen et al., 2011; Liu & Shi, 2013; Tan et al., 2010; Zhou et al., 2006). The simultaneous estimation procedure of MLE for Box-Jenkins with ARCH/GARCH models is built-in, in statistical packages such as EViews and `fGarch` (in R language).

From the analysis conducted in the estimation stage, five of the 20 possible ARIMA-GARCH models show significant results at 5% significance level. The significant models are ARIMA(0,1,0)-GARCH(1,1), ARIMA(0,1,0)-GARCH(1,2), ARIMA(0,1,0)-GARCH(2,1), ARIMA(0,1,0)-GARCH(2,2) and ARIMA(0,1,1)-GARCH(1,4). Table 4.11 shows the empirical results for the significant ARIMA-GARCH models with normal errors assumption on the stationary data of daily gold price using MLE method at 5% significance level. It can be observed that all the significant ARIMA-GARCH models have insignificant constant in the mean equation.

The empirical results of the significant ARIMA-GARCH models indicate that the values of normalised AIC, normalised SIC and the log-likelihood are marginally decreased when the value of r and s is greater than one $(\text{i.e. } r, s > 1)$. Based on Table 4.11, the models of ARIMA(0,1,0)-GARCH(4,4) produced the smallest values for AIC and

SIC, respectively. However, since the criterion values are marginally decreased with ARIMA(0,1,0) - GARCH(1,1), therefore ARIMA(0,1,0) - GARCH(1,1) model is chosen for the next stage due to the principle of parsimony that is well practised in the Box - Jenkins modelling.

Table 4.11 Estimation results of the significant ARIMA-GARCH models

Par	**ARIMA(0,1,0) -GARCH(1,1)**	**ARIMA(0,1,0) - GARCH(1,2)**	**ARIMA(0,1,0) - GARCH(2,1)**	**ARIMA(0,1,0) - GARCH(2,2)**	**ARIMA(0,1,0) -GARCH(4,4)**
c	4.29×10^{-5} (0.7000)	5.79×10^{-5} (0.6030)	7.61×10^{-5} (0.4907)	5.73×10^{-5} (0.6068)	3.21×10^{-5} (0.7694)
α_0	1.27×10^{-7} (0.0000)	1.75×10^{-7} (0.0001)	9.20×10^{-8} (0.0001)	1.16×10^{-8} (0.0198)	3.86×10^{-8} (0.0235)
α_1	0.0428 (0.0000)	0.0620 (0.0000)	0.1128 (0.0000)	0.0970 (0.0000)	0.0943 (0.0000)
α_2	-	-	-0.0775 (0.0000)	-0.0916 (0.0000)	0.0901 (0.0000)
α_3	-	-	-	-	-0.0884 (0.0000)
α_4	-	-	-	-	-0.0776 (0.0000)
β_1	0.9594 (0.0000)	0.4564 (0.0001)	0.9663 (0.0000)	1.7484 (0.0000)	-0.0811 (0.0213)
β_2	-	0.4846 (0.0000)	-	-0.7536 (0.0000)	1.5295 (0.0000)
β_3	-	-	-	-	0.2768 (0.0000)
β_4	-	-	-	-	-0.7428 (0.0000)
AIC	-6.5894	-6.5917	-6.5958	-6.6007	-6.6118
SIC	-6.5837	-6.5846	-6.5887	-6.5922	-6.5976
Log-l	14826.8700	14833.8100	14842.24	14854.3000	14883.3300

* values in parenthesis denotes p-value, Par is abbreaviated for parameter and Log-l is for log-likelihood.

4.2.4.3 Stage III: Diagnostic Checking of ARIMA-GARCH

Figure 4.14 shows the standardised residuals plot for ARIMA(0,1,0)-GARCH(1,1) of the in-sample stationary series. Except for several possible outliers due to volatility clustering in the data series, the standardised residuals look random and reasonable. While the corresponding diagnostic test results for the Box-Jenkins – GARCH model is given by Table 4.12. In the diagnostic checking tests, the DW-test value of approximately two, shows that there is no first-order serial correlation in the residuals for the considered BJ-G model. Even though the LBQ-test on the standardised residuals shows the existence of serial correlation (and the autocorrelation decreases in the higher lag), this is most probably due to small autocorrelations that should not be of practical importance (Ruppert & Matteson, 2015). This situation is not surprising since the sample size used is 4499, which might lead to small autocorrelation due to its large sample size (Ruppert & Matteson, 2015).

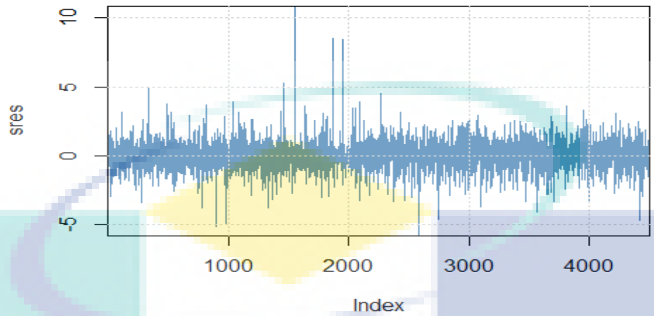


Figure 4.14 Standardised residuals plot for ARIMA(0,1,0)-GARCH(1,1)

Table 4.12 Diagnostic tests on ARIMA(0,1,0)-GARCH(1,1) model

Diagnostic Test	**Value**	***p*-value**
DW-test	2.0125	-
LBQ(10)	9.0971	0.5230
LBQ(15)	17.3830	0.2970
LBQ$^2(10)$	16.8210	0.0780
LBQ$^2(15)$	16.6610	0.2360
ARCH(110)	17.8694	0.0822
ARCH(15)	19.7402	0.2696
JB-test	9221.6300	0.0000

*LBQ(10) represents the LBQ-test on residuals at lag 10, LBQ$^2(10)$ represents the LBQ-test on squared residuals at lag 10 and ARCH(10) represents the ARCH LM-test at lag 10.

To validate the linearity assumption in the mean equation of the model, the Terasvirta test is applied. For the stationary data, the test statistic is $f_{\text{test}} = 4.3915$ with $p-\text{value} = 0.0124$ which indicates that the null hypothesis of the test is rejected at 5% level of significance. Therefore, the result of the test supports the behaviour of random walk in mean model which implies that ARIMA(0,1,0) is correctly specified for the mean model. Otherwise, the *p*-value is insignificant for LBQ-test on the squared standardised residuals which interprets that there is no ARCH in the residuals up to both lag 10 and 15, as supported by the result of LM ARCH. Therefore, the model checking statistics on serial correlations, linearity and heteroscedasticity suggest that the mean and variance equation of the combination model of BJ-G is adequate and correctly specified.

On the other hand, the JB-test of the ARIMA-GARCH model strongly rejects the null hypothesis that the innovations, ε_t is Gaussian. The rejection is supported graphically by the existence of more outliers in the tails of the normal QQ-plot on the ARIMA(0,l,0)-GARCH(1,1) as shown in Figure 4.15. The outliers may be a signal that the conditional variance is not constant which exist when the variance is large.

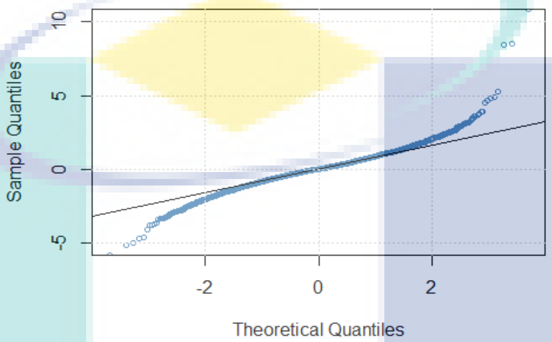


Figure 4.15 The normal QQ-plot of standardised residuals of ARIMA(0,1,0)-GARCH(1,1)

In investigating the characteristics of the innovations of the BJ-G model, the tests on skewness and kurtosis of the model residuals are conducted. Based on Table 4.13, the sample skewness and sample kurtosis of the standardised residuals are 0.4664 and 9.9514, respectively. By considering the hypothesis $H_0 : S_R = 0$ versus $H_0 : S_R \neq 0$, where S_R denotes the skewness of the standardised residuals, the test statistic is 12.7715 with *p*-value zero. The distribution of stationary data looks symmetric as shown in Figure 4.16, however, the hypothesis of zero skewness is rejected at 5% significance level. As for the kurtosis, the hypothesis is $H_0 : K_R - 3 = 0$ versus $H_0 : K_R - 3 \neq 0$, where K_R denotes the kurtosis of the standardised residuals, the test statistic is 95.1754 with *p*-value zero. Therefore, the null hypothesis of zero kurtosis is rejected which indicates that the standardised residuals series is leptokurtic with a higher peak and fatter tails as can be seen in Figure 4.16.

Table 4.13 Descriptive statistics of standardised residuals of ARIMA(0,1,0)-GARCH(1,1)

Min	Max	Mean	Median	Std. dev.	Skewness	Kurtosis	NoO
-5.7572	10.8021	0.0353 (0.0297)	0.0224	1.0000	0.4664 (0.0000)	9.9514 (0.0000)	4499

*Std. dev is abbreviated for standard deviation, the values in parenthesis denotes *p*-value and NoO is abbreviated for number of observation.

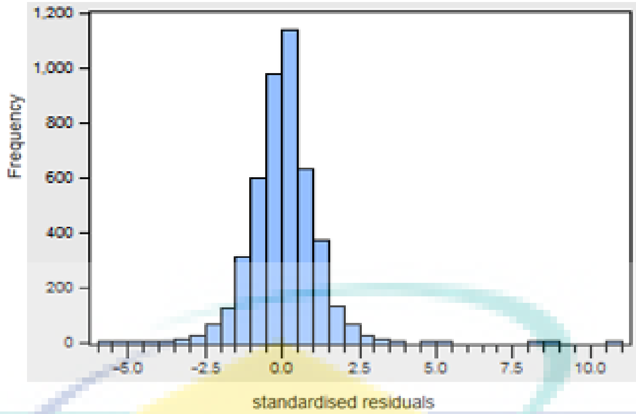


Figure 4.16 Histogram for standardised residuals of ARIMA(0,1,0)-GARCH(1,1)

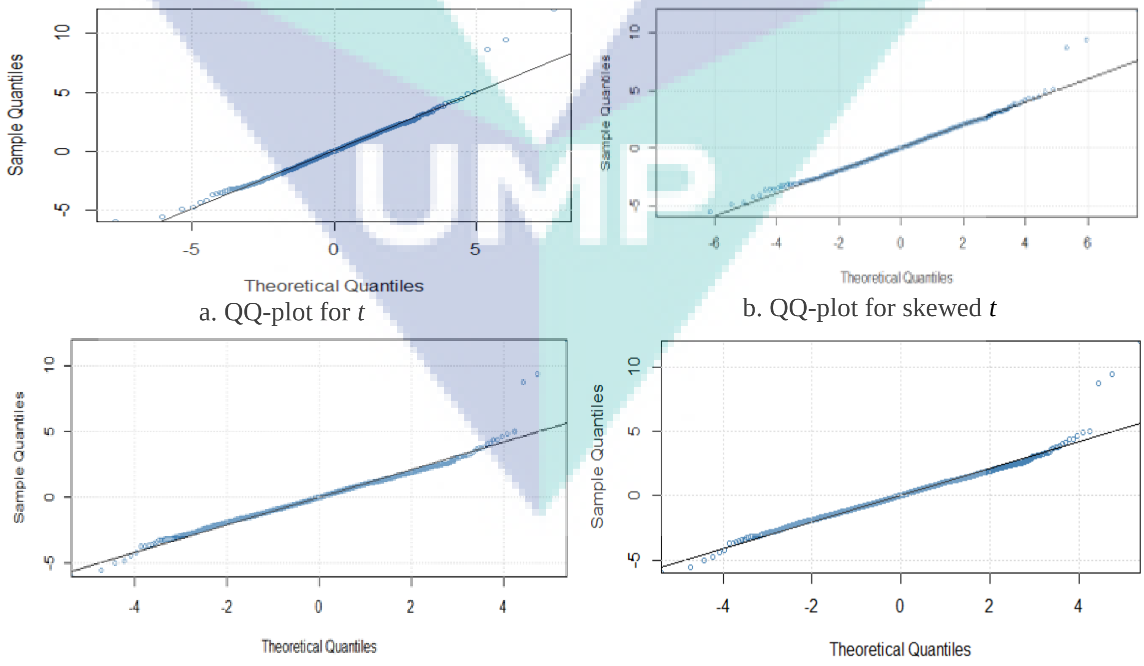
The characteristics of non-normal, heavy-tailed and skewed in the standardised residuals derive the use of *t*, skewed-*t*, GED and SGED distributions to be employed to the innovations for ARIMA(0,1,0)-GARCH(1,1) model. Table 4.14 presents the joint parameter estimation and diagnostic checking for the ARIMA(0,1,0)-GARCH(1,1) with four types of innovations. It is noted that, all models are highly significant and provide a good fit to the data since model checking statistics fail to indicate inadequacy of the model. The results given in the table reveals that ARIMA(0,1,0)-GARCH(1,1) model with *t* innovations is preferred based on the smallest values of normalised AIC and SIC, as well as in line with the principle of parsimony.

The QQ-plot shown in Figure 4.17 supports the decision of preference of *t* innovations for the ARIMA(0,1,0)-GARCH(1,1) model to the stationary series of gold price data. The good fit of the QQ-plot in Figure 4.17(a) that nearly a straight line except for four outliers on the left and right tails, support graphically the use of *t* innovations. The percentage of the outliers is a small fraction of the data, just about 0.09% compared to the in-sample size. It is clear that the heavy-tailed characteristic in the stationary series gives significant impact to the chosen innovations since *t*-related QQ-plot relatively fit better than GED-related QQ-plot. On the other hand, the skewness characteristic shows no impact since the non-skewed-related models have comparatively smaller AIC and SIC than skewed-related models. It is also hard to detect graphically any significant difference in the QQ-plot between skewed and non-skewed models.

Table 4.14 Parameter estimation and diagnostic testing on ARIMA(0,1,0)-GARCH(1,1) model for *t*, skewed-*t*, GED and SGED innovations

Stages	**Innovations**			
	t	**Skewed-*t***	**GED**	**SGED**
	Parameter Estimation			
c	7.58×10^{-5} (0.4227)	3.90×10^{-5} (0.6996)	5.65×10^{-5} (0.6083)	6.10×10^{-5} (0.5621)
α_0	1.92×10^{-7} (0.0000)	1.88×10^{-7} (0.0183)	2.04×10^{-7} (0.0000)	2.04×10^{-7} (0.0000)
α_1	0.0663(0.0000)	0.0663(0.0000)	0.0606(0.0000)	0.0607(0.0000)
β_1	0.9386(0.0000)	0.9387(0.0000)	0.9420(0.0000)	0.9419(0.0000)
v	4.6977(0.0000)	4.6760(0.0000)	1.1400(0.0000)	1.1400(0.0000)
ξ	-	0.9840(0.0000)	-	1.0020(0.0000)
AIC	-6.7078	-6.7073	-6.7023	-6.7019
SIC	-6.7006	-6.6987	-6.6952	-6.6934
Log-l	15094.1500	15094.0600	15081.9100	15081.9300
	Diagnostic Checking			
DW-test	2.0128	1.9396	1.9408	1.9410
LBQ(10)	10.0070(0.4400)	10.1482(0.4276)	10.0333(0.4376)	10.0392(0.4371)
LBQ(15)	18.2380(0.2500)	18.5904(0.2329)	18.5428(0.2352)	18.5399(0.2353)
$\text{LBQ}^2(10)$	6.0308(0.8130)	6.1130(0.8057)	6.8677(0.7379)	6.8565(0.7389)
$\text{LBQ}^2(15)$	8.4942(0.9020)	8.5863(0.8981)	9.2438(0.8644)	9.2325(0.8650)
ARCH(10)	6.1155(0.8055)	6.1148(0.8055)	6.8692(0.7377)	6.8289(0.7415)
ARCH(15)	8.3064(0.9109)	8.3080(0.9108)	8.9089(0.8822)	8.8739(0.8840)

* values in parenthesis denotes *p*-value, Log-l is abbreaviated for log-likelihood, LBQ(10) represents the LBQ-test on residuals at lag 10, $\text{LB}^2(10)$ represents the LBQ-test on squared residuals at lag 10 and ARCH(10) represents the ARCH LM-test at lag 10.



a. QQ-plot for *t*

b. QQ-plot for skewed *t*

c. QQ-plot for GED

d. QQ-plot for SGED

Figure 4.17 The QQ-plot of standardized residuals of ARIMA(0,1,0)-GARCH(1,1) model for innovations of *t*, skewed-*t*, GED and SGED, respectively

4.2.4.4 Stage IV: Forecasting of ARIMA-GARCH

Consequently, the model of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations is used in the forecasting stage. The ARIMA(0,1,0)-GARCH(1,1) model with *t* innovations for the stationary series for gold price data is given by Equation 4.4,

$$\begin{aligned}&\Delta y_t^* = 7.58\times10^{-5} + a_t, \qquad a_t = \sigma_t\,\varepsilon_t \\ &\sigma_t^2 = 1.92\times10^{-7} + 0.06a_{t-1}^2 + 0.93\sigma_{t-1}^2 \text{ and } \varepsilon_t \sim t_{4.70}^*\end{aligned} \tag{4.4}$$

where y_t^* is the log price data at time *t* and $\nabla y_t^{\ *} = y_t^{\ *} - y_{t-1}^{\ *}$ is the differenced log price data at time *t*. The BJ-G model is then applied to the series of out-of-sample data that consists of 500 observations in evaluating the forecasting performance of the model. Since the stationary series or the daily log return price series is in log differenced form, the out-of-sample data must be transformed as well in applying the BJ-G model to obtain the forecast data of one-step ahead and the prediction error.

Referring to Figure 3.12, Equation 4.4 is then modified to be an appropriate model to apply to the original scale in evaluating the performance of the BJ-G model in forecasting gold price. By retransforming the log and rearranging Equation 4.4, the model for daily gold price is given by Equation 4.5, where y_t is the daily gold price and $s_t = \Delta y_t^*$ is the stationary data for the daily gold prices.

$$\begin{aligned}&y_t = y_{t-1}\exp(s_t), \qquad s_t = 7.58\times10^{-5} + a_t, \qquad a_t = \sigma_t\,\varepsilon_t \\ &\sigma_t^2 = 1.92\times10^{-7} + 0.06a_{t-1}^2 + 0.93\sigma_{t-1}^2 \text{ and } \varepsilon_t \sim t_{4.70}^*\end{aligned} \tag{4.5}$$

The forecast evaluations for one-step ahead using the proposed model for stationary series and price series are generated. Table 4.15 shows the joint results of RMSE, MAE and MAPE for 500 stationary forecast data (daily log return price data) and the forecast price data (daily price) using the ARIMA(0,1,0)-GARCH(1,1) model with *t* innovations. There is no result of MAPE for the stationary series since there is zero value of the transformed series due to no changes of the price for two consecutive days.

Table 4.15 Forecast evaluations of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations

Data series	**Forecast Evaluations**		
	RMSE	**MAE**	**MAPE**
Log Return Price Series	0.0124	0.0084	-
Price Series	12.6855	18.3716	0.8402%

The one-step ahead forecast using ARIMA(0,1,0)-GARCH(1,1) model with *t* innovations for daily gold prices from 21st December 2012 to 17th December 2013 is shown in Figure 4.18. In the plot, the dashed line (red colour) presents the forecasted prices with ±2 standard errors whereas the solid line (blue colour) shows the actual gold prices. It is observed that the forecast gold prices fluctuate between USD 1200 and USD 1800 per ounce in the 500-day out-sample period. Graphically, the BJ-G model shows promising performance in forecasting daily gold price series which is the trend of forecast prices that follows closely the actual data for the 500 days out-of-sample period. The comparison between actual daily gold price and its one-step ahead forecast price using the proposed BJ-G model for the last ten days out-of-sample simulation period is given by Table 4.16. In the study, the last ten-day simulation data is sufficient enough to demonstrate the trend of one-day lag (where the second column (forecast) prices can be obtained from the first column (actual) prices by shifting the first column one row downward) in the forecasting part.

Figure 4.18 Graph of the actual and forecast data using ARIMA(0,1,0)-GARCH(1,1) model with *t* innovations for out-of-sample period

Table 4.16 The actual and forecast prices using ARIMA(0,1,0)-GARCH(1,1)

Date	Actual price (USD/Oz)	Forecast price (USD/Oz)	Difference (USD/Oz)
4 Dec 2013	1227.50	1218.18	9.33
5 Dec 2013	1222.50	1228.43	-5.93
6 Dec 2013	1233.00	1223.43	9.57
9 Dec 2013	1237.00	1233.94	3.06
10 Dec 2013	1266.25	1237.94	28.31
11 Dec 2013	1260.75	1267.21	-6.46
12 Dec 2013	1225.25	1261.71	-36.46
13 Dec 2013	1232.00	1226.18	5.82
16 Dec 2013	1234.75	1232.94	1.81
17 Dec 2013	1231.75	1235.69	-3.94

4.2.5 Comparison of the Box-Jenkins, GARCH and the Box-Jenkins – GARCH Model Performance in Forecasting Gold Price

The Box-Jenkins models specifically ARIMA is widely used in research practice for gold price in comparison or forecasting model (Alcidi et al., 2010; Khan, 2013; Miswan, Ping, & Ahmad, 2013). While, Ping et al. (2013) applied GARCH model to forecast Malaysian gold. GARCH as one of widely used models for asset volatility. Volatility is defined as the conditional standard deviation of asset returns (Tsay, 2013). In general, a series of the asset returns is a stationary series in the form of log return. Therefore, in this case study, the performance of the appropriate Box-Jenkins model, GARCH model and the proposed BJ-G model is discussed further. However, the series of gold price in this case study is not a stationary series, therefore the GARCH model is not appropriate to be used to forecast the data and the model would not be considered in the model comparison.

Table 4.17 presents the estimation results for the considered Box-Jenkins model ARIMA(0,1,0) and the BJ-G model (specifically ARIMA(0,1,0)-GARCH(1,1) with *t* innovations) for the daily gold price series. It is found that both the normalised AIC and SIC values as well as the log-likelihood value from the BJ-G model are smaller than that of ARIMA model, hence it shows that ARIMA(0,1,0)-GARCH(1,1) model with *t* innovations is a better model as compared to ARIMA(0,1,0) for estimating daily gold prices. Note that, the single ARIMA model fails to handle the heteroscedasticity that exist in the data series as discussed in section 4.2.3, as well as violating the assumption on the constant variance in the errors. Therefore, it can solely be concluded that ARIMA(0,1,0)-GARCH(1,1) with *t* innovations is appropriate and preferred in forecasting gold price since it reflects its pattern without violating the errors assumptions of the ARIMA model.

Table 4.17 Estimation evaluations for ARIMA and ARIMA-GARCH

Parameter	**ARIMA(0,1,0)**	**ARIMA(0,1,0)-GARCH(1,1)**
c	0.0003(0.0398)	7.58×10^{-5} (0.4227)
α_0	-	1.92×10^{-7}(0.0000)
α_1	-	0.0663(0.0000)
β_1	-	0.9386(0.0000)
v	-	4.6977(0.0000)
AIC	-6.2646	-6.7078
SIC	-6.2632	-6.7006
Log-likelihood	14093.2400	15094.1500

* values in parenthesis denotes *p*-value.

By applying ARIMA(0,1,0)-GARCH(1,1) to the data series, it can be seen that the ARCH and GARCH coefficients in the variance model are statistically significant and the sum of both coefficients is very close to one. This indicates that volatility shocks are quite persistent where this result is often observed in high frequency financial data such as intra-day, daily or weekly data. Based on Table 4.17, The large value of $\beta_1 = 0.9386$ is attributed to a long-term persistence of volatility clustering. Therefore, by applying the BJ-G model to the data series, it can be seen that the combination model is able to explain better about the characteristics of the gold price.

The empirical results using 5000-daily data series indicate that the combination model of Box-Jenkins and GARCH (or BJ-G) is proven as a potential model to analyse and forecast a highly volatile time series data, specifically gold price data. The results of this preliminary study have been published in Yaziz, Azizan, Ahmad, Zakaria, Agrawal and Boland (2015), Yaziz et al., (2014), Yaziz, Azizan, Zakaria and Ahmad (2013). This is in line with other reports that use the BJ-G model for highly volatile data series such as electricity price (Liu & Shi, 2013; Tan et al., 2010), internet traffic (Zhou et al., 2006), traffic flow (C. Chen et al., 2011), stock market (Freedi et al., 2012) and gold price (Ahmad, Ping, Yaziz, & Miswan, 2014; Ahmad et al., 2015).

4.3 The Empirical Results of Gold Price on the Second Proposed Procedure of BJ-G

Given the positive results of the forecasting one-step ahead for the data series using BJ-G model in the preliminary study, the performance of the second proposed procedure of BJ-G as illustrated in Figure 3.14 is then evaluated using the gold price series. The data series of daily world gold price, as applied in section 4.2, is used as the pool of data set in this empirical study. Based on the previous studies, the number of data from 500 to 5000 are usually considered for BJ-G model for daily basis (Babu & Reddy, 2015; Ferenstein and Gasowski, 2004; Gaglianone & Marins, 2017; García-Ferrer et al., 2012; Harrison & Paton, 2004; Koopman et al., 2007; Sohn & Lim, 2007). Note that, 5000 data as considered in the preliminary study is the maximum number of data for daily basis in previous literatures to ensure all significant characteristics related to volatility of data can be captured.

The 5000 data series is then divided into six different data series ranges from 500 to 5000 and each data series is tested using the second proposed procedure of BJ-G as shown in Figure 3.14. Basically, the number of data for each sample is approximately half from the previous duration, with a ratio of estimate to forecast at 90:10. The ratio of estimate to forecast used is the same as in Section 4.2 to maintain the continuity of the study. The details of the classification of sample data series are summarised in Table 4.18.

Table 4.18 Classification of sample data series

Sample	Duration	Number of Data	In-Sample	Out-of-Sample
1	24/11/93 - 17/12/13 (20-year)	5 000	24/11/93 - 20/12/11 (4500 data)	21/12/11 - 17/12/13 (500 data)
2	5/12/03 - 17/12/13 (10-year)	2 500	5/12/03 - 18/12/12 (2250 data)	19/12/12-17/12/13 (250 data)
3	22/12/08 - 17/12/13 (5-year)	1 250	22/12/08 - 24/6/13 (1125 data)	25/6/13 - 17/12/13 (125 data)
4	21/12/09 - 17/12/13 (4-year)	1 000	21/12/09 - 29/7/13 (900 data)	30/7/13 - 17/12/13 (100 data)
5	20/12/10 - 17/12/13 (3-year)	750	20/12/10 - 3/9/13 (675 data)	4/9/13 - 17/12/13 (75 data)
6	21/12/11 - 17/12/13 (2-year)	500	21/12/11- 8/10/13 (450 data)	9/10/13 - 17/12/13 (50 data)

Since a significant modification is done on Stage I of the second proposed procedure of BJ-G (refer to Figure 3.14) where it involves eight steps and will be disussed in detail in-sample data series. The first step is to plot in-sample series for each sample considered, as shown in Figure 4.19. The purpose of this step is to check the occurrence of any trends and seasonality behaviour in the gold price movement graphically. Based on Figure 4.19, it is observed that the gold price series is nonseasonal for all data series. The stationarity behaviour based on the time series plots of all samples can be classified into three groups as follows:

(i) The series for Sample 1 and 2 have the same characteristics, where it exhibits overall upward trend with inconsistent and large variation;

(ii) The series for Sample 3 to 5 have an unclear trend but exhibit inconsistent and large variation;

(iii) The series for Sample 6 has downward trend with consistent and small variation.

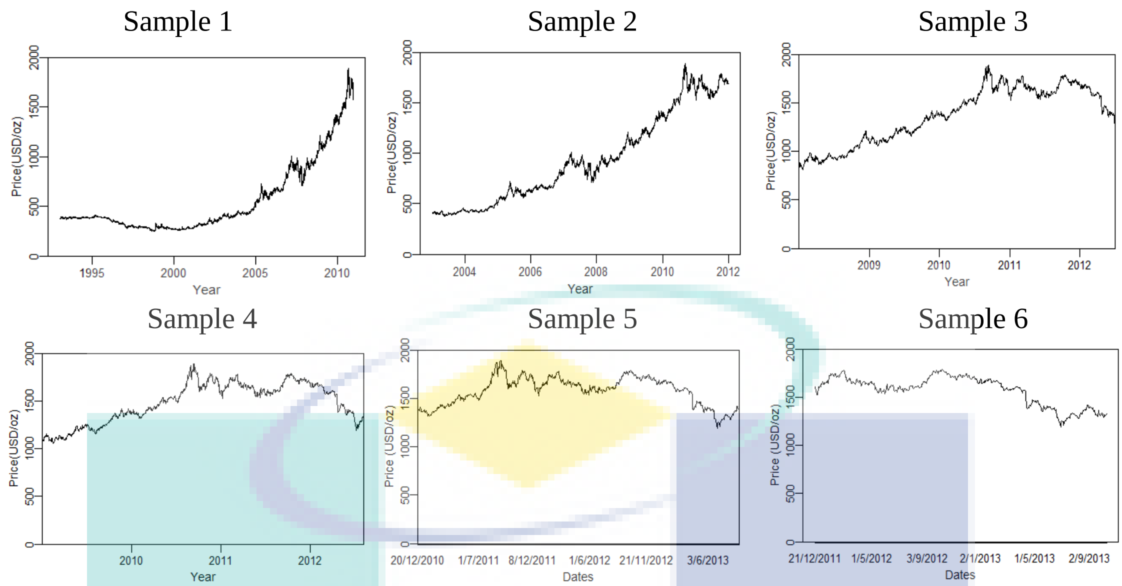


Figure 4.19 In-sample time series plot of original data for Sample 1 to 6

In the second step, the descriptive statistics of all samples is obtained as tabulated in Table 4.19, which support the the behaviour of each sample as described in the first step. The test of hypothesis for skewness imply that the series for all samples is asymmetric. The data series in Sample 1 and 2 are positively skewed, while Sample 3 to 6 are negatively skewed. The JB-test validates the asymmetric characteristics for all samples. Based on Table 4.19, it can be seen that all samples have excess kurtosis, specifically, Sample 1 is leptokurtic while Sample 2 to 6 are platykurtic. The reduction in the kurtosis value from Sample 1 to 6 implies that the peakedness in the data series is decreasing from Sample 1 to 6. The graphical representation of distribution for each sample is shown by histogram in Figure 4.20.

Table 4.19 Descriptive statistics for in-sample original data of Sample 1 to 6

Sample	Number of Data	Mean	Standard deviation	Skewness	Kurtosis	JB - test	Distribution
1	4500	558.7877 (0.0000)	363.4643	1.5925 (0.0000)	1.7121 (0.0000)	2454.4068 (0.0000)	+vely skewed, leptokurtic
2	2250	938.1378 (0.0000)	442.5005	5.0606 (0.0000)	-1.0278 (0.0000)	194.8672 (0.0000)	+vely skewed, platykurtic
3	1125	1375.5310 (0.0000)	289.4461	-2.6388 (0.0003)	-1.2612 (0.0000)	87.2810 (0.0000)	-vely skewed, platykurtic
4	900	1485.9270 (0.0000)	211.4332	-3.2810 (0.0000)	-1.1077 (0.0000)	61.8668 (0.0000)	-vely skewed, platykurtic
5	675	1580.1060 (0.0000)	146.4332	-4.4476 (0.0000)	-6.8583 (0.0003)	35.3177 (0.0000)	-vely skewed, platykurtic
6	450	1573.4772 (0.0000)	150.3654	-0.7256 (0.0003)	-0.7002 (0.0024)	48.6770 (0.0000)	-vely skewed, platykurtic

*Values in parenthesis denotes *p*-value, +vely represents for positively and –vely represent for negatively

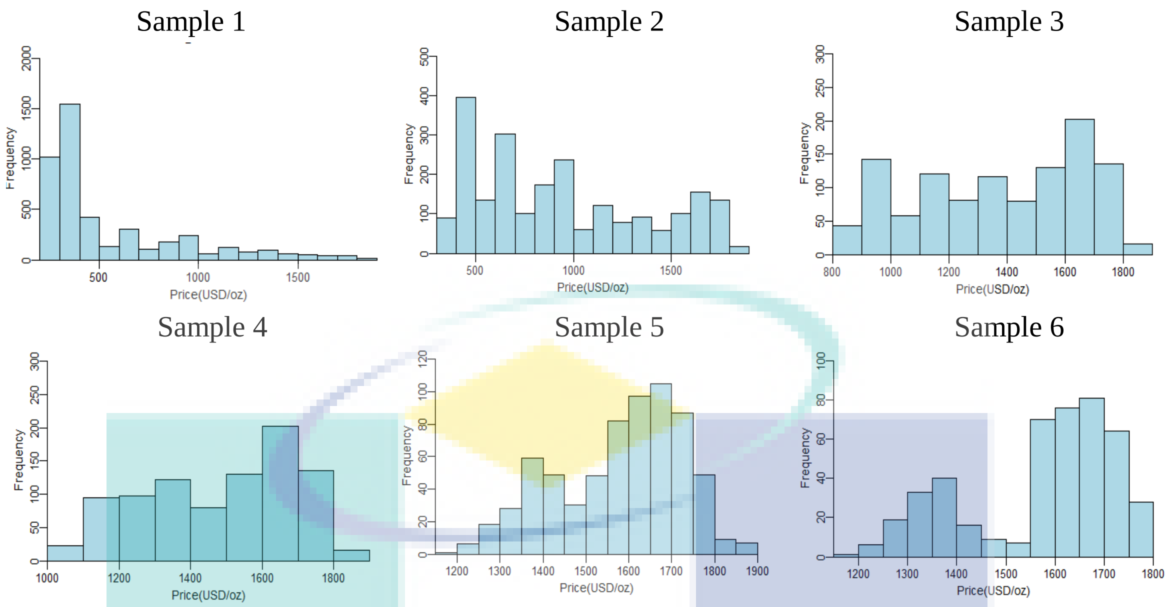


Figure 4.20 Histogram for in-sample original series of Sample 1 to 6

The third step of Stage I in the second proposed procedure of BJ-G is checking the stationarity behavior for the samples. The Box-Cox transformation method is applied first to identify the nonstationary in-variance behaviour in the in-sample data series. The Box-Cox estimated value for the power of estimation, λ_{BC} for each sample with its appropriate transformation is summarised in Table 4.20. It can be seen that, 5 out of 6 samples depict nonstationary in-variance. Hence, the in-sample data series of the samples need to be transformed first in order to stabilise the variance. This indicates the importance of the Box-Cox transformation especially when the time series plot shows inconsistent and large variations, which supports the observation in the first step. Figure 4.21 graphically shows the transformed data series, y_t^* for Sample 1 to 6. Note that the series for Sample 6 does not need to be transformed since its λ_{BC} values is close to 1.

Table 4.20 The transformed and stationary data for Sample 1 to 6

Sample	**Box-Cox Transformation, λ_{BC}**	**Transformed Data, y_t^***
1	$\lambda_{BC} = -0.2147 \rightarrow 0$	$y_t^* = \ln y_t$
2	$\lambda_{BC} = -0.1101 \rightarrow 0$	$y_t^* = \ln y_t$
3	$\lambda_{BC} = 0.0780 \rightarrow 0$	$y_t^* = \ln y_t$
4	$\lambda_{BC} = -0.4217 \rightarrow -0.5$	$y_t^* = \dfrac{1}{\sqrt{y_t}}$
5	$\lambda_{BC} = 0.6421 \rightarrow 0.5$	$y_t^* = \sqrt{y_t}$
6	$\lambda_{BC} = 0.9999 \rightarrow 1$	$y_t^* = y_t$

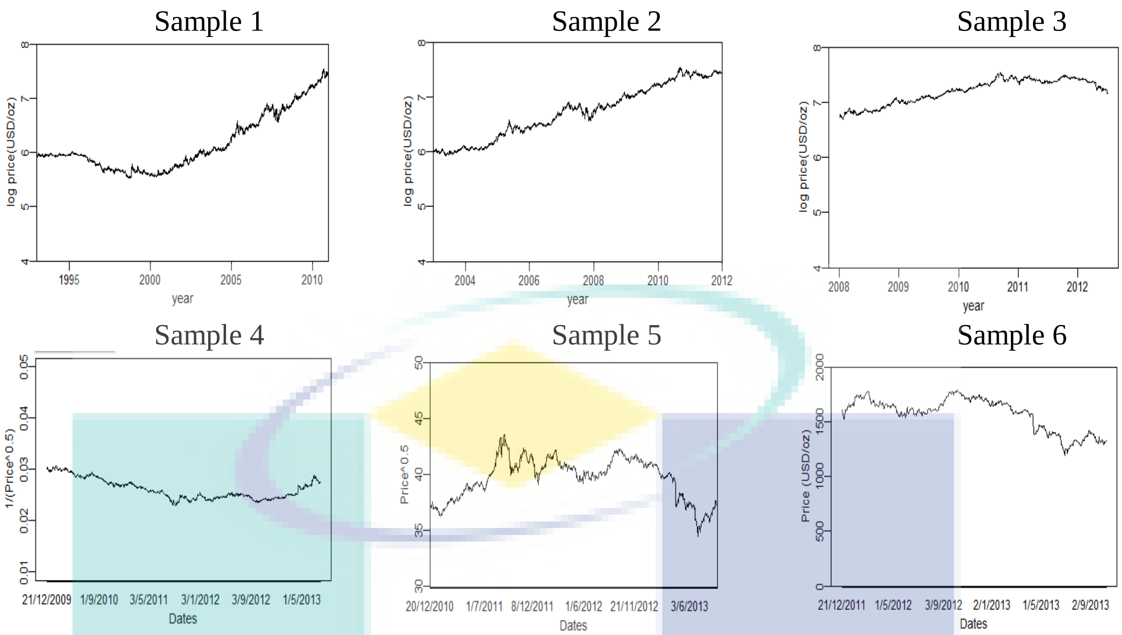


Figure 4.21 In-sample time series plot for transformed data of Sample 1 to 6

The in-sample transformed data of Sample 1 to 5 and in-sample data of Sample 6 are then tested for stationarity in-mean, using autocorrelation method (ACF and PACF) and unit root test (ADF-test). Table 4.21 summarises the behaviour of sample ACF and PACF up to lag $k_{\max}$, where $k_{\max} = 10\log T$ and the ADF-test up to lag $k_{\max}$ where $k_{\max} = 12\left(T/100\right)^{\frac{1}{4}}$ and T is the number of data for in-sample series.

Table 4.21 Checking the stationarity of the transformed series at level (if needed)

Sample	**ACF and PACF**		**ADF-test**		**Stationarity Condition**
	Behaviour	$k_{\max}$	*t*-test	$k_{\max}$	
1	ACF spikes too slowly decays to zero and on one-side, only r_{11} nonzero	36	-1.0016 (0.9422)	31	Not stationary
2	ACF spikes too slowly decays to zero and on one-side, only r_{11} nonzero	34	-3.3392 (0.0603)	26	Not stationary
3	ACF spikes too slowly decays to zero and on one-side, only r_{11} nonzero	31	0.1488 (0.9977)	21	Not stationary
4	ACF spikes too slowly decays to zero and on one-side, only r_{11} nonzero	30	-0.5591 (0.9806)	21	Not stationary
5	ACF spikes slowly decays to zero and on one-side, only r_{11} nonzero	29	-2.0605 (0.5664)	20	Not stationary
6	ACF spikes slowly decays to zero and on one-side, only r_{11} nonzero	27	-1.7495 (0.7274)	18	Not stationary

*Values in parenthesis denotes *p*-value.

Based on the table, all the series are not stationary at level $(d = 0)$, therefore all the considered series need to be differenced in order to remove the trend. The details of the sample ACF and PACF and the ADF-test for the samples considered can be referred to Appendix 3.

The behaviour of the sample ACF and PACF and the ADF-test at the first differenced transformed data are summarised in Table 4.22. The results in the table indicate that series for all samples are stationary after first differenced. It can also be concluded that the ADF-test is useful to confirm numerically the stationarity in a series, since the sample ACF and PACF is based on graphical representation.

Table 4.22 Checking the stationarity at the first differenced series

Sample	ACF and PACF		ADF-test		Stationarity condition
	Behaviour	$k_{\max}$	*t*-test	$k_{\max}$	
1	Only r_1 nonzero, PACF spikes decays to zero exponentially and cuts off at lag 20	36	-13.9454 (0.0000)	31	Stationary
2	ACF spikes dies down to zero from lag 1, PACF spikes is damped sine wave with most of spikes close to zero from lag 1	34	-10.6504 (0.0000)	26	Stationary
3	ACF spikes dies down to zero from lag 1, PACF spikes is damped sine wave with most of spikes close to zero from lag 1	31	-6.8659 (0.0000)	21	Stationary
4	ACF spikes dies down to zero from lag 1, only PACF spikes at lag 21 is nonzero	30	-6.4382 (0.0000)	21	Stationary
5	ACF and PACF spikes dies down to zero from lag 1	29	-4.9654 (0.0002)	20	Stationary
6	ACF spikes dies down to zero from lag 1, only r_{22} and r_{88} are nonzero	27	-5.2250 (0.0001)	18	Stationary

*Values in parenthesis denotes *p*-value.

Figure 4.22 graphically shows the time plot for the stationary series for Sample 1 to 6 where it can be seen that most of the stationary data in all samples are located around the mean of zero. As shown in the figure, there is clear evidence of volatility clustering changes over time in the stationary series which implies that the gold price is one of the volatile data. Since the series found to be volatile, hence it is wise to consider the GARCH model in handling volatility in the series. However, it can be observed that the volatility clustering in a time series decreases as the number of data decreases.

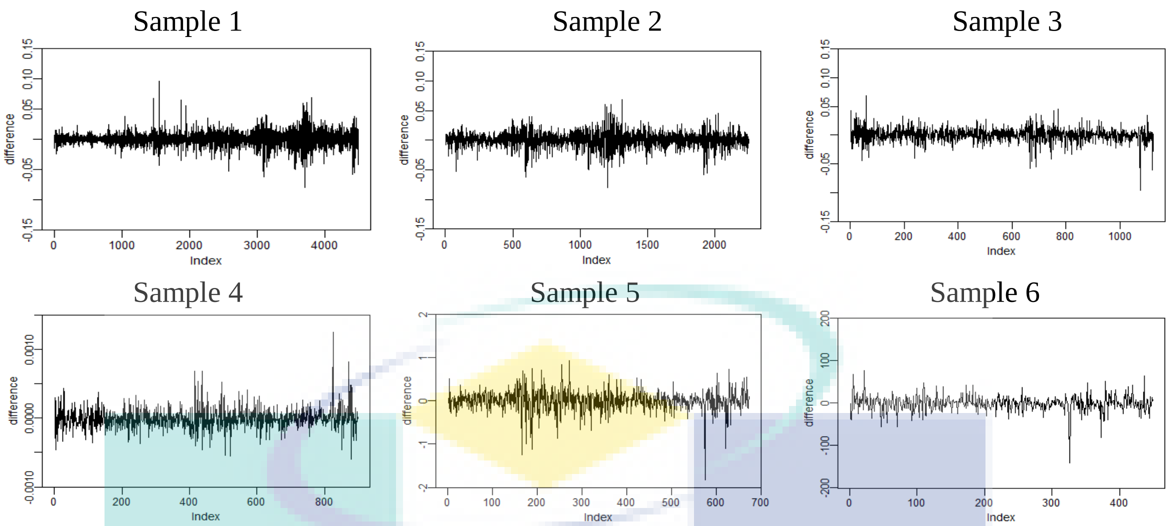


Figure 4.22 Time plot for stationary series of Sample 1 to 6

The descriptive statistics for stationary series, s_t of Sample 1 to 6 is given in Table 4.23. The stationary series for all samples reject the normality assumption and show leptokurtic or heavy-tailed characteristics since the excess kurtosis of the stationary series is significant and positive. This implies that the stationary series of the gold price have more mass on the tails and contains more extreme values.

Based on Table 4.23, it is also observed that the mean of the stationary series for Sample 1 and 2 are significant at 5% significance level while the mean for Sample 3 to 6 are insignificant. The positive value and significance of the mean for Sample 1 and 2, implies that the average return for 9-year and above investment in gold market is positive, as summarised in Table 4.24. This supports the gold market return performs well in the long-term period. The annual simple return of the gold price for the Sample 1 and 2 are 7.17% and 16.14% per annum, respectively, shows that the 9-year investment is the best minimum duration for the gold market.

The fourth and fifth steps are testing the linearity and serial correlation in the stationary series by applying the preliminary of linearity test and Portmanteau test, respectively. These two steps are necessary in satisfying the Box-Jenkins's conditions since the model performs for a linear and serially correlated series. Figure 4.23 shows the plot of the stationary series and its lagged series. Based on the figure, the plot for all samples are nearly straight line, which indicates that the series fits the linear model graphically. This implies that the Box-Jenkins model is appropriate for the data series.

Table 4.23 Descriptive statistics for in-sample stationary series of Sample 1 to 6

Sample	**Stationary series, s_t**	**Mean**	**Std. dev.**	**Skewness**	**Kurtosis**	**JB - test**	**Distribution**
1	$s_t = y_t^* - y_{t-1}^*$	0.0003 (0.0398)	0.0105	0.0402 (0.2710)	6.9142 (0.0000)	8974.3798 (0.0000)	Symmetric, Leptokurtic
2	$s_t = y_t^* - y_{t-1}^*$	0.0006 (0.0174)	0.0127	-0.3374 (0.0000)	3.5419 (0.0000)	1222.1838 (0.0000)	Negatively skewed, Leptokurtic
3	$s_t = y_t^* - y_{t-1}^*$	0.0004 (0.3112)	0.0122	-0.5852 (0.0000)	5.9934 (0.0000)	1755.8600 (0.0000)	Negatively skewed, Leptokurtic
4	$s_t = y_t^* - y_{t-1}^*$	-3×10^{-6} (0.5708)	0.0002	0.9506 (0.0000)	6.7163 (0.0000)	1836.4894 (0.0000)	Positively skewed, Leptokurtic
5	$s_t = y_t^* - y_{t-1}^*$	0.0004 (0.9678)	0.2494	-1.0058 (0.0000)	6.8980 (0.0000)	1461.8425 (0.0000)	Negatively skewed, Leptokurtic
6	$s_t = y_t - y_{t-1}$	-0.6203 (0.4825)	18.698 1	-0.8445 (0.0000)	8.3118 (0.0000)	1361.9921 (0.0000)	Negatively skewed, Leptokurtic

*Values in parenthesis denotes *p*-value and Std. Dev. is abbreviated of standard deviation.

Table 4.24 Annual simple return for Sample 1 and 2

Sample	**Duration**	R_t	**Valuation at the end of duration for $1 investment**
1	18 years (24/11/93 - 20/12/2011)	7.17%	$3.48
2	9 years (5/12/2003 - 18/12/2012)	16.14%	$3.84

Since the linearity is depicted for all stationary series of gold price graphically, then the Portmanteau test of LBQ-test is applied to check whether the series is a serially correlated series statistically. Table 4.25 summarises the results of the *Q*-test for $k_{\max}$ equals to ln *T*, 10 and 15, where *T* is the number of data for in-sample series. The $k_{\max}$ of ln *T*, 10 and 15 for the LB*Q*-test are based on the recommendation for nonseasonal series by Tsay (2013), Hyndman and Anthanasopoulus (2014) and Engle (2001), respectively. Based on the results, it is clear that series in Sample 1 to 3 reject the null hypothesis of no serial correlations in the stationary series for all $k_{\max}$ at 5% significance level, while other samples do not. However, it is found that series in Sample 6 rejects the null hypothesis for $k_{\max} = 15$ at 5% significance level. Therefore, it can be concluded that the stationary series for Sample 1, 2, 3 and 6 are serially correlated, thus, the Box-Jenkins model is appropriate to be applied to the series. While, Sample 4 and 5 are not considered in the next step since the stationary series for the samples are not serially correlated.

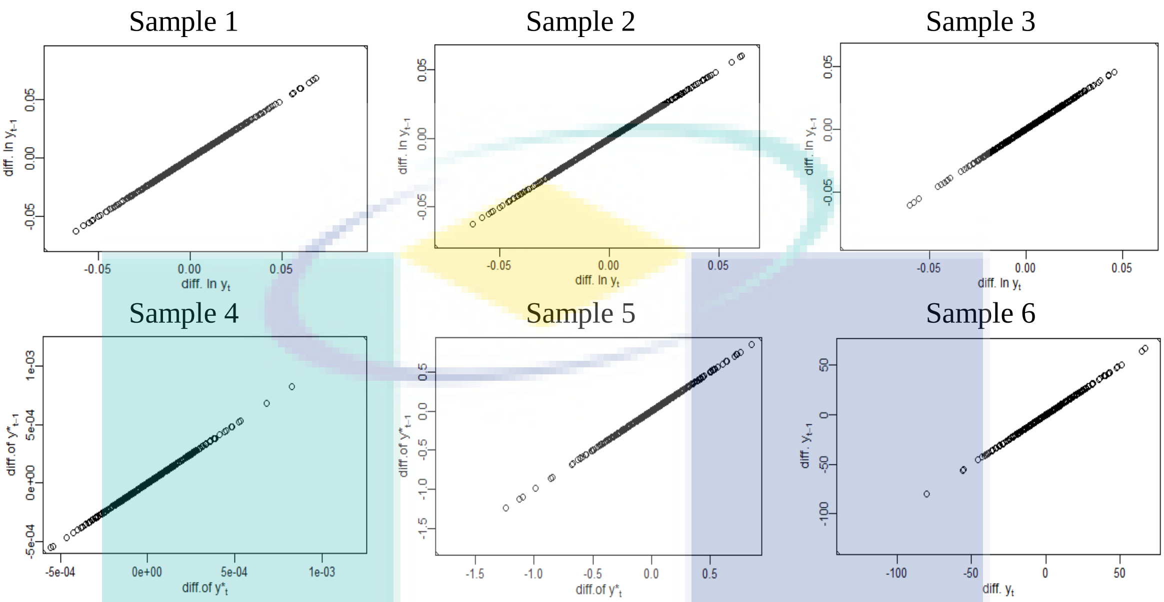


Figure 4.23 Plot of the stationary series and its lagged series

Table 4.25 Portmanteau test of LBQ-test for Sample 1 to 6

Data	***T***	**LBQ-test for** s_t			**Is Serially Correlated at** $\alpha = 0.05$**?**
		$k_{\max} = \ln T$	$k_{\max} = 10$	$k_{\max} = 15$	
1	4500	19.0700 (0.0246)	19.2450 (0.0373)	35.9630 (0.0018)	Yes
2	2250	15.9820 (0.0426)	20.0290 (0.0290)	31.3260 (0.0079)	Yes
3	1125	12.2690 (0.0921)	18.1030 (0.0532)	31.8360 (0.0068)	Yes
4	900	5.7932 (0.5641)	8.1145 (0.6177)	11.9120 (0.6857)	No
5	675	5.6586 (0.5801)	9.0999 (0.5227)	11.8980 (0.6867)	No
6	450	7.3670 (0.2882)	22.5090 (0.0951)	18.7180 (0.0440)	Yes for $k_{\max} = 15$

*Values in parenthesis denotes *p*-value.

The sixth step in Stage I of the proposed procedure is to identify the appropriate Box-Jenkins model for the in-sample series. As discussed in Section 4.2, the model suggested by the EACF method is proven to be the best Box-Jenkins model as compared to other possible models as suggested by the ACF and PACF method. Therefore, in this study, the model chosen by the EACF method will be used for the next step in detecting

ARCH effect in the stationary series. Figure 4.24 shows the EACF table and the corresponding suggested Box-Jenkins model for stationary series of samples 1, 2, 3 and 6. It can be seen that, all the stationary series suggest ARIMA(0,1,0). This means that all stationary series under consideration indicate random walk model. The suggested Box-Jenkins model by the EACF method is observed to be the same as obtained by `forecast` package in R.

Sample 1

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o o x o x o o o o x  x  o
1 x o o x o o o o o o o  x  o
2 x x o o o o o o o o o  o  o
3 x x x o o x o o o o o  o  o
4 x x x x o o o o o o o  o  o
5 x x x x x o o o o o o  o  o
6 x x x x x o o o o o o  o  o
```

ARIMA(0,1,0)

Sample 2

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o o o o x o o o o x  o  o
1 x o o o o x o o o o o  x  o
2 x x o o o x o o o o o  o  o
3 x x x o o x o o o o o  o  o
4 x x x x o x o o o o o  o  o
5 x x x x o o o o o o o  o  o
6 x x x o x x o o o o o  x  o
```

ARIMA(0,1,0)

Sample 3

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o o o o o o o x o o  o  o
1 x o o o o x o o o o o  o  o
2 x x o o o x o o o o o  x  o
3 x x x o o o o o o o o  o  o
4 x x x x o o o o o o o  o  o
5 x x x x x o o o o o o  o  o
6 x x x x x o o o o o o  x  o
```

ARIMA(0,1,0)

Sample 6

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o o o o o o x o o o  o  o
1 o o o o o o o x o o o  o  o
2 x x o o o o o o o o o  o  o
3 x x x o o o o o o o o  o  o
4 x x o x o o o o o o o  o  o
5 x x x x x o o o o o o  o  o
6 x x x x x x o o o o o  o  o
```

ARIMA(0,1,0)

Figure 4.24 EACF table and its Box-Jenkins model for stationary series of Sample 1, 2, 3 and 6

The seventh step in the identification stage of the modified procedure of BJ-G is to detect the existence of volatility clustering in the residuals of the Box-Jenkins model by doing the preliminary of heteroscedasticity test using LBQ-test for squared residuals of the Box-Jenkins model. If the test indicates the presence of ARCH in the residuals up to $k_{\max}$, then there is a need to justify the use of GARCH as compared to ARCH model in handling heteroscedasticity in the residuals of the Box-Jenkins model for the volatile series by examining the value of PACF for the squared residuals of the model.

Table 4.26 shows the LBQ-test for squared residuals of the Box-Jenkins model for the stationary series of the sample considered, at lag 10 and 15. The significant *p*-value of the LBQ-test for samples 1 to 3 at 5% significance level reveals the presence of ARCH in the residuals of the Box-Jenkins model up to lag 20. This indicates the existence of highly volatile characteristic in the series, imply that the variance equation for the Box-

Jenkins model is not correctly specified up to the lag. Since the PACF of the squared residuals for Sample 1 to 3 shows insignificant results at 5% significance level up to lag 17, 20 and 18, respectively, indicates that the GARCH model is parsimony to be used as compared to ARCH in handling the existence of heteroscedasticity in the residuals. While, the LBQ-test for Sample 6 shows that the Box-Jenkins model is sufficient enough to analyse the stationary series, thus, Sample 6 is not considered for the next analysis.

Table 4.26 The preliminary analysis of heteroscedasticity test for the Box-Jenkins model of the stationary series for the samples considered

Sample	**Box-Jenkins Model**	**LBQ-test for a_t^2** $k_{\max}=10$	**LBQ-test for a_t^2** $k_{\max}=15$	$k_{\max}$ **PACF for** a_t^2 at $\alpha=0.05$
1	ARIMA(0,1,0)	809.15000 (0.0000)	1097.9000 (0.0000)	17
2	ARIMA(0,1,0)	393.2200 (0.0000)	601.0300 (0.0000)	20
3	ARIMA(0,1,0)	16.0790 (0.0790)	32.1140 (0.0062)	18
6	ARIMA(0,1,0)	1.8682 (0.9973)	3.2326 (0.9994)	-

*Values in parenthesis denotes *p*-value

The last step in Stage I of the second proposed procedure of BJ-G is to identify the appropriate order of *r* and *s* of GARCH (*r*,*s*) to combine with the appropriate Box-Jenkins model for each sample considered. Figure 4.25 presents the r_k and r_{kk} for the squared residuals of the considered ARIMA model that will be used in identifying the appropriate value of *r* and *s*, respectively. On the assumption that the stationary series for the sample considered are random, the standard error limit of r_k and r_{kk} are the same, that are 0.0298, 0.0422 and 0.0597 for samples 1, 2 and 3, respectively.

Referring to Figure 4.25, it can be seen that the pattern of r_k and r_{kk} for Sample 1 and 2 are similar. By considering the appropriateness of GARCH in the model $(r \neq 0, s \neq 0)$, the suggested order values are $r = 1,2,3,4,5$ and $s = 1,2,3,4$ for the samples. On the other hand, most of r_k and r_{kk} values for Sample 3 are close to zero and it is hard to identify graphically the appropriate order of the GARCH model. Since the pattern for r_k and r_{kk} are similar to ARIMA model in the analysis of Section 4.2, therefore the values of *r* and *s* are suggested to be 1 and 2, respectively. Hence, there are

20 possible model combinations of ARIMA-GARCH that need to be considered for Samples 1 and 2, while four possible ARIMA-GARCH models for Sample 3.

	Sample 1		Sample 2		Sample 3	
	AC	PAC	AC	PAC	AC	PAC
1	0.164	0.164	0.096	0.096	0.041	0.041
2	0.123	0.098	0.120	0.112	0.048	0.046
3	0.164	0.134	0.172	0.155	0.035	0.031
4	0.130	0.080	0.115	0.080	0.004	-0.001
5	0.164	0.115	0.168	0.126	0.011	0.008
6	0.134	0.067	0.137	0.082	0.042	0.040
7	0.127	0.060	0.138	0.079	0.044	0.041
8	0.127	0.054	0.142	0.072	0.034	0.026
9	0.098	0.022	0.107	0.034	0.048	0.040
10	0.085	0.010	0.101	0.021	0.048	0.040
11	0.084	0.012	0.094	0.013	0.014	0.004
12	0.154	0.094	0.203	0.137	0.065	0.056
13	0.112	0.035	0.123	0.047	0.035	0.025
14	0.092	0.021	0.100	0.019	0.021	0.009
15	0.110	0.037	0.128	0.036	0.086	0.075
16	0.077	0.001	0.085	0.002	0.016	0.001
17	0.150	0.083	0.201	0.121	0.016	0.002
18	0.111	0.027	0.122	0.034	0.065	0.052
19	0.092	0.015	0.100	0.010	0.016	0.001
20	0.130	0.048	0.157	0.053	0.060	0.046

Figure 4.25 The sample ACF and the sample PACF for squared residuals of the ARIMA model considered for Sample 1 to 3

The details on how to choose the preferred BJ-G model for Sample 1 can be referred to Section 4.2. The empirical results on the possible BJ-G models for Sample 2 and 3 can be referred to Appendix 3. Based on the empirical results, the similar characteristics as the series in Sample 1 are observed on the series of Sample 2 (refer Section 4.2 for details). Therefore, the same decision as for Sample 1 is decided for the order of r and s for the series of Sample 2, that is the ARIMA model is preferably combined with GARCH(1,1). Meanwhile, for the series of Sample 3, only GARCH(1,1) and GARCH(1,2) are significant to be combined with ARIMA(0,1,0), where GARCH(1,1) is the preferred one based on the principle of parsimony.

The procedure for Stage II to Stage III in the proposed new procedure of BJ-G for the series of Sample 2 and 3 are the same as applied to Sample 1 in Section 4.2. Table 4.27 tabulates the results of the preferred BJ-G model for stationary series of Sample 1 to 3 from the parameter estimation and diagnostic checking results. Based on the table, it can be seen that the series of Sample 1 has the smallest value for both the selection criteria (AIC and SIC) as compared to the series of Sample 2 and 3. However, by applying the parsimonious principle, Sample 3 is more preferred since the estimation results are decreased marginally between the BJ-G models that is adequate to fit the data of the sample considered. The one-step ahead out-of-sample forecast evaluations for daily gold price of Sample 1 to 3, as given in Table 4.28, also support the preference of Sample 3.

Table 4.27 Results from Stage II to III of the proposed framework for the preferred Box-Jenkins – GARCH model for stationary series of Sample 1 to 3

STAGES	**Sample 1**	**Sample 2**	**Sample 3**
	ARIMA(0,1,0) - GARCH(1,1) ***with*** $\varepsilon_t \sim t$	***ARIMA(0,1,0) - GARCH(1,1)*** ***with*** $\varepsilon_t \sim GED$	***ARIMA(0,1,0)- GARCH(1,1)*** ***with*** $\varepsilon_t \sim t$
STAGE II: PARAMETER ESTIMATION			
c	$7.58\times10^{-5}\,(0.4227)$	$0.0008(0.0000)$	$0.0007(0.0223)$
α_0	$1.92\times10^{-7}\,(0.0000)$	$1.19\times10^{-6}\,(0.0166)$	$2.50\times10^{-6}\,(0.0270)$
α_1	$0.0663(0.0000)$	$0.0461(0.0000)$	$0.0345(0.0024)$
β_1	$0.9386(0.0000)$	$0.9466(0.0000)$	$0.9474(0.0000)$
v	$4.6977(0.0000)$	$1.2738(0.0000)$	$4.8148(0.0000)$
AIC	-6.7078	-6.1426	-6.1641
SIC	-6.7006	-6.1299	-6.1417
STAGE III: DIAGNOSTIC CHECKING			
DW-test	2.0128	1.9863	2.0326
LBQ (10)	10.0070(0.4400)	17.346(0.0670)	11.6610(0.3080)
LBQ (15)	18.2380(0.2500)	26.1810(0.0360)	20.6320(0.1490)
LBQ^2 (10)	6.0308(0.8130)	6.3357(0.7860)	1.8660(0.9970)
LBQ^2 (15)	8.4942(0.9020)	12.5020(0.6410)	2.9469(1.0000)
ARCH (10)	6.1155(0.8055)	12.4174(0.6472)	1.8760(0.9972)
ARCH (15)	8.3064(0.9109)	6.2228(0.7962)	2.9419(0.9996)

*Values in parenthesis denotes *p*-value

Table 4.28 Forecast evaluations for the preferred Box-Jenkins – GARCH model for daily gold price series of Sample 1 to 3

Forecasting Evaluation	**Sample 1**	**Sample 2**	**Sample 3**
	ARIMA(0,1,0) - GARCH(1,1) ***with*** $\varepsilon_t \sim t$	***ARIMA(0,1,0) - GARCH(1,1)*** ***with*** $\varepsilon_t \sim GED$	***ARIMA(0,1,0)- GARCH(1,1)*** ***with*** $\varepsilon_t \sim t$
RMSE	18.3716	19.2190	17.8764
MAE	12.6855	12.6869	12.9301
MAPE (%)	0.8402	0.9155	0.9956

Based on the results of Stage II to III for Sample 1 to 3 in Table 4.27, it can be observed that the preferred combination model for the series of all samples are the random walk. The preferred BJ-G model for the series of Sample 1 are totally random walk since the constant of ARIMA model is not significant, while the model for the series of Sample 2 and 3 are random walk with drift. The innovations of all preferred model for the samples considered follow *t* distribution, except for the series of Sample 2, which follows the GED distribution. The standardised residuals of the ARIMA model appear to be random for each sample considered, but their magnitudes exhibit the characteristics of heavy tails, which support the non-normal innovations. The non-normality of the innovations could be related to the heavy tails distribution of the stationary series (see Table 4.21).

The results from Stage II and Stage IV indicate that the optimal number of data to forecast gold price using the second proposed procedure of BJ-G is 1250 data of the 5-year sample. The empirical results of model selection criteria and one-step-ahead forecasting evaluations suggested that the latest 25% of 5000 data is sufficient enough to be employed in the BJ-G model with similar forecasting performance as by using all the data. Consequently, the model of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations for the series of Sample 3 of daily gold price is given by Equation 4.6, where y_t is the daily gold price, s_t is the stationary data for the daily gold prices, a_t is the random error at time *t* period and σ_t^2 is the conditional variance of y_t.

$$\begin{aligned} &y_t = y_{t-1}\exp(s_t), \qquad s_t = 0.0007 + a_t, \qquad a_t = \sigma_t \varepsilon_t \\ &\sigma_t^2 = 2.50\times 10^{-6} + 0.0345 a_{t-1}^2 + 0.9474\sigma_{t-1}^2 \text{ and } \varepsilon_t \sim t_{4.81}^{*} \end{aligned} \qquad 4.6$$

By referring to Equation 4.6, the significant of $c = 0.0007$ in the mean model of ARIMA(0,1,0)-GARCH(1,1) shows the upward trend of the forecast model implies that the expected mean return of the series is about 0.07%, which is positive in long term duration. While, the value of $\left(\hat{a}_1^2 = 0.0345^2\right) < \dfrac{1}{3}$ in the variance model shows that the unconditional fourth moment of the daily log returns of gold price exists (Tsay, 2013). This implies that the distribution of the daily log return tends to contain more extreme values or is said to be leptokurtic which is consistent with the nature of data series (refer to Table 4.23). The extreme values contribute to the existence of highly volatile characeristics in the data series. The large value of $\beta_1 = 0.9474$ in the variance model reflects to a long-term persistence of volatility clustering.

The one-step ahead forecast using the ARIMA(0,1,0)-GARCH(1,1) model with *t* innovations for daily gold prices from 25th June 2013 to 17th December 2013 is graphically shown in Figure 4.26. It is observed that the forecast gold prices (in red dashed line) fluctuate between USD1200 and USD1400 per ounce and follows closely with one-day lag to the actual data (in blue solid line) for the 125-day out-sample period. The characteristics that is reflected from the mean and variance models prove that the ARIMA-GARCH model is able to follow the nature of the highly volatile data series well so that it can be used in forecasting the actual gold price. The comparison values between actual data of daily gold price and its forecast price using the BJ-G model for the last ten days out-of-sample simulation period for the series of Sample 3 is given by Table 4.29.

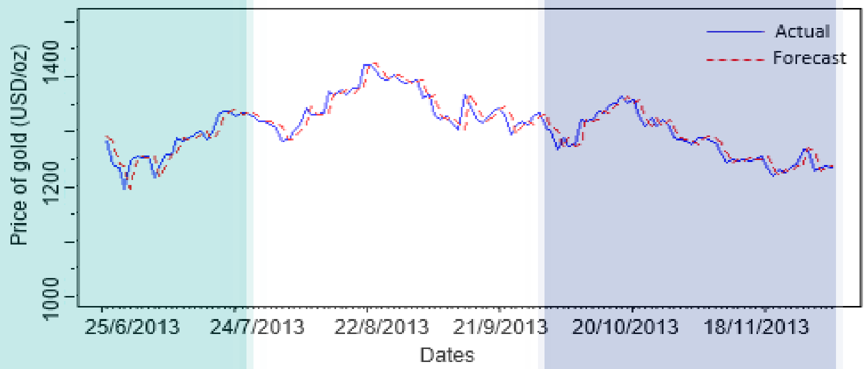


Figure 4.26 Plot of the actual and forecast data using ARIMA(0,1,0)-GARCH(1,1) with *t* innovations for out-of-sample period of the series of Sample 3

Table 4.29 The comparison between actual and forecast gold prices for the last ten days out-of-sample simulation period of Sample 3 using the model of ARIMA(0,1,1)-GARCH(1,1) with *t* innovations

Date	Actual price (USD/Oz)	Forecast price (USD/Oz)	Difference (USD/Oz)
4 Dec 2013	1227.50	1218.07	-9.43
5 Dec 2013	1222.50	1228.33	5.83
6 Dec 2013	1233.00	1223.33	-9.67
9 Dec 2013	1237.00	1233.83	-3.17
10 Dec 2013	1266.25	1237.84	-28.41
11 Dec 2013	1260.75	1267.12	6.37
12 Dec 2013	1225.25	1261.60	36.35
13 Dec 2013	1232.00	1226.08	-5.92
16 Dec 2013	1234.75	1232.83	-1.92
17 Dec 2013	1231.75	1235.58	3.83

4.4 Simulation study on the Multistep Forecasting for Highly Volatile Data using the Third Proposed Procedure of BJ-G

Given the overall positive results at the one-step ahead forecast in the empirical results as described in Section 4.3, therefore the following study is aimed at assessing the forecasting performance of the BJ-G model for higher horizons or at *n*-step-ahead forecast. In evaluating the performance of the multistep ahead forecast, the third proposed procedure as illustrated in Figure 3.15 is employed to daily world gold price series for the last 5-year data (Sample 3), since the series is considered optimal for BJ-G model in the case study. The implementation of the third proposed procedure associated with R codes is explained explicitly in the following step 1 to 5. The Step 1 to 5 is repeated for $h = 2, 3, ..., n$ in order to obtain the multistep ahead forecast evaluations for BJ-G model.

Step 1: Based on the proposed model as in Equation 4.6 for Sample 3 of gold price series, the value of $\hat{s}_{T+h}$ for ARIMA(0,1,0)-GARCH(1,1) using t innovations is obtained through simulation using Equation 4.7.

$$\begin{aligned} &\hat{s}_{T+h} = 0.0007 + \hat{a}_{T+h}, \quad \hat{a}_{T+h} = \hat{\sigma}_{T+h}\,\hat{\varepsilon}_{T+h} \\ &\hat{\sigma}^2_{T+h} = 2.50\times 10^{-6} + 0.0345\hat{a}^2_T + 0.9474\hat{\sigma}^2_T \text{ and } \hat{\varepsilon}_{T+h} \sim t^*_{4.81} \end{aligned} \qquad 4.7$$

The corresponding R codes for the proposed value of $\hat{s}_{T+h}$ are written as follows.

```
spec = garchSpec(model = list(mu=0.0007,omega = 2.5e-6, alpha = 3.45e-
2,beta = 9.474e-1, shape=4.81), cond.dist="std")
st_AG=garchSim(spec, n = 125);st_AG
```

Step 2: For the case study, the forecast data $\hat{y}_{T+h}$ is given by Equation 4.8.

$$\hat{y}_{T+h} = y_T \exp(\hat{s}_{T+h}) \qquad 4.8$$

since the transformed data is in logarithm. Note that $\hat{s}_{T+h}$ is obtained from Step 1. The corresponding R codes of $\hat{y}_{T+h}$ for one-step ahead is written as follows.

```
f_AG_1step<-matrix(0,125,1); f_AG_1step

f_AG_1step[1]=dt[1125]*exp(st_AG[1]);f_AG_1step[1]
for(i in 2:125){
  f_AG_1step[i]=dt_o[i-1]*exp(st_AG[i])
}
f_AG_1step
```

Step 3: Obtain forecasting evaluations of MAE, RMSE and MAPE for *h*-step ahead forecast. The corresponding R codes for the forecasting evaluations of $\hat{y}_{T+h}$ for one-step ahead of the daily gold price is written as follows.

```
forecastevaluation<-function(dt4_o,f_AG_1step)
  T<- 125
error1_AG_1step=matrix(0,T,1);error1_AG_1step
error2_AG_1step=matrix(0,T,1);error2_AG_1step
error3_AG_1step=matrix(0,T,1);error3_AG_1step
error4_AG_1step =matrix(0,T,1); error4_AG_1step
rmse_AG_1step=rep(0,1); rmse_AG_1step
mae_AG_1step=rep(0,1); mae_AG_1step
mape_AG_1step =rep(0,1); mape_AG_1step
for (i in 1:T){
  error1_AG_1step[i]<-dt_o[i]-f_AG_1step[i]
  error2_AG_1step[i]<-abs(error1_AG_1step[i])
  error3_AG_1step[i]<-error1_AG_1step[i]^2
  error4_AG_1step[i]<-abs(error1_AG_1step[i]/dt_o[i])
}
cbind(error1_AG_1step,error2_AG_1step,error3_AG_1step,error4_AG_1step)
mae_AG_1step=sum(error2_AG_1step)/T; mae_AG_1step
rmse_AG_1step=sqrt(sum(error3_AG_1step)/T); rmse_AG_1step
mape_AG_1step=(100/T)*sum(error4_AG_1step); mape_AG_1step
forecastevaluation(dt_o,f_AG_1step)
```

Step 4: Obtain the prediction intervals (PIs) for $\hat{y}_{T+h}$. Since a_t for the series in the case study using the proposed BJ-G model follows *t* distribution with a degrees of freedom $v = 4.81$, therefore the 80% PIs and 95% PIs for *h*-step ahead are given in Equation 4.9 and 4.10, respectively. In obtaining $\mathrm{Var}[e_T(h)]$, Equation 3.64 is applied since the proposed model for the data series is ARIMA(0,1,0)-GARCH(1,1), which is a random walk model. In practice, the $\mathrm{Var}[e_T(h)]$ is the variance of the residual for *h*-step ahead, as can be obtained from basic statistics of the residual for each forecast horizon (refer to Appendix 4).

$$80\%\ \text{PI:}\ \hat{y}_T(h) \pm t_{0.1,4.81}\sqrt{\mathrm{Var}[e_T(h)]} \qquad 4.9$$

$$90\%\ \text{PI:}\ \hat{y}_T(h) \pm t_{0.025,4.81}\sqrt{\mathrm{Var}[e_T(h)]} \qquad 4.10$$

The R codes for PIs of 80% and 95% of $\hat{y}_{T+h}$ in the case study are as follows.

```
resiAG_1step=matrix(0,125,1); resiAG_1step
for(i in 1:125){
  resiAG_1step[i]<-dt_o[i]-f_AG_1step[i]
}
resiAG_1step;basicStats(resiAG4)
v1<-qt(c(.025, .975), df=4.81);v1
```

```
v2<-qt(c(.1, .9), df=4.81);v2
T<- 125
lo95_AG_1step=matrix(0,T,1);lo95_AG_1step
hi95_AG_1step=matrix(0,T,1);hi95_AG_1step
lo80_AG_1step=matrix(0,T,1);lo80_AG_1step
hi80_AG_1step =matrix(0,T,1);hi80_AG_1step
h=matrix(0,T,1);h
for(i in 1:125){
  h[i]=i
  lo95_AG_1step[i]<-f_AG_1step[i]-(2.6014*(sqrt(h[i]*320.3818)))
  hi95_AG_1step[i]<-f_AG_1step[i]+(2.6014*(sqrt(h[i]*320.3818)))
  lo80_AG_1step[i]<-f_AG_1step[i]-(1.4847*(sqrt(h[i]*320.3818)))
  hi80_AG_1step[i]<-f_AG4[i]+(1.4847*(sqrt(h[i]*320.3818)))
}
cbind(dt_o,f_AG_1step,lo95_AG_1step,hi95_AG_1step,lo80_AG_1step,hi80_A
G_1step)
```

Step 5: Plot the graph of actual data in the out-of-sample series, y_{T+h} and the *h*-step ahead forecast, $\hat{y}_{T+h}$ with its prediction intervals. The R codes for PIs of 80% and 95% for one-step ahead forecast in the case study are written as follows.

```
date_AG<- dt[1126:1250,7];date_AG
library(Hmisc)
matplot(dt_o,xaxt="n",type="l",lwd=3,col="blue",xlab="Dates",ylab="Pri
ce(USD/oz)",mgp=c(2,0.4,0),ylim=c(700,1800),cex.lab=1.3,cex.axis=1.3,l
ab=c(4,4,7))
par(new = TRUE)
matplot(dt_o,xaxt="n",type="points",pch=".",cex=5,col="black",xlab="Da
tes",ylab="Price(USD/oz)",mgp=c(2,0.4,0),ylim=c(700,1800),cex.lab=1.3,
cex.axis=1.3,lab=c(4,4,7))
par(new = TRUE)
matplot(f_AG_1step,xaxt="n",type="l",col="red",lty=2,lwd=2,xlab="Dates
",ylab="Price(USD/oz)",mgp=c(2,0.4,0),ylim=c(700,1800),cex.lab=1.3,cex
.axis=1.3,lab=c(4,4,7))
par(new = TRUE)
matplot(f_AG_1step,xaxt="n",col="black",type="points",pch="o",cex=1,xl
ab="Dates",ylab="Price(USD/oz)",mgp=c(2,0.4,0),ylim=c(700,1800),cex.la
b=1.3,cex.axis=1.3,lab=c(4,4,7))
par(new = TRUE)
matplot(cbind(lo95_AG_1step,hi95_AG_1step,lo80_AG_1step,hi80_AG_1step)
,xaxt="n",type="l",lty=2,col=c("black","black","green","green"),xlab"D
ates",ylab="Price(USD/oz)",mgp=c(2,0.4,0),ylim=c(700,1800),cex.lab=1.3
,cex.axis=1.3,lab=c(4,4,7))axis(1.5,at=1:125,labels=date_AG,xaxp=c(2,1
0,124),tck=0,mgp=c(2,0.4,0),xlim=c(1,125),cex.lab=1.3,cex.axis=1.3,lab
=c(4,4,7))
minor.tick(nx=3)
```

The empirical results of the forecasting performance of BJ-G model is based on 1250 daily world gold price series, starting 22 December 2008 to 17 December 2013, that is Sample 3 in Section 4.3. Given the positive results of one-day ahead using the BJ-G model in Section 4.3, the forecasting performance of the model will be assessed at horizons greater than one day $(h > 1)$. For the 5-year data series under study, the first

1125 data (90%) are used to estimate the model while the last 125 data (10%) are defined as the out-of-sample series.

Table 4.30 presents the one-step to 125-step ahead forecast evaluation results with the number of data that lies outside the prediction intervals of 80% and 95% of the forecast value at the forecast origin 1125 for the daily gold price using ARIMA(0,1,0)-GARCH(1,1) with *t* innovations. Referring to Equation 4.6 for the model and Equation 4.8 for the updated point forecast, $\hat{y}_{T+h}$. The details of the analysis can be referred to Appendix 4. Based on Table 4.30, the values of MAE, RMSE and MAPE are increasing as the forecast horizon increases. This is in agreement with common sense that $\hat{y}_{T+2}$ is more uncertain as compared to $\hat{y}_{T+1}$. It can be observed that the forecast evaluations, specifically the values of MAE and RMSE for one-step to seven-step forecast horizons increased gradually. However, there is significant increment of the forecast evaluations from seven-step to ten-step forecast horizons.

On the other hand, by observing the prediction interval for each horizon under consideration, it can be seen that the ten-step ahead forecast results show the lowest number of actual price that lies outside 80% PIs and no actual prices are outside 95% PIs as compared to other multistep ahead forecast horizons. However, it is quite hard to make a decision based on the prediction interval in order to choose the appropriate forecast horizon for the model since the number of data outside the prediction intervals for multistep ahead forecast horizons are different marginally, specifically for 95% PIs.

Table 4.30 Forecast evaluation with prediction intervals for the considered forecast horizon

Forecast Horizon	**Forecast evaluation**			**Number of data outside prediction intervals**	
	MAE	**RMSE**	**MAPE**	**80%**	**95%**
1-step ahead	12.9301	17.8764	0.9956	1 (0.8%)	0 (0%)
2-step ahead	15.7938	21.3297	1.2132	20 (16%)	1 (0.8%)
3-step ahead	18.2953	24.4472	1.4098	25 (20%)	2 (1.6%)
4-step ahead	21.6096	28.3663	1.6716	20 (16%)	1 (0.8%)
5-step ahead	22.8394	28.9304	1.7647	22 (17.6%)	1 (0.8%)
7-step ahead	24.5981	30.1233	1.8941	**17 (13.6%)**	**2 (1.6%)**
10-step ahead	32.2870	40.1970	2.4859	15 (12%)	0 (0%)
15-step ahead	37.6551	46.2091	2.9068	21 (16.8%)	3 (2.4%)
25-step ahead	43.7949	53.0116	3.3840	36 (28.8%)	4 (3.2%)
125-step ahead	59.0288	76.2116	4.6135	23 (18.4%)	2 (1.6%)

*Values in parenthesis denote the percentage for the number of data outside the prediction intervals.

Hence, by considering both results of forecast evaluations and prediction intervals, the seven-step ahead is suggested for practical use since the values of errors are gradually increased from one-step to seven-step ahead forecast and the number of data outside the prediction intervals of 80% and 95% are among the lowest for mustistep forecast horizon. The results indicate that the seven-step ahead forecast perform the best in forecasting as compared to other multistep ahead forecast horizons. However, the performance of seven-step ahead forecast horizon is weaker than the one-step ahead forecast horizon.

Figure 4.27 shows the corresponding out-of-sample forecasting plot of seven-step ahead using the BJ-G model for the daily gold price. The forecast and actual prices are marked by "o" and "●", which linked with red dashed line and blue solid line, respectively. The forecasting plot includes the prediction intervals of 80% and 95% which are presented by the dashed line of green and black, respectively. It can be seen that the forecasting performance of the BJ-G model for up to seven-step ahead forecast is supported graphically by the plot. It is observed that the trend of seven-day ahead forecast price mimics the actual price for the out-of-sample period. Therefore, it can be concluded that ARIMA(0,1,0)-GARCH(1,1) with *t* innovations can be considered for forecast horizons up to seven-day ahead price for five-year data series. For the details of other multistep ahead forecast under consideration, refer to Appendix 4.

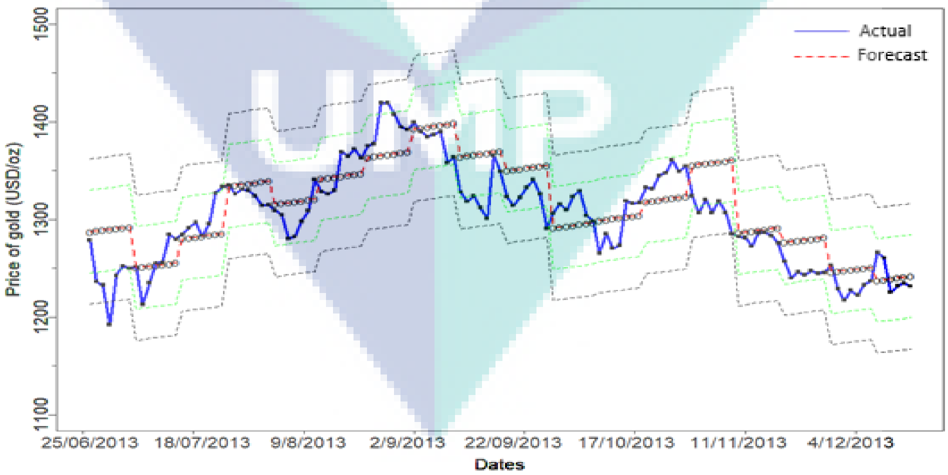


Figure 4.27 Plot of actual data and seven-step ahead forecast using ARIMA(0,1,0)-GARCH(1,1) with 80% (in green dashed line) and 95% (in black dashed line) PIs

Table 4.31 presents the forecast price of the first seven-day out-of-sample period for seven-step ahead forecast using the ARIMA(0,1,0)-GARCH(1,1) with *t* innovations associated with its PIs of 80% and 95% at the forecast origin price of 24 June 2013. Based on the forecast price, only two actual data are not within the 80% PIs as highlighted in Table 4.31, while all actual data are within 95% PIs. This indicates that the proposed model of ARIMA-GARCH is able to follow the trend of actual data up to seven-day ahead, specifically within 95% PIs.

Table 4.31 Actual price and the seven-step ahead forecast price using the model of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations

Date	Actual Price (USD/oz)	Forecast Price (USD/oz)	Prediction Interval	
			80%	95%
25 June 2013	1279.00	1287.62	(1228.12,1347.12)	(1183.37,1391.87)
26 June 2013	1236.25	1288.49	(1228.99,1347.99)	(1184.24,1392.74)
27 June 2013	1232.75	1289.36	(1229.86,1348.86)	(1185.11,1393.61)
28 June 2013	**1192.00**	1290.23	(1230.74,1349.73)	(1185.99,1394.48)
1 July 2013	1242.75	1291.11	(1231.61,1350.61)	(1186.86,1395.36)
2 July 2013	**1252.50**	1291.98	(1232.48,1351.48)	(1187.73,1396.23)
3 July 2013	1292.85	1188.60	(1233.36,1352.35)	(1188.60,1397.10)

4.5 The Empirical Results of the Box-Jenkins with GARCH-type Models using the Fourth Proposed Procedure of BJ-G

Given the promising results of combination model of ARIMA with GARCH in following the trend of actual daily gold price up to seven-day ahead, therefore the fourth proposed procedure as illustrated in Figure 3.16 is then applied to all GARCH-type models that are used in previous studies for highly volatile data including GARCH-M, EGARCH, TGARCH and APARCH (Ahmed, 2017; Ferenstein & Gasowski, 2004; Girish, 2016; Walid et al., 2011; Ord, Koehler, Snyder & Hyndman, 2009; Girish, 2016; Ahmad et al., 2015; Freedi et al., 2012; Liu et al., 2011; Liu & Shi, 2013; Liu et al., 2013). The steps and methods used for Stage I to IV in the fourth proposed procedure of BJ-G are the same as applied to Section 4.3 and 4.4. The ARIMA with standard GARCH or simply called as ARIMA-GARCH that assessed in previous sections (Section 4.3 and 4.4) will be used as a benchmark to other ARIMA-GARCH-type models under consideration. Hence, the empirical results in assessing the performance for the Box-Jenkins with GARCH-type models are based on the same data used in Section 4.4 (or the optimal data series in Section 4.3) to maintain the continuity of the study.

Note that, ARIMA(0,1,0) is the best Box-Jenkins model to analyse the data series as discussed in Section 4.3. Therefore, in this empirical study, the model of

ARIMA(0,1,0) is maintained to be used as the mean model of Box-Jenkins with the considered GARCH-type models. By considering the GARCH-type models applied previously in handling volatility in the highly volatile series that are GARCH-M, EGARCH, TGARCH and APARCH, the estimation results for ARIMA(0,1,0) with the GARCH-type models are presented by Table 4.32. Note that, the modelling procedure of Box-Jenkins with standard GARCH is applied to ARIMA with the GARCH-type models. Therefore, the details on how to choose the order for parameter of the GARCH-type models, specifically *r* and *s*, can be referred to Section 4.2. The details of empirical results on the possible ARIMA-GARCH-type models for the gold price series can be referred to Appendix 5.

By referring to the empirical results, the order of $r = 1$ and $s = 1$ is preferred for all GARCH-type models under consideration based on the principle of parsimony. Table 4.32 displays the estimation results for the parameters of ARIMA(0,1,0)-GARCH(1,1), ARIMA(0,1,0)-EGARCH(1,1), ARIMA(0,1,0)-GARCH(1,1)-M, ARIMA(0,1,0)-APARCH(1,1) and ARIMA(0,1,0)-TGARCH(1,1), respectively, with innovations ε_t follows normal, *t* and GED distributions. The skewed innovations are not considered here due to non-skewed preference for errors distribution of ARIMA-GARCH model to the data series as discussed in Section 4.3. The estimation results for ARIMA-GARCH is represented in this section for benchmark. Note that, ARIMA(0,1,0)-GARCH(1,1) with *t* innovations is chosen as the preferred ARIMA-GARCH model for the data series, as discussed in Section 4.3.

By considering other GARCH-type models in Table 4.32, it can be seen that the use of EGARCH(1,1) and TGARCH(1,1) are highly significant at 5% significance level, specifically for normal innovations. The large values of β_1 in the variance model of the three significant ARIMA-GARCH-type models (i.e. ARIMA(0,1,0)-GARCH(1,1) with *t* innovations, ARIMA(0,1,0)-EGARCH(1,1) with normal innovations, ARIMA(0,1,0)-TGARCH(1,1) with normal innovations) are reflected by the conditional standard deviation processes which demonstrate a relatively long term persistence of volatility clustering. The negative sign of significant leverage effect term g_1 in the asymmetric models of EGARCH and TGARCH with normal innovations implying that negative shocks (or bad news) gives a higher effect on the return of gold price than positive shocks.

Table 4.32 Estimation results for ARIMA(0,1,0) with selected GARCH-type models

Model	Parameter	$\varepsilon_t \sim$ Normal	$\varepsilon_t \sim t$	$\varepsilon_t \sim$ GED
ARIMA(0,1,0)-GARCH(1,1)	c	0.0006(0.0957)	0.0007(0.0223)	0.0006(0.0232)
	α_0	$4.2\text{x}10^{-6}$ (0.0000)	$2.5\text{x}10^{-6}$(0.0270)	$2.8\text{x}10^{-6}$ (0.0100)
	α_1	0.0485(0.0000)	0.0345(0.0024)	0.0372(0.0031)
	β_1	0.9241(0.0000)	0.9474(0.0000)	0.9425(0.0000)
	υ	-	4.8148(0.0000)	1.1682(0.0000)
	AIC	-6.0407	-6.1641	-6.1532
	SIC	-6.0229	-6.1417	-6.1309
ARIMA(0,1,0)-EGARCH(1,1)	c	0.0005(0.1696)	0.0007(0.0148)	0.0006(0.0332)
	α_0	-0.4646(0.0000)	-0.2049(0.0098)	-16.4524(0.0000)
	α_1	0.1411(0.0000)	0.0958(0.0003)	-0.0220(0.5402)
	g_1	-0.0492(0.0000)	0.0126(0.4506)	-0.0728(0.0510)
	β_1	0.9594(0.0000)	0.9849(0.0000)	-0.8604(0.0000)
	υ	-	4.8016(0.0000)	1.0974(0.0000)
	AIC	-6.0457	-6.1634	-6.1075
	SIC	-6.0233	-6.1365	-6.0807
ARIMA(0,1,0)-GARCH(1,1)-M	c	0.0010(0.2547)	0.0011(0.1460)	0.0007(0.2711)
	M	-3.4813(0.5736)	-3.1885(0.5564)	-1.0650(0.8318)
	α_0	$4.1\text{x}10^{-6}$(0.0000)	$2.5\text{x}10^{-6}$(0.0271)	$2.7\text{x}10^{-6}$(0.0100)
	α_1	0.0485(0.0000)	0.0341(0.0023)	0.0370(0.0031)
	β_1	0.9247(0.0000)	0.9480(0.0000)	0.9427(0.0000)
	υ	-	4.8038(0.0000)	1.1687(0.0000)
	AIC	-6.0392	-6.1626	-6.1515
	SIC	-6.0169	-6.1358	-6.1247
ARIMA(0,1,0)-APARCH(1,1)	c	0.0004(0.0000)	0.0007(0.0175)	0.0006(0.0077)
	α_0	0.0065(0.0004)	$4.2\text{x}10^{-5}$(0.0408)	$9.7\text{x}10^{-5}$(0.0294)
	α_1	0.0695(0.0000)	0.0525(0.0002)	0.0589(0.0004)
	g_1	0.5534(0.0023)	-0.1414(0.3793)	0.0324(0.8654)
	β_1	0.9059(0.0000)	0.9446(0.0000)	0.9344(0.0000)
	δ	0.4201(0.1575)	1.3290(0.0066)	1.2130(0.0461)
	υ	-	4.8210(0.0000)	1.1690(0.0000)
	AIC	3.0354	-6.1259	-6.0545
	SIC	3.0354	-6.1259	-6.0546
ARIMA(0,1,0)-TGARCH(1,1)	c	0.0005(0.1743)	0.0007(0.0191)	0.0006(0.0240)
	α_0	$4.5\text{x}10^{-6}$(0.0000)	$2.3\text{x}10^{-6}$(0.0333)	$2.7\text{x}10^{-6}$(0.0103)
	α_1	0.0518(0.0000)	0.0349(0.0022)	0.0349(0.0022)
	g_1	-0.1463(0.0034)	0.1063(0.4126)	0.0374(0.7515)
	β_1	0.9173(0.0000)	0.9489(0.0000)	0.9428(0.0000)
	υ	-	4.8158(0.0000)	1.1679(0.0000)
	AIC	-6.0413	-6.1628	-6.1514
	SIC	-6.0189	-6.1360	-6.1246

*Values in the parenthesis are *p*-values.

The normalised AIC and SIC results for the three significant models indicate that ARIMA-GARCH shows the smallest value for both information criteria as compared to other models, indicating that ARIMA(0,1,0)-GARCH(1,1) with t innovations is the preferred one in modelling the gold price data. However, all the significant ARIMA-GARCH-type models are considered in the next stage of the fourth proposed BJ-G procedure for further investigation.

On the other hand, the models of GARCH(1,1)-M and APARCH(1,1) for all innovations are insignificant. The insignificant of the GARCH-M model due to the estimated risk premium, M are negative and highly insignificant for all innovations. The value of M for the GARCH-M models implying that there are no serial correlations in the stationary series of daily gold price, or in other words, although the extra risk is hold for the asset, the return is indifferent with those who are not taking extra risk. Meanwhile, the insignificant of APARCH model for the stationary gold price series is due to δ and g_1 are highly insignificant for normal innovations and both t and GED innovations, respectively. The insignificant of APARCH models implying that the existence of leverage effect in the gold price series is not significant. Therefore, all models with GARCH-M and APARCH are not considered for Stage III and IV of the fourth proposed procedure of BJ-G in forecasting gold price series.

Table 4.33 presents the joint diagnostic checking for the three significant ARIMA-GARCH-type models. The model checking statistics shows that all considered models are adequate and correctly specified in describing the mean and variance of the stationary series of gold price. Even though the LBQ-test on the standardised residuals shows the existence of serial correlation (and the autocorrelation decreases in the higher lag), this is most probably due to small autocorrelations because of the large number of data used $(T = 1125)$ that should not be of practical importance (Ruppert & Matteson, 2015). This statement supports by the standardised residuals plot for the ARIMA-GARCH-type models to the in-sample stationary series as illustrated in Figure 4.28. Based on the figure, the standardised residuals using the considered models look random and reasonable except for several possible outliers.

Otherwise, the p-value is insignificant for LBQ-test on the squared standardised residuals for all considered models as shown in Table 4.33 which interprets that there is no ARCH in the residuals up to both lag 10 and 15, as supported by the result of LM

ARCH. This demonstrated that the ARIMA-GARCH-type models are able to handle the heteroscedasticity in the stationary series of gold price very well. Regarding the innovations, the good fit of the QQ-plot in Figure 4.29 that nearly a straight line except for not more than five outliers (or small fraction of the data) on the left and right tails, support graphically the use of *t* and normal innovations for the ARIMA-GARCH-type models.

Table 4.33 Model diagnostics for ARIMA with significant GARCH-type models

Diagnostic test	**ARIMA(0,1,0)-GARCH(1,1) with $\varepsilon_t \sim t$**	**ARIMA(0,1,0)-EGARCH(1,1) with $\varepsilon_t \sim$ Normal**	**ARIMA(0,1,0)-TGARCH(1,1) with $\varepsilon_t \sim$ Normal**
DW-test	2.0128	2.0338	2.0337
LBQ(10)	10.0070(0.4400)	11.9030(0.2920)	11.9050(0.2910)
LBQ(15)	18.2380(0.2500)	21.5240(0.1210)	21.7890(0.1130)
$LBQ^2(10)$	6.0308(0.8130)	2.4849(0.9910)	1.8738(0.9970)
$LBQ^2(15)$	8.4942(0.9020)	5.3342(0.9890)	3.7900(0.9980)
ARCH(10)	6.1155(0.8055)	2.6457(0.9886)	1.9639(0.9966)
ARCH(15)	8.3064(0.9109)	5.5423(0.9865)	3.9084(0.9980)

*Values in parenthesis denote *p*-values. *Q*(10) is the Ljung-Box statistics for standardised residuals at lag 10, $Q^2(10)$ is the Ljung-Box statistics for squared standardised residuals at lag 10, ARCH(10) is the Engle's Lagrange Multiplier test for heteroscedasticity at lag 10.

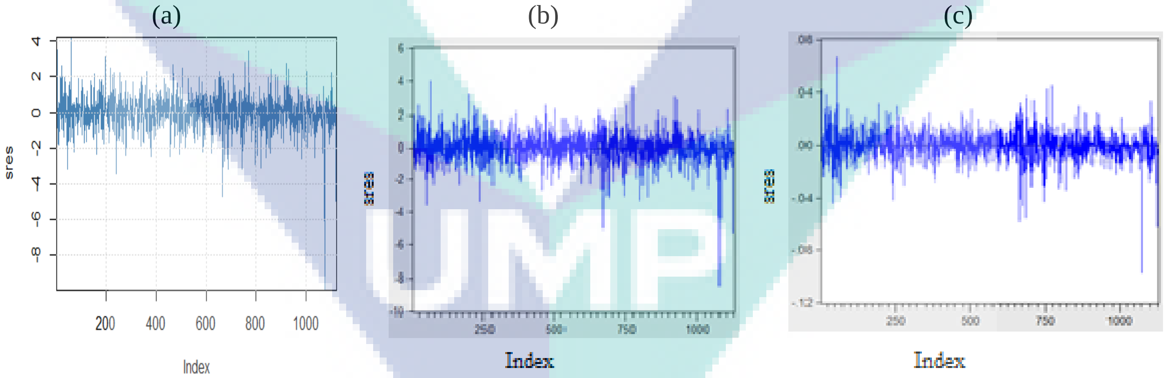


Figure 4.28 Standardised residual plot for in-sample stationary series of gold prices using ARIMA(0,1,0) and (a) GARCH(1,1) with $\varepsilon_t \sim t$, (b) EGARCH(1,1) with $\varepsilon_t \sim$ Normal, (c) TGARCH(1,1) with $\varepsilon_t \sim$ Normal

Forecasting performance results for multistep ahead of the daily gold price using ARIMA with the significant GARCH-type models have been reported in Table 4.34. For the one-step ahead (as highlighted in Table 4.34), all the considered models show similar performance with marginal differences in forecasting, as presented graphically in Figure

4.30. While, for multistep ahead forecast, it is observed that the model of EGARCH and TGARCH have the same characteristics as GARCH, that is the values of MAE and RMSE for one-step to seven-step forecast horizons increased gradually and there is significant increment of the forecast evaluations from seven-step to ten-step forecast horizons. Hence, by using the same consideration as for GARCH model, the seven-step ahead having the best performance of multistep forecasting using EGARCH and TGARCH models.

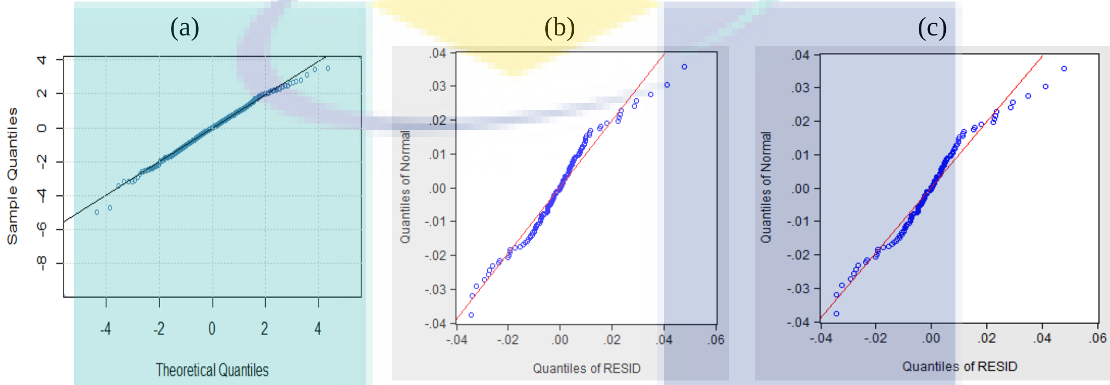


Figure 4.29 QQ-plot for in-sample stationary series of gold prices using ARIMA(0,1,0) and (a) GARCH(1,1) with $\varepsilon_t \sim t$, (b) EGARCH(1,1) with $\varepsilon_t \sim$ Normal, (c) TGARCH(1,1) with $\varepsilon_t \sim$ Normal

Figure 4.31 shows the actual and the seven-step forecast of the daily gold price using the considered ARIMA-GARCH-type models. The forecast and actual prices are marked by “o” and “●”, which linked with red dashed line and blue solid line, respectively. The forecasting plot includes the prediction intervals of 80% and 95% which are presented by the dashed line of green and black, respectively. It can be seen that the forecasting performance of the BJ-G models for up to seven-step ahead forecast is supported graphically by the plot since the trend of seven-day ahead forecast price mimics the actual price for the out-of-sample period.

Therefore, the seven-step ahead forecast evaluations of the ARIMA-GARCH-type models as highlighted in Table 4.34 are used in finding the best model of Box-Jenkins-GARCH-type model in multistep forecasting gold price using the fourth proposed model of BJ-G. It can be observed that all the considered models having similar results of forecast evaluations of RMSE, MAE and MAPE for the seven-step ahead, as the model of EGARCH shows the lowest error evaluations with marginal difference as compared to others. However, by looking at the prediction intervals evaluations, it reveals that GARCH model having the lowest number of data outside the prediction intervals of both 80% and 95%. Surprisingly, the ARIMA-GARCH model shows consistent performance as the lowest number of data outside the prediction intervals for other multistep ahead forecast horizon with significant difference as compared to other models, specifically at 95% PIs.

Table 4.34 Multistep forecast evaluation of ARIMA with significant GARCH-type models under consideration

Model	**Forecast Horizon**	**Forecast evaluation**			**Number of data outside the PIs**	
		MAE	**RMSE**	**MAPE**	**80%**	**95%**
ARIMA(0,1,0)-GARCH(1,1) with $\varepsilon_t \sim \boldsymbol{t}$	**1-step ahead**	**12.9301**	**17.8764**	**0.9956**	**1(0.8%)**	**0(0%)**
	2-step ahead	15.7938	21.3297	1.2132	20(16%)	1(0.8%)
	3-step ahead	18.2953	24.4472	1.4098	25(20%)	2(1.6%)
	4-step ahead	21.6096	28.3663	1.6716	20(16%)	1(0.8%)
	5-step ahead	22.8394	28.9304	1.7647	22(17.6%)	1(0.8%)
	7-step ahead	**24.5981**	**30.1233**	**1.8941**	**17(13.6%)**	**2(1.6%)**
	10-step ahead	32.2870	40.1970	2.4859	15(12%)	0(0%)
	15-step ahead	37.6551	46.2091	2.9068	21(16.8%)	3(2.4%)
	25-step ahead	43.7949	53.0116	3.3840	36(28.8%)	4(3.2%)
	125-step ahead	59.0288	76.2116	4.6135	23(18.4%)	2(1.6%)
ARIMA(0,1,0)-EGARCH(1,1) with $\varepsilon_t \sim$ Normal	**1-step ahead**	**12.9024**	**17.9129**	**0.9930**	**1(0.8%)**	**0(0%)**
	2-step ahead	15.6431	21.4087	1.1990	23(18.4%)	9(7.2%)
	3-step ahead	18.1892	24.4896	1.3988	26(20.8%)	9(7.2%)
	4-step ahead	21.8222	28.4818	1.6812	29(23.2%)	6(4.8%)
	5-step ahead	22.4658	28.3566	1.7296	22(17.6%)	7(5.6%)
	7-step ahead	**23.3542**	**29.2521**	**1.7853**	**27(21.6%)**	**5(4%)**
	10-step ahead	32.0150	41.2452	2.4403	30(24%)	8(6.4%)
	15-step ahead	35.0348	42.5039	2.6747	23(18.4%)	5(4%)
	25-step ahead	35.9671	47.1370	2.7217	23(18.4%)	6(4.8%)
	125-step ahead	140.4588	152.7912	10.7086	91(72.8%)	79(63.2%)
ARIMA(0,1,0)-TGARCH(1,1) with $\varepsilon_t \sim$ Normal	**1-step ahead**	**12.9144**	**17.8579**	**0.9943**	**1(0.8%)**	**0(0%)**
	2-step ahead	15.7330	21.2976	1.2083	25(20%)	11(8.8%)
	3-step ahead	18.2081	24.3872	1.4028	30(24%)	7(5.6%)
	4-step ahead	21.5550	28.2882	1.6668	31(24.8%)	6(4.8%)
	5-step ahead	22.7189	28.7512	1.7549	27(21.6%)	5(4%)
	7-step ahead	**24.3465**	**29.8233**	**1.8736**	**25(20%)**	**4(3.2%)**
	10-step ahead	32.1432	39.9784	2.4727	25(20%)	6(4.8%)
	15-step ahead	37.0640	45.2896	2.8587	30(24%)	8(6.4%)
	25-step ahead	42.2841	51.0354	3.2625	42(33.6%)	14(11.2%)
	125-step ahead	49.0475	63.3683	3.8175	29(23.2%)	10(8%)

*Values in parenthesis denote the percentage for the number of data outside the prediction intervals.

Hence, by considering both results of forecast evaluations and prediction intervals, the model of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations is suggested for practical use. The preference of the ARIMA-GARCH model in forecasting the gold prices is supported by the smallest values of normalised AIC and SIC, as well as in line with the principle of parsimony. Yet, it can be said that world daily gold price can be forecasted accurately using ARIMA(0,1,0)-GARCH(1,1) with MAPE statistic values of less than or around 5% which is considered to be relatively good (Girish, 2016). Therefore, by using the fourth procedure of BJ-G, the ARIMA-GARCH model has the best forecasting performance for the daily gold prices as compared to other models.

(a) (b) (c)

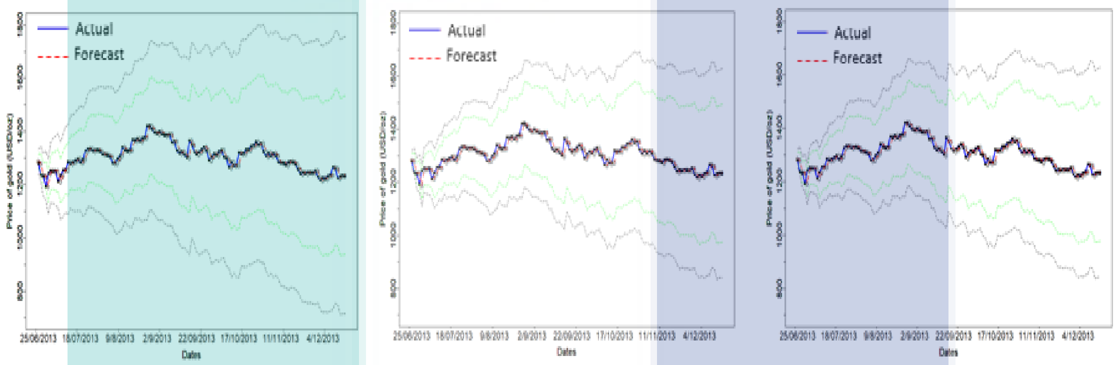


Figure 4.30 Plot of actual data and one-step ahead forecast of gold prices using ARIMA(0,1,0) and (a) GARCH(1,1) with $\varepsilon_t \sim t$, (b) EGARCH(1,1) with $\varepsilon_t \sim$ Normal, (c) TGARCH(1,1) with $\varepsilon_t \sim$ Normal; with 80% (in green dashed line) and 95% (in black dashed line) PIs

(a) (b) (c)

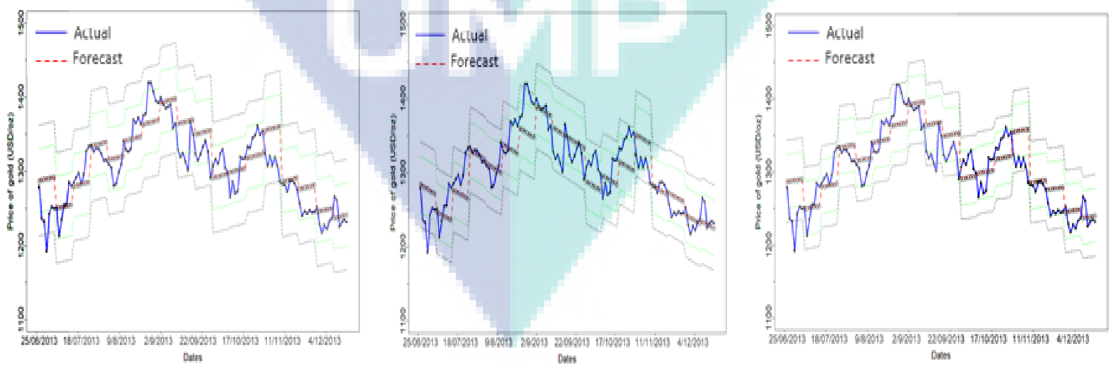


Figure 4.31 Plot of actual data and the seven-step ahead forecast of daily gold price using ARIMA(0,1,0) (a) GARCH(1,1) with $\varepsilon_t \sim t$, (b) EGARCH(1,1) with $\varepsilon_t \sim$ Normal, (c) TGARCH(1,1) with $\varepsilon_t \sim$ Normal; with 80% (in green dashed line) and 95% (in black dashed line) PIs

4.6 Conclusion

This case study evaluates the performance of four proposed procedures of Box-Jenkins – GARCH (or BJ-G) in modelling and forecasting a highly volatile time series data, specifically the world daily gold prices. The first proposed procedure of BJ-G is used to justify and evaluate the performance of BJ-G model using the world gold price which is thoroughly discussed in Section 4.2. Based on the empirical results in Section 4.2, the first proposed procedure of BJ-G shows a promising approach in analysing and forecasting the data series which simultaneously proves that a combination model of BJ-G is reliable in forecasting a highly volatile data. The good performance of BJ-G in forecasting the data series due to its capability to understand the characteristics of a highly volatile data series better without violating the basic assumptions of errors as a Box-Jenkins model does as well as the combination model overcome the weaknesses of GARCH model in dealing with nonstationary series.

Since the first proposed procedure of BJ-G have shown promising approach, then the second proposed procedure of BJ-G is tested on the daily gold price data by emphasizing on the identification of highly volatile characteristic in the data at the early stage before further analysis is conducted since it is focuses on handling the highly volatile time series data specifically. Based on the empirical results of Section 4.3, the second proposed procedure of BJ-G provides a systematic procedure in modelling and forecasting a highly volatile data, being simultaneously practical to be used in determining the optimal number of data while working with any univariate highly volatile data at any frequency. The empirical results of the world daily gold price suggest that the latest 25% (or 1250) is sufficient enough to be employed in the BJ-G model with similar forecasting performance as by using 5000 data.

Given the overall positive results at the one-step ahead forecast in the empirical results in Section 4.3, therefore the case study in Section 4.4 is aimed at assesing the forecasting performance of the BJ-G model for higher horizons by applying the third proposed procedure of BJ-G to the daily world gold price series of Sample 3. Based on the empirical results, the third proposed procedure provides a promising procedure to assess the performance of the BJ-G model, specifically ARIMA(0,1,0)-GARCH(1,1) with *t* innovations, in forecasting up to seven-day ahead gold price. The procedure adds the value of the BJ-G model since it allows the model to follow the nature of the series well and able to explain more about the characteristics of the highly volatile series up to *n*-step-ahead forecast.

In Section 4.5, the fourth proposed procedure of BJ-G is applied to Box-Jenkins with all GARCH-type models under consideration in previous studies on highly volatile data that are GARCH-M, EGARCH, TGARCH and APARCH by employing the same data series in Section 4.4. The results of ARIMA-GARCH model from Section 4.3 and 4.4 are used as benchmark in determining the best Box-Jenkins – GARCH-type model in handling the data series. The empirical results reveal that ARIMA(0,1,0)-GARCH(1,1) with *t* innovations outperforms other ARIMA-GARCH-type models.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter summarises the modelling of highly volatile data using time series model, specifically the combination model of Box-Jenkins with GARCH or BJ-G model. This chapter briefly outlines the major conclusions of this doctoral research and recommendations for further improvement.

5.2 Conclusions

This doctoral research focuses on the modelling and forecasting of the univariate highly volatile time series data with Box-Jenkins as the base model and GARCH-type as the variance model. In this study, four proposed procedures are developed in evaluating the performance of Box Jenkins – GARCH-type model in modelling and forecasting highly volatile time series data: the first procedure is proposed for pre-evaluation performance of Box-Jengkins with standard GARCH (or BJ-G) based on the theoretical Box-Jenkins procedure (refer to Figure 3.5); the second procedure is to emphasize on the identification of highly volatile characteristics in the data at the early stage since it is focuses on handling the highly volatile time series data specifically (refer to Figure 3.14); the third is to evaluate the multistep forecasting performance of BJ-G model (refer to Figure 3.15); and the fourth is the comprehensive procedure of BJ-G to apply the procedure to Box-Jenkins with all GARCH-type models in modelling and forecasting highly volatile data for practical purposes (see Figure 3.16). The steps and methods used for each stage in the proposed procedures are investigated and discussed in detail. All the proposed procedures are illustrated using the daily world gold price.

The contributions of this study can be summarised as follows.

1. A literature review on the models and related studies for highly volatile time series data is identified and it is vital to develop the procedure of reliable model that is able to analyse and forecast data which reflects its pattern and volatility clustering characteristic. Primary criteria for the Box-Jenkins as the base model in the proposed procedure is the model comes from the established forecasting techniques in research practice, its reputation as the benchmark model and the forecasting model as well and its capability to analyse almost any time series data. The GARCH-type model is widely applied to handle volatility in a data series. This study combines the Box-Jenkins and GARCH-type models to achieve optimum forecasting performance for highly volatile time series data.
2. The theoretical Box-Jenkins framework and procedure are used in developing the first proposed procedure of BJ-G. With current practices, Box-Jenkins procedures are quite general and not thorough enough to describe the nature of time series data, specifically for non-constant variance that exist in highly volatile data series. In this study, every method used in each step for every stage specifically in model identification and diagnostic checking stages are thoroughly investigated and explained in the proposed procedure of BJ-G. The preliminary empirical results from the case study in Section 4.2 show that the proposed procedure of BJ-G provides a systematic approach in modelling and forecasting gold price, or in general for highly volatile time series data, as well as justify the capability of BJ-G in handling the data series.
3. In the first proposed procedure of BJ-G, the study suggested new steps and methods, specifically in Stage I (or model identification stage) and Stage III (or diagnostic checking stage). In Stage I, the proposed steps and methods are the step of data descriptive, $k_{\max}$ for ACF and PACF as well as for ADF-test in the data differenced step, the step of prelinearity test, the step of Portmanteau test and the EACF method. In Stage III, the study proposed a system in detecting and handling the heteroscedasticity in the residuals of the Box-Jenkins model which consists of new steps and methods including the DW-test and $k_{\max}$ for LBQ-test in checking on autocorrelation, $k_{\max}$ for ARCH test and LBQ-test for heteroscedasticity test

and introduced the step of linearity test using Terasvirta test in checking the linearity of mean model for BJ-G.

4. The appropriate distribution of innovations of the BJ-G model is also investigated in Stage III of the first proposed procedure of BJ-G to ensure the model fits the data series well. There are five types of widely used innovations ε_t considered in this study that are Normal, *t*, skewed-*t*, GED and Skewed-GED. The steps in considering the appropriate ε_t are discussed thoroughly in the case study of Section 4.2.

5. The promising performance of BJ-G model using the first proposed procedure is lead to the second proposed procedure of BJ-G that focuses on handling the highly volatile time series data specifically, using BJ-G model by emphasizing on the identification of highly volatile characteristics in the data at the early stage. A significant modification is done specifically to Stage I in the second proposed procedure, by proposing the step of preliminary heteroscedasticity test and the identification step of the BJ-G model. This proposed procedure of BJ-G is to accomplish the second objective in the thesis, which is the empirical results are thoroughly discussed in Section 4.3.

6. The empirical results of the world daily gold price in Section 4.3 indicate that the second procedure of BJ-G is more practical than the first propose procedure in modelling highly volatile data using BJ-G model which simultaneously ensures an optimal number of data in dealing with the model to any univariate highly volatile data at any frequency. Hence, it is suggested that the latest 25% of data or 1250 data is sufficient to be employed using BJ-G model with similar forecasting performance as by using all data.

7. This study supports the use of Box-Cox transformation method in the data transformation step to address the issue of nonstationarity in-variance. By applying the method, analyst will choose the appropriate transformation for data series that best suits the nature of data. The importance of the method is proven in the case study of Section 4.3 using the second proposed procedure of BJ-G, where it can be observed that some of the samples considered are not suitable for logarithmic transformation. The implication of using inappropriate transformation data will lead to inaccurate forecasting results.

8. Given the overall positive results at the one-step ahead forecast of BJ-G model using the second proposed procedure, the third proposed procedure of BJ-G is proposed in evaluating the performance of the model at higher horizons. In the proposed procedure, sets of codes in R language are constructed since the software including EViews is only able to provide the results for one-step ahead forecast. The steps are clearly explained in the proposed procedure with consideration of 95% and 80% prediction intervals. This proposed procedure of BJ-G is to cater the third objective in the thesis.

9. The latest 5-year (or 1250 data) world daily gold price is employed to the third proposed procedure in evaluating the multistep forecasting performance of BJ-G model. Based on the empirical results, the model is able to follow the trend of actual data up to seven days ahead, specifically within 95% prediction interval. This indicates that, the third proposed procedure provides a promising procedure in evaluating the forecasting performance of BJ-G model up to *n*-step ahead for a univariate highly volatile time series data.

10. This study contributes to a comprehensive procedure in modelling and forecasting up to *n*-step ahead for highly volatile time series data using Box-Jenkins with all GARCH-type models, as proposed by the fourth proposed procedure of BJ-G. The fourth proposed procedure is a combination of the second proposed procedure of BJ-G and the third proposed procedure of BJ-G. The fourth proposed procedure of BJ-G is not only applicable for Box-Jenkins with standard GARCH model but it can also be applied to Box-Jenkins with all GARCH-type models including GARCH-M, EGARCH, TGARCH and APARCH which heve been widely used in the previous studies. The capability of the fourth proposed procedure of BJ-G in providing an efficient procedure in handling highly volatile data using Box-Jenkins – GARCH-type models is evaluated in Section 4.5. The empirical results in Section 4.5 reveal that ARIMA with standard GARCH, or specifically ARIMA(0,1,0)-GARCH(1,1) with t innovations outperforms other ARIMA-GARCH-type models.

11. This study enhances the capability of standard Box-Jenkins's procedure in handling a highly volatile data by considering GARCH-type model to handle variance without violating the basic assumptions of errors. The proposed procedure of BJ-G adds the value of the Box-Jenkins model since it allows the BJ model with a combination of GARCH to follow the nature of the series well and

understand better the characteristics of the highly volatile series up to *n*-step-ahead forecast.

12. Improving forecasting method is one of the main issues in time series research. Therefore, the comprehensive proposed procedure of BJ-G (or the fourth propose procedure) improves the effectiveness of the forecasting model of Box-Jenkins – GARCH-type in modelling and forecasting a univariate highly volatile time series data. The guidelines given by the proposed procedure of BJ-G package with R codes developed provide a good tool to demonstrate the capability of the model of Box-Jenkins – GARCH-type in handling the highly volatile data systematically and practically.

5.3 Recommendations

There are many possible extensions to enhance the performance of the comprehensive proposed procedure of BJ-G (the fourth proposed procedure). Empirical experience suggests that:

1. The proposed procedure of BJ-G is applicable for any univariate highly volatile time series data such as commodity prices, stock price, temperature data and rainfall data, of any frequencies, i.e. weekly, monthly, quarterly and yearly.
2. The steps and methods used in the proposed procedures of BJ-G are also practical to be used for the Box-Jenkins modelling. It is noted that current practices in the Box-Jenkins modelling are quite general and the steps and methods used are lack of details.
3. The performance of the proposed procedure of BJ-G is suggested to be tested using different ratios of estimate to forecast such as 95:5, 80:20 (as recommended by Hyndman and Athanasopoulus (2014)), 70:30 or 50:50.
4. The proposed procedure of BJ-G is suggested to consider highly volatile time series data with outlier, seasonality effect and missing data.
5. The proposed procedure can be applied to bivariate and multivariate highly volatile time series data. For example, analysis on the comparison between local and world data could be an interesting research in bivariate analysis.

REFERENCES

Ahmad, M. H., Ping, P. Y., Yaziz, S. R., & Miswan, N. H. (2014). A hybrid model for improving Malaysian gold forecast accuracy. *International Journal of Mathematical Analysis*, *8*, 1377–1387.

Ahmad, M. H., Ping, P. Y., Yaziz, S. R., & Miswan, N. H. (2015). Forecasting Malaysian gold using a hybrid of ARIMA and GJR-GARCH models. *Applied Mathematical Sciences*, *9*, 1491–1501.

Ahmed, W. M. A. (2017). On the dynamic interactions between energy and stock markets under structural shifts: Evidence from Egypt. *Research in International Business and Finance*, *42*, 61–74.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Alasali, F., Haben, S., Becerra, V., & Holderbaum, W. (2018). Day-ahead industrial load forecasting for electric RTG cranes. *Journal of Modern Power Systems and Clean Energy*, *6*(2), 223–234.

Alcidi, C., De Grauwe, P., Gros, D., & Oh, Y. (2010). *The Future of the Eurozone and Gold*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1672672

Allen, H., & Taylor, M. P. (1990). Charts, Noise and Fundamentals in the London Foreign Exchange Market. *The Economic Journal*, *100*(400), 49.

Asar, Ö., Ilk, O., & Dag, O. (2017). Estimating Box-Cox power transformation parameter via goodness-of-fit tests. *Communications in Statistics - Simulation and Computation*, *46*(1), 91–105.

Awang, N., Kar Yong, N., & Yin Hoeng, S. (2017). Forecasting ozone concentration levels using Box-Jenkins ARIMA modelling and artificial neural networks: A comparative study. *MATEMATIKA*, *33*(2), 119–130.

Babu, N. C., & Reddy, E. B. (2015). Prediction of selected Indian stock using a partitioning–interpolation based ARIMA–GARCH model. *Applied Computing and Informatics*, *11*(2), 130–143.

Barros, M., & de Medeiros, L. (2017). Electricity spot price forecasting in Brazil using a hybrid neuro-fuzzy system and neural network approach. *International Journal of Energy and Statistics*, *05*(01), 1750004.

Baur, D. G., & Lucey, B. M. (2010). Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold. *Financial Review*, *45*(2), 217–229.

Boland, J. (2008). Time Series Modelling of Solar Radiation. In V. Badescu (Ed.), *Modeling Solar Radiation at the Earth's Surface* (pp. 283–312).

Boland, P. J. (1984). A Biographical Glimpse of William Sealy Gosset. *The American Statistician*, *38*(3), 179–183.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, *26*(2), 211–252.

Box, G. E. P., & Jenkins, G. M. (1968). Some Recent Advances in Forecasting and Control. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *17*(2), 91–109.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control* (Rev. ed). San Francisco: Holden-Day.

Box, G. E. P., Jenkins, G. M., & Bacon, D. W. (1967). *Models for forecasting seasonal and non-seasonal time series* (pp. 271–311).

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Box/Time Series Analysis*.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, *65*(332), 1509–1526.

Box, G. E. P., & Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis: Box/Bayesian*.

Byström, H. N. E. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, *14*(1), 41–55.

Castellano-Méndez, M., González-Manteiga, W., Febrero-Bande, M., Manuel Prada-Sánchez, J., & Lozano-Calderón, R. (2004). Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box–Jenkins and neural networks methods. *Journal of Hydrology*, *296*(1), 38–58.

Chakravarty, S., Mohapatra, P., & Dash, P. K. (2016). Evolutionary extreme learning machine for energy price forecasting. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, *20*(2), 75–96.

Chand, S., Kamal, S., & Ali, I. (2012). Modeling and Volatility Analysis of Share Prices Using ARCH and GARCH Models. *World Applied Sciences Journal*, *19*(1), 77–82.

Chatfield, C. (2001). *Time-series forecasting*. Retrieved from http://www.crcnetbase.com/isbn/9781584880639

Chen, C., Hu, J., Meng, Q., & Zhang, Y. (2011). Short-time traffic flow prediction with ARIMA-GARCH model. *2011 IEEE Intelligent Vehicles Symposium (IV)*, 607–612.

Chen, X., Dong, Z. Y., Meng, K., Xu, Y., Wong, K. P., & Ngan, H. W. (2012). Electricity Price Forecasting With Extreme Learning Machine and Bootstrapping. *IEEE Transactions on Power Systems*, *27*(4), 2055–2062.

Chitsaz, H., Shaker, H., Zareipour, H., Wood, D., & Amjady, N. (2015). Short-term electricity load forecasting of buildings in microgrids. *Energy and Buildings*, *99*, 50–60.

Christodoulos, C., Michalakelis, C., & Varoutas, D. (2010). Forecasting with limited data: Combining ARIMA and diffusion models. *Technological Forecasting and Social Change*, *77*(4), 558–565.

Dalmazo, B. L., Vilela, J. P., & Curado, M. (2017). Performance Analysis of Network Traffic Predictors in the Cloud. *Journal of Network and Systems Management*, *25*(2), 290–320.

Daniels, H. E. (1956). The Approximate Distribution of Serial Correlation Coefficients. *Biometrika*, *43*(1/2), 169–185.

Darekar, A., & Reddy, A. A. A. (2017). Forecasting of Common Paddy Prices in India. *Journal of Rice Research*, *10*(1), 71–75.

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, *22*(3), 443–473.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, *74*(366a), 427–431.

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Dunis, C. L., & Nathani, A. (2007). Quantitative trading of gold and silver using nonlinear models. *Neural Network World*, *17*(2), 93–111.

Durbin, J., & Watson, G. S. (1971). Testing for Serial Correlation in Least Squares Regression. *Biometrika*, *58*(1), 1–19.

Engle, R. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*, *15*(4), 157–168.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007.

Ferenstein, E., & Gasowski, M. (2004). Modelling stock returns with AR-GARCH processes. *SORT*, *28*(1), 55–68.

Fouli, H., Fouli, R., Bashir, B., & Loni, O. A. (2017). Seasonal forecasting of rainfall and runoff volumes in Riyadh region, KSA. *KSCE Journal of Civil Engineering*, 1–11.

Fox, J., Weisberg, S., Adler, D., & Bates, D. (2015). Package ‘car’ (Version 2.1-4).

Francq, C., & Zakoïan, J.-M. (2010). Inconsistency of the MLE and inference based on weighted LS for LARCH models. *Journal of Econometrics*, *159*(1), 151–165.

Freedi, A. A., Shamiri, A., & Isa, Z. (2012). A Study on the Behavior of Volatility in Saudi Arabia Stock Market Using Symmetric and Asymmetric GARCH Models. *Journal of Mathematics and Statistics*, *8*(1), 98–106.

Gaglianone, W. P., & Marins, J. T. M. (2017). Evaluation of exchange rate point and density forecasts: An application to Brazil. *International Journal of Forecasting*, *33*(3), 707–728.

García-Ferrer, A., González-Prieto, E., & Peña, D. (2012). A conditionally heteroskedastic independent factor model with an application to financial stock returns. *International Journal of Forecasting*, *28*(1), 70–93.

Gel, Y. R., Lyubchich, V., & Ahmed, S. E. (2016). Catching Uncertainty of Wind: A Blend of Sieve Bootstrap and Regime Switching Models for Probabilistic Short-Term Forecasting of Wind Speed. In W. K. Li, D. A. Stanford, & H. Yu (Eds.), *Advances in Time Series Methods and Applications* (Vol. 78, pp. 279–293).

Giddy, I. H., & Dufey, G. (1975). The Random Behavior of Flexible Exchange Rates: Implications for Forecasting. *Journal of International Business Studies*, *6*(1), 1–32.

Giller, G. L. (2005). *A Generalized Error Distribution* (SSRN Scholarly Paper No. ID 2265027). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2265027

Girish, G. P. (2016). Spot electricity price forecasting in Indian electricity market using autoregressive-GARCH models. *Energy Strategy Reviews*, *11–12*, 52–57.

Gonçalves, S., & Meddahi, N. (2011). Box–Cox transforms for realized volatility. *Journal of Econometrics*, *160*(1), 129–144.

Granger, C. W. J., & Andersen, A. (1978). On the invertibility of time series models. *Stochastic Processes and Their Applications*, *8*(1), 87–92.

Hammoudeh, S. M., Yuan, Y., McAleer, M., & Thompson, M. A. (2010). Precious metals-exchange rate volatility transmissions and hedging strategies. *International Review of Economics and Finance*, *19*(4), 633–647.

Hammoudeh, S., Malik, F., & McAleer, M. (2011). Risk management of precious metals. *Quarterly Review of Economics and Finance*, *51*(4), 435–441.

Hammoudeh, S., & Yuan, Y. (2008). Metal volatility in presence of oil and interest rate shocks. *Energy Economics*, *30*(2), 606–620.

Hanke, J. E., Reitsch, A. G., & Wichern, D. W. (2001). *Business forecasting* (7th ed). Upper Saddle River, N.J: Prentice Hall.

Harrison, B., & Paton, D. (2004). Transition, the Evolution of Stock Market Efficiency and Entry into EU: The Case of Romania. *Economic Change and Restructuring*, *37*(3–4), 203–223.

Higgins, M. L., & Bera, A. K. (1992). A Class of Nonlinear Arch Models. *International Economic Review*, *33*(1), 137–158.

Ho, S. ., Xie, M., & Goh, T. . (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, *42*(2–4), 371–375.

Hu, J., Wang, J., & Zeng, G. (2013). A hybrid forecasting approach applied to wind speed time series. *Renewable Energy*, *60*, 185–194.

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OText.

Hyndman, R. J., & Athanasopoulos, G. (2017). *Forecasting: principles and practice*. OText.

Hyndman, R. J., Athanasopoulos, G., Razbash, S., Schmidt, D., Zhou, Z., Khan, Y., & Wang, E. (2015). forecast: Forecasting functions for time series and linear models. R package (Version 6(6),7).

Hyndman, R. J., & Kostenko, A. V. (2017). Minimum sample size requirements for seasonal forecasting models. *Foresight: The International Journal of Applied Forecasting*, *6*, 12–15.

Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review / Revue Internationale de Statistique*, *55*(2), 163–172.

Jenkins, G. M. (1961). General Considerations in the Analysis of Spectra. *Technometrics*, *3*(2), 133–166.

Johnston, J., & DiNardo, J. (1997). *Econometric methods* (4th ed). New York: McGraw-Hill.

Jun, W., Lingyu, T., Yuyan, L., & Peng, G. (2017). A weighted EMD-based prediction model based on TOPSIS and feed forward neural network for noised time series. *Knowledge-Based Systems*, *132*, 167–178.

Khan, M. M. A. (2013). Forecasting of Gold Prices (Box Jenkins Approach). *International Journal of Emerging Technology and Advanced Engineering*, *3*(3), 662–670.

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, *37*(1), 479–489.

Konishi, S., & Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Retrieved from //www.springer.com/gp/book/9780387718866

Koopman, S. J., Ooms, M., & Carnero, M. A. (2007). Periodic Seasonal Reg-ARFIMA–GARCH Models for Daily Electricity Spot Prices. *Journal of the American Statistical Association*, *102*(477), 16–27.

Lee, M. H., Sadaei, H. J., & Suhartono. (2013). Improving TAIEX forecasting using fuzzy time series with Box–Cox power transformation. *Journal of Applied Statistics*, *40*(11), 2407–2422.

Liu, G., Tarasick, D. W., Fioletov, V. E., Sioris, C. E., & Rochon, Y. J. (2009). Ozone correlation lengths and measurement uncertainties from analysis of historical ozonesonde data in North America and Europe. *Journal of Geophysical Research*, *114*(D4).

Liu, H., Erdem, E., & Shi, J. (2011). Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. *Applied Energy*, *88*(3), 724–732.

Liu, H., & Shi, J. (2013). Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, *37*, 152–166.

Liu, Heping, & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, *37*, 152–166.

Liu, Heping, Shi, J., & Qu, X. (2013). Empirical investigation on using wind speed volatility to estimate the operation probability and power output of wind turbines. *Energy Conversion and Management*, *67*, 8–17.

Liu, W., Chung, C. Y., & Wen, F. (2014). Multifractal based return interval approach for short-term electricity price volatility risk estimation. *IET Generation, Transmission & Distribution*, *8*(9), 1550–1560.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Loi, T. S. A., & Ng, J. L. (2018). Anticipating electricity prices for future needs – Implications for liberalised retail markets. *Applied Energy*, *212*, 244–264.

Luetkepohl, H., & Xu, F. (2011). Forecasting Annual Inflation with Seasonal Monthly Data: Using Levels versus Logs of the Underlying Price Index. *Journal of Time Series Econometrics*, *3*(1).

MacKinnon, J. G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, *11*(6), 601–618.

McLeod, A. I., & Li, W. K. (1983). Diagnostic Checking Arma Time Series Models Using Squared-Residual Autocorrelations. *Journal of Time Series Analysis*, *4*(4), 269–273.

Michis, A. A. (2015). A wavelet smoothing method to improve conditional sales forecasting. *Journal of the Operational Research Society*, *66*(5), 832–844.

Miswan, N. H., Ping, P. Y., & Ahmad, M. H. (2013). On parameter estimation for malaysian gold prices modelling and forecasting. *International Journal of Mathematical Analysis*, *7*(21–24), 1059–1068.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100.

Naser, H. (2016). Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach. *Energy Economics*, *56*, 75–87.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, *59*(2), 347.

Olivier, J., & Norberg, M. M. (2010). Positively Skewed Data: Revisiting the Box-Cox Power Transformation. *International Journal of Psychological Research*, *3*(1), 68–75.

Ord, J. K., Koehler, A. B., Snyder, R. D., & Hyndman, R. J. (2009). Monitoring processes with changing variances. *International Journal of Forecasting*, *25*(3), 518–525.

Osborne, J. W. (2010). *Improving your data transformations: Applying the Box-Cox transformation*. *15*(12), 1–9.

Panapakidis, I. P. (2016). Application of hybrid computational intelligence models in short-term bus load forecasting. *Expert Systems with Applications*, *54*, 105–120.

Pandey, T. N., Jagadev, A. K., Dehuri, S., & Cho, S.-B. (2019). A review and empirical analysis of neural networks based exchange rate prediction. *Intelligent Decision Technologies*, *12*(4), 423–439.

Pham, H. T., & Yang, B.-S. (2010). Estimation and forecasting of machine health condition using ARMA/GARCH model. *Mechanical Systems and Signal Processing*, *24*(2), 546–558.

Polydoras, G. N., Anagnostopoulos, J. S., & Bergeles, G. C. (1998). Air quality predictions: dispersion model vs Box-Jenkins stochastic models. An implementation and comparison for Athens, Greece. *Applied Thermal Engineering*, *18*(11), 1037–1048.

Proietti, T., & Lütkepohl, H. (2013). Does the Box–Cox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting*, *29*(1), 88–99.

Qadan, M., & Yagil, J. (2012). Fear sentiments and gold price: Testing causality in-mean and in-variance. *Applied Economics Letters*, *19*(4), 363–366.

Quenouille, M. H. (1949). Approximate Tests of Correlation in Time-Series. *Journal of the Royal Statistical Society. Series B (Methodological)*, *11*(1), 68–84.

Ripley, B., Venables, B., Bates, D. M., & Hornik, K. (2017). Package 'MASS' (Version 7.3-49).

Robeson, S. M., & Steyn, D. G. (1990). Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment. Part B. Urban Atmosphere*, *24*(2), 303–312.

Ruppert, D., & Matteson, D. S. (2015). GARCH Models. In *Springer Texts in Statistics. Statistics and Data Analysis for Financial Engineering* (pp. 405–452).

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*(3), 599–607.

Sakia, R. M. (1992). The Box-Cox Transformation Technique: A Review. *The Statistician*, *41*(2), 169.

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

Schwert, G. W. (1989). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, *44*(5), 1115–1153.

Shafiee, S., & Topal, E. (2010). An overview of global gold market and gold price forecasting. *Resources Policy*, *35*(3), 178–189.

Singh, U. P., & Jain, S. (2018). Optimization of neural network for nonlinear discrete time system using modified quaternion firefly algorithm: case study of Indian currency exchange rate prediction. *Soft Computing*, *22*(8), 2667–2681.

Smith, G. (2002). Tests of the random walk hypothesis for London gold prices. *Applied Economics Letters*, *9*(10), 671–674.

Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed). Ames, Iowa: Iowa State University Press.

Sohn, S. Y., & Lim, M. (2007). Hierarchical forecasting based on AR-GARCH model in a coherent structure. *European Journal of Operational Research*, *176*(2), 1033–1040.

Subbotin, M. T. (1923). On the Law of Frequency of Error. *Matematicheskii Sbornik*, *31*(2), 296–301.

Tan, Z., Zhang, J., Wang, J., & Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy*, *87*(11), 3606–3610.

Taneja, K., Ahmad, S., Ahmad, K., & Attri, S. D. (2016). Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach. *Atmospheric Pollution Research*, *7*(4), 585–596.

Teräsvirta, T. (1994). Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, *89*(425), 208–218.

Teräsvirta, T., Lin, C.-F., & Granger, C. W. J. (1993). Power of the Neural Network Linearity Test. *Journal of Time Series Analysis*, *14*(2), 209–220.

Theodossiou, P. (2008). *Financial Data and the Skewed Generalized T Distribution* (SSRN Scholarly Paper No. ID 65037). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=65037

Trück, S., & Liang, K. (2012). Modelling and forecasting volatility in the gold market. *International Journal of Banking and Finance*, *9*(1). Retrieved from http://epublications.bond.edu.au/ijbf/vol9/iss1/3

Tsay, R. S. (2005). *Analysis of Financial Time Series, 2nd ed.* John Wiley & Sons, Inc.

Tsay, R. S. (2013). *An introduction to analysis of financial data with R*. Hoboken, N.J: Wiley.

Tsay, R. S., & Tiao, G. C. (1984). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. *Journal of the American Statistical Association*, *79*(385), 84–96.

Tukey, J. W. (1957). On the Comparative Anatomy of Transformations. *The Annals of Mathematical Statistics*, *28*(3), 602–632.

Tzagkarakis, G., Caicedo-Llano, J., & Dionysopoulos, T. (2015). Sparse modeling of volatile financial time series via low-dimensional patterns over learned dictionaries. *Algorithmic Finance*, *4*(3–4), 139–158.

Verbeek, J. J. (2004). *Mixture models for clustering and dimension reduction*., S.l.

Villegas, M. A., Pedregal, D. J., & Trapero, J. R. (2018). A support vector machine for model selection in demand forecasting applications. *Computers & Industrial Engineering*, *121*, 1–7.

Walid, C., Chaker, A., Masood, O., & Fry, J. (2011). Stock market volatility and exchange rates in emerging countries: A Markov-state switching approach. *Emerging Markets Review*, *12*(3), 272–292.

Wang, B., Huang, H., & Wang, X. (2012). A novel text mining approach to financial time series forecasting. *Neurocomputing*, *83*, 136–145.

Wang, T. Y., & Huang, C. Y. (2007). Improving forecasting performance by employing the Taguchi method. *European Journal of Operational Research*, *176*(2), 1052–1065.

Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K., & Ma, J. (2005). Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes. *Nonlinear Processes in Geophysics*, *12*(1), 55–66.

Weerathunga, H. P. S. ., & Silva, A. T. P. (2018). DRNN-ARIMA Approach to Short-term Trend Forecasting in Forex Market. *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)*, 287–293.

Wei, W. W. S. (2006). *Time series analysis: univariate and multivariate methods* (2nd ed). Boston: Pearson Addison Wesley.

Yang, B., Zhang, W., & Wang, H. (2019). Stock Market Forecasting Using Restricted Gene Expression Programming. *Computational Intelligence and Neuroscience*, *2019*, 1–14.

Yaziz, S. R., Ahmad, M. H., Nian, L. C., & Muhammad, N. (2011). A comparative study on Box-Jenkins and Garch models in forecasting crude oil prices. *Journal of Applied Sciences*, *11*(7), 1129–1135.

Yaziz, S. R., Azizan, N. A., Ahmad, M. H., & Zakaria, R. (2016). Modelling gold price using ARIMA-TGARCH. *Applied Mathematical Sciences*, *10*, 1391–1402.

Yaziz, S. R., Azizan, N. A., Ahmad, M. H., Zakaria, R., Agrawal, M., & Boland, J. (2014). Innovations in the ARIMA-GARCH Modeling in Forecasting Gold Price. *PROCEEDINGS OF THE 10TH IMT-GT INTERNATIONAL CONFERENCE ON MATHEMATICS, STATISTICS AND ITS APPLICATIONS 2014 (ICMSA2014)*, 650–658.

Yaziz, S. R., Azizan, N. A., Zakaria, R., & Ahmad, M. H. (2013, December 1). *The performance of hybrid ARIMA-GARCH modeling in forecasting gold price*. Presented at the 20th International Congress on Modelling and Simulation, Adelaide, Australia.

Yu, Y., Choi, T.-M., & Hui, C.-L. (2012). An Intelligent Quick Prediction Algorithm With Applications in Industrial Control and Loading Problems. *IEEE Transactions on Automation Science and Engineering*, *9*(2), 276–287.

Yule, G. U. (1927). On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans. R. Soc. Lond. A*, *226*(636–646), 267–298.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.

Zhou, B., He, D., & Sun, Z. (2006). Traffic predictability based on ARIMA/GARCH model. *2006 2nd Conference on Next Generation Internet Design and Engineering, NGI 2006*, 200–207.

Živkov, D., Njegić, J., Momčilović, M., & Milenković, I. (2016). Exchange Rate Volatility and Uncovered Interest Rate Parity in the European Emerging Economies. *Prague Economic Papers*, *25*(3), 253–270.

Zivot, E., & Wang, J. (2003). Rolling Analysis of Time Series. In *Modeling Financial Time Series with S-Plus®* (pp. 299–346).

APPENDIX 1
RESEARCH ACHIEVEMENT

PATENT/COPYRIGHT

1. Patent of "Method and System for Forecasting Commodity Prices", **Yaziz, S. R.**, Zakaria, R. Patent Application PI2018000655, Dated 7 May 2018.

2. Copyright of "Forecasting Malaysia Gold Price using Hybrid ARIMA with Symmetric GARCH Modeling (Backward ARIMA-GARCH in Forecasting Gold Price for Malaysia Market)", **Yaziz, S. R.**, Zakaria, R. Azizan, N. A., Ahmad, M. H., Satari, S. Z., Dated 12 May 2015.

PUBLICATIONS (JOURNAL AND INDEXED PROCEEDINGS)

1. **Yaziz, S.R.**, Zakaria, R. and Boland, J., Multistep Forecasting for Highly Volatile Data using New Procedure of Box-Jenkins and GARCH, ASM Science Journal. (Scopus Indexed). Submitted in October 2018 (in review)
2. **Yaziz, S. R.**, Zakaria, R. and Ahmad, M. H. (2017). Determination of sample size for higher volatile data using new framework of Box-Jenkins model with GARCH: A case study on gold price, *IOP Conf. Series: Journal of Physics: Conf. Series 890*, 012161. (Scopus Indexed)
3. **Yaziz, S. R.**, Azizan, N. A., Ahmad, M. H. and Zakaria, R. (2016). Modeling gold price using ARIMA – TGARCH", *Applied Mathematical Sciences, 10(28)*, 1391-1402. (Scopus Indexed) – 3 citations (based on google scholar on 10 May 2019)
4. **Yaziz, S.R.**, Azizan, N.A., Ahmad, M.H., Zakaria, R., Agrawal, M. and Boland, J. (2015). Preliminary analysis on hybrid Box-Jenkins – GARCH modeling in forecasting gold price, *AIP Conference Proceedings 1643*, 289-297. (ISI Indexed)
   - 3 citations (based on google scholar on 10 May 2019)
5. Ahmad, M. H, Pung, Y. P., **Yaziz, S. R**. and Miswan, N. H. (2015). Forecasting Malaysian gold using a hybrid of ARIMA and GJR-GARCH models, *Applied Mathematical Sciences, 9(30)*, 1491-1501. (Scopus Indexed)
6. Ahmad, M. H, Pung, Y. P., **Yaziz, S. R**. and Miswan, N. H. (2014). A hybrid model for improving Malaysian gold forecast accuracy, *International Journal of Mathematical Analysis, 8(28)*, 1377-1387. (Scopus Indexed)
7. **Yaziz, S.R.**, Azizan, N.A., Ahmad, M.H., Zakaria, R., Agrawal, M. and Boland, J. (2014). Innovations in the ARIMA-GARCH modelling in forecasting gold price, *Proceedings of The 10th IMT-GT International Conference on Mathematics, Statistics and Its Applications 2014 (ICMSA2014)*, e-ISBN 978-967-0524-67-2, 650-658.
8. **Yaziz, S. R.**, Azizan, N. A., Zakaria, R. and Ahmad, M. H. (2013). The performance of hybrid ARIMA – GARCH modelling in forecasting gold price, *Proceedings of The 20th International Congress on Modelling and Simulation (MODSIM2013)*, 1201-1207. (ISI Indexed)- **17 citations** (based on google scholar on 10 May 2019)

MANUSCRIPT REVIEWER

1. Reviewer for manuscript (on Oct 2017) entitled "A Comparative Study of Series ARIMA/MLP Hybrid Models for Stock Price Forecasting" for Journal of Statistical Computation and Simulation (IF 0.869, Q3), Publisher: Taylor & Francis.
2. Reviewer for manuscript (on July 2018) entitled "Forecasting Electricity Consumption Using Time Series Model" for The 4th International Conference on the Applications of Science and Mathematics 2018 (SCIEMATHIC2018).
3. Reviewer for manuscript (on July 2018) entitled "Fuzzy time series forecasting model based on frequency density and similarity measure approach" for The 4th International Conference on the Applications of Science and Mathematics 2018 (SCIEMATHIC2018).
4. Reviewer for manuscript (on May 2019) entitled "Different Time Series Models for Forecasting Prices of Coconut Exports in Sri Lanka" for International Conference on Applied & Industrial Mathematics and Statistics 2019 (ICoAIMS 2019).

AWARDS

1. **Silver Medal**, in the ITEX 2018 for "Reliable Gold Price Predictor", Kuala Lumpur.
2. **Gold Medal**, in CITREX 2018 for "Forecasting Gold Price based on Box-Jenkins – GARCH's Procedures", UMP.
3. **Bronze Medal**, in CITREX 2017 for "Determination of Sample Size for Volatile Data using New Box-Jenkins-GARCH Framework: Gold Price Forecasting", UMP.
4. **Bronze Medal**, in MTE 2016 for "Forecasting Gold Price using Hybrid of Backward ARIMA – GARCH for Malaysia Market", Kuala Lumpur.
5. **Gold Medal**, in the CITREX 2015 for "Forecasting Malaysia Gold Price using Hybrid ARIMA with Symmetric GARCH Modeling", UMP.
6. **Bronze Medal**, in the ITEX 2014 for "Modeling Gold Price using Hybrid of Box-Jenkins – GARCH", Kuala Lumpur.
7. **Gold Medal**, in the CITREX 2014 for "Modeling Gold Price using Hybrid of Box-Jenkins – GARCH", UMP.

CONFERENCES/NON-INDEXED PROCEEDINGS

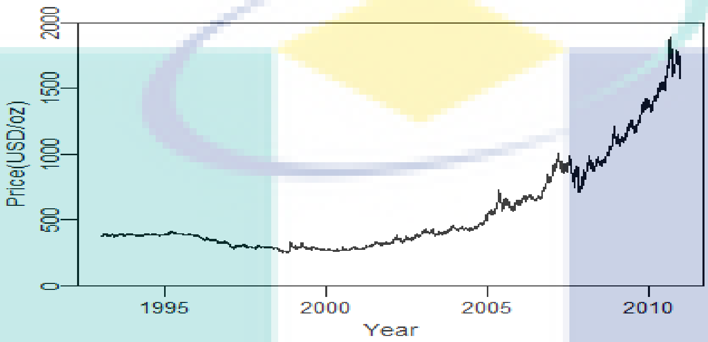
1. **Yaziz, S.R.,** Zakaria, R. and Suhartono, "ARIMA and Symmetric GARCH-type Models for Forecasting Malaysia Gold Price", International Conference on Applied & Industrial Mathematics and Statistics 2019 (ICoAIMS 2019), 8-10 Aug 2017, Kuantan (**Presenter**)

2. **Yaziz, S.R.**, Zakaria, R. and Boland, J., “Multistep Forecasting for Highly Volatile Data using New Procedure of Box-Jenkins and GARCH”, Simposium Kebangsaan Sains Matematik ke-26 (SKSM26), 28-29 Nov 2018, Kota Kinabalu, Malaysia. (**Presenter**)
3. **Yaziz, S. R**., Zakaria, R. and Ahmad, M. H. “Determination of sample size for higher volatile data using new framework of Box-Jenkins model with GARCH: A case study on gold price”, 1st International Conference on Applied & Industrial Mathematics and Statistics 2017 (ICoAIMS 2017), 8-10 Aug 2017, Kuantan, Malaysia. (**Presenter**)
4. **Yaziz, S.R**., Azizan, N.A., Zakaria, R. and Ahmad, M. H. “Modeling Malaysia Gold Price using hybrid of ARIMA and Symmetric GARCH-type models”, 4th International Conference on Computer Engineering & Mathematical Sciences (ICCEMS 2014), 4-5 Dec 2014, Langkawi, Malaysia. (**Presenter**)
5. **Yaziz, S. R.**, Azizan, N. A., Ahmad, M. H. and Zakaria, R. “Modeling Gold Price using ARIMA – TGARCH”, Simposium Kebangsaan Sains Matematik ke-22 (SKSM22), 24-26 Nov 2014, Shah Alam, Malaysia. (**Presenter**)
6. **Yaziz, S.R**., Azizan, N.A., Ahmad, M.H., Zakaria, R., Agrawal, M. and Boland, J. “Innovations in the ARIMA-GARCH modeling in forecasting gold price”, The 10th IMT-GT International Conference on Mathematics, Statistics and Its Applications 2014 (ICMSA2014), 14-16 Oct 2014, Terengganu, Malaysia. (**Presenter**)
7. **Yaziz, S.R.**, Azizan, N.A., Ahmad, M.H., Zakaria, R., Agrawal, M. and Boland, J. “Preliminary Analysis on Hybrid Box-Jenkins – GARCH Modeling in Forecasting Gold Price”, The 2nd ISM International Statistical Conference 2014 with Applications in Sciences and Engineering (ISM-II), 12-14 Aug 2014, Kuantan, Malaysia. (**Presenter**)
8. **Yaziz, S. R.**, Azizan, N. A., Zakaria, R. and Ahmad, M. H. “The performance of hybrid ARIMA – GARCH modeling in forecasting gold price”, The 20th International Congress on Modelling and Simulation (MODSIM2013), 1-6 Dec 2013, Adelaide, Australia. (**Presenter**)

APPENDIX 2
ANALYSIS OF CHAPTER 4 SECTION 4.2

A. Data for Preliminary Study (24 Nov 1993 – 17 Dec 2013)

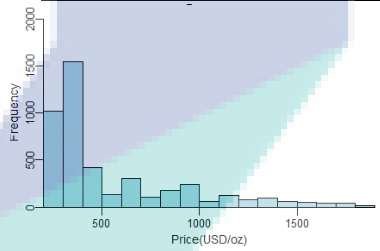
1. Time plot for estimate data (24 Nov 1993 – 20 Dec 2011)



2. Descriptive Statistics for original data

```
nobs          4.500000e+03
NAs           0.000000e+00
Minimum       2.528000e+02
Maximum       1.895000e+03
1. Quartile   3.106000e+02
3. Quartile   6.720000e+02
Mean          5.587877e+02
Median        3.871250e+02
Sum           2.514545e+06
SE Mean       5.418206e+00
LCL Mean      5.481654e+02
UCL Mean      5.694101e+02
Variance      1.321063e+05
Stdev         3.634643e+02
Skewness      1.592544e+00
Kurtosis      1.712147e+00
```

test	test statistic	*p*-value
mean	103.1316	0.0000
skewness	43.6124	0.0000
kurtosis	23.4439	0.0000
normality	2454.4068	0.0000

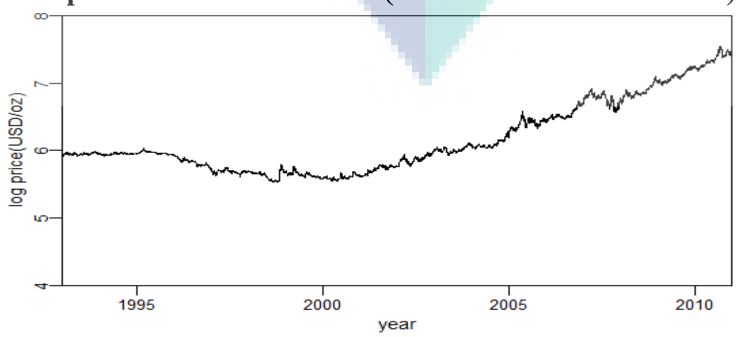


3. Data Stationarity

i. Nonstationary in-Variance: Box-Cox Transformation

```
> lambda <- BoxCox.lambda(dt2, method=c("guerrero","loglik
"), lower=-1, upper=1);lambda # to get the value of lambda
[1] -0.2146852
```

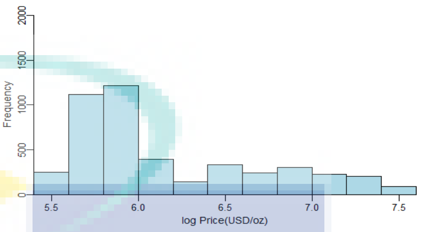
4. Time plot for ln data- estimate (24 Nov 1993 – 20 Dec 2011)



5. Descriptive Statistics for ln data

nobs	4500.000000
NAs	0.000000
Minimum	5.532599
Maximum	7.546974
1. Quartile	5.738506
3. Quartile	6.510258
Mean	6.165210
Median	5.958748
Sum	27743.447032
SE Mean	0.007923
LCL Mean	6.149678
UCL Mean	6.180743
Variance	0.282450
Stdev	0.531460
Skewness	0.895939
Kurtosis	-0.410726

test	test statistic	*p*-value
mean	778.1264	0.0000
skewness	24.5352	0.0000
kurtosis	-5.6237	0.0000
normality	633.8854	0.0000

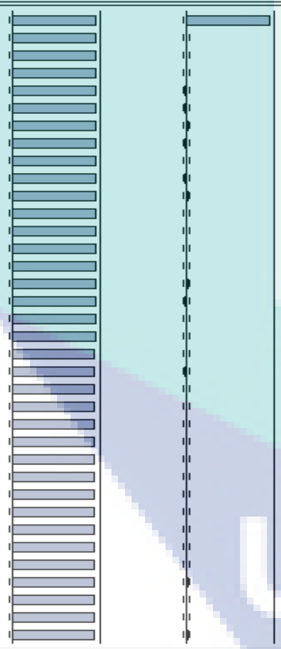


6. Data Stationarity: in- Mean

i. Analysis on ACF and PACF for Log Data

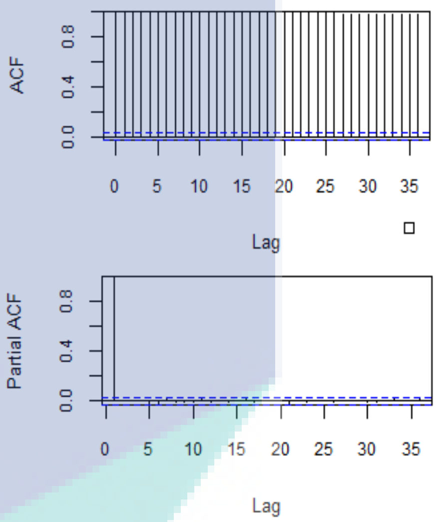
using EViews

Sample: 1 4500
Included observations: 4500



	AC	PAC	Q-Stat	Prob
1	0.999	0.999	4495.7	0.000
2	0.998	0.007	8985.4	0.000
3	0.998	-0.001	13469.	0.000
4	0.997	0.005	17946.	0.000
5	0.996	-0.023	22417.	0.000
6	0.995	-0.023	26882.	0.000
7	0.994	0.013	31339.	0.000
8	0.993	-0.013	35791.	0.000
9	0.993	-0.004	40235.	0.000
10	0.992	-0.011	44672.	0.000
11	0.991	0.012	49103.	0.000
12	0.990	-0.003	53527.	0.000
13	0.989	0.007	57945.	0.000
14	0.988	-0.003	62356.	0.000
15	0.987	-0.001	66760.	0.000
16	0.987	0.015	71157.	0.000
17	0.986	-0.011	75548.	0.000
18	0.985	0.008	79933.	0.000
19	0.984	-0.002	84311.	0.000
20	0.983	0.003	88683.	0.000
21	0.982	-0.015	93048.	0.000
22	0.982	0.003	97406.	0.000
23	0.981	-0.006	101758	0.000
24	0.980	-0.002	106103	0.000
25	0.979	0.007	110442	0.000
26	0.978	-0.003	114774	0.000
27	0.977	0.006	119099	0.000
28	0.976	0.006	123419	0.000
29	0.976	0.004	127731	0.000
30	0.975	-0.004	132038	0.000
31	0.974	-0.002	136338	0.000
32	0.973	0.005	140631	0.000
33	0.972	0.010	144919	0.000
34	0.972	0.004	149200	0.000
35	0.971	0.003	153476	0.000
36	0.970	0.011	157745	0.000

using R



ii. ADF test

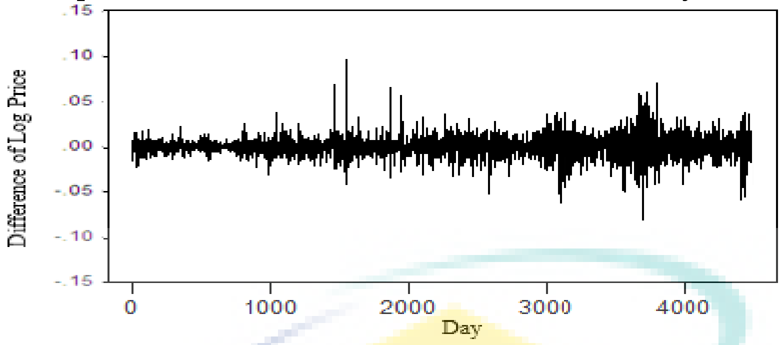
Null Hypothesis: LDT2_ESTIMATE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 31 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.001607	0.9422
Test critical values:	1% level	-3.960092	
	5% level	-3.410811	
	10% level	-3.127201	

*MacKinnon (1996) one-sided p-values.

```
Augmented Dickey-Fuller Test

data:  ldt2
Dickey-Fuller = -1.0016, Lag order = 31, p-value = 0.9385
alternative hypothesis: stationary
```

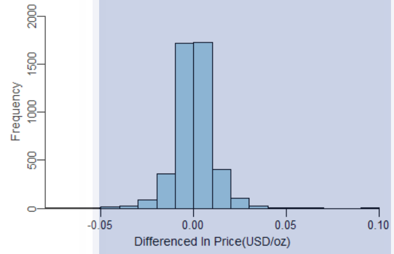
7. **Time plot for the first differenced of ln data** - Stationary data



8. **Descriptive Statistics for the first differenced of ln data- Stationary series**

```
nobs          4499.000000
NAs              0.000000
Minimum         -0.079719
Maximum          0.096416
1. Quartile     -0.004261
3. Quartile      0.005289
Mean             0.000324
Median           0.000178
Sum              1.455641
SE Mean          0.000157
LCL Mean         0.000015
UCL Mean         0.000632
Variance         0.000111
Stdev            0.010553
Skewness         0.040233
Kurtosis         6.914227
```

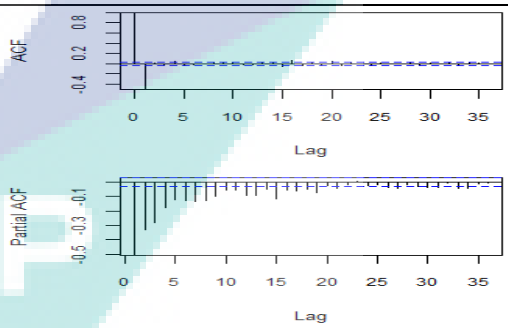
test	test statistic	*p*-value
mean	2.0565	0.0398
skewness	1.1008	0.2710
kurtosis	94.6661	0.0000
normality	8974.3798	0.0000



9. **Data Stationarity for the first differenced of ln data: in- Mean**

i. **Analysis on ACF and PACF for Log Data**

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.2247	0.635
		2	0.006	0.006	0.3854	0.825
		3	0.002	0.002	0.4010	0.940
		4	0.046	0.046	9.7870	0.044
		5	-0.009	-0.009	10.179	0.070
		6	-0.037	-0.038	16.498	0.011
		7	-0.017	-0.018	17.838	0.013
		8	0.005	0.003	17.958	0.022
		9	0.016	0.017	19.070	0.025
		10	-0.006	-0.003	19.245	0.037
		11	-0.042	-0.041	27.100	0.004
		12	-0.032	-0.035	31.745	0.002
		13	0.008	0.005	32.020	0.002
		14	0.010	0.012	32.494	0.003
		15	-0.028	-0.022	35.963	0.002
		16	0.049	0.051	46.834	0.000
		17	-0.001	-0.005	46.843	0.000
		18	0.008	0.003	47.163	0.000
		19	-0.007	-0.005	47.400	0.000
		20	0.029	0.027	51.287	0.000
		21	-0.021	-0.020	53.282	0.000
		22	-0.004	-0.004	53.344	0.000
		23	-0.022	-0.023	55.541	0.000
		24	-0.046	-0.049	64.928	0.000
		25	-0.025	-0.024	67.831	0.000
		26	-0.029	-0.028	71.667	0.000
		27	-0.013	-0.010	72.435	0.000
		28	-0.006	0.001	72.585	0.000
		29	-0.022	-0.023	74.712	0.000
		30	-0.007	-0.011	74.954	0.000
		31	-0.006	-0.004	75.108	0.000
		32	0.003	-0.001	75.147	0.000
		33	-0.031	-0.031	79.460	0.000
		34	0.009	0.006	79.829	0.000
		35	0.011	0.008	80.361	0.000
		36	-0.011	-0.022	80.917	0.000



ii. **ADF test**

Null Hypothesis: D(LDT2_ESTIMATE) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 31 (Fixed)

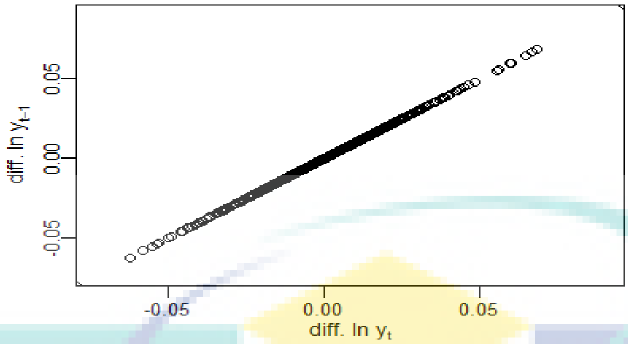
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.94541	0.0000
Test critical values:	1% level	-3.960093	
	5% level	-3.410811	
	10% level	-3.127201	

*MacKinnon (1996) one-sided p-values.

```
Augmented Dickey-Fuller Test

data:  dldt2
Dickey-Fuller = -13.945, Lag order = 31, p-value = 0.01
alternative hypothesis: stationary
```

10. Preliminary of Linearity Test

- Plot of stationary data vs lagged stationary data



11. Portmanteau Test

$k = \ln T$ (Tsay's suggestion):
`X-squared = 19.07, df = 9, p-value = 0.0246`
$k = 10$ (Hyndman's suggestion)
`X-squared = 19.245, df = 10, p-value = 0.03726`
$k = 15$ (Engle's suggestion)
`X-squared = 35.963, df = 15, p-value = 0.00179`

MODELLING GOLD PRICE USING BOX-JENKINS MODEL

12. Stage I: BJ Model Identification

i. Method 1: ACF and PACF

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.2247	0.635
		2	0.006	0.006	0.3854	0.825
		3	0.002	0.002	0.4010	0.940
		4	0.046	0.046	9.7870	0.044
		5	-0.009	-0.009	10.179	0.070
		6	-0.037	-0.038	16.498	0.011
		7	-0.017	-0.018	17.838	0.013
		8	0.005	0.003	17.958	0.022
		9	0.016	0.017	19.070	0.025
		10	-0.006	-0.003	19.245	0.037
		11	-0.042	-0.041	27.100	0.004
		12	-0.032	-0.035	31.745	0.002
		13	0.008	0.005	32.020	0.002
		14	0.010	0.012	32.494	0.003
		15	-0.028	-0.022	35.963	0.002
		16	0.049	0.051	46.834	0.000
		17	-0.001	-0.005	46.843	0.000
		18	0.008	0.003	47.163	0.000
		19	-0.007	-0.005	47.400	0.000
		20	0.029	0.027	51.287	0.000
		21	-0.021	-0.020	53.282	0.000
		22	-0.004	-0.004	53.344	0.000
		23	-0.022	-0.023	55.541	0.000
		24	-0.046	-0.049	64.928	0.000
		25	-0.025	-0.024	67.831	0.000
		26	-0.029	-0.028	71.667	0.000
		27	-0.013	-0.010	72.435	0.000
		28	-0.006	0.001	72.585	0.000
		29	-0.022	-0.023	74.712	0.000
		30	-0.007	-0.011	74.954	0.000
		31	-0.006	-0.004	75.108	0.000
		32	0.003	-0.001	75.147	0.000
		33	-0.031	-0.031	79.460	0.000
		34	0.009	0.006	79.829	0.000
		35	0.011	0.008	80.361	0.000
		36	-0.011	-0.022	80.917	0.000

ii. Method 2: EACF

Data: dldt2

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12
0	o	o	o	x	o	x	o	o	o	o	x	x	o
1	x	o	o	x	o	o	o	o	o	o	o	x	o
2	x	x	o	o	o	o	o	o	o	o	o	o	o
3	x	x	x	o	o	x	o	o	o	o	o	o	o
4	x	x	x	x	o	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o
6	x	x	x	x	x	o	o	o	o	o	o	o	o

13. Stage II: BJ Parameter Estimation

i. Method 1: Ordinary Least Square (OLS)

BJ Models	Estimation Results

ARIMA(0,1,0)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 16:47
Sample (adjusted): 2 4500
Included observations: 4499 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000324	0.000157	2.056480	0.0398

R-squared	0.000000	Mean dependent var	0.000324
Adjusted R-squared	0.000000	S.D. dependent var	0.010553
S.E. of regression	0.010553	Akaike info criterion	-6.264609
Sum squared resid	0.500914	Schwarz criterion	-6.263184
Log likelihood	14093.24	Hannan-Quinn criter.	-6.264107
Durbin-Watson stat	2.013938		

ARIMA(0,1,1)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 16:50
Sample (adjusted): 2 4500
Included observations: 4499 after adjustments
Convergence achieved after 3 iterations
MA Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000324	0.000156	2.070636	0.0384
MA(1)	-0.006983	0.014913	-0.468246	0.6396

R-squared	0.000049	Mean dependent var	0.000324
Adjusted R-squared	-0.000173	S.D. dependent var	0.010553
S.E. of regression	0.010554	Akaike info criterion	-6.264214
Sum squared resid	0.500890	Schwarz criterion	-6.261364
Log likelihood	14093.35	Hannan-Quinn criter.	-6.263209
F-statistic	0.221870	Durbin-Watson stat	1.999891
Prob(F-statistic)	0.637642		

ARIMA(0,1,2)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 16:52
Sample (adjusted): 2 4500
Included observations: 4499 after adjustments
Convergence achieved after 5 iterations
MA Backcast: 0 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000324	0.000157	2.059271	0.0395
MA(1)	-0.007053	0.014915	-0.472900	0.6363
MA(2)	0.005504	0.014915	0.369038	0.7121

R-squared	0.000082	Mean dependent var	0.000324
Adjusted R-squared	-0.000362	S.D. dependent var	0.010553
S.E. of regression	0.010555	Akaike info criterion	-6.263802
Sum squared resid	0.500873	Schwarz criterion	-6.259527
Log likelihood	14093.42	Hannan-Quinn criter.	-6.262296
F-statistic	0.185243	Durbin-Watson stat	1.999781
Prob(F-statistic)	0.830909		

ARIMA(1,1,0)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 16:49
Sample (adjusted): 3 4500
Included observations: 4498 after adjustments
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000323	0.000156	2.066070	0.0389
AR(1)	-0.007066	0.014915	-0.473776	0.6357

R-squared	0.000050	Mean dependent var	0.000323
Adjusted R-squared	-0.000172	S.D. dependent var	0.010554
S.E. of regression	0.010555	Akaike info criterion	-6.264010
Sum squared resid	0.500880	Schwarz criterion	-6.261159
Log likelihood	14089.76	Hannan-Quinn criter.	-6.263005
F-statistic	0.224464	Durbin-Watson stat	1.999704
Prob(F-statistic)	0.635683		

ARIMA(1,1,1)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 16:58
Sample (adjusted): 3 4500
Included observations: 4498 after adjustments
Convergence achieved after 23 iterations
MA Backcast: 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000323	0.000157	2.062202	0.0392
AR(1)	-0.949915	0.048752	-19.48449	0.0000
MA(1)	0.941225	0.052630	17.88396	0.0000

R-squared	0.000769	Mean dependent var	0.000323
Adjusted R-squared	0.000325	S.D. dependent var	0.010554
S.E. of regression	0.010552	Akaike info criterion	-6.264285
Sum squared resid	0.500520	Schwarz criterion	-6.260009
Log likelihood	14091.38	Hannan-Quinn criter.	-6.262778
F-statistic	1.730768	Durbin-Watson stat	1.995018
Prob(F-statistic)	0.177266		

ARIMA(1,1,2)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 17:03
Sample (adjusted): 3 4500
Included observations: 4498 after adjustments
Convergence achieved after 31 iterations
MA Backcast: 1 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000323	0.000157	2.056656	0.0398
AR(1)	-0.948769	0.050704	-18.71180	0.0000
MA(1)	0.942483	0.052860	17.82970	0.0000
MA(2)	0.002639	0.015434	0.171005	0.8642

R-squared	0.000776	Mean dependent var	0.000323
Adjusted R-squared	0.000109	S.D. dependent var	0.010554
S.E. of regression	0.010553	Akaike info criterion	-6.263847
Sum squared resid	0.500517	Schwarz criterion	-6.258145
Log likelihood	14091.39	Hannan-Quinn criter.	-6.261838
F-statistic	1.163617	Durbin-Watson stat	1.999805
Prob(F-statistic)	0.322076		

ARIMA(2,1,0)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 17:07
Sample (adjusted): 4 4500
Included observations: 4497 after adjustments
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000323	0.000157	2.056469	0.0398
AR(1)	-0.007010	0.014918	-0.469925	0.6384
AR(2)	0.005924	0.014918	0.397125	0.6913

R-squared	0.000085	Mean dependent var	0.000323
Adjusted R-squared	-0.000360	S.D. dependent var	0.010555
S.E. of regression	0.010557	Akaike info criterion	-6.263388
Sum squared resid	0.500858	Schwarz criterion	-6.259111
Log likelihood	14086.23	Hannan-Quinn criter.	-6.261881
F-statistic	0.190605	Durbin-Watson stat	1.999508
Prob(F-statistic)	0.826466		

ARIMA(2,1,1)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 17:06
Sample (adjusted): 4 4500
Included observations: 4497 after adjustments
Convergence achieved after 15 iterations
MA Backcast: 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000327	0.000160	2.042679	0.0411
AR(1)	0.512899	0.677528	0.757016	0.4491
AR(2)	0.014383	0.014921	0.963884	0.3352
MA(1)	-0.520036	0.677605	-0.767462	0.4428

R-squared	0.000363	Mean dependent var	0.000323
Adjusted R-squared	-0.000304	S.D. dependent var	0.010555
S.E. of regression	0.010557	Akaike info criterion	-6.263222
Sum squared resid	0.500718	Schwarz criterion	-6.257519
Log likelihood	14086.85	Hannan-Quinn criter.	-6.261213
F-statistic	0.544494	Durbin-Watson stat	1.999788
Prob(F-statistic)	0.651848		

ARIMA(2,1,2)

Dependent Variable: D(LDT2)
Method: Least Squares
Date: 01/26/19 Time: 17:05
Sample (adjusted): 4 4500
Included observations: 4497 after adjustments
Convergence achieved after 25 iterations
MA Backcast: 2 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000325	0.000159	2.046624	0.0408
AR(1)	-0.312166	1.141721	-0.273417	0.7845
AR(2)	0.603937	1.092842	0.552629	0.5805
MA(1)	0.308241	1.147441	0.268634	0.7882
MA(2)	-0.593204	1.089705	-0.544371	0.5862

R-squared	0.000844	Mean dependent var	0.000323
Adjusted R-squared	-0.000045	S.D. dependent var	0.010555
S.E. of regression	0.010555	Akaike info criterion	-6.263258
Sum squared resid	0.500477	Schwarz criterion	-6.256130
Log likelihood	14087.94	Hannan-Quinn criter.	-6.260747
F-statistic	0.949085	Durbin-Watson stat	2.004295
Prob(F-statistic)	0.434377		

ii. **Method 2: Maximum Likelihood Estimator (MLE)**

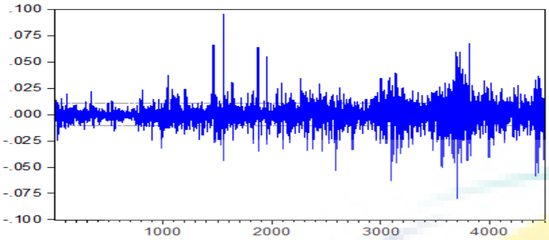
```
> fit1_bj<-Arima(dldt2, order=c(0,0,0),method=c("ML"));fit1_bj
Series: dldt2
ARIMA(0,0,0) with non-zero mean

Coefficients:
       mean
      3e-04
s.e.  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.24
AIC=-28182.48   AICc=-28182.47   BIC=-28169.65
> fit2_bj<-Arima(dldt2, order=c(0,0,1),method=c("ML"));fit2_bj
Series: dldt2
ARIMA(0,0,1) with non-zero mean

Coefficients:
          ma1   mean
      -0.0070  3e-04
s.e.   0.0148  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.35
AIC=-28180.7   AICc=-28180.69   BIC=-28161.46
> fit3_bj<-Arima(dldt2, order=c(0,0,2),method=c("ML"));fit3_bj
Series: dldt2
ARIMA(0,0,2) with non-zero mean

Coefficients:
          ma1     ma2   mean
      -0.0071  0.0055  3e-04
s.e.   0.0149  0.0143  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.42
AIC=-28178.85   AICc=-28178.84   BIC=-28153.2
> fit4_bj<-Arima(dldt2, order=c(1,0,0),method=c("ML"));fit4_bj
Series: dldt2
ARIMA(1,0,0) with non-zero mean

Coefficients:
          ar1   mean
      -0.0071  3e-04
s.e.   0.0149  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.35
AIC=-28180.7   AICc=-28180.7   BIC=-28161.47
> fit5_bj<-Arima(dldt2, order=c(1,0,1),method=c("ML"));fit5_bj
Series: dldt2
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1      ma1   mean
      -0.0036  -0.0034  3e-04
s.e.   2.1684   2.1556  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.35
AIC=-28178.7   AICc=-28178.69   BIC=-28153.05
```

```
> fit6_bj<-Arima(dldt2, order=c(1,0,2),method=c("ML"));fit6_bj
Series: dldt2
ARIMA(1,0,2) with non-zero mean

Coefficients:
          ar1      ma1     ma2   mean
      -0.0035  -0.0035  0.0055  3e-04
s.e.      NaN      NaN  0.0128  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.42
AIC=-28176.85   AICc=-28176.83   BIC=-28144.79
Warning message:
In sqrt(diag(x$var.coef)) : NaNs produced
> fit7_bj<-Arima(dldt2, order=c(2,0,0),method=c("ML"));fit7_bj
Series: dldt2
ARIMA(2,0,0) with non-zero mean

Coefficients:
          ar1     ar2   mean
      -0.0070  0.0059  3e-04
s.e.   0.0149  0.0149  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.43
AIC=-28178.86   AICc=-28178.85   BIC=-28153.21
> fit8_bj<-Arima(dldt2, order=c(2,0,1),method=c("ML"));fit8_bj
Series: dldt2
ARIMA(2,0,1) with non-zero mean

Coefficients:
          ar1     ar2      ma1   mean
      -0.0035  0.0060  -0.0035  3e-04
s.e.      NaN  0.0136      NaN  2e-04

sigma^2 estimated as 0.0001114:  log likelihood=14093.43
AIC=-28176.86   AICc=-28176.85   BIC=-28144.8
Warning message:
In sqrt(diag(x$var.coef)) : NaNs produced
> fit9_bj<-Arima(dldt2, order=c(2,0,2),method=c("ML"));fit9_bj
Series: dldt2
ARIMA(2,0,2) with non-zero mean

Coefficients:
          ar1     ar2      ma1     ma2   mean
      -0.0034  0.0032  -0.0035  0.0027  3e-04
s.e.      NaN     NaN      NaN     NaN  2e-04

sigma^2 estimated as 0.0001115:  log likelihood=14093.43
AIC=-28174.85   AICc=-28174.83   BIC=-28136.38
Warning message:
In sqrt(diag(x$var.coef)) : NaNs produced
```

14. Stage III: BJ Diagnostic Checking

i. ARIMA(0,1,0) using OLS

a. Residual Plot



b. LBQ-statistic Test for Residuals

Correlogram of Residuals

Date: 01/26/19 Time: 18:13
Sample: 2 4500
Included observations: 4499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.2247	0.635
		2	0.006	0.006	0.3854	0.825
		3	0.002	0.002	0.4010	0.940
		4	0.046	0.046	9.7870	0.044
		5	-0.009	-0.009	10.179	0.070
		6	-0.037	-0.038	16.498	0.011
		7	-0.017	-0.018	17.838	0.013
		8	0.005	0.003	17.958	0.022
		9	0.016	0.017	19.070	0.025
		10	-0.006	-0.003	19.245	0.037
		11	-0.042	-0.041	27.100	0.004
		12	-0.032	-0.035	31.745	0.002
		13	0.008	0.005	32.020	0.002
		14	0.010	0.012	32.494	0.003
		15	-0.028	-0.022	35.963	0.002
		16	0.049	0.051	46.834	0.000
		17	-0.001	-0.005	46.843	0.000
		18	0.008	0.003	47.163	0.000
		19	-0.007	-0.005	47.400	0.000
		20	0.029	0.027	51.287	0.000

c. Durbin-Watson Test

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.224509	Prob. F(1,4497)	0.6356
Obs*R-squared	0.224598	Prob. Chi-Square(1)	0.6356

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 01/26/19 Time: 18:11
Sample: 2 4500
Included observations: 4499
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.47E-08	0.000157	-9.31E-05	0.9999
RESID(-1)	-0.007066	0.014911	-0.473876	0.6356

R-squared	0.000050	Mean dependent var	1.58E-17
Adjusted R-squared	0.000050	S.D. dependent var	0.010553
S.E. of regression	0.010553	Akaike info criterion	-6.264214
Sum squared resid	0.500889	Schwarz criterion	-6.261364
Log likelihood	14093.35	Hannan-Quinn criter.	-6.263210
F-statistic	0.224559	Durbin-Watson stat	1.999724
Prob(F-statistic)	0.635611		

d. LBQ-statistic Test on Squared Residuals

Correlogram of Residuals Squared

Date: 01/26/19 Time: 18:17
Sample: 2 4500
Included observations: 4499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.164	0.164	121.09	0.000
		2	0.123	0.098	188.85	0.000
		3	0.164	0.134	309.41	0.000
		4	0.130	0.080	385.09	0.000
		5	0.164	0.115	506.19	0.000
		6	0.134	0.067	587.08	0.000
		7	0.127	0.060	659.79	0.000
		8	0.127	0.054	732.73	0.000
		9	0.098	0.022	776.22	0.000
		10	0.085	0.010	809.15	0.000
		11	0.084	0.012	840.74	0.000
		12	0.154	0.094	948.24	0.000
		13	0.112	0.035	1004.9	0.000
		14	0.092	0.021	1043.0	0.000
		15	0.110	0.037	1097.9	0.000
		16	0.077	0.001	1124.6	0.000
		17	0.150	0.083	1226.7	0.000
		18	0.111	0.027	1282.0	0.000
		19	0.092	0.015	1320.5	0.000
		20	0.130	0.048	1397.3	0.000

e. ARCH Test

Heteroskedasticity Test: ARCH

F-statistic	44.97267	Prob. F(9,4480)	0.0000
Obs*R-squared	372.0444	Prob. Chi-Square(9)	0.0000

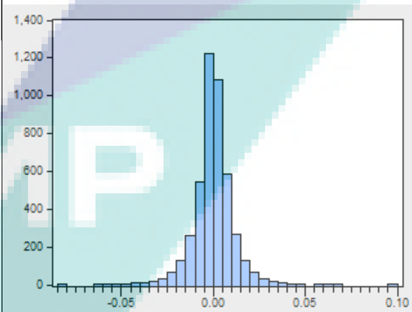
Heteroskedasticity Test: ARCH

F-statistic	40.51851	Prob. F(10,4478)	0.0000
Obs*R-squared	372.4773	Prob. Chi-Square(10)	0.0000

Heteroskedasticity Test: ARCH

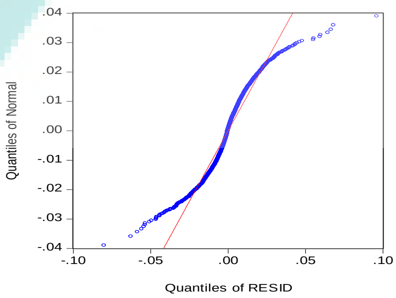
F-statistic	30.88411	Prob. F(15,4468)	0.0000
Obs*R-squared	421.2442	Prob. Chi-Square(15)	0.0000

f. Descriptive statistics for the residuals

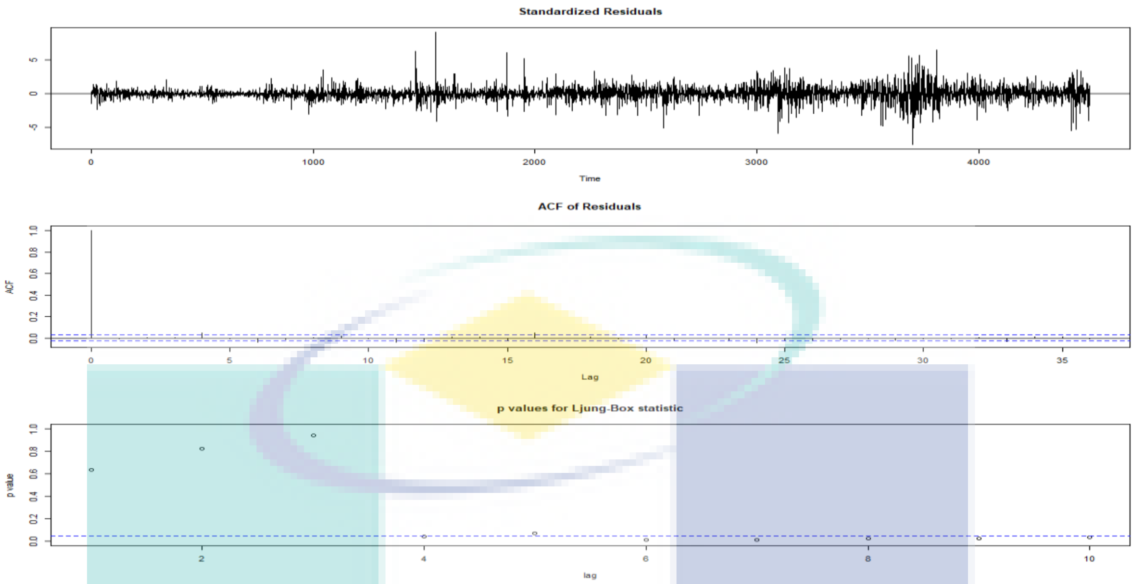


Series: Residuals	
Sample 2 4500	
Observations 4499	
Mean	1.58e-17
Median	-0.000146
Maximum	0.096093
Minimum	-0.080042
Std. Dev.	0.010553
Skewness	0.040247
Kurtosis	9.918636
Jarque-Bera	8974.380
Probability	0.000000

g. Normal QQ-plot

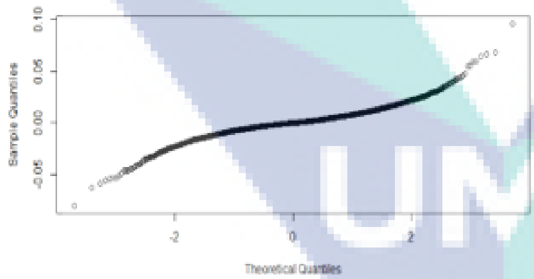


ii. ARIMA(0,1,0) using MLE



Standardized residuals: to examine IID assumption and to detect possible outliers
ACF of Residuals: ideally,should be within the limit of 2 standard errors
P-values: Ljung-Box Q-stat for residuals, P-value>0.05 up to lag 5

```
> basicStats(residuals(fit1_bj))
               residuals.fit1_bj
nobs                 4499.000000
NAs                     0.000000
Minimum                -0.080042
Maximum                 0.096093
1. Quartile            -0.004584
3. Quartile             0.004966
Mean                    0.000000
Median                 -0.000146
Sum                     0.000000
SE Mean                 0.000157
LCL Mean               -0.000308
UCL Mean                0.000308
Variance                0.000111
Stdev                   0.010553
Skewness                0.040233
Kurtosis                6.914227
> qqnorm(residuals(fit1_bj)) #normal
```




```
> Box.test(residuals(fit1_bj),lag=10,type="Ljung") #Ljun

        Box-Ljung test

data:  residuals(fit1_bj)
X-squared = 19.245, df = 10, p-value = 0.03726

> Box.test(((residuals(fit1_bj))^2),lag=10,type="Ljung")

        Box-Ljung test

data:  ((residuals(fit1_bj))^2)
X-squared = 809.15, df = 10, p-value < 2.2e-16

> normalTest(residuals(fit1_bj), method='jb')#Jarque-Ber

Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 8974.3798
  P VALUE:
    Asymptotic p Value: < 2.2e-16
```

B. MODELLING GOLD PRICE USING BOX-JENKINS - GARCH MODEL

1. Stage I: BJ-GARCH Model Identification

i. Mean Model: ARIMA(0,1,0)

ii. Variance Model

ACF and PACF on Squared Residuals

Correlogram of Residuals Squared

Date: 01/26/19 Time: 18:17
Sample: 2 4500
Included observations: 4499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.164	0.164	121.09	0.000
		2	0.123	0.098	188.85	0.000
		3	0.164	0.134	309.41	0.000
		4	0.130	0.080	385.09	0.000
		5	0.164	0.115	506.19	0.000
		6	0.134	0.067	587.08	0.000
		7	0.127	0.060	659.79	0.000
		8	0.127	0.054	732.73	0.000
		9	0.098	0.022	776.22	0.000
		10	0.085	0.010	809.15	0.000
		11	0.084	0.012	840.74	0.000
		12	0.154	0.094	948.24	0.000
		13	0.112	0.035	1004.9	0.000
		14	0.092	0.021	1043.0	0.000
		15	0.110	0.037	1097.9	0.000
		16	0.077	0.001	1124.6	0.000
		17	0.150	0.083	1226.7	0.000
		18	0.111	0.027	1282.0	0.000
		19	0.092	0.015	1320.5	0.000
		20	0.130	0.048	1397.3	0.000

2. Stage II: BJ-GARCH Parameter Estimation (using MLE)

1. ARIMA (0,1,0)-GARCH (1,1)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	4.29E-05	0.000111	0.385310	0.7000
		Variance Equation		
C	1.27E-07	2.97E-08	4.274505	0.0000
RESID(-1)^2	0.042750	0.001711	24.98154	0.0000
GARCH(-1)	0.959363	0.001120	856.8717	0.0000

R-squared	-0.000707	Mean dependent var	0.000324
Adjusted R-squared	-0.000707	S.D. dependent var	0.010553
S.E. of regression	0.010557	Akaike info criterion	-6.589408
Sum squared resid	0.501269	Schwarz criterion	-6.583708
Log likelihood	14826.87	Hannan-Quinn criter.	-6.587400
Durbin-Watson stat	2.012515		

Heteroskedasticity Test: ARCH

F-statistic	1.667833	Prob. F(10,4478)	0.0822
Obs*R-squared	16.65726	Prob. Chi-Square(10)	0.0823

2. ARIMA (0,1,0)-GARCH (1,2)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.79E-05	0.000111	0.520100	0.6030
		Variance Equation		
C	1.75E-07	4.50E-08	3.888283	0.0001
RESID(-1)^2	0.061963	0.005376	11.52598	0.0000
GARCH(-1)	0.456451	0.117310	3.890967	0.0001
GARCH(-2)	0.484601	0.112835	4.294787	0.0000

R-squared	-0.000634	Mean dependent var	0.000324
Adjusted R-squared	-0.000634	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.591689
Sum squared resid	0.501232	Schwarz criterion	-6.584563
Log likelihood	14833.00	Hannan-Quinn criter.	-6.589178
Durbin-Watson stat	2.012662		

Heteroskedasticity Test: ARCH

F-statistic	1.111560	Prob. F(10,4478)	0.3489
Obs*R-squared	11.11532	Prob. Chi-Square(10)	0.3486

3. ARIMA (0,1,0)-GARCH (1,3)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	7.70E-05	0.000110	0.699343	0.4843
Variance Equation				
C	2.54E-07	6.56E-08	3.867270	0.0001
RESID(-1)^2	0.094978	0.005156	18.42236	0.0000
GARCH(-1)	-0.016153	0.058538	-0.275947	0.7826
GARCH(-2)	0.464139	0.060262	7.702047	0.0000
GARCH(-3)	0.461739	0.066419	6.951906	0.0000

R-squared	-0.000546	Mean dependent var	0.000324
Adjusted R-squared	-0.000546	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.593671
Sum squared resid	0.501188	Schwarz criterion	-6.585120
Log likelihood	14838.46	Hannan-Quinn criter.	-6.590658
Durbin-Watson stat	2.012839		

Heteroskedasticity Test: ARCH

F-statistic	0.781379	Prob. F(10,4478)	0.6470
Obs*R-squared	7.819337	Prob. Chi-Square(10)	0.6465

4. ARIMA (0,1,0)-GARCH (1,4)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	6.15E-05	0.000112	0.549955	0.5824
Variance Equation				
C	2.25E-07	6.17E-08	3.648231	0.0003
RESID(-1)^2	0.079755	0.007935	10.05044	0.0000
GARCH(-1)	0.503297	0.121961	4.126717	0.0000
GARCH(-2)	0.083382	0.185763	0.448863	0.6535
GARCH(-3)	0.135360	0.155407	0.871003	0.3838
GARCH(-4)	0.201870	0.090822	2.222700	0.0262

R-squared	-0.000617	Mean dependent var	0.000324
Adjusted R-squared	-0.000617	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.592764
Sum squared resid	0.501223	Schwarz criterion	-6.582789
Log likelihood	14837.42	Hannan-Quinn criter.	-6.589249
Durbin-Watson stat	2.012697		

Heteroskedasticity Test: ARCH

F-statistic	0.921538	Prob. F(10,4478)	0.5119
Obs*R-squared	9.219046	Prob. Chi-Square(10)	0.5114

5. ARIMA (0,1,0)-GARCH (2,1)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	7.61E-05	0.000110	0.689193	0.4907
Variance Equation				
C	9.20E-08	2.40E-08	3.827081	0.0001
RESID(-1)^2	0.112859	0.010820	10.43038	0.0000
RESID(-2)^2	-0.077518	0.010627	-7.294480	0.0000
GARCH(-1)	0.966284	0.001168	827.4891	0.0000

R-squared	-0.000550	Mean dependent var	0.000324
Adjusted R-squared	-0.000550	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.595796
Sum squared resid	0.501190	Schwarz criterion	-6.588670
Log likelihood	14842.24	Hannan-Quinn criter.	-6.593285
Durbin-Watson stat	2.012831		

Heteroskedasticity Test: ARCH

F-statistic	0.693568	Prob. F(10,4478)	0.7314
Obs*R-squared	6.941967	Prob. Chi-Square(10)	0.7309

6. ARIMA (0,1,0)-GARCH (2,2)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.73E-05	0.000111	0.514690	0.6068
Variance Equation				
C	1.16E-08	4.99E-09	2.330626	0.0198
RESID(-1)^2	0.096978	0.007939	12.21587	0.0000
RESID(-2)^2	-0.091608	0.007359	-12.44903	0.0000
GARCH(-1)	1.748431	0.038313	45.63544	0.0000
GARCH(-2)	-0.753558	0.037066	-20.33018	0.0000

R-squared	-0.000637	Mean dependent var	0.000324
Adjusted R-squared	-0.000637	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.600711
Sum squared resid	0.501233	Schwarz criterion	-6.592160
Log likelihood	14854.30	Hannan-Quinn criter.	-6.597698
Durbin-Watson stat	2.012656		

Heteroskedasticity Test: ARCH

F-statistic	0.427855	Prob. F(10,4478)	0.9338
Obs*R-squared	4.284965	Prob. Chi-Square(10)	0.9336

7. ARIMA (0,1,0)-GARCH (2,3)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.96E-05	0.000112	0.533460	0.5937
Variance Equation				
C	1.15E-08	5.13E-09	2.239985	0.0251
RESID(-1)^2	0.099911	0.011374	8.784145	0.0000
RESID(-2)^2	-0.094571	0.010744	-8.802133	0.0000
GARCH(-1)	1.684175	0.133115	12.65203	0.0000
GARCH(-2)	-0.627320	0.243279	-2.578601	0.0099
GARCH(-3)	-0.061955	0.113358	-0.546542	0.5847

R-squared	-0.000626	Mean dependent var	0.000324
Adjusted R-squared	-0.000626	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.600285
Sum squared resid	0.501228	Schwarz criterion	-6.590310
Log likelihood	14854.34	Hannan-Quinn criter.	-6.596770
Durbin-Watson stat	2.012678		

Heteroskedasticity Test: ARCH

F-statistic	0.424856	Prob. F(10,4478)	0.9353
Obs*R-squared	4.254957	Prob. Chi-Square(10)	0.9351

8. ARIMA (0,1,0)-GARCH (2,4)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.96E-05	0.000112	0.532609	0.5943
Variance Equation				
C	1.14E-08	5.23E-09	2.174552	0.0297
RESID(-1)^2	0.100134	0.011412	8.774352	0.0000
RESID(-2)^2	-0.094835	0.010762	-8.811991	0.0000
GARCH(-1)	1.687334	0.167754	10.05840	0.0000
GARCH(-2)	-0.643629	0.406315	-1.584064	0.1132
GARCH(-3)	-0.037894	0.376036	-0.100772	0.9197
GARCH(-4)	-0.010871	0.134732	-0.080686	0.9357

R-squared	-0.000626	Mean dependent var	0.000324
Adjusted R-squared	-0.000626	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.599842
Sum squared resid	0.501228	Schwarz criterion	-6.588441
Log likelihood	14854.34	Hannan-Quinn criter.	-6.595825
Durbin-Watson stat	2.012679		

Heteroskedasticity Test: ARCH

F-statistic	0.426483	Prob. F(10,4478)	0.9345
Obs*R-squared	4.271239	Prob. Chi-Square(10)	0.9343

9. ARIMA (0,1,0)-GARCH (3,1)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.95E-05	0.000111	0.536371	0.5917
		Variance Equation		
C	8.47E-08	2.24E-08	3.777134	0.0002
RESID(-1)^2	0.105819	0.011301	9.363299	0.0000
RESID(-2)^2	-0.030033	0.017271	-1.738951	0.0820
RESID(-3)^2	-0.042619	0.012043	-3.538826	0.0004
GARCH(-1)	0.968318	0.001249	775.3966	0.0000

R-squared	-0.000626	Mean dependent var	0.000324
Adjusted R-squared	-0.000626	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.596428
Sum squared resid	0.501228	Schwarz criterion	-6.587877
Log likelihood	14844.66	Hannan-Quinn criter.	-6.593415
Durbin-Watson stat	2.012678		

Heteroskedasticity Test: ARCH

F-statistic	0.721362	Prob. F(10,4478)	0.7051
Obs*R-squared	7.219713	Prob. Chi-Square(10)	0.7046

10. ARIMA (0,1,0)-GARCH (3,2)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000188	0.000108	1.746501	0.0807
		Variance Equation		
C	8.01E-08	4.75E-08	1.686576	0.0917
RESID(-1)^2	0.103189	0.010555	9.776752	0.0000
RESID(-2)^2	-0.084712	0.057434	-1.474954	0.1402
RESID(-3)^2	0.006499	0.035485	0.183142	0.8547
GARCH(-1)	1.360870	0.471266	2.887691	0.0039
GARCH(-2)	-0.384646	0.452087	-0.850823	0.3949

R-squared	-0.000164	Mean dependent var	0.000324
Adjusted R-squared	-0.000164	S.D. dependent var	0.010553
S.E. of regression	0.010554	Akaike info criterion	-6.594743
Sum squared resid	0.500997	Schwarz criterion	-6.584767
Log likelihood	14841.87	Hannan-Quinn criter.	-6.591228
Durbin-Watson stat	2.013608		

Heteroskedasticity Test: ARCH

F-statistic	0.440616	Prob. F(10,4478)	0.9271
Obs*R-squared	4.412639	Prob. Chi-Square(10)	0.9268

11. ARIMA (0,1,0)-GARCH (3,3)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.64E-05	0.000112	0.504780	0.6137
		Variance Equation		
C	2.11E-09	7.35E-09	0.287209	0.7740
RESID(-1)^2	0.098745	0.009985	9.889448	0.0000
RESID(-2)^2	-0.171865	0.060149	-2.857309	0.0043
RESID(-3)^2	0.074128	0.058271	1.272133	0.2033
GARCH(-1)	2.522538	0.695298	3.627998	0.0003
GARCH(-2)	-2.099647	1.229494	-1.707732	0.0877
GARCH(-3)	0.576147	0.537481	1.071940	0.2837

R-squared	-0.000641	Mean dependent var	0.000324
Adjusted R-squared	-0.000641	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.599857
Sum squared resid	0.501235	Schwarz criterion	-6.588456
Log likelihood	14854.38	Hannan-Quinn criter.	-6.595840
Durbin-Watson stat	2.012648		

Heteroskedasticity Test: ARCH

F-statistic	0.428127	Prob. F(10,4478)	0.9337
Obs*R-squared	4.287683	Prob. Chi-Square(10)	0.9334

12. ARIMA (0,1,0)-GARCH (3,4)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	6.74E-05	0.000110	0.614601	0.5388
		Variance Equation		
C	1.68E-08	9.38E-09	1.792725	0.0730
RESID(-1)^2	0.117103	0.009837	11.90384	0.0000
RESID(-2)^2	0.018517	0.006547	2.828555	0.0047
RESID(-3)^2	-0.125924	0.008448	-14.90547	0.0000
GARCH(-1)	0.243932	0.029418	8.291940	0.0000
GARCH(-2)	1.346024	0.045136	29.82157	0.0000
GARCH(-3)	0.029908	0.038149	0.783977	0.4331
GARCH(-4)	-0.629064	0.039302	-16.00576	0.0000

R-squared	-0.000589	Mean dependent var	0.000324
Adjusted R-squared	-0.000589	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.604844
Sum squared resid	0.501209	Schwarz criterion	-6.592018
Log likelihood	14866.60	Hannan-Quinn criter.	-6.600324
Durbin-Watson stat	2.012753		

Heteroskedasticity Test: ARCH

F-statistic	0.236453	Prob. F(10,4478)	0.9927
Obs*R-squared	2.369086	Prob. Chi-Square(10)	0.9926

13. ARIMA (0,1,0)-GARCH (4,1)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.52E-05	0.000111	0.497786	0.6186
		Variance Equation		
C	6.98E-08	2.19E-08	3.183778	0.0015
RESID(-1)^2	0.103129	0.011667	8.839029	0.0000
RESID(-2)^2	-0.026468	0.018276	-1.448232	0.1476
RESID(-3)^2	-0.006812	0.020200	-0.337213	0.7360
RESID(-4)^2	-0.038733	0.012351	-3.136068	0.0017
GARCH(-1)	0.970404	0.001319	735.4973	0.0000

R-squared	-0.000647	Mean dependent var	0.000324
Adjusted R-squared	-0.000647	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.597379
Sum squared resid	0.501238	Schwarz criterion	-6.587403
Log likelihood	14847.80	Hannan-Quinn criter.	-6.593863
Durbin-Watson stat	2.012637		

Heteroskedasticity Test: ARCH

F-statistic	0.607839	Prob. F(10,4478)	0.8085
Obs*R-squared	6.085062	Prob. Chi-Square(10)	0.8081

14. ARIMA (0,1,0)-GARCH (4,2)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.98E-05	0.000112	0.533948	0.5934
		Variance Equation		
C	1.06E-08	5.45E-09	1.944813	0.0518
RESID(-1)^2	0.100013	0.011409	8.766137	0.0000
RESID(-2)^2	-0.102830	0.024943	-4.122579	0.0000
RESID(-3)^2	0.008271	0.028516	0.290059	0.7718
RESID(-4)^2	-0.000516	0.016762	-0.030792	0.9754
GARCH(-1)	1.764809	0.057882	30.49001	0.0000
GARCH(-2)	-0.769524	0.056253	-13.67961	0.0000

R-squared	-0.000625	Mean dependent var	0.000324
Adjusted R-squared	-0.000625	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.599841
Sum squared resid	0.501227	Schwarz criterion	-6.588440
Log likelihood	14854.34	Hannan-Quinn criter.	-6.595824
Durbin-Watson stat	2.012680		

Heteroskedasticity Test: ARCH

F-statistic	0.424816	Prob. F(10,4478)	0.9354
Obs*R-squared	4.254564	Prob. Chi-Square(10)	0.9351

15. ARIMA (0,1,0)-GARCH (4,3)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	7.13E-05	0.000112	0.639126	0.5227
Variance Equation				
C	1.75E-08	9.74E-09	1.796153	0.0725
RESID(-1)^2	0.099511	0.011185	8.896539	0.0000
RESID(-2)^2	-0.023654	0.047742	-0.495451	0.6203
RESID(-3)^2	-0.082072	0.047078	-1.743302	0.0813
RESID(-4)^2	0.013871	0.013966	0.993173	0.3206
GARCH(-1)	0.998083	0.527577	1.891825	0.0585
GARCH(-2)	0.623112	0.937111	0.664929	0.5061
GARCH(-3)	-0.628540	0.414179	-1.517558	0.1291

R-squared	-0.000571	Mean dependent var	0.000324
Adjusted R-squared	-0.000571	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.599858
Sum squared resid	0.501201	Schwarz criterion	-6.587032
Log likelihood	14855.38	Hannan-Quinn criter.	-6.595339
Durbin-Watson stat	2.012788		

Heteroskedasticity Test: ARCH

F-statistic	0.410686	Prob. F(10,4478)	0.9423
Obs*R-squared	4.113174	Prob. Chi-Square(10)	0.9421

16. ARIMA (0,1,0)-GARCH (4,4)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	3.21E-05	0.000110	0.293118	0.7694
Variance Equation				
C	3.86E-08	1.71E-08	2.265480	0.0235
RESID(-1)^2	0.094346	0.007786	12.11670	0.0000
RESID(-2)^2	0.090065	0.008967	10.04460	0.0000
RESID(-3)^2	-0.088451	0.006248	-14.15764	0.0000
RESID(-4)^2	-0.077605	0.008470	-9.161910	0.0000
GARCH(-1)	-0.081088	0.035211	-2.302892	0.0213
GARCH(-2)	1.529531	0.032893	46.49949	0.0000
GARCH(-3)	0.276761	0.029289	9.449232	0.0000
GARCH(-4)	-0.742764	0.034858	-21.30850	0.0000

R-squared	-0.000763	Mean dependent var	0.000324
Adjusted R-squared	-0.000763	S.D. dependent var	0.010553
S.E. of regression	0.010557	Akaike info criterion	-6.611840
Sum squared resid	0.501296	Schwarz criterion	-6.597589
Log likelihood	14883.33	Hannan-Quinn criter.	-6.606818
Durbin-Watson stat	2.012403		

Heteroskedasticity Test: ARCH

F-statistic	0.404963	Prob. F(10,4478)	0.9450
Obs*R-squared	4.055907	Prob. Chi-Square(10)	0.9448

17. ARIMA (0,1,0)-GARCH (5,1)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.95E-05	0.000111	0.537143	0.5912
Variance Equation				
C	6.28E-08	2.04E-08	3.070882	0.0021
RESID(-1)^2	0.101598	0.011511	8.826466	0.0000
RESID(-2)^2	-0.024978	0.018082	-1.381375	0.1672
RESID(-3)^2	-0.009872	0.020071	-0.491867	0.6228
RESID(-4)^2	-0.001748	0.016434	-0.106368	0.9153
RESID(-5)^2	-0.036382	0.011367	-3.200722	0.0014
GARCH(-1)	0.972723	0.001446	672.8030	0.0000

R-squared	-0.000626	Mean dependent var	0.000324
Adjusted R-squared	-0.000626	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.598479
Sum squared resid	0.501228	Schwarz criterion	-6.587078
Log likelihood	14851.28	Hannan-Quinn criter.	-6.594462
Durbin-Watson stat	2.012678		

Heteroskedasticity Test: ARCH

F-statistic	0.409065	Prob. F(10,4478)	0.9431
Obs*R-squared	4.096960	Prob. Chi-Square(10)	0.9429

18. ARIMA (0,1,0)-GARCH (5,2)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.61E-05	0.000111	0.505809	0.6130
Variance Equation				
C	9.43E-08	3.64E-08	2.587537	0.0097
RESID(-1)^2	0.106461	0.011895	8.949989	0.0000
RESID(-2)^2	0.043952	0.015232	2.885533	0.0039
RESID(-3)^2	-0.029673	0.015827	-1.874871	0.0608
RESID(-4)^2	-0.001594	0.012662	-0.125876	0.8998
RESID(-5)^2	-0.066918	0.011062	-6.049140	0.0000
GARCH(-1)	0.231661	0.081856	2.830106	0.0047
GARCH(-2)	0.718965	0.078628	9.143863	0.0000

R-squared	-0.000643	Mean dependent var	0.000324
Adjusted R-squared	-0.000643	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.601034
Sum squared resid	0.501236	Schwarz criterion	-6.588208
Log likelihood	14858.03	Hannan-Quinn criter.	-6.596514
Durbin-Watson stat	2.012645		

Heteroskedasticity Test: ARCH

F-statistic	0.278569	Prob. F(10,4478)	0.9860
Obs*R-squared	2.790796	Prob. Chi-Square(10)	0.9859

19. ARIMA (0,1,0)-GARCH (5,3)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	4.95E-05	0.000111	0.446175	0.6555
Variance Equation				
C	1.83E-07	6.53E-08	2.803722	0.0051
RESID(-1)^2	0.109972	0.012116	9.076771	0.0000
RESID(-2)^2	0.106725	0.017541	6.084435	0.0000
RESID(-3)^2	0.004034	0.014159	0.284901	0.7757
RESID(-4)^2	-0.052046	0.014164	-3.674524	0.0002
RESID(-5)^2	-0.074273	0.009585	-7.748923	0.0000
GARCH(-1)	-0.349688	0.101221	-3.454712	0.0006
GARCH(-2)	0.708149	0.044333	15.97331	0.0000
GARCH(-3)	0.552277	0.095432	5.787111	0.0000

R-squared	-0.000675	Mean dependent var	0.000324
Adjusted R-squared	-0.000675	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.601953
Sum squared resid	0.501252	Schwarz criterion	-6.587702
Log likelihood	14861.09	Hannan-Quinn criter.	-6.596932
Durbin-Watson stat	2.012581		

Heteroskedasticity Test: ARCH

F-statistic	0.320696	Prob. F(10,4478)	0.9761
Obs*R-squared	3.212541	Prob. Chi-Square(10)	0.9760

20. ARIMA (0,1,0)-GARCH (5,4)

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	4.57E-05	0.000112	0.409245	0.6824
Variance Equation				
C	2.18E-08	1.13E-08	1.924034	0.0544
RESID(-1)^2	0.101760	0.011995	8.483355	0.0000
RESID(-2)^2	-0.020465	0.016672	-1.227502	0.2196
RESID(-3)^2	0.021071	0.007426	2.837639	0.0045
RESID(-4)^2	-0.121954	0.011561	-10.54834	0.0000
RESID(-5)^2	0.030364	0.014249	2.130987	0.0331
GARCH(-1)	0.813762	0.039916	20.38667	0.0000
GARCH(-2)	0.079904	0.028054	2.848164	0.0044
GARCH(-3)	0.837738	0.034947	23.97151	0.0000
GARCH(-4)	-0.741660	0.041209	-17.99758	0.0000

R-squared	-0.000693	Mean dependent var	0.000324
Adjusted R-squared	-0.000693	S.D. dependent var	0.010553
S.E. of regression	0.010557	Akaike info criterion	-6.602141
Sum squared resid	0.501262	Schwarz criterion	-6.586465
Log likelihood	14862.52	Hannan-Quinn criter.	-6.596617
Durbin-Watson stat	2.012543		

Heteroskedasticity Test: ARCH

F-statistic	0.361853	Prob. F(10,4478)	0.9629
Obs*R-squared	3.624491	Prob. Chi-Square(10)	0.9627

3. Stage III: BJ Diagnostic Checking for ARIMA(0,1,0)-GARCH(1,1)

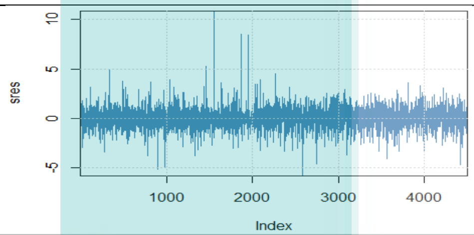
i. ARIMA (0,1,0)-GARCH (1,1) with Normal distribution

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 4500

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	4.29E-05	0.000111	0.385310	0.7000
		Variance Equation		
C	1.27E-07	2.97E-08	4.274505	0.0000
RESID(-1)^2	0.042750	0.001711	24.98154	0.0000
GARCH(-1)	0.959363	0.001120	856.8717	0.0000

R-squared	-0.000707	Mean dependent var	0.000324
Adjusted R-squared	-0.000707	S.D. dependent var	0.010553
S.E. of regression	0.010557	Akaike info criterion	-6.589408
Sum squared resid	0.501269	Schwarz criterion	-6.583708
Log likelihood	14826.87	Hannan-Quinn criter.	-6.587400
Durbin-Watson stat	2.012515		

```
Teraesvirta Neural Network Test

data:  dlnx2
F = 4.3915, df1 = 2, df2 = 4496, p-value = 0.01243
```



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 02:07
Sample: 2 4500
Included observations: 4499

	AC	PAC	Q-Stat	Prob
1	0.011	0.011	0.5292	0.467
2	0.026	0.026	3.6261	0.163
3	0.007	0.007	3.8595	0.277
4	0.002	0.001	3.8826	0.422
5	-0.009	-0.010	4.2832	0.509
6	-0.008	-0.008	4.5855	0.598
7	-0.022	-0.022	6.8012	0.450
8	0.015	0.016	7.8226	0.451
9	0.008	0.009	8.1156	0.523
10	0.015	0.014	9.0971	0.523
11	-0.023	-0.024	11.387	0.411
12	-0.032	-0.033	15.931	0.194
13	-0.003	-0.002	15.981	0.250
14	0.003	0.005	16.019	0.312
15	-0.017	-0.016	17.383	0.297
16	0.033	0.033	22.169	0.138
17	0.015	0.014	23.155	0.144
18	0.014	0.010	23.983	0.156
19	0.014	0.011	24.850	0.166
20	0.028	0.028	28.412	0.100

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 02:10
Sample: 2 4500
Included observations: 4499

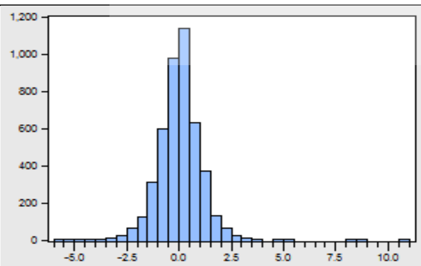
	AC	PAC	Q-Stat	Prob
1	0.050	0.050	11.449	0.001
2	0.011	0.008	11.971	0.003
3	0.018	0.017	13.367	0.004
4	0.015	0.013	14.357	0.006
5	-0.005	-0.007	14.467	0.013
6	0.013	0.013	15.206	0.019
7	-0.012	-0.013	15.818	0.027
8	0.002	0.003	15.844	0.045
9	-0.007	-0.007	16.054	0.066
10	-0.013	-0.012	16.821	0.078
11	-0.012	-0.010	17.473	0.095
12	-0.008	-0.007	17.775	0.123
13	-0.008	-0.006	18.091	0.154
14	-0.005	-0.004	18.198	0.198
15	-0.009	-0.007	18.530	0.236
16	-0.022	-0.021	20.750	0.188
17	-0.011	-0.009	21.308	0.213
18	-0.003	-0.001	21.345	0.262
19	0.002	0.003	21.358	0.317
20	0.009	0.009	21.718	0.356

Heteroskedasticity Test: ARCH

F-statistic	1.668213	Prob. F(10,4478)	0.0821
Obs*R-squared	16.66104	Prob. Chi-Square(10)	0.0822

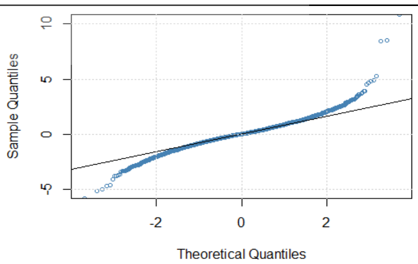
Heteroskedasticity Test: ARCH

F-statistic	1.191793	Prob. F(15,4468)	0.2697
Obs*R-squared	17.86942	Prob. Chi-Square(15)	0.2696



Series: Standardized Residuals
Sample 2 4500
Observations 4499

Mean	0.035292
Median	0.022435
Maximum	10.80213
Minimum	-5.757208
Std. Dev.	1.000114
Skewness	0.466439
Kurtosis	9.951453
Jarque-Bera	9221.630
Probability	0.000000



```
> # estimate model ARIMA(1,1,1)-GARCH(1,1) with cond.dist=normal
> fit21<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt2, con
d.dist="norm", include.mean = TRUE,trace=F) # for model with constant
> sum21<-summary(fit21)
Title: GARCH Modelling
Call: garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt2, cond.dist = "norm",inc
lude.mean = TRUE, trace = F)
Mean and Variance Equation: data ~ arma(0, 0) + garch(1, 1)
 [data = dldt2]
Conditional Distribution: norm
Coefficient(s):
        mu        omega       alpha1        beta1
4.4334e-05  1.2762e-07  4.2918e-02  9.5914e-01
Std. Errors:based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     4.433e-05   1.072e-04    0.414 0.679161
omega  1.276e-07   3.799e-08    3.359 0.000782 ***
alpha1 4.292e-02   3.787e-03   11.332  < 2e-16 ***
beta1  9.591e-01   3.206e-03  299.142  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Log Likelihood:
 14826.16    normalized:  3.295435
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  9453.627   0
 Shapiro-Wilk Test  R    W      0.9546877 0
 Ljung-Box Test     R    Q(10)  9.327292  0.5013495
 Ljung-Box Test     R    Q(15)  17.69191  0.2792092
 Ljung-Box Test     R    Q(20)  29.1451   0.08494138
 Ljung-Box Test     R^2  Q(10)  16.81045  0.07866499
 Ljung-Box Test     R^2  Q(15)  18.55147  0.2347828
 Ljung-Box Test     R^2  Q(20)  21.85274  0.3485547
 LM Arch Test       R    TR^2   17.17721  0.143052

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.589091 -6.583391 -6.589093 -6.587083

> #to calculate Durbin-Watson Test: Method 2
> x21<-residuals(fit21,standardize=TRUE)
> f21<-acf(x21,lag=40);f21
Autocorrelations of series 'x21', by lag
1      2      3      4      5      6      7      8      9      10     11     12
13     14     15     16     17     18     19     20
 0.011 0.027 0.008 0.003 -0.010 -0.008 -0.022 0.015 0.008  0.015 -0.023 -0.032
-0.003  0.003 -0.016  0.033  0.014  0.014  0.015  0.029
21     22     23     24     25     26     27     28     29     30     31     32
33     34     35     36     37     38     39     40
 -0.004 -0.002 -0.020 -0.019 -0.030 -0.011 -0.008 -0.014 -0.008 -0.010 -0.010
-0.002 0.007 -0.002 0.020 0.011 -0.017 0.014 0.014 -0.002
> dw21<-2*(1-f21$acf[2]);dw21 #use formula dw=2(1-r1)
[1] 1.946512

> basicStats(x21) #to get the descriptive of the standardized residuals
```

```
 nobs          4499.000000
 NAs              0.000000
 Minimum         -5.771694
 Maximum         10.827922
 1. Quartile     -0.511818
 3. Quartile      0.578078
 Mean             0.029693
 Median           0.016326
 Sum            133.588760
 SE Mean          0.014892
 LCL Mean         0.000498
 UCL Mean         0.058888
 Variance         0.997699
 Stdev            0.998849
 Skewness         0.477820
 Kurtosis         7.032348
```

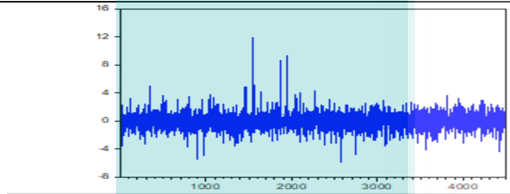
```
> ##Test of Hypotheses for standardised residuals##
> t.test(x21) #test of hypotesis of mean
        One Sample t-test
data:  x21
t = 1.9939, df = 4498, p-value = 0.04622
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.0004981102 0.0588878691
sample estimates: mean of x
0.02969299
> s_x21=0.4664/sqrt(6/4499);s_x21 # Hnull:S(x)=0
[1] 12.77147
> pvs_x21=2*(1-pnorm(s_x21));pvs_x21 #pvalue for the skewness
[1] 0
> k_x21=(6.9514)/sqrt(24/4499);k_x21 # Hnull:K(x)-3=0
[1] 95.17539
> pvk_x21=2*(1-pnorm(k_x21));pvk_x21 #p-value for the kurtosis
[1] 0
```

ii. **ARIMA (0,1,0)-GARCH (1,1) with *t* distribution**

Dependent Variable: D(LDT2)
Method: ML - ARCH (Marquardt) - Student's t distribution
Date: 07/28/17 Time: 14:36
Sample (adjusted): 2 4500
Included observations: 4499 after adjustments
Convergence achieved after 15 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	7.58E-05	9.46E-05	0.801690	0.4227
		Variance Equation		
C	1.92E-07	8.21E-08	2.335608	0.0195
RESID(-1)^2	0.066284	0.007208	9.195282	0.0000
GARCH(-1)	0.938563	0.005795	161.9728	0.0000
T-DIST. DOF	4.697706	0.317229	14.80854	0.0000

R-squared	-0.000551	Mean dependent var	0.000324
Adjusted R-squared	-0.000551	S.D. dependent var	0.010553
S.E. of regression	0.010556	Akaike info criterion	-6.707778
Sum squared resid	0.501190	Schwarz criterion	-6.700652
Log likelihood	15094.15	Hannan-Quinn criter.	-6.705267
Durbin-Watson stat	2.012829		



Correlogram of Standardized Residuals

Date: 07/28/17 Time: 14:58
Sample: 2 4500
Included observations: 4499

	AC	PAC	Q-Stat	Prob
1	0.011	0.011	0.5574	0.455
2	0.030	0.030	4.5724	0.102
3	0.008	0.008	4.8967	0.180
4	-0.001	-0.002	4.8982	0.298
5	-0.011	-0.012	5.4603	0.362
6	-0.008	-0.007	5.7191	0.455
7	-0.023	-0.022	8.0702	0.326
8	0.012	0.014	8.7626	0.363
9	0.007	0.008	8.9557	0.441
10	0.015	0.015	10.007	0.440
11	-0.022	-0.024	12.253	0.345
12	-0.032	-0.033	16.823	0.156
13	-0.003	-0.001	16.865	0.206
14	0.003	0.006	16.912	0.261
15	-0.017	-0.016	18.238	0.250
16	0.034	0.035	23.608	0.098
17	0.015	0.015	24.651	0.103
18	0.015	0.011	25.723	0.106
19	0.016	0.013	26.928	0.106
20	0.027	0.026	30.217	0.066

Correlogram of Standardized Residuals Squared

Date: 07/28/17 Time: 15:00
Sample: 2 4500
Included observations: 4499

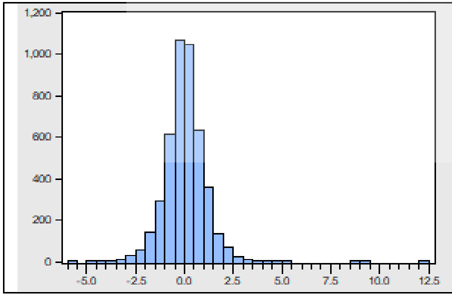
	AC	PAC	Q-Stat	Prob
1	0.026	0.026	3.0594	0.080
2	0.000	-0.001	3.0594	0.217
3	0.004	0.004	3.1418	0.370
4	0.003	0.003	3.1942	0.526
5	-0.008	-0.008	3.4974	0.624
6	0.005	0.006	3.6162	0.728
7	-0.015	-0.015	4.6080	0.708
8	-0.001	0.000	4.6093	0.798
9	-0.010	-0.010	5.0259	0.832
10	-0.015	-0.014	6.0308	0.813
11	-0.012	-0.012	6.7299	0.821
12	-0.012	-0.011	7.3432	0.834
13	-0.010	-0.010	7.8398	0.854
14	-0.007	-0.006	8.0357	0.887
15	-0.010	-0.010	8.4942	0.902
16	-0.020	-0.020	10.334	0.849
17	-0.013	-0.012	11.071	0.853
18	-0.005	-0.005	11.203	0.886
19	-0.001	-0.001	11.208	0.917
20	0.003	0.002	11.241	0.940

Heteroskedasticity Test: ARCH

F-statistic	0.610881	Prob. F(10,4478)	0.8059
Obs*R-squared	6.115469	Prob. Chi-Square(10)	0.8055

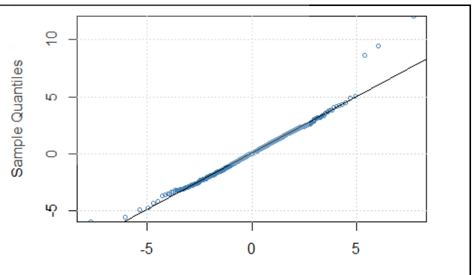
Heteroskedasticity Test: ARCH

F-statistic	0.552810	Prob. F(15,4468)	0.9113
Obs*R-squared	8.306430	Prob. Chi-Square(15)	0.9109



Series: Standardized Residuals
Sample 2 4500
Observations 4499

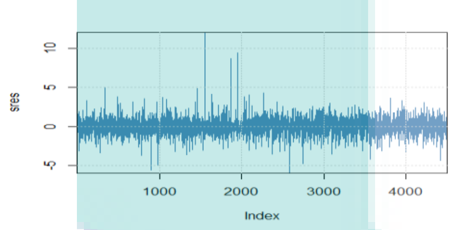
Mean	0.027268
Median	0.012820
Maximum	12.04160
Minimum	-5.933470
Std. Dev.	1.006737
Skewness	0.619809
Kurtosis	12.00534
Jarque-Bera	15490.19
Probability	0.000000



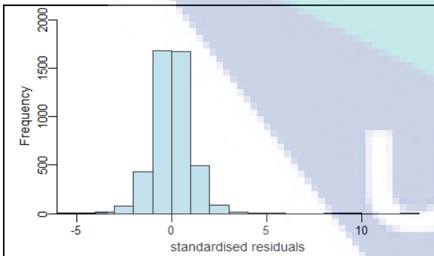
iii. ARIMA (0,1,0)-GARCH (1,1) with skewed-*t* distribution

```
> fit23_r<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt2, cond.
dist="sstd", include.mean = FALSE,trace=F) # for ARIMA without constant
> sum23<-summary(fit23_r)
```

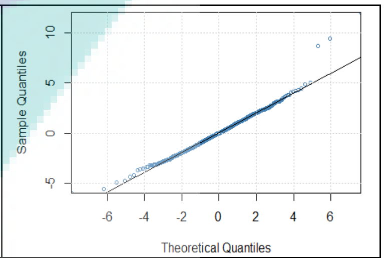
```
Title:GARCH Modelling
Call:garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt2, cond.dist
= "sstd",include.mean = FALSE, trace = F)
Mean and Variance Equation:data ~ arma(0, 0) + garch(1, 1)
<environment: 0x15403994>
 [data = dldt2]
Conditional Distribution:sstd
Coefficient(s):
     omega        alpha1        beta1         skew         shape
1.8641e-07   6.6367e-02   9.3870e-01   9.8110e-01   4.6733e+00
Std. Errors:based on Hessian
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  1.864e-07   7.912e-08    2.356   0.0185 *
alpha1 6.637e-02   7.784e-03    8.526   <2e-16 ***
beta1  9.387e-01   6.334e-03  148.209   <2e-16 ***
skew   9.811e-01   1.711e-02   57.337   <2e-16 ***
shape  4.673e+00   3.380e-01   13.826   <2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 15093.98    normalized:  3.354964
Standardised Residuals Tests:
                                Statistic p-value
 Jarque-Bera Test   R    Chi^2  15615.52  0
 Shapiro-Wilk Test  R    W      0.949175  0
 Ljung-Box Test     R    Q(10)  10.118    0.4302023
 Ljung-Box Test     R    Q(15)  18.65161  0.2299577
 Ljung-Box Test     R    Q(20)  30.56298  0.06123051
 Ljung-Box Test     R^2  Q(10)  6.104533  0.8064057
 Ljung-Box Test     R^2  Q(15)  8.581091  0.8983834
 Ljung-Box Test     R^2  Q(20)  11.40161  0.9351353
 LM Arch Test       R    TR^2   7.274739  0.8389364
Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.707705 -6.700579 -6.707707 -6.705194
```




```
> ArchTest (x23, lags=10, demean = FALSE) #arch test at lag=10

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x23
Chi-squared = 6.1051, df = 10, p-value = 0.8064

> ArchTest (x23, lags=15, demean = FALSE)

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x23
Chi-squared = 8.3006, df = 15, p-value = 0.9112

> dw23<-2*(1-f23$acf[2]);dw23 #use formula dw=2(1-r1)
[1] 1.939688
```

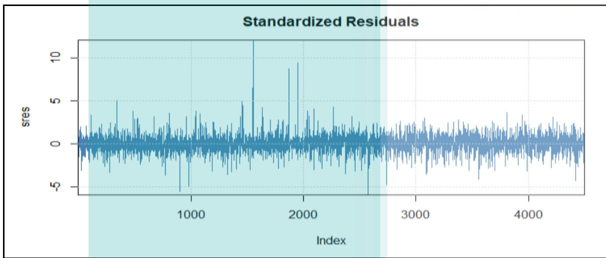



```
                      x23
nobs          4499.000000
NAs              0.000000
Minimum         -5.910399
Maximum         12.025272
1. Quartile     -0.505570
3. Quartile      0.585161
Mean             0.036466
Median           0.023095
Sum            164.060187
SE Mean          0.014964
LCL Mean         0.007128
UCL Mean         0.065804
Variance         1.007476
Stdev            1.003731
Skewness         0.624484
Kurtosis         9.035683
```




```
> # estimate model ARIMA(0,1,1)-GARCH(1,1) with cond.dist=skewed t
> fit23<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt2, cond.dist="
sstd", include.mean = TRUE,trace=F) # for model with constant
> sum23<-summary(fit23)
Title:GARCH Modelling
Call:garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt2, cond.dist = "sstd",
    include.mean = TRUE, trace = F)
Mean and Variance Equation:data ~ arma(0, 0) + garch(1, 1)
[data = dldt2]
Conditional Distribution: sstd
Coefficient(s):
        mu       omega       alpha1        beta1         skew        shape
3.8961e-05  1.8769e-07  6.6340e-02  9.3867e-01  9.8397e-01  4.6757e+00
Std. Errors:based on Hessian
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     3.896e-05   1.010e-04    0.386   0.6996
omega  1.877e-07   7.953e-08    2.360   0.0183 *
```

```
alpha1 6.634e-02   7.780e-03    8.527   <2e-16 ***
beta1  9.387e-01   6.337e-03  148.130   <2e-16 ***
skew   9.840e-01   1.870e-02   52.618   <2e-16 ***
shape  4.676e+00   3.381e-01   13.830   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 15094.06     normalized:  3.35498

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  15634.1   0
 Shapiro-Wilk Test  R    W      0.9491605 0
 Ljung-Box Test     R    Q(10)  10.14822  0.4275869
 Ljung-Box Test     R    Q(15)  18.59038  0.2328994
 Ljung-Box Test     R    Q(20)  30.64627  0.06003686
 Ljung-Box Test     R^2  Q(10)  6.113017  0.8056803
 Ljung-Box Test     R^2  Q(15)  8.586274  0.8981381
 Ljung-Box Test     R^2  Q(20)  11.40931  0.9349096
 LM Arch Test       R    TR^2   7.280071  0.8385639

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.707293 -6.698742 -6.707297 -6.704280
```

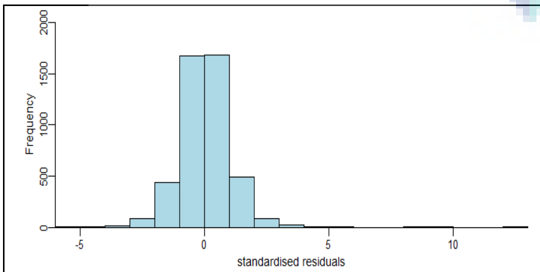


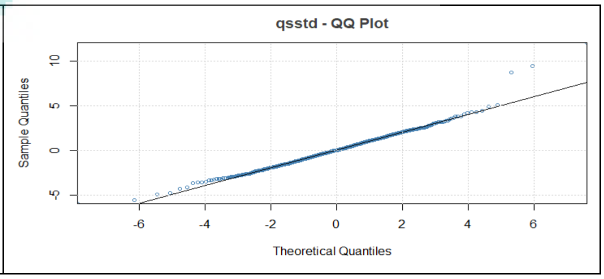

```
> ArchTest (x23, lags=10, demean = FALSE) #arch test at lag=10
            ARCH LM-test; Null hypothesis: no ARCH effects
data:  x23
Chi-squared = 6.1148, df = 10, p-value = 0.8055
> ArchTest (x23, lags=15, demean = FALSE)
            ARCH LM-test; Null hypothesis: no ARCH effects
data:  x23
Chi-squared = 8.308, df = 15, p-value = 0.9108
```

```
> #to calculate Durbin-Watson Test: Method 2
> x23<-residuals(fit23,standardize=TRUE)
> f23<-acf(x23,lag=40);f23
Autocorrelations of series 'x23', by lag
1      2      3      4      5      6      7      8      9      10     11     12     13     14     15
   16     17     18     19     20
0.011  0.030  0.009 -0.001 -0.011 -0.008 -0.023  0.012  0.006  0.016 -0.023 -0.032 -0.003  0.003 -0.0
17  0.034  0.014  0.015  0.017  0.028
21     22     23     24     25     26     27     28     29     30     31     32     33     34     35
    36     37     38     39     40
 -0.003 -0.002 -0.020 -0.022 -0.029 -0.011 -0.007 -0.017 -0.007 -0.009 -0.010 -0.001 0.007 -0.002  0.
021  0.011 -0.015  0.012 0.012 -0.002

> dw23<-2*(1-f23$acf[2]);dw23 #use formula dw=2(1-r1)
[1] 1.939565
> basicStats(x23) #to get the descriptive of the standardized residuals
```

```
 nobs          4499.000000
 NAs              0.000000
 Minimum         -5.920441
 Maximum         12.029642
 1. Quartile     -0.511359
 3. Quartile      0.580443
 Mean             0.031644
 Median           0.018047
 Sum            142.368133
 SE Mean          0.014972
 LCL Mean         0.002292
 UCL Mean         0.060997
 Variance         1.008508
 Stdev            1.004245
 Skewness         0.625185
 Kurtosis         9.040968
```

```
> ##Test of Hypotheses for standardised residuals##
> t.test(x23) #test of hypotesis of mean
          One Sample t-test
data:  x23
t = 2.1136, df = 4498, p-value = 0.03461
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.002291785 0.060997005
sample estimates: mean of x
0.0316444
> s_x23=0.6252/sqrt(6/4499);s_x23 # Hnull:S(x)=0
[1] 17.1199
> pvs_x23=2*(1-pnorm(s_x23));pvs_x23 #pvalue for the skewness
[1] 0
> k_x23=(9.0410)/sqrt(24/4499);k_x23 # Hnull:K(x)-3=0
[1] 123.7852
> pvk_x23=2*(1-pnorm(k_x23));pvk_x23 #p-value for the kurtosis
[1] 0
```



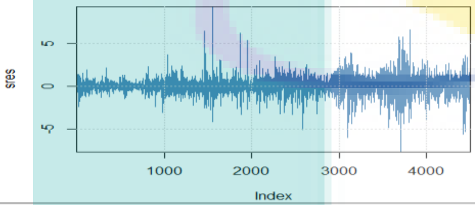


iv. **ARIMA (0,1,0)-GARCH (1,1) with GED distribution**

Dependent Variable: D(LDT2_ESTIMATE)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Date: 03/13/17 Time: 01:38
Sample (adjusted): 2 4500
Included observations: 4499 after adjustments
Convergence achieved after 18 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
	Variance Equation			
C	2.02E-07	8.36E-08	2.418886	0.0156
RESID(-1)^2	0.060560	0.006485	9.338634	0.0000
GARCH(-1)	0.942147	0.005399	174.5087	0.0000
GED PARAMETER	1.139490	0.022071	51.62895	0.0000

R-squared	-0.000940	Mean dependent var	0.000324
Adjusted R-squared	-0.000718	S.D. dependent var	0.010553
S.E. of regression	0.010557	Akaike info criterion	-6.702930
Sum squared resid	0.501385	Schwarz criterion	-6.697230
Log likelihood	15082.24	Hannan-Quinn criter.	-6.700922
Durbin-Watson stat	2.012046		



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 03:33
Sample: 2 4500
Included observations: 4499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.011	0.011	0.5176	0.472
		2	0.029	0.029	4.3217	0.115
		3	0.008	0.007	4.6150	0.202
		4	-0.000	-0.001	4.6150	0.329
		5	-0.011	-0.012	5.1747	0.395
		6	-0.008	-0.008	5.4671	0.485
		7	-0.023	-0.022	7.8534	0.346
		8	0.013	0.014	8.5880	0.378
		9	0.007	0.008	8.7845	0.457
		10	0.015	0.014	9.8091	0.457
		11	-0.023	-0.024	12.120	0.355
		12	-0.032	-0.033	16.811	0.157
		13	-0.003	-0.002	16.859	0.206
		14	0.003	0.005	16.899	0.262
		15	-0.018	-0.016	18.310	0.247
		16	0.034	0.034	23.560	0.100
		17	0.015	0.015	24.550	0.105
		18	0.015	0.011	25.563	0.110
		19	0.016	0.013	26.663	0.113
		20	0.027	0.026	29.912	0.071

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 03:45
Sample: 2 4500
Included observations: 4499

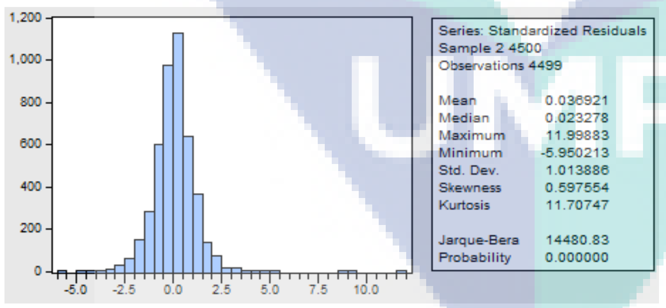
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.029	0.029	3.8015	0.051
		2	0.001	0.001	3.8109	0.149
		3	0.006	0.006	3.9665	0.265
		4	0.005	0.005	4.0740	0.396
		5	-0.008	-0.008	4.3465	0.501
		6	0.006	0.007	4.5148	0.607
		7	-0.014	-0.015	5.4399	0.606
		8	-0.000	0.001	5.4400	0.710
		9	-0.009	-0.009	5.8264	0.757
		10	-0.015	-0.014	6.7988	0.744
		11	-0.012	-0.011	7.4976	0.757
		12	-0.011	-0.011	8.0811	0.779
		13	-0.010	-0.009	8.5547	0.806
		14	-0.006	-0.006	8.7393	0.847
		15	-0.010	-0.009	9.1709	0.868
		16	-0.020	-0.020	11.062	0.806
		17	-0.012	-0.012	11.766	0.814
		18	-0.005	-0.005	11.884	0.853
		19	-0.001	-0.001	11.886	0.890
		20	0.003	0.003	11.939	0.918

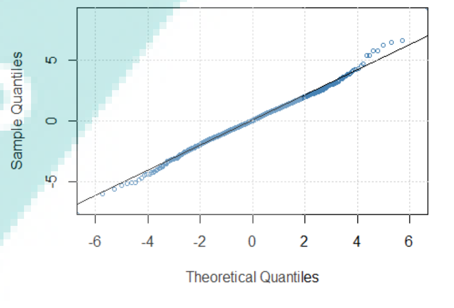
Heteroskedasticity Test: ARCH

F-statistic	0.688251	Prob. F(10,4478)	0.7364
Obs*R-squared	6.888826	Prob. Chi-Square(10)	0.7359

Heteroskedasticity Test: ARCH

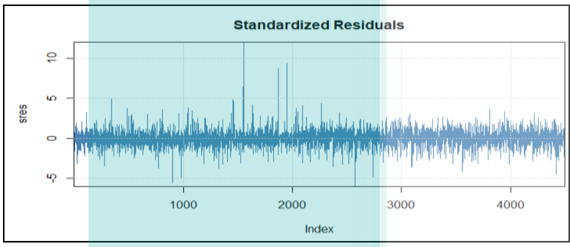
F-statistic	0.594459	Prob. F(15,4468)	0.8816
Obs*R-squared	8.930992	Prob. Chi-Square(15)	0.8811






```
> # estimate model ARIMA(0,1,0)-GARCH(1,1) with cond.dist=GED
> fit24<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt2,algorithm ="
lbfgsb+nm",hessian = "ropt",cond.dist="ged", include.mean = TRUE,trace=F)
> sum24<-summary(fit24)
Title:GARCH Modelling
Call:garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt2, cond.dist = "ged",
    include.mean = TRUE, trace = F, algorithm = "lbfgsb+nm", hessian = "ropt")
Mean and Variance Equation: data ~ arma(0, 0) + garch(1, 1)
[data = dldt2]
Conditional Distribution:ged
Coefficient(s):
         mu        omega        alpha1        beta1        shape
5.6492e-05  2.0368e-07  6.0634e-02  9.4196e-01  1.1401e+00
Std. Errors:based on Hessian
```

```
Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      5.649e-05   1.102e-04    0.513   0.6083
omega   2.037e-07   8.171e-08    2.493   0.0127 *
alpha1  6.063e-02   7.884e-03    7.691 1.47e-14 ***
beta1   9.420e-01   6.851e-03  137.484  < 2e-16 ***
shape   1.140e+00   3.081e-02   37.008  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 15081.91    normalized:  3.352281

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  14719.6    0
 Shapiro-Wilk Test  R    W      0.949981  0
 Ljung-Box Test     R    Q(10)  10.0333   0.4375771
 Ljung-Box Test     R    Q(15)  18.54279  0.2352048
 Ljung-Box Test     R    Q(20)  30.58091  0.0609718
 Ljung-Box Test     R^2  Q(10)  6.867687  0.7378743
 Ljung-Box Test     R^2  Q(15)  9.24379   0.8644292
 Ljung-Box Test     R^2  Q(20)  12.09128  0.9128982
 LM Arch Test       R    TR^2   7.971098  0.7873849

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.702339 -6.695214 -6.702342 -6.699828
```

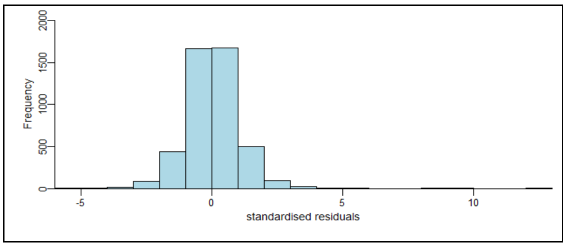


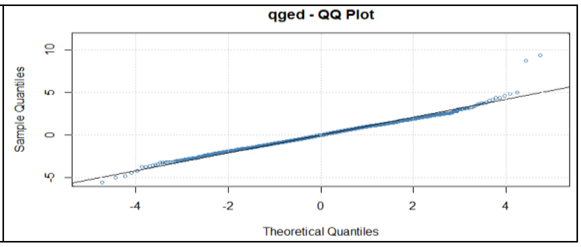

```
> x24<-residuals(fit24,standardize=TRUE)
> ArchTest (x24, lags=10, demean = FALSE)
ARCH LM-test; Null hypothesis: no ARCH effects
data:  x24
Chi-squared = 6.8692, df = 10, p-value = 0.7377
> ArchTest (x24, lags=15, demean = FALSE)
ARCH LM-test; Null hypothesis: no ARCH effects
Chi-squared = 8.9089, df = 15, p-value = 0.8822
```

```
> #to calculate Durbin-Watson Test: Method 2
> x24<-residuals(fit24,standardize=TRUE)
> f24<-acf(x24,lag=40);f24
Autocorrelations of series 'x24', by lag
1      2      3      4      5      6      7      8      9      10     11     12     13     14     15
   16     17     18     19     20
0.011  0.030  0.009  0.000 -0.011 -0.008 -0.023  0.013  0.007  0.016 -0.023 -0.033 -0.003  0.003 -0.0
17  0.034  0.014  0.015  0.016 0.028
21     22     23     24     25     26     27     28     29     30     31     32     33     34     35
    36     37     38     39     40
 -0.003 -0.002 -0.020 -0.021 -0.029 -0.011 -0.007 -0.017 -0.007 -0.009 -0.010 -0.001  0.007 -0.001  0
.021  0.011 -0.016  0.012 0.013 -0.002
> dw24<-2*(1-f24$acf[2]);dw24 #use formula dw=2(1-r1)
[1] 1.940839
> basicStats(x24) #to get the descriptive of the standardized residuals
```

nobs	4499.000000
NAs	0.000000
Minimum	-5.968045
Maximum	12.015778
1. Quartile	-0.521731
3. Quartile	0.584364
Mean	0.029719
Median	0.015725
Sum	133.706183
SE Mean	0.015108
LCL Mean	0.000101
UCL Mean	0.059337
Variance	1.026845
Stdev	1.013334
Skewness	0.605946
Kurtosis	8.772710

```
> ##Test of Hypotheses for standardised residuals##
> t.test(x24) #test of hypotesis of mean
          One Sample t-test
data:  x24
t = 1.9672, df = 4498, p-value = 0.04923
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.0001008398 0.0593373389
sample estimates: mean of x
0.02971909
> s_x24=0.6059/sqrt(6/4499);s_x24 # Hnull:S(x)=0
[1] 16.59141
> pvs_x24=2*(1-pnorm(s_x24));pvs_x24 #pvalue for the skewness
[1] 0
> k_x24=(8.7727)/sqrt(24/4499);k_x24 # Hnull:K(x)-3=0
[1] 120.1118
> pvk_x24=2*(1-pnorm(k_x24));pvk_x24 #p-value for the kurtosis
[1] 0
```

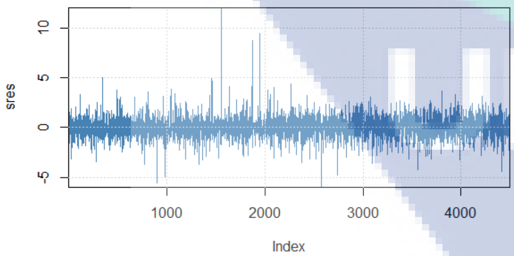




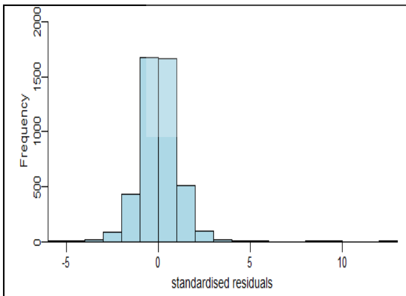
v. ARIMA (0,1,0)-GARCH (1,1) with skewed-GED distribution

```
>fit25<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt2, proc
edure ="lbfgsb+nm",hessian = "ropt",cond.dist="sged", include.mean = T
RUE,trace=F)#ARIMA with constant
> sum25<-summary(fit25)
Title:GARCH Modelling
Call:garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt2, cond.dist = "sged",incl
ude.mean = TRUE, trace = F, procedure = "lbfgsb+nm", hessian = "ropt")
Mean and Variance Equation:data ~ arma(0, 0) + garch(1, 1)
[data = dldt2]
Conditional Distribution: sged
Coefficient(s):
        mu        omega       alpha1        beta1         skew        shape
6.0983e-05   2.0440e-07   6.0677e-02   9.4191e-01   1.0025e+00   1.1398e+00

Std. Errors:based on Hessian
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     6.098e-05   1.052e-04    0.580   0.5621
omega  2.044e-07   8.194e-08    2.494   0.0126 *
alpha1 6.068e-02   7.887e-03    7.693 1.44e-14 ***
beta1  9.419e-01   6.857e-03  137.355  < 2e-16 ***
skew   1.002e+00   1.049e-02   95.559  < 2e-16 ***
shape  1.140e+00   3.086e-02   36.938  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 15081.93    normalized:  3.352286

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  14735.41  0
 Shapiro-Wilk Test  R    W      0.9499696 0
 Ljung-Box Test     R    Q(10)  10.03918  0.4370627
 Ljung-Box Test     R    Q(15)  18.53988  0.2353464
 Ljung-Box Test     R    Q(20)  30.59644  0.06074859
 Ljung-Box Test     R^2  Q(10)  6.856489  0.7389203
 Ljung-Box Test     R^2  Q(15)  9.232515  0.8650494
 Ljung-Box Test     R^2  Q(20)  12.07892  0.9133329
 LM Arch Test       R    TR^2   7.960352  0.788221

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.701905 -6.693354 -6.701908 -6.698892
```

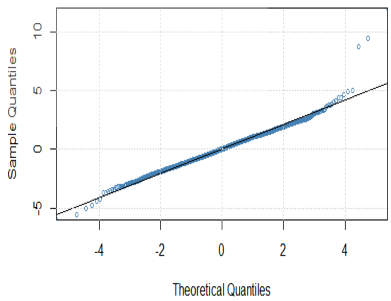
Standardized residuals:




```
> ArchTest (x25, lags=10, demean = FALSE) #arch test at lag=10

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  x25
Chi-squared = 6.8289, df = 10, p-value = 0.7415

> ArchTest (x25, lags=15, demean = FALSE)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  x25
Chi-squared = 8.8739, df = 15, p-value = 0.884

> dw25<-2*(1-f25$acf[2]);dw25 #use formula dw=2(1-r1)
[1] 1.941
```

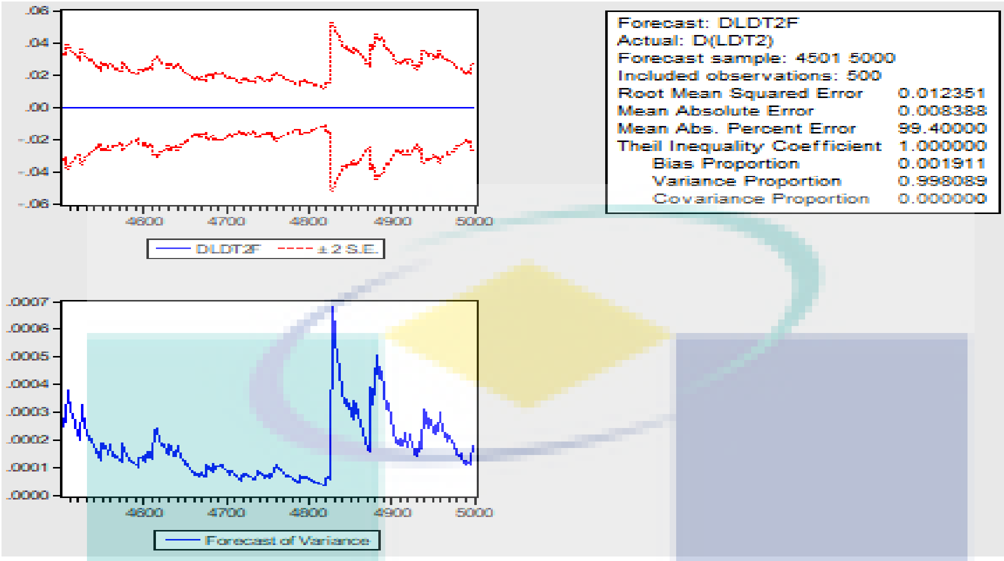



```
                      x25
nobs          4499.000000
NAs              0.000000
Minimum         -5.956316
Maximum         12.018184
1. Quartile     -0.512561
3. Quartile      0.590265
Mean             0.036770
Median           0.023307
Sum            165.429385
SE Mean          0.015101
LCL Mean         0.007165
UCL Mean         0.066376
Variance         1.025946
Stdev            1.012890
Skewness         0.605487
Kurtosis         8.773310
```



4. STAGE IV: BJ-GARCH FORECASTING (ARIMA(0,1,0)-GARCH(1,1) with *t*)

i. For Stationary Data (Daily Log Return Gold Price)




```
> mae_st_AG2=sum(error2st_AG2)/T; mae_st_AG2; [1] 0.008388001
> rmse_st_AG2=sqrt(sum(error3st_AG2)/T); rmse_st_AG2; [1] 0.01235061
> #mape_st_AG2=(100/T)*sum(error4s_AG); mape_st_AG2 #cannot calculate mape because of ex
ist of "0" data in stationary data
```

ii. For Daily Gold Price

```
> mae_AG2=sum(error2_AG2)/T; mae_AG2;[1] 12.6855
> rmse_AG2=sqrt(sum(error3_AG2)/T); rmse_AG2;[1] 18.37164
> mape_AG2=(100/T)*sum(error4_AG2); mape_AG2;[1] 0.8401619
```

Actual and Forecast data for the last 10 days (model ARIMA(0,1,0)-GARCH(1,1) with *t*):

```
         [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]     [,9]    [,10]
[1,] 1614.726 1706.295 1555.181 1618.979 1775.348 1695.287 1580.950 1369.790 1330.760 1330.510
[2,] 1609.222 1670.268 1590.708 1623.232 1762.588 1666.265 1575.196 1355.779 1325.156 1304.991
[3,] 1607.721 1678.775 1593.710 1600.215 1770.344 1651.754 1580.700 1361.784 1315.499 1299.487
[4,] 1572.194 1691.284 1583.702 1598.213 1768.092 1652.755 1582.952 1409.570 1315.999 1266.462
[5,] 1532.163 1688.782 1550.177 1603.217 1737.319 1656.758 1580.200 1381.549 1310.245 1286.477
[6,] 1532.163 1698.790 1569.692 1611.223 1747.827 1658.759 1595.211 1391.306 1305.741 1271.465
[7,] 1599.214 1691.284 1570.693 1612.224 1750.329 1695.037 1590.458 1377.546 1281.473 1274.468
[8,] 1614.226 1645.499 1575.796 1614.476 1744.324 1680.776 1587.205 1383.550 1283.474 1320.252
[9,] 1600.215 1649.252 1580.700 1616.227 1738.320 1649.252 1596.712 1414.574 1299.236 1317.500
[10] 1617.728 1659.260 1541.170 1619.730 1728.062 1646.500 1604.969 1395.560 1309.995 1318.501
[11] 1616.227 1662.762 1559.184 1623.733 1712.300 1657.258 1611.974 1403.566 1342.019 1334.013
[12] 1638.244 1658.009 1607.220 1598.964 1707.797 1659.010 1608.721 1400.563 1329.509 1332.262
[13] 1635.742 1650.503 1636.242 1602.967 1716.803 1676.273 1614.976 1405.067 1327.508 1345.772
[14] 1662.262 1636.743 1607.220 1605.719 1717.304 1658.759 1608.972 1401.064 1330.760 1348.774
[15] 1636.743 1665.264 1577.698 1615.977 1708.297 1667.766 1600.465 1387.053 1370.290 1362.034
[16] 1642.247 1681.527 1585.204 1616.227 1711.299 1681.777 1599.214 1384.301 1366.037 1350.275
[17] 1657.258 1693.286 1604.718 1640.746 1720.306 1677.524 1604.218 1375.294 1373.543 1355.779
[18] 1648.251 1677.273 1620.731 1643.248 1717.554 1676.273 1599.464 1383.801 1364.036 1325.006
[19] 1656.257 1658.759 1614.726 1666.515 1686.280 1689.783 1584.703 1386.052 1376.545 1307.743
[20] 1654.256 1663.763 1628.486 1668.267 1684.779 1688.782 1575.947 1392.307 1378.547 1321.503
[21] 1676.773 1678.775 1616.727 1669.267 1692.285 1691.784 1547.675 1385.802 1420.328 1308.243
[22] 1666.765 1677.524 1626.735 1661.261 1716.553 1691.534 1569.191 1367.788 1420.579 1320.002
[23] 1651.254 1622.232 1602.216 1661.762 1718.305 1672.270 1576.197 1373.793 1408.820 1308.243
[24] 1728.312 1632.239 1583.202 1649.753 1739.571 1661.261 1578.448 1293.482 1395.810 1286.477
[25] 1727.311 1645.249 1566.689 1692.785 1736.568 1657.759 1576.197 1296.234 1393.308 1283.474
[26] 1730.314 1659.260 1571.193 1698.289 1727.562 1664.764 1566.189 1287.728 1400.563 1282.224
[27] 1745.325 1669.768 1577.197 1691.284 1727.562 1678.775 1536.667 1279.972 1391.056 1273.467
[28] 1741.322 1667.766 1574.696 1702.292 1711.299 1666.015 1396.060 1237.189 1386.052 1286.977
[29] 1752.330 1654.256 1559.684 1729.313 1714.802 1670.268 1381.049 1233.687 1388.054 1288.228
[30] 1735.318 1636.743 1599.715 1733.316 1731.815 1667.266 1393.058 1192.906 1391.056 1284.475
[31] 1720.306 1645.249 1593.210 1738.070 1733.566 1674.772 1394.809 1243.694 1359.282 1276.719
[32] 1725.310 1651.254 1618.729 1738.320 1725.310 1675.522 1406.568 1253.452 1364.786 1257.955
[33] 1747.327 1642.747 1614.726 1734.567 1732.315 1669.267 1425.582 1250.950 1329.009 1240.942
[34] 1749.328 1630.238 1605.219 1776.849 1735.818 1669.518 1409.070 1252.701 1319.502 1247.197
```

```
[35] 1712.800 1650.753 1588.206 1771.345 1751.830 1653.255 1429.585 1213.671 1325.006 1243.944
[36] 1721.307 1638.994 1586.204 1770.844 1747.577 1648.752 1452.102 1236.189 1313.247 1248.448
[37] 1723.308 1654.756 1596.462 1768.092 1709.298 1646.250 1472.618 1256.454 1301.989 1245.946
[38] 1734.317 1664.764 1578.198 1759.836 1726.311 1647.251 1468.615 1256.954 1366.538 1246.446
[39] 1714.302 1652.505 1557.432 1785.856 1727.311 1613.475 1470.116 1285.976 1350.275 1253.952
[40] 1724.309 1665.264 1596.712 1763.839 1721.307 1611.974 1455.855 1280.722 1324.005 1230.434
[41] 1734.317 1649.252 1590.958 1772.846 1699.040 1608.972 1470.366 1285.726 1315.249 1218.175
[42] 1749.328 1638.994 1586.455 1746.076 1695.287 1589.707 1470.366 1292.481 1323.755 1228.433
[43] 1753.331 1644.999 1576.447 1764.340 1695.537 1578.198 1445.347 1298.236 1334.013 1223.429
[44] 1778.350 1603.718 1585.204 1777.349 1702.793 1577.698 1469.115 1284.225 1342.019 1233.937
[45] 1778.851 1583.702 1577.448 1788.358 1713.801 1587.455 1466.614 1296.735 1327.508 1237.940
[46] 1773.346 1599.715 1573.445 1776.849 1711.299 1591.708 1427.584 1328.008 1291.731 1267.212
[47] 1782.353 1584.203 1584.453 1776.599 1717.554 1605.469 1431.837 1334.513 1307.243 1261.708
[48] 1771.345 1559.684 1602.216 1793.111 1694.036 1589.707 1434.839 1336.014 1317.000 1226.181
[49] 1715.302 1557.683 1619.229 1785.356 1697.539 1583.452 1411.071 1327.008 1310.745 1232.936
[50] 1708.297 1549.677 1619.480 1774.848 1697.038 1575.446 1382.049 1332.011 1324.506 1235.688
```

UMP

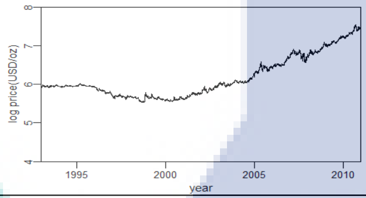
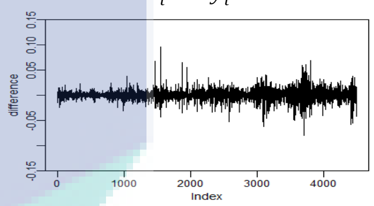
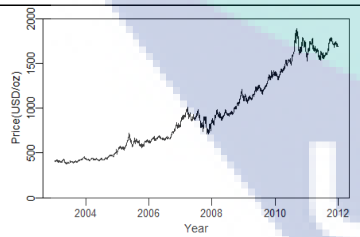
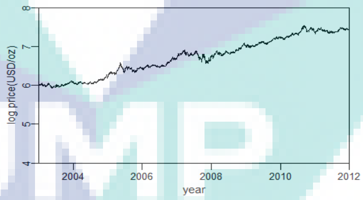
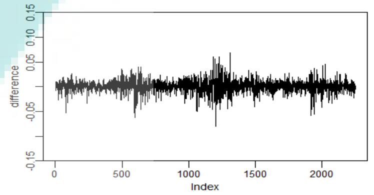
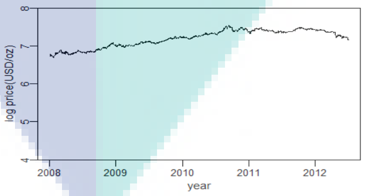
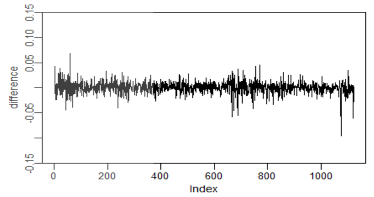
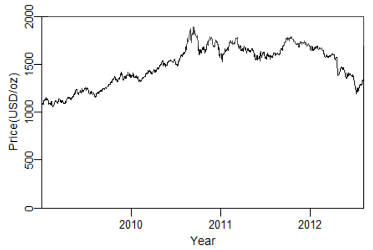
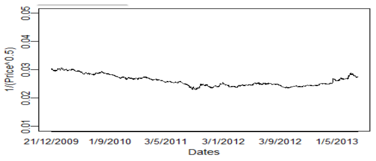
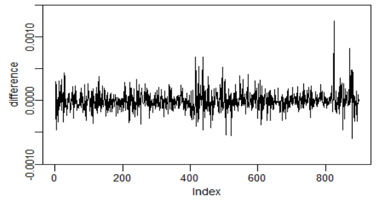
APPENDIX 3
ANALYSIS OF CHAPTER 4 SECTION 4.3

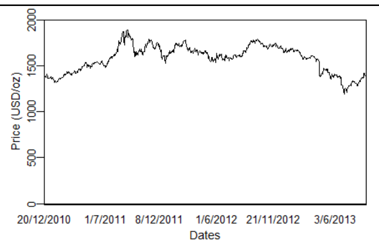
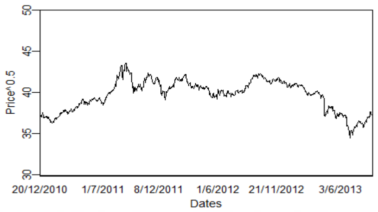
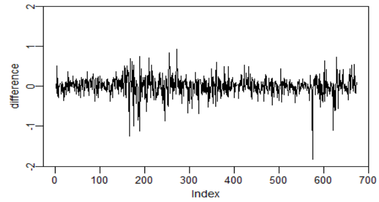
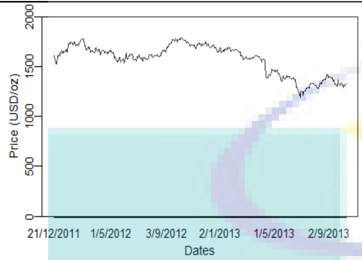
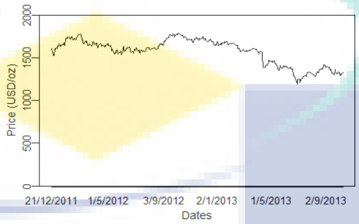
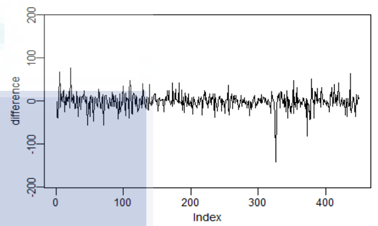
1. Data of Study

Sample	Duration	Number of Data	In-Sample Data	Out-of-Sample Data
1	24/11/1993 - 17/12/2013 (20-year)	5 000	24/11/1993 - 20/12/2011 (4500 data)	21/12/2011 - 17/12/2013 (500 data)
2	5/12/2003 - 17/12/2013 (10-year)	2 500	5/12/2003 - 18/12/2012 (2250 data)	19/12/2012-17/12/2013 (250 data)
3	22/12/2008 - 17/12/2013 (5-year)	1 250	22/12/2008 - 24/6/2013 (1125 data)	25/6/2013 - 17/12/2013 (125 data)
4	21/12/2009 - 17/12/2013 (4-year)	1 000	21/12/2009 - 29/7/2013 (900 data)	30/7/2013 - 17/12/2013 (100 data)
5	20/12/2010 - 17/12/2013 (3-year)	750	20/12/2010 - 3/9/2013 (675 data)	4/9/2013 - 17/12/2013 (75 data)
6	21/12/2011 - 17/12/2013 (2-year)	500	21/12/2011- 8/10/2013 (450 data)	9/10/2013 - 17/12/2013 (50 data)

STAGE I: MODEL IDENTIFICATION

1. Time Plot

Data	**Original Data, *y***	**Transformed Data,** y_t^*	**Stationary Data,** s_t
1		$\lambda_{BC} = -0.2147 \rightarrow 0 : y_t^* = \ln y_t$ 	$s_t = \Delta y_t^*$ 
2		$\lambda_{BC} = -0.1101 \rightarrow 0 : y_t^* = \ln y_t$ 	$s_t = \Delta y_t^*$ 
3		$\lambda_{BC} = 0.0780 \rightarrow 0 : y_t^* = \ln y_t$ 	$s_t = \Delta y_t^*$ 
4		$\lambda_{BC} = -0.4217 \rightarrow -0.5 : y_t^* = \frac{1}{\sqrt{y_t}}$ 	$s_t = \Delta y_t^*$ 

5		$\lambda_{BC} = 0.6421 \rightarrow 0.5 : y_t^* = \sqrt{y_t}$ 	$s_t = \Delta y_t^*$ 
6		$\lambda_{BC} = 0.9999 \rightarrow 1 : y_t^* = y$ 	$s_t = \Delta y$ 

2. Descriptive Statistics for original data, transformed data and stationary data

Data	Original Data	Transformed Data	Stationary Data

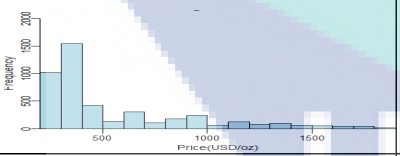
Data 1

Original Data

y

```
nobs          4.500000e+03
NAs           0.000000e+00
Minimum       2.528000e+02
Maximum       1.895000e+03
1. Quartile   3.106000e+02
3. Quartile   6.720000e+02
Mean          5.587877e+02
Median        3.871250e+02
Sum           2.514545e+06
SE Mean       5.418206e+00
LCL Mean      5.481654e+02
UCL Mean      5.694101e+02
Variance      1.321063e+05
Stdev         3.634643e+02
Skewness      1.592544e+00
Kurtosis      1.712147e+00
```

test	test statistic	*p*-value
mean	103.1316	0.0000
skewness	43.6124	0.0000
kurtosis	23.4439	0.0000
normality	2454.4068	0.0000



Transformed Data

Box-Cox Transformation:

$\lambda = -0.2147 \rightarrow 0 : y^* = \ln y$

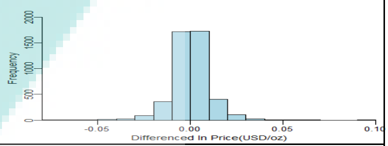
```
nobs           4500.000000
NAs               0.000000
Minimum           5.532599
Maximum           7.546974
1. Quartile       5.738506
3. Quartile       6.510258
Mean              6.165210
Median            5.958748
Sum           27743.447032
SE Mean           0.007923
LCL Mean          6.149678
UCL Mean          6.180743
Variance          0.282450
Stdev             0.531460
Skewness          0.895939
Kurtosis         -0.410726
```

Stationary Data

Stationary data: Differenced ln y

```
nobs          4499.000000
NAs              0.000000
Minimum         -0.079719
Maximum          0.096416
1. Quartile     -0.004261
3. Quartile      0.005289
Mean             0.000324
Median           0.000178
Sum              1.455641
SE Mean          0.000157
LCL Mean         0.000015
UCL Mean         0.000632
Variance         0.000111
Stdev            0.010553
Skewness         0.040233
Kurtosis         6.914227
```

test	test statistic	*p*-value
mean	2.0565	0.0398
skewness	1.1008	0.2710
kurtosis	94.6661	0.0000
normality	8974.3798	0.0000



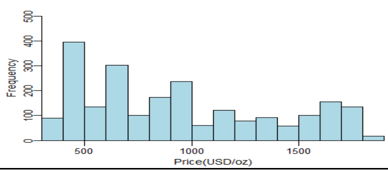
Data 2

Original Data

y

```
                   dt3
nobs          2.250000e+03
NAs           0.000000e+00
Minimum       3.750000e+02
Maximum       1.895000e+03
1. Quartile   5.688125e+02
3. Quartile   1.287875e+03
Mean          9.381378e+02
Median        8.768750e+02
Sum           2.110810e+06
SE Mean       9.328730e+00
LCL Mean      9.198440e+02
UCL Mean      9.564316e+02
Variance      1.958067e+05
Stdev         4.425005e+02
Skewness      5.060550e-01
Kurtosis     -1.027839e+00
```

test	test statistic	*p*-value
mean	100.5644	0.0000
skewness	9.8006	0.0000
kurtosis	-9.9516	0.0000
normality	194.8672	0.0000



Transformed Data

Box-Cox Transformation:

$\lambda = -0.1101 \rightarrow 0 : \ln y$

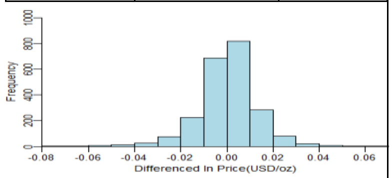
```
                      ldt3
nobs           2250.000000
NAs               0.000000
Minimum           5.926926
Maximum           7.546974
1. Quartile       6.343551
3. Quartile       7.160741
Mean              6.728727
Median            6.776364
Sum           15139.635509
SE Mean           0.010260
LCL Mean          6.708606
UCL Mean          6.748847
Variance          0.236861
Stdev             0.486683
Skewness         -0.023734
Kurtosis         -1.247232
```

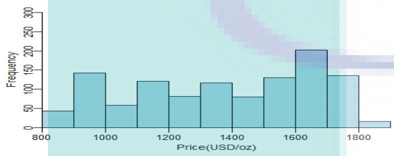
Stationary Data

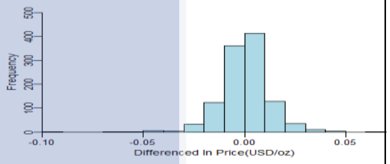
Stationary data: Differenced ln y

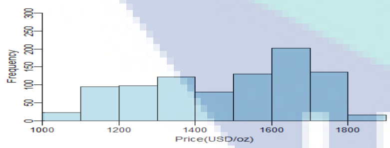
```
                  dldt3
nobs          2249.000000
NAs              0.000000
Minimum         -0.079719
Maximum          0.068414
1. Quartile     -0.005349
3. Quartile      0.007424
Mean             0.000639
Median           0.000785
Sum              1.437401
SE Mean          0.000269
LCL Mean         0.000113
UCL Mean         0.001166
Variance         0.000162
Stdev            0.012734
Skewness        -0.337412
Kurtosis         3.541914
```

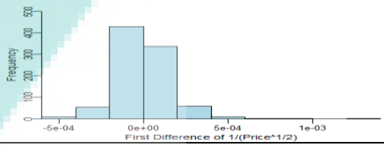
test	test statistic	*p*-value
mean	2.3803	0.0174
skewness	-6.5323	0.0000
kurtosis	34.2867	0.0000
normality	1222.1838	0.0000

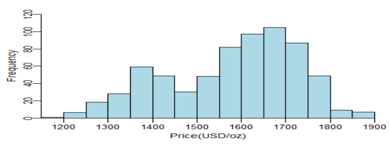


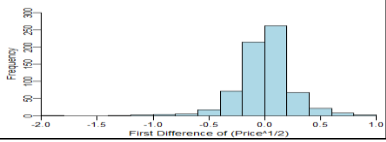
<table>
<tr><th>Data</th><th>Original Data</th><th>Transformed Data</th><th>Stationary Data</th></tr>
<tr>
<td>3</td>
<td>
<table>
<tr><td></td><td>dt4</td></tr>
<tr><td>nobs</td><td>1.125000e+03</td></tr>
<tr><td>NAs</td><td>0.000000e+00</td></tr>
<tr><td>Minimum</td><td>8.100000e+02</td></tr>
<tr><td>Maximum</td><td>1.895000e+03</td></tr>
<tr><td>1. Quartile</td><td>1.123750e+03</td></tr>
<tr><td>3. Quartile</td><td>1.637750e+03</td></tr>
<tr><td>Mean</td><td>1.375531e+03</td></tr>
<tr><td>Median</td><td>1.401000e+03</td></tr>
<tr><td>Sum</td><td>1.547472e+06</td></tr>
<tr><td>SE Mean</td><td>8.629615e+00</td></tr>
<tr><td>LCL Mean</td><td>1.358599e+03</td></tr>
<tr><td>UCL Mean</td><td>1.392463e+03</td></tr>
<tr><td>Variance</td><td>8.377903e+04</td></tr>
<tr><td>Stdev</td><td>2.894461e+02</td></tr>
<tr><td>Skewness</td><td>-2.638760e-01</td></tr>
<tr><td>Kurtosis</td><td>-1.261161e+00</td></tr>
</table>
<table>
<tr><th>test</th><th>test statistic</th><th>p-value</th></tr>
<tr><td>mean</td><td>159.3966</td><td>0.0000</td></tr>
<tr><td>skewness</td><td>-3.6136</td><td>0.0003</td></tr>
<tr><td>kurtosis</td><td>-8.6348</td><td>0.0000</td></tr>
<tr><td>normality</td><td>87.2810</td><td>0.0000</td></tr>
</table>


</td>
<td>
Box-Cox Transformation:
$\lambda = 0.0780 \rightarrow 0 : \ln y$
<table>
<tr><td></td><td>ldt4</td></tr>
<tr><td>nobs</td><td>1125.000000</td></tr>
<tr><td>NAs</td><td>0.000000</td></tr>
<tr><td>Minimum</td><td>6.697034</td></tr>
<tr><td>Maximum</td><td>7.546974</td></tr>
<tr><td>1. Quartile</td><td>7.024427</td></tr>
<tr><td>3. Quartile</td><td>7.401079</td></tr>
<tr><td>Mean</td><td>7.202654</td></tr>
<tr><td>Median</td><td>7.244942</td></tr>
<tr><td>Sum</td><td>8102.985371</td></tr>
<tr><td>SE Mean</td><td>0.006663</td></tr>
<tr><td>LCL Mean</td><td>7.189579</td></tr>
<tr><td>UCL Mean</td><td>7.215728</td></tr>
<tr><td>Variance</td><td>0.049951</td></tr>
<tr><td>Stdev</td><td>0.223498</td></tr>
<tr><td>Skewness</td><td>-0.492670</td></tr>
<tr><td>Kurtosis</td><td>-1.058125</td></tr>
</table>
</td>
<td>
Stationary data: Differenced log y
<table>
<tr><td></td><td>dldt4</td></tr>
<tr><td>nobs</td><td>1124.000000</td></tr>
<tr><td>NAs</td><td>0.000000</td></tr>
<tr><td>Minimum</td><td>-0.095962</td></tr>
<tr><td>Maximum</td><td>0.068414</td></tr>
<tr><td>1. Quartile</td><td>-0.005854</td></tr>
<tr><td>3. Quartile</td><td>0.006969</td></tr>
<tr><td>Mean</td><td>0.000370</td></tr>
<tr><td>Median</td><td>0.000607</td></tr>
<tr><td>Sum</td><td>0.415816</td></tr>
<tr><td>SE Mean</td><td>0.000365</td></tr>
<tr><td>LCL Mean</td><td>-0.000347</td></tr>
<tr><td>UCL Mean</td><td>0.001086</td></tr>
<tr><td>Variance</td><td>0.000150</td></tr>
<tr><td>Stdev</td><td>0.012243</td></tr>
<tr><td>Skewness</td><td>-0.585243</td></tr>
<tr><td>Kurtosis</td><td>5.993798</td></tr>
</table>
<table>
<tr><th>test</th><th>test statistic</th><th>p-value</th></tr>
<tr><td>mean</td><td>1.0131</td><td>0.3112</td></tr>
<tr><td>skewness</td><td>-8.0096</td><td>0.0000</td></tr>
<tr><td>kurtosis</td><td>41.0178</td><td>0.0000</td></tr>
<tr><td>normality</td><td>1755.86</td><td>0.0000</td></tr>
</table>


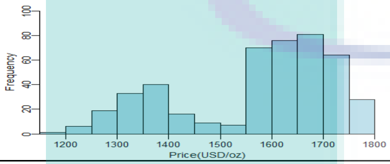
</td>
</tr>
<tr>
<td>4</td>
<td>
<table>
<tr><td></td><td>dt5</td></tr>
<tr><td>nobs</td><td>9.000000e+02</td></tr>
<tr><td>NAs</td><td>0.000000e+00</td></tr>
<tr><td>Minimum</td><td>1.058000e+03</td></tr>
<tr><td>Maximum</td><td>1.895000e+03</td></tr>
<tr><td>1. Quartile</td><td>1.327750e+03</td></tr>
<tr><td>3. Quartile</td><td>1.661000e+03</td></tr>
<tr><td>Mean</td><td>1.485927e+03</td></tr>
<tr><td>Median</td><td>1.537750e+03</td></tr>
<tr><td>Sum</td><td>1.337334e+06</td></tr>
<tr><td>SE Mean</td><td>7.047775e+00</td></tr>
<tr><td>LCL Mean</td><td>1.472095e+03</td></tr>
<tr><td>UCL Mean</td><td>1.499759e+03</td></tr>
<tr><td>Variance</td><td>4.470402e+04</td></tr>
<tr><td>Stdev</td><td>2.114332e+02</td></tr>
<tr><td>Skewness</td><td>-3.280950e-01</td></tr>
<tr><td>Kurtosis</td><td>-1.107732e+00</td></tr>
</table>
<table>
<tr><th>test</th><th>test statistic</th><th>p-value</th></tr>
<tr><td>mean</td><td>210.8364</td><td>0.0000</td></tr>
<tr><td>skewness</td><td>-4.0184</td><td>0.0000</td></tr>
<tr><td>kurtosis</td><td>-6.7832</td><td>0.0000</td></tr>
<tr><td>normality</td><td>61.8668</td><td>0.0000</td></tr>
</table>


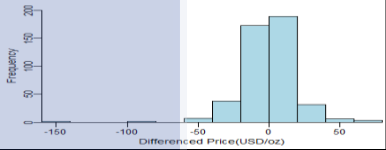
</td>
<td>
Box-Cox Transformation:
$\lambda = -0.4217 \rightarrow -0.5$
$$y_t^* = \frac{1}{\sqrt{y_t}}$$
<table>
<tr><td></td><td>tdt5</td></tr>
<tr><td>nobs</td><td>900.000000</td></tr>
<tr><td>NAs</td><td>0.000000</td></tr>
<tr><td>Minimum</td><td>0.022972</td></tr>
<tr><td>Maximum</td><td>0.030744</td></tr>
<tr><td>1. Quartile</td><td>0.024537</td></tr>
<tr><td>3. Quartile</td><td>0.027444</td></tr>
<tr><td>Mean</td><td>0.026153</td></tr>
<tr><td>Median</td><td>0.025501</td></tr>
<tr><td>Sum</td><td>23.537276</td></tr>
<tr><td>SE Mean</td><td>0.000066</td></tr>
<tr><td>LCL Mean</td><td>0.026024</td></tr>
<tr><td>UCL Mean</td><td>0.026281</td></tr>
<tr><td>Variance</td><td>0.000004</td></tr>
<tr><td>Stdev</td><td>0.001970</td></tr>
<tr><td>Skewness</td><td>0.591838</td></tr>
<tr><td>Kurtosis</td><td>-0.843198</td></tr>
</table>
</td>
<td>
Stationary data: First Differenced of transformed data
<table>
<tr><td></td><td>dtdt5</td></tr>
<tr><td>nobs</td><td>899.000000</td></tr>
<tr><td>NAs</td><td>0.000000</td></tr>
<tr><td>Minimum</td><td>-0.000598</td></tr>
<tr><td>Maximum</td><td>0.001254</td></tr>
<tr><td>1. Quartile</td><td>-0.000089</td></tr>
<tr><td>3. Quartile</td><td>0.000071</td></tr>
<tr><td>Mean</td><td>-0.000003</td></tr>
<tr><td>Median</td><td>-0.000010</td></tr>
<tr><td>Sum</td><td>-0.002653</td></tr>
<tr><td>SE Mean</td><td>0.000005</td></tr>
<tr><td>LCL Mean</td><td>-0.000013</td></tr>
<tr><td>UCL Mean</td><td>0.000007</td></tr>
<tr><td>Variance</td><td>0.000000</td></tr>
<tr><td>Stdev</td><td>0.000156</td></tr>
<tr><td>Skewness</td><td>0.950646</td></tr>
<tr><td>Kurtosis</td><td>6.716339</td></tr>
</table>
<table>
<tr><th>test</th><th>test statistic</th><th>p-value</th></tr>
<tr><td>mean</td><td>-0.5671</td><td>0.5708</td></tr>
<tr><td>skewness</td><td>11.6360</td><td>0.0000</td></tr>
<tr><td>kurtosis</td><td>41.1059</td><td>0.0000</td></tr>
<tr><td>normality</td><td>1836.4894</td><td>0.0000</td></tr>
</table>


</td>
</tr>
<tr>
<td>5</td>
<td>
<table>
<tr><td></td><td>dt6</td></tr>
<tr><td>nobs</td><td>6.750000e+02</td></tr>
<tr><td>NAs</td><td>0.000000e+00</td></tr>
<tr><td>Minimum</td><td>1.192000e+03</td></tr>
<tr><td>Maximum</td><td>1.895000e+03</td></tr>
<tr><td>1. Quartile</td><td>1.466625e+03</td></tr>
<tr><td>3. Quartile</td><td>1.690750e+03</td></tr>
<tr><td>Mean</td><td>1.580106e+03</td></tr>
<tr><td>Median</td><td>1.606000e+03</td></tr>
<tr><td>Sum</td><td>1.066572e+06</td></tr>
<tr><td>SE Mean</td><td>5.636216e+00</td></tr>
<tr><td>LCL Mean</td><td>1.569039e+03</td></tr>
<tr><td>UCL Mean</td><td>1.591173e+03</td></tr>
<tr><td>Variance</td><td>2.144268e+04</td></tr>
<tr><td>Stdev</td><td>1.464332e+02</td></tr>
<tr><td>Skewness</td><td>-4.447560e-01</td></tr>
<tr><td>Kurtosis</td><td>-6.858280e-01</td></tr>
</table>
<table>
<tr><th>test</th><th>test statistic</th><th>p-value</th></tr>
<tr><td>mean</td><td>280.3487</td><td>0.0000</td></tr>
<tr><td>skewness</td><td>-4.7178</td><td>0.0000</td></tr>
<tr><td>kurtosis</td><td>-3.6370</td><td>0.0003</td></tr>
<tr><td>normality</td><td>35.3177</td><td>0.0000</td></tr>
</table>


</td>
<td>
Box-Cox Transformation:
$\lambda = 6421 \rightarrow 0.5$
$$y_t^* = \sqrt{y_t}$$
<table>
<tr><td></td><td>tdt6</td></tr>
<tr><td>nobs</td><td>675.000000</td></tr>
<tr><td>NAs</td><td>0.000000</td></tr>
<tr><td>Minimum</td><td>34.525353</td></tr>
<tr><td>Maximum</td><td>43.531598</td></tr>
<tr><td>1. Quartile</td><td>38.296539</td></tr>
<tr><td>3. Quartile</td><td>41.118730</td></tr>
<tr><td>Mean</td><td>39.706772</td></tr>
<tr><td>Median</td><td>40.074930</td></tr>
<tr><td>Sum</td><td>26802.070816</td></tr>
<tr><td>SE Mean</td><td>0.071837</td></tr>
<tr><td>LCL Mean</td><td>39.565720</td></tr>
<tr><td>UCL Mean</td><td>39.847823</td></tr>
<tr><td>Variance</td><td>3.483377</td></tr>
<tr><td>Stdev</td><td>1.866381</td></tr>
<tr><td>Skewness</td><td>-0.524267</td></tr>
<tr><td>Kurtosis</td><td>-0.614317</td></tr>
</table>
</td>
<td>
Stationary data: First Differenced of transformed data
<table>
<tr><td></td><td>dtdt6</td></tr>
<tr><td>nobs</td><td>674.000000</td></tr>
<tr><td>NAs</td><td>0.000000</td></tr>
<tr><td>Minimum</td><td>-1.835758</td></tr>
<tr><td>Maximum</td><td>0.936997</td></tr>
<tr><td>1. Quartile</td><td>-0.109765</td></tr>
<tr><td>3. Quartile</td><td>0.128327</td></tr>
<tr><td>Mean</td><td>0.000388</td></tr>
<tr><td>Median</td><td>0.012316</td></tr>
<tr><td>Sum</td><td>0.261540</td></tr>
<tr><td>SE Mean</td><td>0.009607</td></tr>
<tr><td>LCL Mean</td><td>-0.018474</td></tr>
<tr><td>UCL Mean</td><td>0.019250</td></tr>
<tr><td>Variance</td><td>0.062201</td></tr>
<tr><td>Stdev</td><td>0.249401</td></tr>
<tr><td>Skewness</td><td>-1.005759</td></tr>
<tr><td>Kurtosis</td><td>6.898007</td></tr>
</table>
<table>
<tr><th>test</th><th>test statistic</th><th>p-value</th></tr>
<tr><td>mean</td><td>0.0404</td><td>0.9678</td></tr>
<tr><td>skewness</td><td>-10.6602</td><td>0.0000</td></tr>
<tr><td>kurtosis</td><td>36.5551</td><td>0.0000</td></tr>
<tr><td>normality</td><td>1461.8425</td><td>0.0000</td></tr>
</table>


</td>
</tr>
</table>

<table>
<tr><th>Data</th><th>Original Data</th><th>Transformed Data</th><th>Stationary Data</th></tr>
<tr>
<td>6</td>
<td>
<pre>
                     dt7
nobs   450.000000
NAs   0.000000
Minimum   1192.000000
Maximum   1791.750000
1. Quartile 1431.500000
3. Quartile 1686.875000
Mean   1573.477222
Median   1613.125000
Sum   708064.750000
SE Mean   7.088292
LCL Mean   1559.546875
UCL Mean   1587.407569
Variance   22609.747548
Stdev   150.365380
Skewness   -0.725606
Kurtosis   -0.700177
</pre>
<table>
<tr><td>test</td><td>test statistic</td><td>p-value</td></tr>
<tr><td>mean</td><td>221.9825</td><td>0.0000</td></tr>
<tr><td>skewness</td><td>-6.2839</td><td>0.0003</td></tr>
<tr><td>kurtosis</td><td>-3.0320</td><td>0.0024</td></tr>
<tr><td>normality</td><td>48.6770</td><td>0.0000</td></tr>
</table>


</td>
<td>
Box-Cox Transformation:
$\lambda = 0.9999 \rightarrow 1: y$
No transformation
<pre>
                     dt7
nobs   450.000000
NAs   0.000000
Minimum   1192.000000
Maximum   1791.750000
1. Quartile 1431.500000
3. Quartile 1686.875000
Mean   1573.477222
Median   1613.125000
Sum   708064.750000
SE Mean   7.088292
LCL Mean   1559.546875
UCL Mean   1587.407569
Variance   22609.747548
Stdev   150.365380
Skewness   -0.725606
Kurtosis   -0.700177
</pre>
</td>
<td>
Stationary data: Differenced y
<pre>
                    ddt7
nobs   449.000000
NAs   0.000000
Minimum   -140.500000
Maximum   77.000000
1. Quartile   -9.500000
3. Quartile   8.750000
Mean   -0.620267
Median   0.250000
Sum   -278.500000
SE Mean   0.882420
LCL Mean   -2.354464
UCL Mean   1.113929
Variance   349.620693
Stdev   18.698147
Skewness   -0.844473
Kurtosis   8.311841
</pre>
<table>
<tr><td>test</td><td>test statistic</td><td>p-value</td></tr>
<tr><td>mean</td><td>-0.7029</td><td>0.4825</td></tr>
<tr><td>skewness</td><td>-7.3054</td><td>0.0000</td></tr>
<tr><td>kurtosis</td><td>35.9511</td><td>0.0000</td></tr>
<tr><td>normality</td><td>1361.9921</td><td>0.0000</td></tr>
</table>


</td>
</tr>
</table>

Average Annual Return for Sample 1 – Sample 3 (significant mean)

Sample 1

```
> t2=prod(dldt2+1);t2 #product
[1] 3.336908
> r2=t2^(250/4499)-1;r2 #annual log retur
n
[1] 0.06925463
> R2=exp(r2)-1;R2 #annual simple return
[1] 0.07170906
> fv2=1*((R2+1)^18);fv2#compound return(f
v)=pv*[(Rt+1)^T]
[1] 3.478438
```

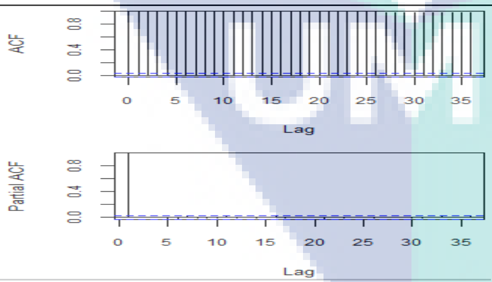
Sample 2

```
> t3=prod(dldt3+1);t3 #product
[1] 3.505386
> r3=t3^(250/2249)-1;r3 #annual log retur
           n
[1] 0.1496169
> R3=exp(r3)-1;R3 #annual simple return
[1] 0.1613892
> fv3=1*((R3+1)^9);fv3#compound return(fv
           =pv*[(Rt+1)^T]
[1] 3.844147
```

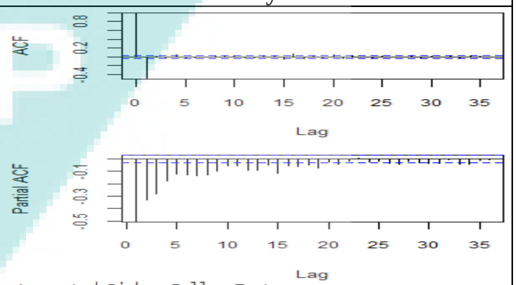
3. **Nonstationary in-mean**: (i) using sample ACF and sample PACF; (ii) using ADF-test

<table>
<tr><th>Data</th><th>Transformed Data</th><th>Stationary Data</th></tr>
<tr>
<td>
1

$k_{\max} = 36$

$k_{\max} = 31$
</td>
<td>


<pre>
Augmented Dickey-Fuller Test

data: ldt2
Dickey-Fuller = -1.0016, Lag order = 31, p-value = 0.9385
alternative hypothesis: stationary
</pre>
Null Hypothesis: LDT2_ESTIMATE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 31 (Fixed)
<table>
<tr><td></td><td></td><td>t-Statistic</td><td>Prob.*</td></tr>
<tr><td>Augmented Dickey-Fuller test statistic</td><td></td><td>-1.001607</td><td>0.9422</td></tr>
<tr><td>Test critical values:</td><td>1% level</td><td>-3.960092</td><td></td></tr>
<tr><td></td><td>5% level</td><td>-3.410811</td><td></td></tr>
<tr><td></td><td>10% level</td><td>-3.127201</td><td></td></tr>
</table>
*MacKinnon (1996) one-sided p-values.
</td>
<td>


<pre>
Augmented Dickey-Fuller Test

data: dldt2
Dickey-Fuller = -13.945, Lag order = 31, p-value = 0.01
alternative hypothesis: stationary
</pre>
Null Hypothesis: D(LDT2_ESTIMATE) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 31 (Fixed)
<table>
<tr><td></td><td></td><td>t-Statistic</td><td>Prob.*</td></tr>
<tr><td>Augmented Dickey-Fuller test statistic</td><td></td><td>-13.94541</td><td>0.0000</td></tr>
<tr><td>Test critical values:</td><td>1% level</td><td>-3.960093</td><td></td></tr>
<tr><td></td><td>5% level</td><td>-3.410811</td><td></td></tr>
<tr><td></td><td>10% level</td><td>-3.127201</td><td></td></tr>
</table>
*MacKinnon (1996) one-sided p-values.
</td>
</tr>
</table>

Sample: 1 4500
Included observations: 4500

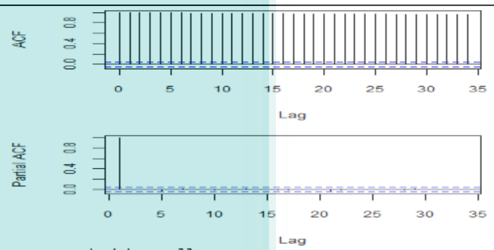
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.999	0.999	4495.7	0.000
		2	0.998	0.007	8985.4	0.000
		3	0.998	-0.001	13469.	0.000
		4	0.997	0.005	17946.	0.000
		5	0.996	-0.023	22417.	0.000
		6	0.995	-0.023	26882.	0.000
		7	0.994	0.013	31339.	0.000
		8	0.993	-0.013	35791.	0.000
		9	0.993	-0.004	40235.	0.000
		10	0.992	-0.011	44672.	0.000
		11	0.991	0.012	49103.	0.000
		12	0.990	-0.003	53527.	0.000
		13	0.989	0.007	57945.	0.000
		14	0.988	-0.003	62356.	0.000
		15	0.987	-0.001	66760.	0.000
		16	0.987	0.015	71157.	0.000
		17	0.986	-0.011	75548.	0.000
		18	0.985	0.008	79933.	0.000
		19	0.984	-0.002	84311.	0.000
		20	0.983	0.003	88683.	0.000
		21	0.982	-0.015	93048.	0.000
		22	0.982	0.003	97406.	0.000
		23	0.981	-0.006	101758	0.000
		24	0.980	-0.002	106103	0.000
		25	0.979	0.007	110442	0.000
		26	0.978	-0.003	114774	0.000
		27	0.977	0.006	119099	0.000
		28	0.976	0.006	123419	0.000
		29	0.976	0.004	127731	0.000
		30	0.975	-0.004	132038	0.000
		31	0.974	-0.002	136338	0.000
		32	0.973	0.005	140631	0.000
		33	0.972	0.010	144919	0.000
		34	0.972	0.004	149200	0.000
		35	0.971	0.003	153476	0.000
		36	0.970	0.011	157745	0.000

Sample: 1 4500
Included observations: 4499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.2247	0.635
		2	0.006	0.006	0.3854	0.825
		3	0.002	0.002	0.4010	0.940
		4	0.046	0.046	9.7870	0.044
		5	-0.009	-0.009	10.179	0.070
		6	-0.037	-0.038	16.498	0.011
		7	-0.017	-0.018	17.838	0.013
		8	0.005	0.003	17.958	0.022
		9	0.016	0.017	19.070	0.025
		10	-0.006	-0.003	19.245	0.037
		11	-0.042	-0.041	27.100	0.004
		12	-0.032	-0.035	31.745	0.002
		13	0.008	0.005	32.020	0.002
		14	0.010	0.012	32.494	0.003
		15	-0.028	-0.022	35.963	0.002
		16	0.049	0.051	46.834	0.000
		17	-0.001	-0.005	46.843	0.000
		18	0.008	0.003	47.163	0.000
		19	-0.007	-0.005	47.400	0.000
		20	0.029	0.027	51.287	0.000
		21	-0.021	-0.020	53.282	0.000
		22	-0.004	-0.004	53.344	0.000
		23	-0.022	-0.023	55.541	0.000
		24	-0.046	-0.049	64.928	0.000
		25	-0.025	-0.024	67.831	0.000
		26	-0.029	-0.028	71.667	0.000
		27	-0.013	-0.010	72.435	0.000
		28	-0.006	0.001	72.585	0.000
		29	-0.022	-0.023	74.712	0.000
		30	-0.007	-0.011	74.954	0.000
		31	-0.006	-0.004	75.108	0.000
		32	0.003	-0.001	75.147	0.000
		33	-0.031	-0.031	79.460	0.000
		34	0.009	0.006	79.829	0.000
		35	0.011	0.008	80.361	0.000
		36	-0.011	-0.022	80.917	0.000

2

$k_{\max} = 34$




```
Augmented Dickey-Fuller Test

data:  ldt3
Dickey-Fuller = -3.3392, Lag order = 26, p-value = 0.06391
alternative hypothesis: stationary
```


```
Augmented Dickey-Fuller Test

data:  dldt3
Dickey-Fuller = -10.65, Lag order = 26, p-value = 0.01
alternative hypothesis: stationary
```

$k_{\max} = 26$

Sample: 1 2250
Included observations: 2250

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.999	0.999	2247.1	0.000
		2	0.997	0.001	4489.3	0.000
		3	0.996	0.009	6726.7	0.000
		4	0.995	0.005	8959.5	0.000
		5	0.994	-0.022	11187.	0.000
		6	0.992	-0.001	13410.	0.000
		7	0.991	0.011	15628.	0.000
		8	0.990	0.005	17841.	0.000
		9	0.988	0.006	20050.	0.000
		10	0.987	-0.009	22254.	0.000
		11	0.986	0.001	24453.	0.000
		12	0.985	0.001	26648.	0.000
		13	0.983	0.011	28838.	0.000
		14	0.982	-0.001	31023.	0.000
		15	0.981	-0.000	33204.	0.000
		16	0.980	-0.002	35380.	0.000
		17	0.978	-0.010	37552.	0.000
		18	0.977	0.001	39719.	0.000
		19	0.976	-0.000	41881.	0.000
		20	0.974	0.010	44039.	0.000
		21	0.973	-0.019	46192.	0.000
		22	0.972	-0.010	48340.	0.000
		23	0.971	0.004	50483.	0.000
		24	0.969	0.009	52622.	0.000
		25	0.968	0.009	54756.	0.000
		26	0.967	0.005	56886.	0.000
		27	0.966	0.006	59011.	0.000
		28	0.964	-0.004	61132.	0.000
		29	0.963	0.014	63249.	0.000
		30	0.962	0.001	65361.	0.000
		31	0.961	0.001	67469.	0.000
		32	0.960	0.002	69573.	0.000
		33	0.958	-0.000	71673.	0.000
		34	0.957	-0.001	73769.	0.000

Null Hypothesis: LDT3_ESTIMATE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 26 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.339164	0.0603
Test critical values:	1% level	-3.962174	
	5% level	-3.411830	
	10% level	-3.127806	

*MacKinnon (1996) one-sided p-values.

Sample: 1 2250
Included observations: 2249

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.007	0.007	0.0976	0.755
		2	-0.032	-0.032	2.3714	0.306
		3	-0.007	-0.007	2.4933	0.477
		4	0.037	0.037	5.6636	0.226
		5	0.016	0.015	6.2207	0.285
		6	-0.061	-0.060	14.732	0.022
		7	-0.010	-0.008	14.964	0.036
		8	-0.021	-0.026	15.982	0.043
		9	0.041	0.039	19.834	0.019
		10	-0.009	-0.007	20.028	0.029
		11	-0.043	-0.038	24.132	0.012
		12	-0.041	-0.042	27.905	0.006
		13	0.023	0.018	29.114	0.006
		14	0.021	0.015	30.159	0.007
		15	-0.023	-0.015	31.325	0.008
		16	0.046	0.051	36.200	0.003
		17	-0.006	-0.011	36.273	0.004
		18	0.008	0.002	36.432	0.006
		19	-0.024	-0.023	37.740	0.006
		20	0.032	0.034	40.014	0.005
		21	-0.021	-0.022	41.008	0.006
		22	-0.012	-0.008	41.351	0.007
		23	-0.025	-0.030	42.731	0.007
		24	-0.056	-0.054	49.839	0.001
		25	-0.023	-0.027	50.999	0.002
		26	-0.043	-0.041	55.296	0.001
		27	-0.007	-0.009	55.414	0.001
		28	-0.010	-0.003	55.643	0.001
		29	-0.029	-0.036	57.595	0.001
		30	-0.007	-0.013	57.713	0.002
		31	-0.008	-0.011	57.874	0.002
		32	0.008	0.001	58.009	0.003
		33	-0.050	-0.051	63.828	0.001
		34	0.024	0.020	65.160	0.001

Null Hypothesis: D(LDT3_ESTIMATE) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 26 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-10.65042	0.0000
Test critical values:	1% level	-3.962176	
	5% level	-3.411831	
	10% level	-3.127806	

*MacKinnon (1996) one-sided p-values.

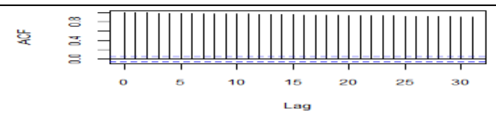
Data	Transformed Data	Stationary Data

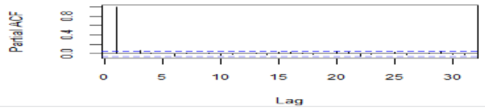
3

$k_{\max} = 31$

$k_{\max} = 22$

Transformed Data





```
Augmented Dickey-Fuller Test

data:  ldt4
Dickey-Fuller = 0.14876, Lag order = 22, p-value = 0.99
alternative hypothesis: stationary
```

Sample: 1 1125
Included observations: 1125

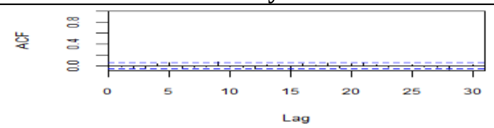
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.997	0.997	1120.4	0.000
		2	0.993	-0.002	2234.1	0.000
		3	0.990	0.069	3342.4	0.000
		4	0.987	0.009	4445.2	0.000
		5	0.985	-0.003	5542.6	0.000
		6	0.981	-0.052	6633.9	0.000
		7	0.978	0.013	7719.1	0.000
		8	0.975	0.001	8798.5	0.000
		9	0.972	0.017	9872.4	0.000
		10	0.969	-0.043	10940.	0.000
		11	0.966	-0.017	12001.	0.000
		12	0.962	0.002	13056.	0.000
		13	0.959	0.011	14105.	0.000
		14	0.956	-0.030	15147.	0.000
		15	0.953	0.033	16184.	0.000
		16	0.950	0.034	17215.	0.000
		17	0.947	0.015	18240.	0.000
		18	0.944	-0.009	19261.	0.000
		19	0.941	-0.001	20275.	0.000
		20	0.938	0.034	21285.	0.000
		21	0.936	0.028	22291.	0.000
		22	0.933	-0.045	23291.	0.000
		23	0.930	-0.006	24286.	0.000
		24	0.927	-0.000	25276.	0.000
		25	0.925	0.025	26262.	0.000
		26	0.922	-0.020	27242.	0.000
		27	0.919	0.003	28218.	0.000
		28	0.917	0.014	29189.	0.000
		29	0.914	0.045	30156.	0.000
		30	0.912	-0.017	31119.	0.000
		31	0.909	-0.019	32077.	0.000

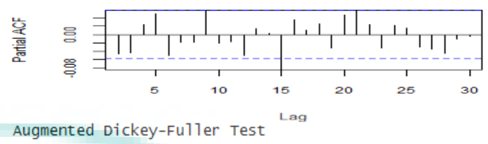
Null Hypothesis: LDT4_ESTIMATE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 22 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.148758	0.9977
Test critical values:	1% level	-3.966397	
	5% level	-3.413895	
	10% level	-3.129030	

*MacKinnon (1996) one-sided p-values.

Stationary Data





```
Augmented Dickey-Fuller Test

data:  dldt4
Dickey-Fuller = -6.8659, Lag order = 22, p-value = 0.01
alternative hypothesis: stationary
```

Sample: 1 1125
Included observations: 1124

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.017	-0.017	0.3349	0.563
		2	-0.046	-0.046	2.6991	0.259
		3	-0.042	-0.044	4.6891	0.196
		4	0.028	0.024	5.5654	0.234
		5	0.053	0.050	8.6938	0.122
		6	-0.051	-0.049	11.659	0.070
		7	-0.023	-0.019	12.269	0.092
		8	-0.017	-0.018	12.589	0.127
		9	0.067	0.059	17.707	0.039
		10	-0.019	-0.020	18.103	0.053
		11	-0.026	-0.017	18.890	0.063
		12	-0.052	-0.050	22.023	0.037
		13	0.023	0.015	22.618	0.047
		14	0.016	0.003	22.904	0.062
		15	-0.088	-0.084	31.836	0.007
		16	0.032	0.036	32.974	0.007
		17	0.012	0.011	33.140	0.011
		18	0.041	0.026	35.104	0.009
		19	-0.040	-0.032	36.915	0.008
		20	0.034	0.047	38.265	0.008
		21	0.062	0.059	42.608	0.004
		22	0.027	0.024	43.450	0.004
		23	-0.034	-0.032	44.795	0.004
		24	0.001	0.022	44.795	0.006
		25	0.027	0.016	45.643	0.007
		26	-0.022	-0.030	46.205	0.009
		27	-0.028	-0.035	47.116	0.010
		28	-0.063	-0.045	51.641	0.004
		29	-0.001	-0.011	51.642	0.006
		30	0.016	-0.004	51.922	0.008
		31	-0.022	-0.022	52.493	0.009

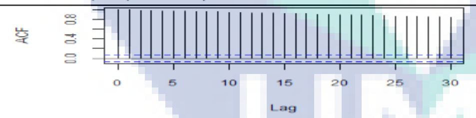
Null Hypothesis: D(LDT4_ESTIMATE) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 22 (Fixed)

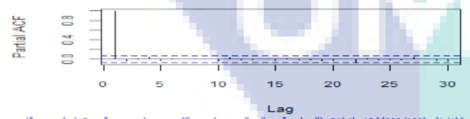
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.865871	0.0000
Test critical values:	1% level	-3.966404	
	5% level	-3.413899	
	10% level	-3.129032	

*MacKinnon (1996) one-sided p-values.

4

Transformed Data





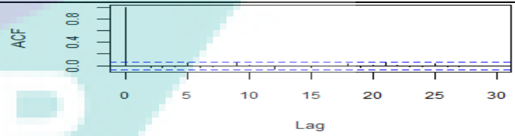
```
> adf.test(tdt5, alternative = c("stationary", "explosive"),21)#k=12((900/100)^(1/4))=20.78
46

 Augmented Dickey-Fuller Test

data:  tdt5
Dickey-Fuller = -0.55911, Lag order = 21, p-value = 0.9791
alternative hypothesis: stationary

> adf.test(tdt5, alternative = c("stationary", "explosive"),0)#k=0

 Augmented Dickey-Fuller Test

data:  tdt5
Dickey-Fuller = -0.86428, Lag order = 0, p-value = 0.9559
alternative hypothesis: stationary
```

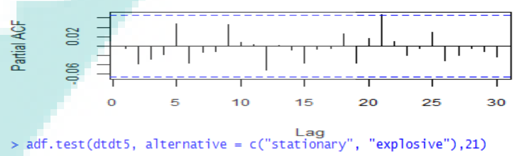
Null Hypothesis: TDT5_ESTIMATE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 21 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.559105	0.9806
Test critical values:	1% level	-3.968536	
	5% level	-3.414941	
	10% level	-3.129650	

*MacKinnon (1996) one-sided p-values.

Stationary Data





```
> adf.test(dtdt5, alternative = c("stationary", "explosive"),21)

 Augmented Dickey-Fuller Test

data:  dtdt5
Dickey-Fuller = -6.4382, Lag order = 21, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(dtdt5, alternative = c("stationary", "explosive"), 21) :
  p-value smaller than printed p-value
```

Null Hypothesis: D(TDT5_ESTIMATE) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 21 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.438224	0.0000
Test critical values:	1% level	-3.968548	
	5% level	-3.414947	
	10% level	-3.129653	

*MacKinnon (1996) one-sided p-values.

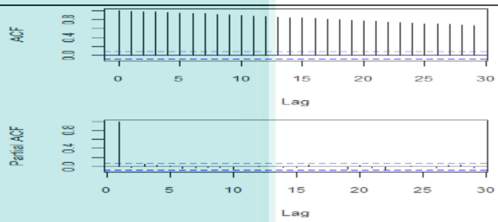
Sample: 1 900
Included observations: 900

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.994	0.994	892.97	0.000
		2	0.989	-0.030	1776.4	0.000
		3	0.983	0.019	2650.8	0.000
		4	0.978	0.043	3517.1	0.000
		5	0.973	-0.022	4374.9	0.000
		6	0.967	0.017	5224.8	0.000
		7	0.962	0.009	6066.8	0.000
		8	0.958	0.009	6901.3	0.000
		9	0.953	0.011	7728.5	0.000
		10	0.948	-0.036	8547.9	0.000
		11	0.943	0.026	9359.9	0.000
		12	0.938	-0.007	10164.	0.000
		13	0.933	-0.004	10962.	0.000
		14	0.929	0.006	11752.	0.000
		15	0.924	-0.019	12534.	0.000
		16	0.919	0.021	13309.	0.000
		17	0.914	-0.012	14077.	0.000
		18	0.909	-0.001	14838.	0.000
		19	0.904	-0.035	15592.	0.000
		20	0.899	-0.020	16337.	0.000
		21	0.894	0.004	17074.	0.000
		22	0.888	-0.067	17803.	0.000
		23	0.882	0.014	18522.	0.000
		24	0.876	-0.003	19233.	0.000
		25	0.870	-0.002	19936.	0.000
		26	0.865	-0.001	20631.	0.000
		27	0.860	0.044	21318.	0.000
		28	0.855	0.011	21998.	0.000
		29	0.849	-0.015	22670.	0.000
		30	0.844	-0.033	23335.	0.000

Sample: 1 900
Included observations: 899

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.005	-0.005	0.0270	0.869
		2	-0.038	-0.038	1.3485	0.510
		3	-0.027	-0.028	2.0297	0.566
		4	-0.016	-0.018	2.2532	0.689
		5	0.049	0.047	4.4340	0.489
		6	-0.036	-0.037	5.5885	0.471
		7	-0.015	-0.013	5.7932	0.564
		8	-0.012	-0.012	5.9152	0.657
		9	0.048	0.046	7.9730	0.537
		10	0.012	0.008	8.1145	0.618
		11	-0.003	0.003	8.1243	0.702
		12	-0.053	-0.051	10.710	0.554
		13	-0.000	0.002	10.710	0.635
		14	-0.000	-0.009	10.710	0.709
		15	-0.036	-0.037	11.912	0.686
		16	-0.006	-0.007	11.940	0.748
		17	-0.008	-0.005	11.996	0.800
		18	0.034	0.026	13.081	0.787
		19	-0.032	-0.036	14.012	0.783
		20	0.013	0.016	14.163	0.822
		21	0.065	0.067	18.071	0.645
		22	0.008	0.010	18.132	0.698
		23	-0.019	-0.019	18.473	0.731
		24	-0.015	-0.005	18.692	0.768
		25	0.029	0.030	19.497	0.773
		26	-0.025	-0.032	20.066	0.788
		27	-0.017	-0.020	20.326	0.817
		28	-0.011	-0.006	20.433	0.848
		29	-0.010	-0.012	20.534	0.876
		30	-0.012	-0.023	20.660	0.898

5




```
> adf.test(tdt6, alternative = c("stationary", "explosive"),20)#k=12((675/100)^(1/4))=19.34
22

 Augmented Dickey-Fuller Test

data:  tdt6
Dickey-Fuller = -2.0242, Lag order = 20, p-value = 0.5681
alternative hypothesis: stationary

> adf.test(tdt6, alternative = c("stationary", "explosive"),0)#k=0

 Augmented Dickey-Fuller Test

data:  tdt6
Dickey-Fuller = -1.9083, Lag order = 0, p-value = 0.6171
alternative hypothesis: stationary
```

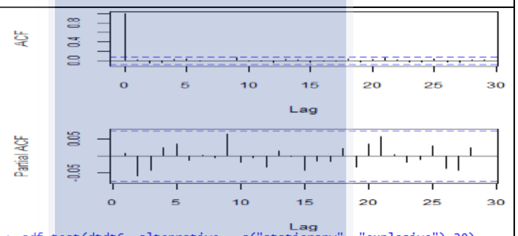
Sample: 1 675
Included observations: 675

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.992	0.992	667.70	0.000
		2	0.985	0.016	1326.6	0.000
		3	0.978	0.019	1977.1	0.000
		4	0.972	0.039	2620.1	0.000
		5	0.965	-0.039	3255.0	0.000
		6	0.957	-0.072	3880.3	0.000
		7	0.950	0.041	4497.2	0.000
		8	0.942	-0.022	5105.3	0.000
		9	0.934	-0.023	5704.4	0.000
		10	0.926	-0.067	6293.2	0.000
		11	0.917	0.017	6872.3	0.000
		12	0.909	-0.009	7441.9	0.000
		13	0.902	0.057	8003.0	0.000
		14	0.893	-0.050	8554.7	0.000
		15	0.885	0.009	9097.3	0.000
		16	0.877	-0.008	9630.7	0.000
		17	0.869	-0.036	10155.	0.000
		18	0.860	-0.007	10669.	0.000
		19	0.851	-0.024	11174.	0.000
		20	0.843	0.027	11669.	0.000
		21	0.834	-0.051	12155.	0.000
		22	0.824	-0.062	12629.	0.000
		23	0.814	-0.009	13093.	0.000
		24	0.804	0.019	13547.	0.000
		25	0.795	0.022	13992.	0.000
		26	0.786	0.010	14427.	0.000
		27	0.777	0.012	14852.	0.000
		28	0.769	0.046	15270.	0.000
		29	0.761	0.036	15680.	0.000

Null Hypothesis: TDT6_ESTIMATE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 20 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.060466	0.5664
Test critical values:	1% level	-3.972148	
	5% level	-3.416704	
	10% level	-3.130694	

*MacKinnon (1996) one-sided p-values.




```
> adf.test(dtdt6, alternative = c("stationary", "explosive"),20)

 Augmented Dickey-Fuller Test

data:  dtdt6
Dickey-Fuller = -5.5404, Lag order = 20, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(dtdt6, alternative = c("stationary", "explosive"), 20) :
  p-value smaller than printed p-value
```

Sample: 1 675
Included observations: 674

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.017	-0.017	0.2044	0.651
		2	-0.078	-0.078	4.3438	0.114
		3	-0.047	-0.050	5.8292	0.120
		4	0.036	0.028	6.7151	0.152
		5	0.047	0.041	8.2160	0.145
		6	-0.059	-0.055	10.614	0.101
		7	-0.001	0.006	10.615	0.156
		8	0.019	0.014	10.854	0.210
		9	0.100	0.094	17.708	0.039
		10	-0.014	-0.006	17.845	0.058
		11	-0.013	0.007	17.958	0.083
		12	-0.071	-0.070	21.397	0.045
		13	0.060	0.051	23.858	0.032
		14	-0.008	-0.023	23.899	0.047
		15	-0.036	-0.023	24.791	0.053
		16	0.026	0.029	25.262	0.065
		17	-0.002	-0.006	25.264	0.089
		18	0.061	0.044	27.848	0.064
		19	-0.054	-0.041	29.904	0.053
		20	0.055	0.064	32.013	0.043
		21	0.064	0.071	34.862	0.029
		22	-0.012	-0.015	34.969	0.039
		23	-0.065	-0.052	37.924	0.026
		24	-0.036	-0.030	38.830	0.028
		25	0.010	-0.013	38.895	0.038
		26	0.020	0.008	39.188	0.047
		27	-0.075	-0.083	43.185	0.025
		28	-0.053	-0.038	45.134	0.021
		29	-0.029	-0.065	45.745	0.025

Null Hypothesis: D(TDT6_ESTIMATE) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 20 (Fixed)

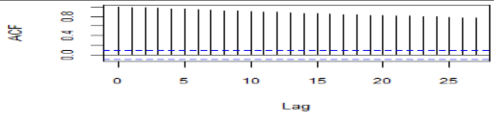
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.965426	0.0002
Test critical values:	1% level	-3.972169	
	5% level	-3.416715	
	10% level	-3.130700	

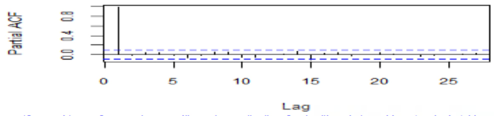
*MacKinnon (1996) one-sided p-values.

Data	Transformed Data	Stationary Data
6		

Data 6, $k_{max} = 27$

Transformed Data






```
> adf.test(dt7, alternative = c("stationary", "explosive"),18)#k=12((450/100)^(1/4))=18

 Augmented Dickey-Fuller Test

data:  dt7
Dickey-Fuller = -1.7495, Lag order = 18, p-value = 0.6837
alternative hypothesis: stationary
```

$k_{max} = 18$

```
> adf.test(dt7, alternative = c("stationary", "explosive"),0)#k=0

 Augmented Dickey-Fuller Test

data:  dt7
Dickey-Fuller = -2.1778, Lag order = 0, p-value = 0.5026
alternative hypothesis: stationary
```

Sample: 1 450
Included observations: 450

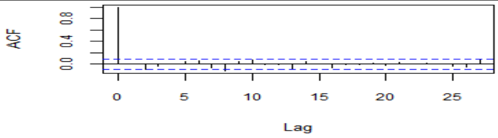
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.989	0.989	443.35	0.000
		2	0.978	-0.016	877.91	0.000
		3	0.969	0.050	1304.8	0.000
		4	0.960	0.038	1724.9	0.000
		5	0.951	-0.016	2137.9	0.000
		6	0.940	-0.061	2542.9	0.000
		7	0.930	-0.013	2939.7	0.000
		8	0.920	0.055	3329.5	0.000
		9	0.913	0.059	3713.8	0.000
		10	0.904	-0.046	4091.6	0.000
		11	0.894	-0.054	4461.9	0.000
		12	0.884	0.008	4825.0	0.000
		13	0.875	0.011	5181.4	0.000
		14	0.868	0.066	5532.6	0.000
		15	0.858	-0.099	5876.9	0.000
		16	0.849	0.027	6214.4	0.000
		17	0.840	0.050	6546.2	0.000
		18	0.832	-0.028	6872.3	0.000
		19	0.824	0.005	7193.0	0.000
		20	0.817	0.043	7508.7	0.000
		21	0.810	0.021	7819.6	0.000
		22	0.803	0.005	8126.1	0.000
		23	0.796	-0.040	8427.8	0.000
		24	0.788	-0.012	8724.6	0.000
		25	0.781	0.007	9016.2	0.000
		26	0.774	0.016	9303.4	0.000
		27	0.767	0.028	9586.6	0.000

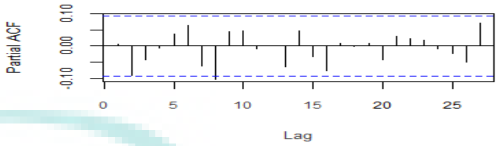
Exogenous: Constant, Linear Trend
Lag Length: 18 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.749541	0.7274
Test critical values:	1% level	-3.979493	
	5% level	-3.420283	
	10% level	-3.132811	

*MacKinnon (1996) one-sided p-values.

Stationary Data






```
 Augmented Dickey-Fuller Test

data:  ddt7
Dickey-Fuller = -5.225, Lag order = 18, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(ddt7, alternative = c("stationary", "explosive"), 18) :
  p-value smaller than printed p-value
```

Sample: 1 450
Included observations: 449

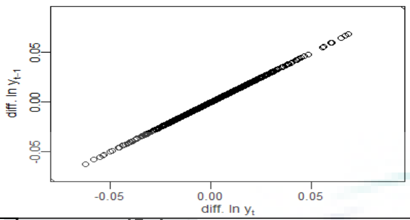
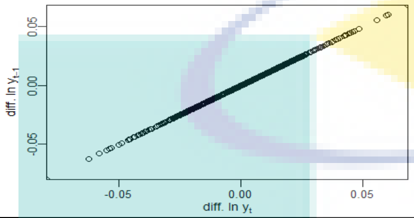
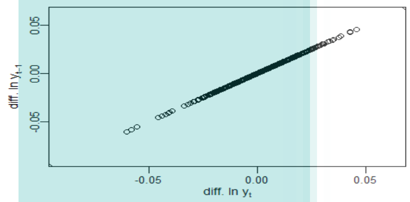
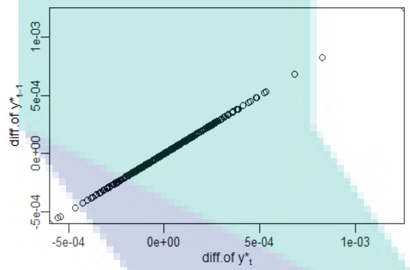
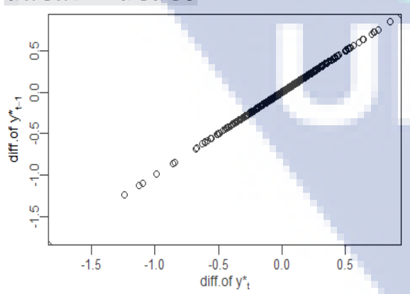
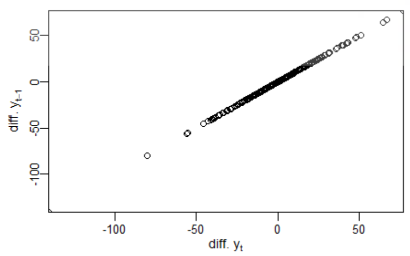
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.005	0.005	0.0117	0.914
		2	-0.091	-0.091	3.7321	0.155
		3	-0.044	-0.043	4.6074	0.203
		4	0.002	-0.006	4.6099	0.330
		5	0.044	0.037	5.4954	0.358
		6	0.064	0.062	7.3670	0.288
		7	-0.068	-0.062	9.4886	0.219
		8	-0.114	-0.101	15.460	0.051
		9	0.049	0.044	16.544	0.056
		10	0.069	0.046	18.718	0.044
		11	-0.003	-0.009	18.722	0.066
		12	-0.011	0.002	18.778	0.094
		13	-0.081	-0.064	21.798	0.059
		14	0.039	0.045	22.506	0.069
		15	0.002	-0.032	22.509	0.095
		16	-0.062	-0.076	24.315	0.083
		17	-0.011	0.007	24.372	0.110
		18	-0.004	-0.002	24.381	0.143
		19	0.013	0.008	24.463	0.179
		20	-0.024	-0.043	24.744	0.211
		21	0.031	0.029	25.194	0.239
		22	0.003	0.022	25.197	0.288
		23	0.022	0.018	25.423	0.329
		24	0.009	-0.008	25.461	0.381
		25	-0.034	-0.023	26.006	0.407
		26	-0.055	-0.050	27.470	0.385
		27	0.072	0.071	29.939	0.317

Exogenous: Constant, Linear Trend
Lag Length: 18 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.225023	0.0001
Test critical values:	1% level	-3.979543	
	5% level	-3.420308	
	10% level	-3.132826	

*MacKinnon (1996) one-sided p-values.

4. Preliminary of Linearity Test and Portmanteau Test

Data	Linearity Test	Portmanteau Test: Ljung-Box *Q*-test
1	data: dldt2 	$k = \ln T$ (Tsay's suggestion): X-squared = 19.07, df = 9, p-value = 0.0246 k = 10 (Hyndman's suggestion) X-squared = 19.245, df = 10, p-value = 0.03726 k =15 (Engle's suggestion) X-squared = 35.963, df = 15, p-value = 0.00179
2	data: dldt3 	$k = \ln T$ (Tsay's suggestion) X-squared = 15.982, df = 8, p-value = 0.04263 k = 10 (Hyndman's suggestion) X-squared = 20.029, df = 10, p-value = 0.02898 k =15 (Engle's suggestion) X-squared = 31.326, df = 15, p-value = 0.00794 3
3	data: dldt4 	$k = \ln T$ (Tsay's suggestion) X-squared = 12.269, df = 7, p-value = 0.09206 k = 10 (Hyndman's suggestion) X-squared = 18.103, df = 10, p-value = 0.05325 k =15 (Engle's suggestion) X-squared = 31.836, df = 15, p-value = 0.00677 7
4	data: dtdt5 	> Box.test(dtdt5,lag=7,type='Ljung')#Tsay's suggestion,k=ln 899=6.8013 Box-Ljung test data: dtdt5 X-squared = 5.7932, df = 7, p-value = 0.5641 > Box.test(dtdt5,lag=10,type='Ljung')#hyndman's suggestion k=10 Box-Ljung test data: dtdt5 X-squared = 8.1145, df = 10, p-value = 0.6177 > Box.test(dtdt5,lag=15,type='Ljung')#Engle's suggestion Box-Ljung test data: dtdt5 X-squared = 11.912, df = 15, p-value = 0.6857
5	data: dtdt6 	> Box.test(dtdt6,lag=7,type='Ljung')#Tsay's suggestion,k=ln 674=6.5132 Box-Ljung test data: dtdt6 X-squared = 5.6586, df = 7, p-value = 0.5801 > Box.test(dtdt6,lag=10,type='Ljung')#hyndman's suggestion k=10 Box-Ljung test data: dtdt6 X-squared = 9.0999, df = 10, p-value = 0.5227 > Box.test(dtdt6,lag=15,type='Ljung')#Engle's suggestion Box-Ljung test data: dtdt6 X-squared = 11.898, df = 15, p-value = 0.6867
6	data: ddt7 	$k = \ln T$ (Tsay's suggestion) X-squared = 7.367, df = 6, p-value = 0.2882 k = 10 (Hyndman's suggestion) X-squared = 22.509, df = 15, p-value = 0.09514 k =15 (Engle's suggestion) X-squared = 18.718, df = 10, p-value = 0.04399 k = 10 (Hyndman's suggestion) X-squared = 7.2201, df = 10, p-value = 0.7045 k =15 (Engle's suggestion) X-squared = 9.9381, df = 15, p-value = 0.8236

5. Possible Box-Jenkins model and Preliminary of Heteroscedasticity test

Data	Possible Model	Heteroscedasticity test

1

Possible Model

Data: dldt2

AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	o	o	o	x	o	x	o	o	o	o	x	x	o
1	x	o	o	x	o	o	o	o	o	o	o	x	o
2	x	x	o	o	o	o	o	o	o	o	o	o	o
3	x	x	x	o	o	x	o	o	o	o	o	o	o
4	x	x	x	x	o	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o
6	x	x	x	x	x	o	o	o	o	o	o	o	o

ARIMA(0,1,0)

Dependent Variable: D(LDT2_ESTIMATE)
Method: Least Squares
Date: 03/13/17 Time: 01:20
Sample (adjusted): 2 4500
Included observations: 4499 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000324	0.000157	2.056480	0.0398

R-squared	0.000000	Mean dependent var	0.000324
Adjusted R-squared	0.000000	S.D. dependent var	0.010553
S.E. of regression	0.010553	Akaike info criterion	-6.264609
Sum squared resid	0.500914	Schwarz criterion	-6.263184
Log likelihood	14093.24	Hannan-Quinn criter.	-6.264107
Durbin-Watson stat	2.013938		

Heteroscedasticity test

Correlogram of Residuals Squared

Date: 03/13/17 Time: 01:23
Sample: 2 4500
Included observations: 4499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.164	0.164	121.09	0.000
		2	0.123	0.098	188.85	0.000
		3	0.164	0.134	309.41	0.000
		4	0.130	0.080	385.09	0.000
		5	0.164	0.115	506.19	0.000
		6	0.134	0.067	587.08	0.000
		7	0.127	0.060	659.79	0.000
		8	0.127	0.054	732.73	0.000
		9	0.098	0.022	776.22	0.000
		10	0.085	0.010	809.15	0.000
		11	0.084	0.012	840.74	0.000
		12	0.154	0.094	948.24	0.000
		13	0.112	0.035	1004.9	0.000
		14	0.092	0.021	1043.0	0.000
		15	0.110	0.037	1097.9	0.000
		16	0.077	0.001	1124.6	0.000
		17	0.150	0.083	1226.7	0.000
		18	0.111	0.027	1282.0	0.000
		19	0.092	0.015	1320.5	0.000
		20	0.130	0.048	1397.3	0.000

```
> Box.test(r2^2,lag=10,type='Ljung') #Ljung-Box test for squared residuals

 Box-Ljung test

data:  r2^2
X-squared = 809.15, df = 10, p-value < 2.2e-16

> Box.test(r2^2,lag=15,type='Ljung')

 Box-Ljung test

data:  r2^2
X-squared = 1097.9, df = 15, p-value < 2.2e-16
```

2

Possible Model

Data: dldt3

AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	o	o	o	o	o	x	o	o	o	o	x	o	o
1	x	o	o	o	o	x	o	o	o	o	o	x	o
2	x	x	o	o	o	x	o	o	o	o	o	o	o
3	x	x	x	o	o	x	o	o	o	o	o	o	o
4	x	x	x	x	o	x	o	o	o	o	o	o	o
5	x	x	x	x	o	o	o	o	o	o	o	o	o
6	x	x	x	o	x	x	o	o	o	o	o	x	o

ARIMA(0,1,0)

Dependent Variable: D(LDT3)
Method: Least Squares
Date: 03/13/17 Time: 04:24
Sample (adjusted): 2 2250
Included observations: 2249 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000639	0.000269	2.380272	0.0174

R-squared	0.000000	Mean dependent var	0.000639
Adjusted R-squared	0.000000	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-5.888675
Sum squared resid	0.364510	Schwarz criterion	-5.886133
Log likelihood	6622.815	Hannan-Quinn criter.	-5.887747
Durbin-Watson stat	1.986621		

Heteroscedasticity test

Correlogram of Residuals Squared

Date: 03/13/17 Time: 04:26
Sample: 2 2250
Included observations: 2249

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.096	0.096	20.906	0.000
		2	0.120	0.112	53.368	0.000
		3	0.172	0.155	120.34	0.000
		4	0.115	0.080	149.98	0.000
		5	0.168	0.126	213.39	0.000
		6	0.137	0.082	255.73	0.000
		7	0.138	0.079	298.89	0.000
		8	0.142	0.072	344.34	0.000
		9	0.107	0.034	370.18	0.000
		10	0.101	0.021	393.22	0.000
		11	0.094	0.013	413.15	0.000
		12	0.203	0.137	506.51	0.000
		13	0.123	0.047	540.98	0.000
		14	0.100	0.019	563.63	0.000
		15	0.128	0.036	601.03	0.000
		16	0.085	0.002	617.59	0.000
		17	0.201	0.121	709.39	0.000
		18	0.122	0.034	743.37	0.000
		19	0.100	0.010	766.29	0.000
		20	0.157	0.053	822.26	0.000

```
> Box.test(r3^2,lag=10,type='Ljung') #Ljung-Box test for squared residuals

 Box-Ljung test

data:  r3^2
X-squared = 393.22, df = 10, p-value < 2.2e-16

> Box.test(r3^2,lag=15,type='Ljung')

 Box-Ljung test

data:  r3^2
X-squared = 601.03, df = 15, p-value < 2.2e-16
```

3

Possible Model

Data: dldt4

AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	o	o	o	o	o	o	o	o	x	o	o	o	o
1	x	o	o	o	o	x	o	o	o	o	o	o	o
2	x	x	o	o	o	x	o	o	o	o	o	x	o
3	x	x	x	o	o	o	o	o	o	o	o	o	o
4	x	x	x	x	o	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o
6	x	x	x	x	x	o	o	o	o	o	o	x	o

ARIMA(0,1,0)

Dependent Variable: D(LDT4)
Method: Least Squares
Date: 03/13/17 Time: 06:05
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000370	0.000365	1.013073	0.3112

R-squared	0.000000	Mean dependent var	0.000370
Adjusted R-squared	0.000000	S.D. dependent var	0.012243
S.E. of regression	0.012243	Akaike info criterion	-5.966886
Sum squared resid	0.168319	Schwarz criterion	-5.962415
Log likelihood	3354.390	Hannan-Quinn criter.	-5.965196
Durbin-Watson stat	2.033911		

Heteroscedasticity test

Correlogram of Residuals Squared

Date: 03/13/17 Time: 06:06
Sample: 2 1125
Included observations: 1124

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.041	0.041	1.9049	0.168
		2	0.048	0.046	4.4952	0.106
		3	0.035	0.031	5.8701	0.118
		4	0.004	-0.001	5.8848	0.208
		5	0.011	0.008	6.0164	0.305
		6	0.042	0.040	7.9749	0.240
		7	0.044	0.041	10.208	0.177
		8	0.034	0.026	11.485	0.176
		9	0.048	0.040	14.142	0.117
		10	0.048	0.040	16.795	0.079
		11	0.014	0.004	17.005	0.108
		12	0.065	0.056	21.765	0.040
		13	0.035	0.025	23.187	0.039
		14	0.021	0.009	23.676	0.050
		15	0.086	0.075	32.114	0.006
		16	0.016	0.001	32.409	0.009
		17	0.016	0.002	32.709	0.012
		18	0.065	0.052	37.612	0.004
		19	0.016	0.001	37.893	0.006
		20	0.060	0.046	41.954	0.003

```
> Box.test(r4^2,lag=10,type='Ljung') #Ljung-Box test for squared residuals

 Box-Ljung test

data:  r4^2
X-squared = 16.795, df = 10, p-value = 0.07903

> Box.test(r4^2,lag=12,type='Ljung')

 Box-Ljung test

data:  r4^2
X-squared = 21.765, df = 12, p-value = 0.04024

> Box.test(r4^2,lag=15,type='Ljung')

 Box-Ljung test

data:  r4^2
X-squared = 32.114, df = 15, p-value = 0.006212
```

Data	Possible Model	Heteroscedacity test

Data 4

Possible Model

```
data:dtdt5
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o o o o o o o o o o  o  o
1 x o o o o o o o o o o  o  o
2 x x o o o o o o o o o  o  o
3 x x x o o o o o o o o  o  o
4 x x o x o o o o o o o  o  o
5 x x o x x o o o o o o  o  o
6 x x x o x o o o o o o  o  o
```

ARIMA(0,1,0)

Dependent Variable: D(TDT5_ESTIMATE)
Method: Least Squares
Date: 07/21/17 Time: 12:33
Sample (adjusted): 2 900
Included observations: 899 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.95E-06	5.20E-06	-0.567079	0.5708

R-squared	0.000000	Mean dependent var	-2.95E-06
Adjusted R-squared	0.000000	S.D. dependent var	0.000156
S.E. of regression	0.000156	Akaike info criterion	-14.69188
Sum squared resid	2.19E-05	Schwarz criterion	-14.68654
Log likelihood	6605.002	Hannan-Quinn criter.	-14.68984
Durbin-Watson stat	2.006831		

Heteroscedacity test

```
> Box.test(r5^2,lag=10,type='Ljung') #Ljung-Box test for squared residuals

 Box-Ljung test

data:  r5^2
X-squared = 17.393, df = 10, p-value = 0.0661

> Box.test(r5^2,lag=15,type='Ljung')

 Box-Ljung test

data:  r5^2
X-squared = 20.181, df = 15, p-value = 0.1651
```

Data 5

Possible Model

```
data:  dtdt6
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o o o o o o o o o o  o  o
1 x o o o o o o o o o o  o  o
2 x x o o o o o o o o o  o  o
3 x x x o o o o o o o o  o  o
4 x x x x o o o o o o o  o  o
5 x o x x o o o o o o o  o  o
6 x x o x o x o o o o o  o  o
```

ARIMA(0,1,0)

Dependent Variable: D(TDT6_ESTIMATE)
Method: Least Squares
Date: 07/21/17 Time: 14:02
Sample (adjusted): 2 675
Included observations: 674 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006431	0.008761	0.734002	0.4632

R-squared	0.000000	Mean dependent var	0.006431
Adjusted R-squared	0.000000	S.D. dependent var	0.227448
S.E. of regression	0.227448	Akaike info criterion	-0.122306
Sum squared resid	34.81609	Schwarz criterion	-0.115610
Log likelihood	42.21715	Hannan-Quinn criter.	-0.119713
Durbin-Watson stat	2.034393		

Heteroscedacity test

```
> Box.test(r6^2,lag=10,type='Ljung') #Ljung-Box test for squared residuals

 Box-Ljung test

data:  r6^2
X-squared = 16.84, df = 10, p-value = 0.07798

> Box.test(r6^2,lag=15,type='Ljung')

 Box-Ljung test

data:  r6^2
X-squared = 20.899, df = 15, p-value = 0.1401
```

Data 6

Possible Model

```
data:  ddt7
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o o o o o o x o o o  o  o
1 o o o o o o o x o o o  o  o
2 x x o o o o o o o o o  o  o
3 x x x o o o o o o o o  o  o
4 x x o x o o o o o o o  o  o
5 x x x x x o o o o o o  o  o
6 x x x x x x o o o o o  o  o
```

ARIMA(1,1,0) or **(0,1,0)**

Method: Least Squares
Date: 03/02/17 Time: 18:27
Sample (adjusted): 3 450
Included observations: 448 after adjustments
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.618218	0.889902	-0.694703	0.4876
AR(1)	0.005093	0.047357	0.107550	0.9144

R-squared	0.000026	Mean dependent var	-0.618304
Adjusted R-squared	-0.002216	S.D. dependent var	18.71900
S.E. of regression	18.73973	Akaike info criterion	8.703624
Sum squared resid	156625.2	Schwarz criterion	8.721949
Log likelihood	-1947.612	Hannan-Quinn criter.	8.710847
F-statistic	0.011567	Durbin-Watson stat	1.991429
Prob(F-statistic)	0.914401		

Method: Least Squares
Date: 03/02/17 Time: 18:29
Sample (adjusted): 2 450
Included observations: 449 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.620267	0.882420	-0.702916	0.4825

R-squared	0.000000	Mean dependent var	-0.620267
Adjusted R-squared	0.000000	S.D. dependent var	18.69815
S.E. of regression	18.69815	Akaike info criterion	8.696951
Sum squared resid	156630.1	Schwarz criterion	8.706098
Log likelihood	-1951.465	Hannan-Quinn criter.	8.700556
Durbin-Watson stat	1.989532		

Heteroscedacity test

Based on (0,1,0)

```
> Box.test(r9^2,lag=10,type='Ljung') #Ljung-Box test for squared residuals

 Box-Ljung test

data:  r9^2
X-squared = 1.8682, df = 10, p-value = 0.9973

> Box.test(r9^2,lag=15,type='Ljung')

 Box-Ljung test

data:  r9^2
X-squared = 3.2326, df = 15, p-value = 0.9994
```

Correlogram of Residuals Squared

Date: 03/02/17 Time: 18:30
Sample: 2 450
Included observations: 449

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.002	-0.002	0.0018	0.966
		2	-0.008	-0.008	0.0344	0.983
		3	-0.019	-0.019	0.1964	0.978
		4	-0.011	-0.011	0.2510	0.993
		5	-0.002	-0.003	0.2536	0.998
		6	0.010	0.009	0.2993	1.000
		7	0.051	0.051	1.4892	0.983
		8	0.008	0.008	1.5189	0.992
		9	0.011	0.012	1.5740	0.997
		10	-0.025	-0.023	1.8682	0.997
		11	-0.012	-0.010	1.9320	0.999
		12	-0.032	-0.032	2.4053	0.998
		13	-0.029	-0.031	2.7906	0.999
		14	-0.024	-0.029	3.0659	0.999
		15	0.019	0.016	3.2326	0.999
		16	-0.008	-0.011	3.2653	1.000
		17	0.026	0.027	3.5717	1.000
		18	0.026	0.028	3.8836	1.000
		19	-0.008	-0.003	3.9148	1.000
		20	0.006	0.011	3.9293	1.000

STAGE II: BOX-JENKINS – GARCH PARAMETER ESTIMATION

SAMPLE 1: Refer to Chapter 4.2 and Appendix 2

SAMPLE 2

1. ARIMA (0,1,0)-GARCH (1,1)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000586	0.000217	2.696675	0.0070
Variance Equation				
C	1.15E-06	3.15E-07	3.648057	0.0003
RESID(-1)^2	0.047859	0.005482	8.729874	0.0000
GARCH(-1)	0.945348	0.006267	150.8374	0.0000

R-squared	-0.000017	Mean dependent var	0.000639
Adjusted R-squared	-0.000017	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.091298
Sum squared resid	0.364517	Schwarz criterion	-6.081128
Log likelihood	6853.665	Hannan-Quinn criter.	-6.087586
Durbin-Watson stat	1.986586		

Heteroskedasticity Test: ARCH

F-statistic	0.668990	Prob. F(10,2228)	0.7542
Obs*R-squared	6.702799	Prob. Chi-Square(10)	0.7532

2. ARIMA (0,1,0)-GARCH (1,2)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000512	0.000214	2.394022	0.0167
Variance Equation				
C	3.57E-07	1.18E-07	3.016582	0.0026
RESID(-1)^2	0.014239	0.003869	3.680771	0.0002
GARCH(-1)	1.729060	0.077562	22.29267	0.0000
GARCH(-2)	-0.745368	0.073510	-10.13972	0.0000

R-squared	-0.000100	Mean dependent var	0.000639
Adjusted R-squared	-0.000100	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.094498
Sum squared resid	0.364547	Schwarz criterion	-6.081785
Log likelihood	6858.263	Hannan-Quinn criter.	-6.089858
Durbin-Watson stat	1.986423		

Heteroskedasticity Test: ARCH

F-statistic	0.391313	Prob. F(10,2228)	0.9510
Obs*R-squared	3.925556	Prob. Chi-Square(10)	0.9506

3. ARIMA (0,1,0)-GARCH (1,3)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000510	0.000214	2.385342	0.0171
Variance Equation				
C	3.15E-07	2.34E-07	1.347890	0.1777
RESID(-1)^2	0.012622	0.009016	1.399979	0.1615
GARCH(-1)	1.901347	0.781948	2.431551	0.0150
GARCH(-2)	-1.058490	1.398439	-0.756908	0.4491
GARCH(-3)	0.142698	0.627506	0.227406	0.8201

R-squared	-0.000102	Mean dependent var	0.000639
Adjusted R-squared	-0.000102	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.093644
Sum squared resid	0.364548	Schwarz criterion	-6.078388
Log likelihood	6858.302	Hannan-Quinn criter.	-6.088075
Durbin-Watson stat	1.986417		

Heteroskedasticity Test: ARCH

F-statistic	0.378886	Prob. F(10,2228)	0.9562
Obs*R-squared	3.801105	Prob. Chi-Square(10)	0.9559

4. ARIMA (0,1,0)-GARCH (1,4)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000518	0.000214	2.417106	0.0156
Variance Equation				
C	3.43E-07	9.78E-08	3.507108	0.0005
RESID(-1)^2	0.013232	0.002464	5.369955	0.0000
GARCH(-1)	2.554376	0.044009	58.04164	0.0000
GARCH(-2)	-3.148347	0.077746	-40.49514	0.0000
GARCH(-3)	2.356992	0.078852	29.89153	0.0000
GARCH(-4)	-0.778264	0.042724	-18.21602	0.0000

R-squared	-0.000090	Mean dependent var	0.000639
Adjusted R-squared	-0.000090	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.096483
Sum squared resid	0.364543	Schwarz criterion	-6.078685
Log likelihood	6862.495	Hannan-Quinn criter.	-6.089987
Durbin-Watson stat	1.986441		

Heteroskedasticity Test: ARCH

F-statistic	0.356648	Prob. F(10,2228)	0.9647
Obs*R-squared	3.578357	Prob. Chi-Square(10)	0.9644

5. ARIMA (0,1,0)-GARCH (2,1)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000487	0.000212	2.291986	0.0219
Variance Equation				
C	1.33E-06	3.53E-07	3.756726	0.0002
RESID(-1)^2	-0.001775	0.009037	-0.196418	0.8443
RESID(-2)^2	0.057162	0.011040	5.177781	0.0000
GARCH(-1)	0.936857	0.007345	127.5569	0.0000

R-squared	-0.000143	Mean dependent var	0.000639
Adjusted R-squared	-0.000143	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.094441
Sum squared resid	0.364562	Schwarz criterion	-6.081728
Log likelihood	6858.199	Hannan-Quinn criter.	-6.089801
Durbin-Watson stat	1.986336		

Heteroskedasticity Test: ARCH

F-statistic	0.354045	Prob. F(10,2228)	0.9656
Obs*R-squared	3.552286	Prob. Chi-Square(10)	0.9653

6. ARIMA (0,1,0)-GARCH (2,2)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000486	0.000213	2.286534	0.0222
Variance Equation				
C	8.44E-07	3.45E-07	2.447858	0.0144
RESID(-1)^2	-0.002187	0.009171	-0.238449	0.8115
RESID(-2)^2	0.037319	0.016956	2.200952	0.0277
GARCH(-1)	1.346376	0.229560	5.865027	0.0000
GARCH(-2)	-0.386390	0.216344	-1.785995	0.0741

R-squared	-0.000144	Mean dependent var	0.000639
Adjusted R-squared	-0.000144	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.094011
Sum squared resid	0.364563	Schwarz criterion	-6.078755
Log likelihood	6858.715	Hannan-Quinn criter.	-6.088442
Durbin-Watson stat	1.986335		

Heteroskedasticity Test: ARCH

F-statistic	0.300661	Prob. F(10,2228)	0.9812
Obs*R-squared	3.017386	Prob. Chi-Square(10)	0.9810

7. ARIMA (0,1,0)-GARCH (2,3)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000488	0.000213	2.285090	0.0223
Variance Equation				
C	8.49E-07	4.21E-07	2.015688	0.0438
RESID(-1)^2	-0.002333	0.009135	-0.255394	0.7984
RESID(-2)^2	0.037675	0.019361	1.945879	0.0517
GARCH(-1)	1.358051	0.650846	2.086595	0.0369
GARCH(-2)	-0.414850	1.068258	-0.388343	0.6978
GARCH(-3)	0.016552	0.447829	0.036960	0.9705

R-squared	-0.000141	Mean dependent var	0.000639
Adjusted R-squared	-0.000141	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.093122
Sum squared resid	0.364562	Schwarz criterion	-6.075324
Log likelihood	6858.716	Hannan-Quinn criter.	-6.086626
Durbin-Watson stat	1.986340		

Heteroskedasticity Test: ARCH

F-statistic	0.299873	Prob. F(10,2228)	0.9814
Obs*R-squared	3.009486	Prob. Chi-Square(10)	0.9812

8. ARIMA (0,1,0)-GARCH (2,4)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000489	0.000213	2.292605	0.0219
Variance Equation				
C	8.94E-07	4.77E-07	1.871768	0.0612
RESID(-1)^2	0.002819	0.010119	0.278530	0.7806
RESID(-2)^2	0.033412	0.016869	1.980660	0.0476
GARCH(-1)	0.716549	0.600927	1.192407	0.2331
GARCH(-2)	0.774657	0.551333	1.405062	0.1600
GARCH(-3)	-0.440752	0.731182	-0.602794	0.5466
GARCH(-4)	-0.091848	0.505959	-0.181533	0.8559

R-squared	-0.000139	Mean dependent var	0.000639
Adjusted R-squared	-0.000139	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.092418
Sum squared resid	0.364561	Schwarz criterion	-6.072077
Log likelihood	6858.924	Hannan-Quinn criter.	-6.084993
Durbin-Watson stat	1.986346		

Heteroskedasticity Test: ARCH

F-statistic	0.322091	Prob. F(10,2228)	0.9756
Obs*R-squared	3.232137	Prob. Chi-Square(10)	0.9754

9. ARIMA (0,1,0)-GARCH (3,1)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000491	0.000213	2.308445	0.0210
Variance Equation				
C	1.39E-06	3.68E-07	3.763185	0.0002
RESID(-1)^2	-0.002819	0.008994	-0.313447	0.7539
RESID(-2)^2	0.034139	0.018502	1.845177	0.0650
RESID(-3)^2	0.027049	0.017981	1.504372	0.1325
GARCH(-1)	0.933632	0.007988	116.8839	0.0000

R-squared	-0.000135	Mean dependent var	0.000639
Adjusted R-squared	-0.000135	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.094059
Sum squared resid	0.364559	Schwarz criterion	-6.078804
Log likelihood	6858.770	Hannan-Quinn criter.	-6.088491
Durbin-Watson stat	1.986354		

Heteroskedasticity Test: ARCH

F-statistic	0.291744	Prob. F(10,2228)	0.9832
Obs*R-squared	2.928013	Prob. Chi-Square(10)	0.9831

10. ARIMA (0,1,0)-GARCH (3,2)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000495	0.000213	2.325158	0.0201
Variance Equation				
C	1.35E-06	9.91E-07	1.362115	0.1732
RESID(-1)^2	-0.002649	0.009041	-0.292979	0.7695
RESID(-2)^2	0.033983	0.018870	1.800895	0.0717
RESID(-3)^2	0.025357	0.048155	0.526561	0.5985
GARCH(-1)	0.964134	0.732399	1.316405	0.1880
GARCH(-2)	-0.028609	0.685947	-0.041707	0.9667

R-squared	-0.000128	Mean dependent var	0.000639
Adjusted R-squared	-0.000128	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.093171
Sum squared resid	0.364557	Schwarz criterion	-6.075373
Log likelihood	6858.770	Hannan-Quinn criter.	-6.086674
Durbin-Watson stat	1.986367		

Heteroskedasticity Test: ARCH

F-statistic	0.291568	Prob. F(10,2228)	0.9833
Obs*R-squared	2.926249	Prob. Chi-Square(10)	0.9831

11. ARIMA (0,1,0)-GARCH (3,3)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000498	0.000214	2.323407	0.0202
Variance Equation				
C	1.40E-06	1.24E-06	1.125849	0.2602
RESID(-1)^2	-0.002713	0.009006	-0.301306	0.7632
RESID(-2)^2	0.034039	0.019258	1.767537	0.0771
RESID(-3)^2	0.027598	0.060118	0.459073	0.6462
GARCH(-1)	0.980322	1.224108	0.800846	0.4232
GARCH(-2)	-0.106149	1.546798	-0.068625	0.9453
GARCH(-3)	0.058850	0.445423	0.132122	0.8949

R-squared	-0.000124	Mean dependent var	0.000639
Adjusted R-squared	-0.000124	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.092287
Sum squared resid	0.364555	Schwarz criterion	-6.071947
Log likelihood	6858.777	Hannan-Quinn criter.	-6.084863
Durbin-Watson stat	1.986375		

Heteroskedasticity Test: ARCH

F-statistic	0.290109	Prob. F(10,2228)	0.9836
Obs*R-squared	2.911622	Prob. Chi-Square(10)	0.9834

12. ARIMA (0,1,0)-GARCH (3,4)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000532	0.000211	2.525188	0.0116
Variance Equation				
C	9.87E-07	3.39E-07	2.914576	0.0036
RESID(-1)^2	0.014692	0.007239	2.029502	0.0424
RESID(-2)^2	-0.002053	0.009300	-0.220796	0.8253
RESID(-3)^2	0.036770	0.008782	4.187042	0.0000
GARCH(-1)	0.512522	0.068953	7.432912	0.0000
GARCH(-2)	0.377200	0.029274	12.88506	0.0000
GARCH(-3)	0.763463	0.022483	33.95731	0.0000
GARCH(-4)	-0.707516	0.067260	-10.51907	0.0000

R-squared	-0.000071	Mean dependent var	0.000639
Adjusted R-squared	-0.000071	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.097588
Sum squared resid	0.364536	Schwarz criterion	-6.074705
Log likelihood	6865.737	Hannan-Quinn criter.	-6.089235
Durbin-Watson stat	1.986480		

Heteroskedasticity Test: ARCH

F-statistic	0.368952	Prob. F(10,2228)	0.9601
Obs*R-squared	3.701606	Prob. Chi-Square(10)	0.9598

13. ARIMA (0,1,0)-GARCH (4,1)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000495	0.000213	2.324481	0.0201
Variance Equation				
C	1.39E-06	3.82E-07	3.651472	0.0003
RESID(-1)^2	-0.002685	0.009034	-0.297231	0.7663
RESID(-2)^2	0.033900	0.018535	1.829027	0.0674
RESID(-3)^2	0.026017	0.026294	0.989453	0.3224
RESID(-4)^2	0.001319	0.019711	0.066933	0.9466
GARCH(-1)	0.933419	0.008327	112.0927	0.0000

R-squared	-0.000128	Mean dependent var	0.000639
Adjusted R-squared	-0.000128	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.093172
Sum squared resid	0.364557	Schwarz criterion	-6.075374
Log likelihood	6858.771	Hannan-Quinn criter.	-6.086675
Durbin-Watson stat	1.986366		

Heteroskedasticity Test: ARCH

F-statistic	0.291484	Prob. F(10,2228)	0.9833
Obs*R-squared	2.925407	Prob. Chi-Square(10)	0.9831

14. ARIMA (0,1,0)-GARCH (4,2)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000499	0.000214	2.334015	0.0196
Variance Equation				
C	2.47E-06	7.16E-07	3.447850	0.0006
RESID(-1)^2	-0.000545	0.009443	-0.057707	0.9540
RESID(-2)^2	0.028515	0.016711	1.706355	0.0879
RESID(-3)^2	0.038530	0.023596	1.632867	0.1025
RESID(-4)^2	0.038400	0.017070	2.249512	0.0245
GARCH(-1)	0.173886	0.328890	0.528705	0.5970
GARCH(-2)	0.707079	0.309801	2.282363	0.0225

R-squared	-0.000121	Mean dependent var	0.000639
Adjusted R-squared	-0.000121	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.092810
Sum squared resid	0.364554	Schwarz criterion	-6.072470
Log likelihood	6859.365	Hannan-Quinn criter.	-6.085386
Durbin-Watson stat	1.986380		

Heteroskedasticity Test: ARCH

F-statistic	0.244684	Prob. F(10,2228)	0.9916
Obs*R-squared	2.456226	Prob. Chi-Square(10)	0.9915

15. ARIMA (0,1,0)-GARCH (4,3)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000513	0.000210	2.447341	0.0144
Variance Equation				
C	3.39E-06	1.05E-06	3.228001	0.0012
RESID(-1)^2	-0.002886	0.008902	-0.324187	0.7458
RESID(-2)^2	0.037085	0.008600	4.312171	0.0000
RESID(-3)^2	0.077700	0.009507	8.173139	0.0000
RESID(-4)^2	0.064934	0.011087	5.856681	0.0000
GARCH(-1)	-0.268060	0.018613	-14.40164	0.0000
GARCH(-2)	0.177385	0.016741	10.59604	0.0000
GARCH(-3)	0.896402	0.017453	51.36080	0.0000

R-squared	-0.000099	Mean dependent var	0.000639
Adjusted R-squared	-0.000099	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.097739
Sum squared resid	0.364546	Schwarz criterion	-6.074856
Log likelihood	6865.908	Hannan-Quinn criter.	-6.089387
Durbin-Watson stat	1.986425		

Heteroskedasticity Test: ARCH

F-statistic	0.231087	Prob. F(10,2228)	0.9933
Obs*R-squared	2.319870	Prob. Chi-Square(10)	0.9932

16. ARIMA (0,1,0)-GARCH (4,4)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000516	0.000209	2.472259	0.0134
Variance Equation				
C	5.37E-06	1.62E-06	3.311219	0.0009
RESID(-1)^2	0.004662	0.009604	0.485487	0.6273
RESID(-2)^2	0.062700	0.012694	4.939330	0.0000
RESID(-3)^2	0.110906	0.014451	7.674529	0.0000
RESID(-4)^2	0.097262	0.011967	8.127572	0.0000
GARCH(-1)	-0.917708	0.120822	-7.595527	0.0000
GARCH(-2)	0.032043	0.040286	0.795391	0.4264
GARCH(-3)	0.970409	0.037621	25.79415	0.0000
GARCH(-4)	0.611538	0.113957	5.366366	0.0000

R-squared	-0.000093	Mean dependent var	0.000639
Adjusted R-squared	-0.000093	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.097491
Sum squared resid	0.364544	Schwarz criterion	-6.072066
Log likelihood	6866.629	Hannan-Quinn criter.	-6.088211
Durbin-Watson stat	1.986436		

Heteroskedasticity Test: ARCH

F-statistic	0.167719	Prob. F(10,2228)	0.9983
Obs*R-squared	1.684205	Prob. Chi-Square(10)	0.9982

17. ARIMA (0,1,0)-GARCH (5,1)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000495	0.000214	2.310023	0.0209
Variance Equation				
C	1.37E-06	3.83E-07	3.588362	0.0003
RESID(-1)^2	-0.002586	0.009032	-0.286370	0.7746
RESID(-2)^2	0.033645	0.018523	1.816370	0.0693
RESID(-3)^2	0.026389	0.026259	1.004947	0.3149
RESID(-4)^2	0.006291	0.023950	0.262661	0.7928
RESID(-5)^2	-0.005892	0.016347	-0.360413	0.7185
GARCH(-1)	0.934222	0.009024	103.5304	0.0000

R-squared	-0.000128	Mean dependent var	0.000639
Adjusted R-squared	-0.000128	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.092313
Sum squared resid	0.364557	Schwarz criterion	-6.071973
Log likelihood	6858.806	Hannan-Quinn criter.	-6.084889
Durbin-Watson stat	1.986367		

Heteroskedasticity Test: ARCH

F-statistic	0.284285	Prob. F(10,2228)	0.9848
Obs*R-squared	2.853243	Prob. Chi-Square(10)	0.9847

18. ARIMA (0,1,0)-GARCH (5,2)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000512	0.000216	2.371608	0.0177
Variance Equation				
C	2.21E-06	7.17E-07	3.076076	0.0021
RESID(-1)^2	0.000666	0.009448	0.070468	0.9438
RESID(-2)^2	0.027400	0.016543	1.656301	0.0977
RESID(-3)^2	0.050106	0.022802	2.197458	0.0280
RESID(-4)^2	0.044000	0.017375	2.532426	0.0113
RESID(-5)^2	-0.025242	0.022167	-1.138711	0.2548
GARCH(-1)	0.234615	0.272100	0.862238	0.3886
GARCH(-2)	0.655996	0.253461	2.588152	0.0096

R-squared	-0.000100	Mean dependent var	0.000639
Adjusted R-squared	-0.000100	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.092407
Sum squared resid	0.364547	Schwarz criterion	-6.069524
Log likelihood	6859.912	Hannan-Quinn criter.	-6.084055
Durbin-Watson stat	1.986421		

Heteroskedasticity Test: ARCH

F-statistic	0.196713	Prob. F(10,2228)	0.9966
Obs*R-squared	1.975098	Prob. Chi-Square(10)	0.9965

19. ARIMA (0,1,0)-GARCH (5,3)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000513	0.000215	2.380960	0.0173
Variance Equation				
C	3.02E-06	7.90E-07	3.823726	0.0001
RESID(-1)^2	-0.000290	0.010896	-0.026575	0.9788
RESID(-2)^2	0.040720	0.017629	2.309901	0.0209
RESID(-3)^2	0.025257	0.014914	1.693529	0.0904
RESID(-4)^2	0.041770	0.015922	2.623497	0.0087
RESID(-5)^2	0.014687	0.017112	0.858298	0.3907
GARCH(-1)	0.658617	0.007645	86.14478	0.0000
GARCH(-2)	-0.738503	0.002232	-330.8747	0.0000
GARCH(-3)	0.939490	0.007689	122.1920	0.0000

R-squared	-0.000099	Mean dependent var	0.000639
Adjusted R-squared	-0.000099	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.107254
Sum squared resid	0.364546	Schwarz criterion	-6.081828
Log likelihood	6877.607	Hannan-Quinn criter.	-6.097973
Durbin-Watson stat	1.986425		

Heteroskedasticity Test: ARCH

F-statistic	0.333834	Prob. F(10,2228)	0.9722
Obs*R-squared	3.349807	Prob. Chi-Square(10)	0.9720

20. ARIMA (0,1,0)-GARCH (5,4)

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 2250

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000556	0.000213	2.608191	0.0091
Variance Equation				
C	6.09E-06	3.20E-06	1.904396	0.0569
RESID(-1)^2	0.002382	0.009385	0.253843	0.7996
RESID(-2)^2	0.037906	0.021196	1.788333	0.0737
RESID(-3)^2	0.100264	0.046629	2.150224	0.0315
RESID(-4)^2	0.107588	0.071584	1.502977	0.1328
RESID(-5)^2	0.037049	0.042927	0.863072	0.3881
GARCH(-1)	-1.236278	0.727763	-1.698737	0.0894
GARCH(-2)	0.498015	0.682975	0.729185	0.4659
GARCH(-3)	1.148969	0.570154	2.015193	0.0439
GARCH(-4)	0.270583	0.674328	0.401264	0.6882

R-squared	-0.000043	Mean dependent var	0.000639
Adjusted R-squared	-0.000043	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.097376
Sum squared resid	0.364526	Schwarz criterion	-6.069408
Log likelihood	6867.500	Hannan-Quinn criter.	-6.087168
Durbin-Watson stat	1.986535		

Heteroskedasticity Test: ARCH

F-statistic	0.234321	Prob. F(10,2228)	0.9929
Obs*R-squared	2.352304	Prob. Chi-Square(10)	0.9929

SAMPLE 3

1. ARIMA (0,1,0)-GARCH (1,1)

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000562	0.000337	1.665885	0.0957
Variance Equation				
C	4.16E-06	4.79E-07	8.674440	0.0000
RESID(-1)^2	0.048512	0.008067	6.013365	0.0000
GARCH(-1)	0.924102	0.007995	115.5896	0.0000

R-squared	-0.000247	Mean dependent var	0.000370
Adjusted R-squared	-0.000247	S.D. dependent var	0.012243
S.E. of regression	0.012244	Akaike info criterion	-6.040748
Sum squared resid	0.168361	Schwarz criterion	-6.022866
Log likelihood	3398.900	Hannan-Quinn criter.	-6.033990
Durbin-Watson stat	2.033410		

Heteroskedasticity Test: ARCH

F-statistic	0.183795	Prob. F(10,1103)	0.9974
Obs*R-squared	1.853196	Prob. Chi-Square(10)	0.9973

2. ARIMA (0,1,0)-GARCH (1,2)

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000593	0.000344	1.720520	0.0853
Variance Equation				
C	5.36E-06	7.70E-07	6.966283	0.0000
RESID(-1)^2	0.068871	0.014326	4.807503	0.0000
GARCH(-1)	0.471007	0.266033	1.770482	0.0766
GARCH(-2)	0.425200	0.253875	1.674841	0.0940

R-squared	-0.000331	Mean dependent var	0.000370
Adjusted R-squared	-0.000331	S.D. dependent var	0.012243
S.E. of regression	0.012245	Akaike info criterion	-6.040165
Sum squared resid	0.168375	Schwarz criterion	-6.017813
Log likelihood	3399.573	Hannan-Quinn criter.	-6.031718
Durbin-Watson stat	2.033238		

Heteroskedasticity Test: ARCH

F-statistic	0.200106	Prob. F(10,1103)	0.9963
Obs*R-squared	2.017361	Prob. Chi-Square(10)	0.9962

3. ARIMA (0,1,0)-GARCH (2,1)

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000650	0.000346	1.878698	0.0603
Variance Equation				
C	2.90E-06	7.15E-07	4.050449	0.0001
RESID(-1)^2	0.107685	0.018391	5.855285	0.0000
RESID(-2)^2	-0.067210	0.018723	-3.589721	0.0003
GARCH(-1)	0.940772	0.009192	102.3431	0.0000

R-squared	-0.000525	Mean dependent var	0.000370
Adjusted R-squared	-0.000525	S.D. dependent var	0.012243
S.E. of regression	0.012246	Akaike info criterion	-6.041812
Sum squared resid	0.168408	Schwarz criterion	-6.019460
Log likelihood	3400.498	Hannan-Quinn criter.	-6.033365
Durbin-Watson stat	2.032844		

Heteroskedasticity Test: ARCH

F-statistic	0.236809	Prob. F(10,1103)	0.9926
Obs*R-squared	2.386587	Prob. Chi-Square(10)	0.9924

4. ARIMA (0,1,0)-GARCH (2,2)

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000651	0.000346	1.882021	0.0598
Variance Equation				
C	2.82E-06	1.27E-06	2.221816	0.0263
RESID(-1)^2	0.107448	0.018389	5.843127	0.0000
RESID(-2)^2	-0.068233	0.024483	-2.786927	0.0053
GARCH(-1)	0.967346	0.351774	2.749905	0.0060
GARCH(-2)	-0.024833	0.328786	-0.075529	0.9398

R-squared	-0.000529	Mean dependent var	0.000370
Adjusted R-squared	-0.000529	S.D. dependent var	0.012243
S.E. of regression	0.012246	Akaike info criterion	-6.040037
Sum squared resid	0.168408	Schwarz criterion	-6.013215
Log likelihood	3400.501	Hannan-Quinn criter.	-6.029901
Durbin-Watson stat	2.032835		

Heteroskedasticity Test: ARCH

F-statistic	0.236608	Prob. F(10,1103)	0.9926
Obs*R-squared	2.384560	Prob. Chi-Square(10)	0.9925

STAGE II - STAGE III: BJ - G model's Parameter Estimation and Diagnostic Checking

SAMPLE 1: Refer to Appendix 2

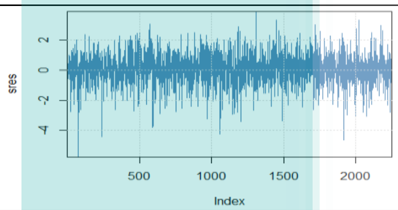
SAMPLE 2

i. ARIMA (0,1,0)-GARCH (1,1) with Normal distribution

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 03/13/17 Time: 04:29
Sample (adjusted): 2 2250
Included observations: 2249 after adjustments
Convergence achieved after 9 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000586	0.000217	2.696675	0.0070
		Variance Equation		
C	1.15E-06	3.15E-07	3.648057	0.0003
RESID(-1)^2	0.047859	0.005482	8.729874	0.0000
GARCH(-1)	0.945348	0.006267	150.8374	0.0000

R-squared	-0.000017	Mean dependent var	0.000639
Adjusted R-squared	-0.000017	S.D. dependent var	0.012734
S.E. of regression	0.012734	Akaike info criterion	-6.091298
Sum squared resid	0.364517	Schwarz criterion	-6.081128
Log likelihood	6853.665	Hannan-Quinn criter.	-6.087586
Durbin-Watson stat	1.986586		



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 04:30
Sample: 2 2250
Included observations: 2249

	AC	PAC	Q-Stat	Prob
1	0.021	0.021	1.0262	0.311
2	-0.027	-0.028	2.6980	0.259
3	-0.008	-0.007	2.8488	0.416
4	0.004	0.004	2.8893	0.577
5	0.024	0.023	4.1446	0.529
6	-0.048	-0.049	9.3654	0.154
7	-0.020	-0.017	10.296	0.172
8	-0.006	-0.008	10.380	0.239
9	0.053	0.052	16.764	0.053
10	0.013	0.010	17.133	0.071
11	-0.031	-0.027	19.303	0.056
12	-0.047	-0.046	24.299	0.019
13	0.016	0.015	24.850	0.024
14	0.009	0.002	25.017	0.034
15	-0.020	-0.016	25.957	0.038
16	0.030	0.036	27.976	0.032
17	-0.005	-0.006	28.024	0.045
18	0.006	-0.001	28.107	0.060
19	-0.010	-0.013	28.355	0.077
20	0.028	0.033	30.137	0.068

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 04:32
Sample: 2 2250
Included observations: 2249

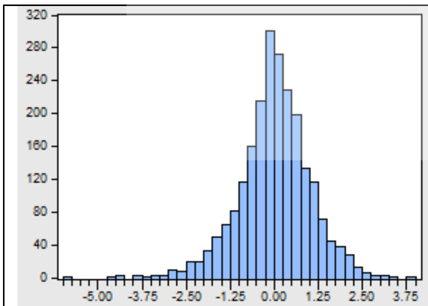
	AC	PAC	Q-Stat	Prob
1	-0.038	-0.038	3.2456	0.072
2	-0.011	-0.012	3.5135	0.173
3	0.006	0.005	3.5817	0.310
4	0.010	0.011	3.8284	0.430
5	-0.024	-0.023	5.1572	0.397
6	0.016	0.015	5.7547	0.451
7	-0.008	-0.008	5.9123	0.550
8	0.008	0.008	6.0741	0.639
9	0.013	0.013	6.4341	0.696
10	-0.013	-0.013	6.8232	0.742
11	0.003	0.003	6.8402	0.812
12	0.019	0.019	7.6965	0.808
13	-0.002	-0.000	7.7063	0.862
14	0.025	0.026	9.1202	0.823
15	0.041	0.042	12.933	0.607
16	-0.029	-0.026	14.904	0.532
17	0.021	0.020	15.873	0.533
18	-0.014	-0.014	16.307	0.571
19	-0.003	-0.003	16.332	0.635
20	0.035	0.036	19.140	0.513

Heteroskedasticity Test: ARCH

F-statistic	0.668990	Prob. F(10,2228)	0.7542
Obs*R-squared	6.702799	Prob. Chi-Square(10)	0.7532

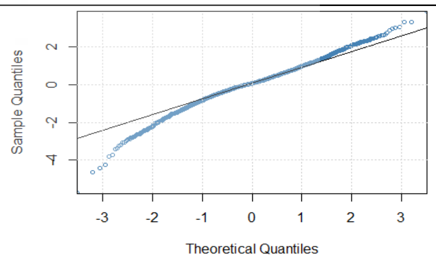
Heteroskedasticity Test: ARCH

F-statistic	0.857075	Prob. F(15,2218)	0.6134
Obs*R-squared	12.87424	Prob. Chi-Square(15)	0.6120



Series: Standardized Residuals
Sample 2 2250
Observations 2249

Mean	0.012271
Median	0.019709
Maximum	3.875859
Minimum	-5.864775
Std. Dev.	1.000445
Skewness	-0.396411
Kurtosis	4.623967
Jarque-Bera	306.0359
Probability	0.000000

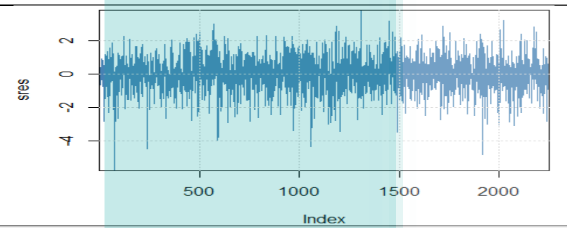


ii. ARIMA (0,1,0)-GARCH (1,1) with *t* distribution

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Student's t distribution
Date: 03/13/17 Time: 04:36
Sample (adjusted): 2 2250
Included observations: 2249 after adjustments
Convergence achieved after 14 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000956	0.000209	4.578737	0.0000
Variance Equation				
C	1.28E-06	4.91E-07	2.600062	0.0093
RESID(-1)^2	0.045899	0.008480	5.412727	0.0000
GARCH(-1)	0.946957	0.009413	100.6033	0.0000
T-DIST. DOF	5.882421	0.810737	7.255647	0.0000

R-squared	-0.000621	Mean dependent var	0.000639
Adjusted R-squared	-0.000621	S.D. dependent var	0.012734
S.E. of regression	0.012738	Akaike info criterion	-6.134982
Sum squared resid	0.364737	Schwarz criterion	-6.122269
Log likelihood	6903.787	Hannan-Quinn criter.	-6.130342
Durbin-Watson stat	1.985388		



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 04:41
Sample: 2 2250
Included observations: 2249

	AC	PAC	Q-Stat	Prob
1	0.021	0.021	0.9961	0.318
2	-0.029	-0.029	2.8540	0.240
3	-0.009	-0.008	3.0401	0.385
4	0.004	0.004	3.0824	0.544
5	0.023	0.022	4.2770	0.510
6	-0.049	-0.050	9.7636	0.135
7	-0.021	-0.018	10.771	0.149
8	-0.006	-0.008	10.864	0.210
9	0.053	0.051	17.182	0.046
10	0.012	0.009	17.491	0.064
11	-0.031	-0.027	19.736	0.049
12	-0.047	-0.046	24.680	0.016
13	0.015	0.014	25.202	0.022
14	0.008	0.002	25.363	0.031
15	-0.021	-0.016	26.330	0.035
16	0.030	0.035	28.333	0.029
17	-0.005	-0.007	28.387	0.041
18	0.006	-0.001	28.462	0.055
19	-0.011	-0.013	28.729	0.070
20	0.028	0.033	30.560	0.061

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 04:44
Sample: 2 2250
Included observations: 2249

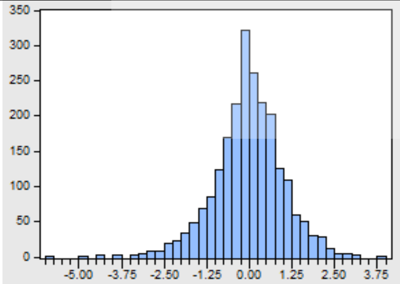
	AC	PAC	Q-Stat	Prob
1	-0.032	-0.032	2.2669	0.132
2	-0.009	-0.010	2.4567	0.293
3	0.008	0.007	2.5909	0.459
4	0.014	0.014	3.0185	0.555
5	-0.022	-0.021	4.1501	0.528
6	0.018	0.017	4.8670	0.561
7	-0.006	-0.005	4.9435	0.667
8	0.009	0.010	5.1430	0.742
9	0.016	0.016	5.6909	0.770
10	-0.013	-0.013	6.0752	0.809
11	0.003	0.003	6.0988	0.867
12	0.021	0.020	7.0642	0.853
13	0.001	0.003	7.0665	0.899
14	0.024	0.026	8.4217	0.866
15	0.041	0.042	12.257	0.659
16	-0.027	-0.024	13.902	0.606
17	0.021	0.020	14.888	0.604
18	-0.013	-0.014	15.272	0.643
19	-0.003	-0.003	15.292	0.704
20	0.037	0.038	18.469	0.557

Heteroskedasticity Test: ARCH

F-statistic	0.596131	Prob. F(10,2228)	0.8183
Obs*R-squared	5.974759	Prob. Chi-Square(10)	0.8174

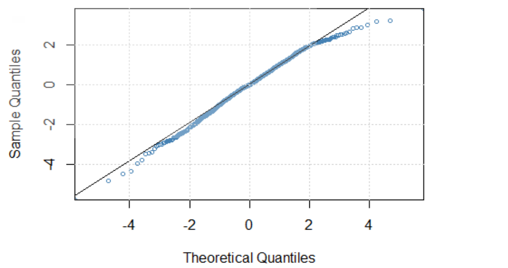
Heteroskedasticity Test: ARCH

F-statistic	0.809047	Prob. F(15,2218)	0.6686
Obs*R-squared	12.15673	Prob. Chi-Square(15)	0.6671



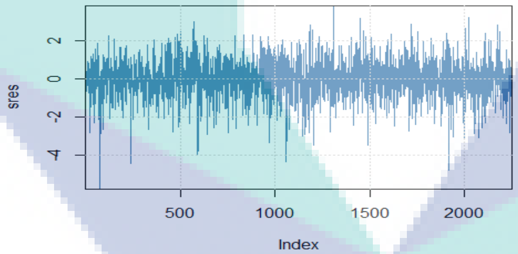
Series: Standardized Residuals
Sample 2 2250
Observations 2249

Mean	-0.021360
Median	-0.015170
Maximum	3.841162
Minimum	-5.824776
Std. Dev.	0.992419
Skewness	-0.400985
Kurtosis	4.652023
Jarque-Bera	316.0157
Probability	0.000000

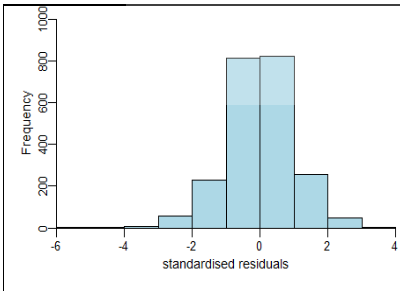


iii. ARIMA (0,1,0)-GARCH (1,1) with skewed-*t* distribution

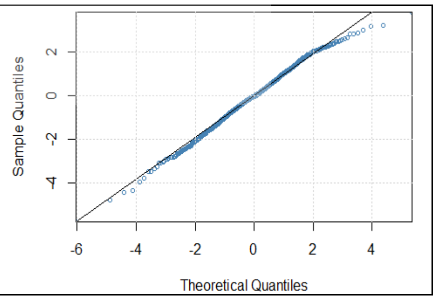
```
> fit33<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt3, cond.di
st="sstd", include.mean = TRUE,trace=F) # for ARIMA with constant
> sum33<-summary(fit33)
Title:GARCH Modelling
Call:garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt3, cond.dist
= "sstd",include.mean = TRUE, trace = F)
Mean and Variance Equation:data ~ arma(0, 0) + garch(1, 1)
<environment: 0x14efe1d4>
 [data = dldt3]
Conditional Distribution:sstd
Coefficient(s):
        mu       omega       alpha1        beta1         skew        shape
7.4405e-04  1.1544e-06  4.5162e-02  9.4838e-01  9.3725e-01  5.9387e+00

Std. Errors:based on Hessian
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     7.440e-04   2.230e-04    3.337 0.000847 ***
omega  1.154e-06   4.421e-07    2.611 0.009023 **
alpha1 4.516e-02   7.346e-03    6.147 7.87e-10 ***
beta1  9.484e-01   7.781e-03  121.890  < 2e-16 ***
skew   9.372e-01   2.614e-02   35.856  < 2e-16 ***
shape  5.939e+00   8.066e-01    7.362 1.81e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Log Likelihood:
 6905.012    normalized:  3.070259
Standardised Residuals Tests:
                                Statistic p-value
 Jarque-Bera Test   R    Chi^2  304.5085  0
 Shapiro-Wilk Test  R    W      0.9829788 9.389528e-16
 Ljung-Box Test     R    Q(10)  17.37142  0.06653892
 Ljung-Box Test     R    Q(15)  26.01236  0.037893
 Ljung-Box Test     R    Q(20)  30.39286  0.0637331
 Ljung-Box Test     R^2  Q(10)  6.692161  0.754152
 Ljung-Box Test     R^2  Q(15)  13.55087  0.5598278
 Ljung-Box Test     R^2  Q(20)  19.79902  0.4705635
 LM Arch Test       R    TR^2   7.881679  0.794307
Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.135182 -6.119926 -6.135196 -6.129613
```




```
> ArchTest (x33, lags=10, demean = FALSE) #arch test at lag=10

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x33
Chi-squared = 6.6036, df = 10, p-value = 0.7623

> ArchTest (x33, lags=15, demean = FALSE)

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x33
Chi-squared = 13.617, df = 15, p-value = 0.5548

> dw33<-2*(1-f33$acf[2]);dw33 #use formula dw=2(1-r1)
[1] 2.056946
```



	x33
nobs	2249.000000
NAs	0.000000
Minimum	-5.754884
Maximum	3.841975
1. Quartile	-0.534216
3. Quartile	0.585469
Mean	-0.002353
Median	0.004645
Sum	-5.292293
SE Mean	0.020853
LCL Mean	-0.043246
UCL Mean	0.038539
Variance	0.977941
Stdev	0.988909
Skewness	-0.386499
Kurtosis	1.624139

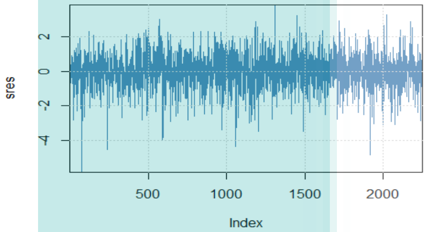


iv. ARIMA (0,1,0)-GARCH (1,1) with GED distribution

Dependent Variable: D(LDT3)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Date: 03/13/17 Time: 05:11
Sample (adjusted): 2 2250
Included observations: 2249 after adjustments
Convergence achieved after 12 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000802	0.000198	4.060084	0.0000
Variance Equation				
C	1.19E-06	4.99E-07	2.396478	0.0166
RESID(-1)^2	0.046085	0.008681	5.308623	0.0000
GARCH(-1)	0.946606	0.009879	95.82158	0.0000
GED PARAMETER	1.273843	0.049894	25.53103	0.0000

R-squared	-0.000164	Mean dependent var	0.000639
Adjusted R-squared	-0.000164	S.D. dependent var	0.012734
S.E. of regression	0.012735	Akaike info criterion	-6.142602
Sum squared resid	0.364570	Schwarz criterion	-6.129889
Log likelihood	6912.356	Hannan-Quinn criter.	-6.137962
Durbin-Watson stat	1.986294		



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 05:19
Sample: 2 2250
Included observations: 2249

	AC	PAC	Q-Stat	Prob
1	0.021	0.021	1.0118	0.314
2	-0.028	-0.029	2.8010	0.246
3	-0.009	-0.007	2.9716	0.396
4	0.004	0.004	3.0125	0.556
5	0.023	0.023	4.2340	0.516
6	-0.049	-0.050	9.6088	0.142
7	-0.021	-0.017	10.586	0.158
8	-0.006	-0.008	10.673	0.221
9	0.053	0.052	17.013	0.049
10	0.012	0.009	17.346	0.067
11	-0.031	-0.027	19.565	0.052
12	-0.047	-0.046	24.530	0.017
13	0.015	0.014	25.062	0.023
14	0.008	0.002	25.225	0.032
15	-0.021	-0.016	26.181	0.036
16	0.030	0.035	28.187	0.030
17	-0.005	-0.007	28.239	0.042
18	0.006	-0.001	28.316	0.057
19	-0.011	-0.013	28.576	0.073
20	0.028	0.033	30.383	0.064

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 05:21
Sample: 2 2250
Included observations: 2249

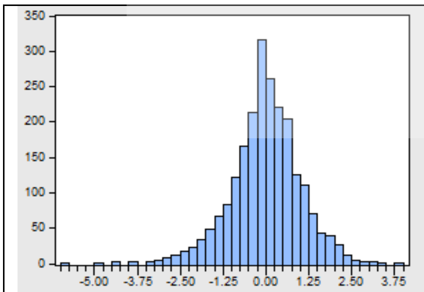
	AC	PAC	Q-Stat	Prob
1	-0.034	-0.034	2.6292	0.105
2	-0.010	-0.011	2.8430	0.241
3	0.007	0.006	2.9516	0.399
4	0.013	0.013	3.3053	0.508
5	-0.023	-0.022	4.5115	0.478
6	0.017	0.016	5.1890	0.520
7	-0.007	-0.006	5.2942	0.624
8	0.009	0.009	5.4808	0.705
9	0.014	0.015	5.9522	0.745
10	-0.013	-0.013	6.3357	0.786
11	0.003	0.003	6.3580	0.848
12	0.020	0.019	7.2894	0.838
13	-0.000	0.002	7.2895	0.887
14	0.025	0.026	8.6733	0.851
15	0.041	0.042	12.502	0.641
16	-0.028	-0.025	14.282	0.578
17	0.021	0.020	15.259	0.577
18	-0.013	-0.014	15.664	0.616
19	-0.003	-0.003	15.687	0.678
20	0.036	0.037	18.712	0.541

Heteroskedasticity Test: ARCH

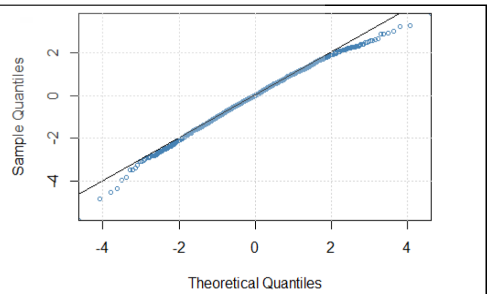
F-statistic	0.826490	Prob. F(15,2218)	0.6486
Obs*R-squared	12.41737	Prob. Chi-Square(15)	0.6472

Heteroskedasticity Test: ARCH

F-statistic	0.620944	Prob. F(10,2228)	0.7972
Obs*R-squared	6.222750	Prob. Chi-Square(10)	0.7962



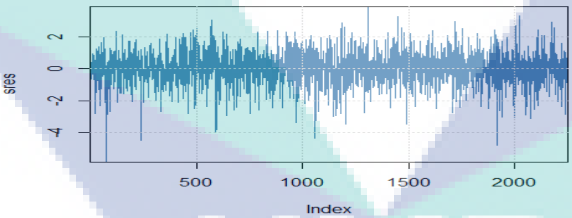
Series: Standardized Residuals	
Sample 2 2250	
Observations 2249	
Mean	-0.007488
Median	-0.001609
Maximum	3.870287
Minimum	-5.870238
Std. Dev.	1.000480
Skewness	-0.399336
Kurtosis	4.640747
Jarque-Bera	312.0419
Probability	0.000000



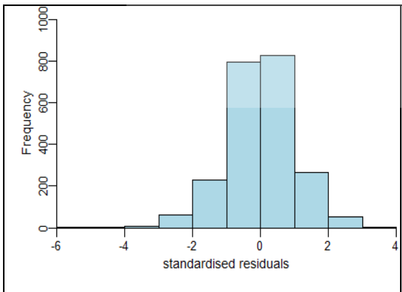
v. **ARIMA (0,1,0)-GARCH (1,1) with skewed-GED distribution**

```
> fit35<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt3,cond.dis
t="sged", include.mean = TRUE,trace=F) # for ARIMA with constant
> sum35<-summary(fit35)
Title:GARCH Modelling
Call:
 garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt3, cond.dist = "
sged", include.mean = TRUE, trace = F)
Mean and Variance Equation:data ~ arma(0, 0) + garch(1, 1)
<environment: 0x15568a60>[data = dldt3]
Conditional Distribution:sged
Coefficient(s):
        mu       omega      alpha1        beta1         skew        shape
5.9970e-04  1.0357e-06  4.4840e-02  9.4857e-01  9.5576e-01  1.2890e+00
Std. Errors:based on Hessian
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      5.997e-04   2.185e-04    2.745  0.00605 **
omega   1.036e-06   4.286e-07    2.416  0.01568 *
alpha1  4.484e-02   7.271e-03    6.167 6.98e-10 ***
beta1   9.486e-01   7.863e-03  120.634  < 2e-16 ***
skew    9.558e-01   2.161e-02   44.222  < 2e-16 ***
shape   1.289e+00   5.566e-02   23.157  < 2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Log Likelihood:
 6913.182    normalized:  3.073892
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  300.9763  0
 Shapiro-Wilk Test  R    W      0.9830998 1.087622e-15
 Ljung-Box Test     R    Q(10)  17.22897  0.06944851
 Ljung-Box Test     R    Q(15)  25.84524  0.03968119
 Ljung-Box Test     R    Q(20)  30.19396  0.06677117
 Ljung-Box Test     R^2  Q(10)  6.941965  0.7309122
 Ljung-Box Test     R^2  Q(15)  13.74797  0.5447233
 Ljung-Box Test     R^2  Q(20)  20.00254  0.4577707
 LM Arch Test       R    TR^2   8.116195  0.7759855

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.142447 -6.127192 -6.142462 -6.136879
```

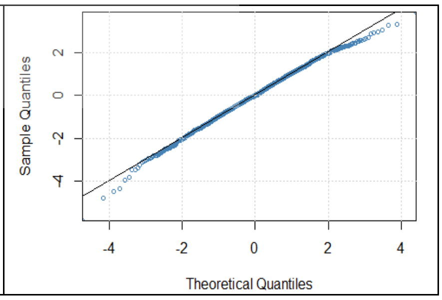
Standardised residuals:




```
> dw35<-2*(1-f35$acf[2]);dw35 #use formula dw=2(1-r1)
[1] 2.056001
> ArchTest (x35, lags=10, demean = FALSE) #arch test at lag=10

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x35
Chi-squared = 6.8539, df = 10, p-value = 0.7392

> ArchTest (x35, lags=15, demean = FALSE)

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x35
Chi-squared = 13.857, df = 15, p-value = 0.5364
```



	x35
nobs	2249.000000
NAs	0.000000
Minimum	-5.814602
Maximum	3.872066
1. Quartile	-0.527398
3. Quartile	0.604548
Mean	0.010835
Median	0.018960
Sum	24.367257
SE Mean	0.021074
LCL Mean	-0.030492
UCL Mean	0.052161
Variance	0.998819
Stdev	0.999409
Skewness	-0.384969
Kurtosis	1.613994



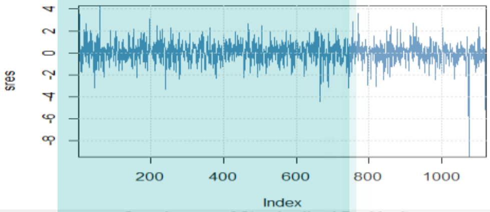
SAMPLE 3

i. ARIMA (0,1,0)-GARCH (1,1) with Normal distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 03/13/17 Time: 06:09
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 14 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
	Variance Equation			
C	4.17E-06	4.56E-07	9.144686	0.0000
RESID(-1)^2	0.045807	0.008057	5.685143	0.0000
GARCH(-1)	0.926391	0.008346	111.0018	0.0000

R-squared	-0.000914	Mean dependent var	0.000370
Adjusted R-squared	-0.000023	S.D. dependent var	0.012243
S.E. of regression	0.012243	Akaike info criterion	-6.040033
Sum squared resid	0.168473	Schwarz criterion	-6.026622
Log likelihood	3397.498	Hannan-Quinn criter.	-6.034965
Durbin-Watson stat	2.032054		



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 06:13
Sample: 2 1125
Included observations: 1124

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.005	0.005	0.0290	0.865
		2	-0.034	-0.034	1.2993	0.522
		3	-0.022	-0.021	1.8350	0.607
		4	0.003	0.002	1.8463	0.764
		5	0.045	0.043	4.1093	0.534
		6	-0.038	-0.039	5.7490	0.452
		7	-0.013	-0.009	5.9313	0.548
		8	-0.013	-0.014	6.1265	0.633
		9	0.067	0.065	11.227	0.260
		10	0.002	-0.002	11.233	0.340
		11	-0.013	-0.006	11.428	0.408
		12	-0.048	-0.046	14.050	0.298
		13	0.019	0.020	14.469	0.342
		14	0.024	0.013	15.099	0.371
		15	-0.067	-0.064	20.249	0.163
		16	0.018	0.023	20.612	0.194
		17	-0.003	-0.002	20.623	0.244
		18	0.041	0.031	22.530	0.209
		19	-0.007	-0.009	22.588	0.256
		20	0.018	0.028	22.943	0.292

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 06:15
Sample: 2 1125
Included observations: 1124

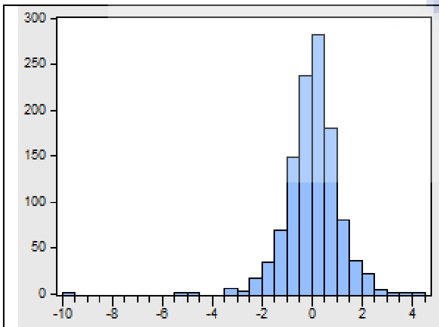
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.002	0.002	0.0069	0.934
		2	-0.006	-0.006	0.0506	0.975
		3	-0.013	-0.013	0.2367	0.971
		4	-0.018	-0.018	0.6092	0.962
		5	-0.018	-0.018	0.9907	0.963
		6	0.009	0.009	1.0826	0.982
		7	0.018	0.017	1.4572	0.984
		8	-0.005	-0.006	1.4897	0.993
		9	0.008	0.008	1.5573	0.997
		10	0.001	0.001	1.5581	0.999
		11	-0.015	-0.014	1.8207	0.999
		12	-0.003	-0.002	1.8299	1.000
		13	-0.011	-0.011	1.9624	1.000
		14	-0.014	-0.014	2.1803	1.000
		15	0.027	0.026	3.0093	1.000
		16	-0.016	-0.018	3.3021	1.000
		17	-0.014	-0.014	3.5213	1.000
		18	0.014	0.014	3.7503	1.000
		19	-0.016	-0.016	4.0344	1.000
		20	0.007	0.008	4.0953	1.000

Heteroskedasticity Test: ARCH

F-statistic	0.159351	Prob. F(10,1103)	0.9986
Obs*R-squared	1.607076	Prob. Chi-Square(10)	0.9986

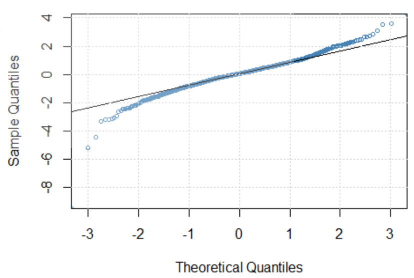
Heteroskedasticity Test: ARCH

F-statistic	0.201945	Prob. F(15,1093)	0.9996
Obs*R-squared	3.065030	Prob. Chi-Square(15)	0.9995



Series: Standardized Residuals
Sample 2 1125
Observations 1124

Mean	0.033670
Median	0.057964
Maximum	4.307814
Minimum	-9.506190
Std. Dev.	0.999945
Skewness	-0.975202
Kurtosis	11.60952
Jarque-Bera	3649.627
Probability	0.000000

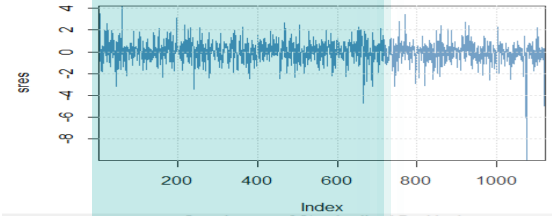


ii. ARIMA (0,1,0)-GARCH (1,1) with *t* distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Date: 03/13/17 Time: 06:20
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 13 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000676	0.000296	2.285321	0.0223
Variance Equation				
C	2.50E-06	1.13E-06	2.211212	0.0270
RESID(-1)^2	0.034497	0.011345	3.040572	0.0024
GARCH(-1)	0.947368	0.015875	59.67712	0.0000
T-DIST. DOF	4.814805	0.634583	7.587356	0.0000

R-squared	-0.000626	Mean dependent var	0.000370
Adjusted R-squared	-0.000626	S.D. dependent var	0.012243
S.E. of regression	0.012247	Akaike info criterion	-6.164056
Sum squared resid	0.168425	Schwarz criterion	-6.141704
Log likelihood	3469.199	Hannan-Quinn criter.	-6.155609
Durbin-Watson stat	2.032639		



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 22:39
Sample: 2 1125
Included observations: 1124

	AC	PAC	Q-Stat	Prob
1	0.009	0.009	0.0828	0.774
2	-0.038	-0.038	1.7364	0.420
3	-0.023	-0.023	2.3470	0.504
4	0.001	-0.000	2.3480	0.672
5	0.044	0.042	4.5160	0.478
6	-0.039	-0.040	6.2216	0.399
7	-0.014	-0.010	6.4379	0.490
8	-0.010	-0.011	6.5517	0.586
9	0.067	0.065	11.651	0.234
10	0.003	-0.001	11.661	0.308
11	-0.015	-0.007	11.919	0.370
12	-0.049	-0.047	14.624	0.263
13	0.019	0.020	15.043	0.305
14	0.021	0.011	15.569	0.340
15	-0.067	-0.063	20.632	0.149
16	0.016	0.022	20.933	0.181
17	-0.003	-0.003	20.945	0.229
18	0.039	0.029	22.693	0.203
19	-0.007	-0.009	22.746	0.249
20	0.017	0.027	23.060	0.286

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 14:59
Sample: 2 1125
Included observations: 1124

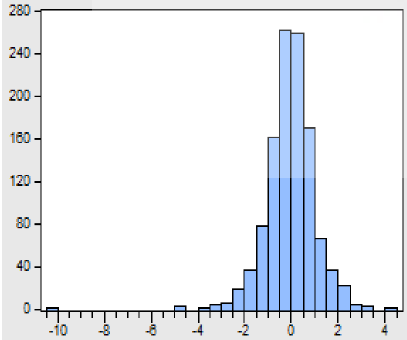
	AC	PAC	Q-Stat	Prob
1	0.012	0.012	0.1759	0.675
2	-0.003	-0.004	0.1890	0.910
3	-0.010	-0.010	0.3094	0.958
4	-0.016	-0.016	0.6083	0.962
5	-0.017	-0.017	0.9524	0.966
6	0.011	0.011	1.0907	0.982
7	0.023	0.022	1.6964	0.975
8	-0.004	-0.005	1.7104	0.989
9	0.012	0.012	1.8658	0.993
10	-0.000	-0.000	1.8660	0.997
11	-0.012	-0.011	2.0406	0.998
12	-0.003	-0.002	2.0493	0.999
13	-0.008	-0.008	2.1144	1.000
14	-0.013	-0.013	2.3092	1.000
15	0.024	0.023	2.9469	1.000
16	-0.015	-0.016	3.1874	1.000
17	-0.015	-0.015	3.4575	1.000
18	0.012	0.013	3.6344	1.000
19	-0.015	-0.016	3.9016	1.000
20	0.004	0.006	3.9224	1.000

Heteroskedasticity Test: ARCH

F-statistic	0.186065	Prob. F(10,1103)	0.9973
Obs*R-squared	1.876045	Prob. Chi-Square(10)	0.9972

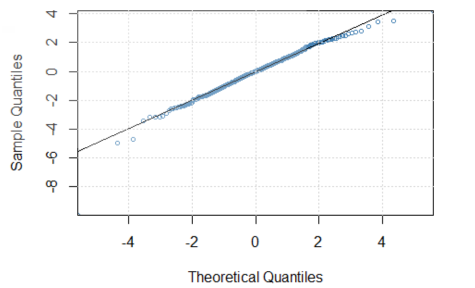
Heteroskedasticity Test: ARCH

F-statistic	0.193810	Prob. F(15,1093)	0.9997
Obs*R-squared	2.941880	Prob. Chi-Square(15)	0.9996



Series: Standardized Residuals
Sample 2 1125
Observations 1124

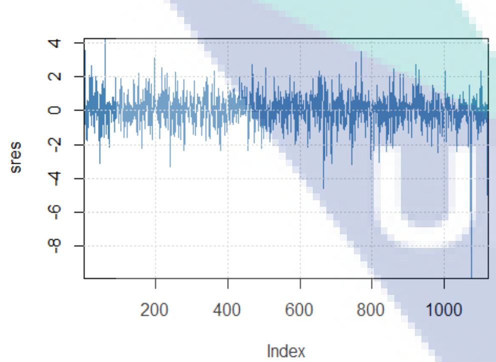
Mean	-0.024592
Median	-0.006513
Maximum	4.184246
Minimum	-10.07516
Std. Dev.	1.014638
Skewness	-1.054650
Kurtosis	12.63098
Jarque-Bera	4552.433
Probability	0.000000



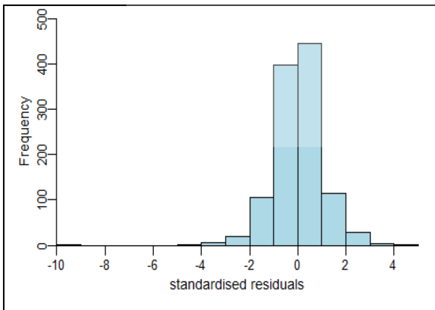
iii. ARIMA (0,1,0)-GARCH (1,1) with skewed-*t* distribution

```
> fit43_r<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt4, cond.
dist="sstd", include.mean = FALSE,trace=F) # for ARIMA without constant
> sum43<-summary(fit43_r)
Title:GARCH Modelling
Call:garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt4, cond.dist
= "sstd",include.mean = FALSE, trace = F)
Mean and Variance Equation:data ~ arma(0, 0) + garch(1, 1)
<environment: 0x13de4310>[data = dldt4]
Conditional Distribution:sstd
Coefficient(s):
     omega        alpha1         beta1          skew         shape
2.8875e-06   4.2102e-02   9.3935e-01   9.2790e-01   4.7647e+00
Std. Errors:based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  2.887e-06   1.255e-06    2.302  0.02136 *
alpha1 4.210e-02   1.204e-02    3.497  0.00047 ***
beta1  9.393e-01   1.546e-02   60.777  < 2e-16 ***
skew   9.279e-01   3.329e-02   27.871  < 2e-16 ***
shape  4.765e+00   6.946e-01    6.860 6.89e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Log Likelihood:
 3466.126    normalized:  3.083742

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  4223.067  0
 Shapiro-Wilk Test  R    W      0.9435526 0
 Ljung-Box Test     R    Q(10)  10.68823  0.3823176
 Ljung-Box Test     R    Q(15)  19.39194  0.1965191
 Ljung-Box Test     R    Q(20)  22.61217  0.3082366
 Ljung-Box Test     R^2  Q(10)  1.50635   0.9989154
 Ljung-Box Test     R^2  Q(15)  2.646951  0.9998168
 Ljung-Box Test     R^2  Q(20)  3.686727  0.9999763
 LM Arch Test       R    TR^2   1.820182  0.9996358

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.158587 -6.136236 -6.158627 -6.150140
```

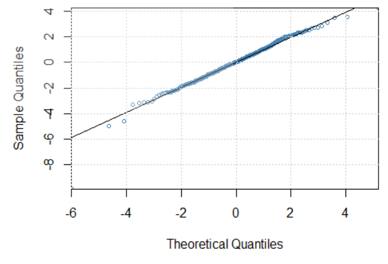
Standardised residuals




```
> dw43<-2*(1-f43$acf[2]);dw43 #use formula dw=2(1-r1)
[1] 2.065749
> ArchTest (x43, lags=10, demean = FALSE) #arch test at lag=10

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x43
Chi-squared = 1.5911, df = 10, p-value = 0.9986

> ArchTest (x43, lags=15, demean = FALSE)

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x43
Chi-squared = 2.7622, df = 15, p-value = 0.9998
```




```
                     x43
nobs         1124.000000
NAs             0.000000
Minimum        -9.865254
Maximum         4.246764
1. Quartile    -0.497426
3. Quartile     0.589705
Mean            0.035489
Median          0.058873
Sum            39.889306
SE Mean         0.030149
LCL Mean       -0.023666
UCL Mean        0.094643
Variance        1.021667
Stdev           1.010775
Skewness       -0.981293
Kurtosis        9.268469
```

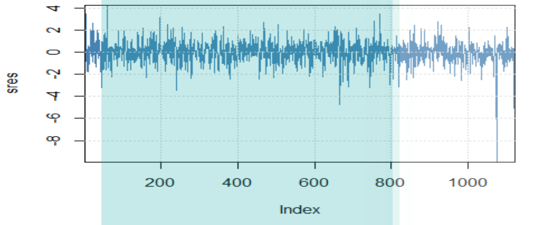


iv. ARIMA (0,1,0)-GARCH (1,1) with GED distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Date: 03/13/17 Time: 15:02
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 23 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000623	0.000275	2.269519	0.0232
Variance Equation				
C	2.75E-06	1.07E-06	2.576444	0.0100
RESID(-1)^2	0.037180	0.012565	2.959012	0.0031
GARCH(-1)	0.942461	0.016372	57.56708	0.0000
GED PARAMETER	1.168244	0.044418	26.30106	0.0000

R-squared	-0.000428	Mean dependent var	0.000370
Adjusted R-squared	-0.000428	S.D. dependent var	0.012243
S.E. of regression	0.012245	Akaike info criterion	-6.153220
Sum squared resid	0.168391	Schwarz criterion	-6.130868
Log likelihood	3463.110	Hannan-Quinn criter.	-6.144773
Durbin-Watson stat	2.033042		



Correlogram of Standardized Residuals

Date: 03/13/17 Time: 15:06
Sample: 2 1125
Included observations: 1124

	AC	PAC	Q-Stat	Prob
1	0.008	0.008	0.0681	0.794
2	-0.037	-0.037	1.6270	0.443
3	-0.023	-0.022	2.2107	0.530
4	0.001	0.000	2.2123	0.697
5	0.044	0.042	4.3918	0.494
6	-0.039	-0.040	6.0735	0.415
7	-0.014	-0.010	6.2811	0.507
8	-0.011	-0.012	6.4124	0.601
9	0.067	0.065	11.498	0.243
10	0.003	-0.002	11.507	0.319
11	-0.014	-0.007	11.744	0.383
12	-0.049	-0.047	14.441	0.273
13	0.019	0.020	14.861	0.316
14	0.022	0.011	15.414	0.350
15	-0.067	-0.063	20.503	0.153
16	0.017	0.022	20.815	0.186
17	-0.004	-0.003	20.829	0.234
18	0.039	0.029	22.600	0.206
19	-0.007	-0.009	22.649	0.253
20	0.017	0.027	22.962	0.291

Correlogram of Standardized Residuals Squared

Date: 03/13/17 Time: 15:05
Sample: 2 1125
Included observations: 1124

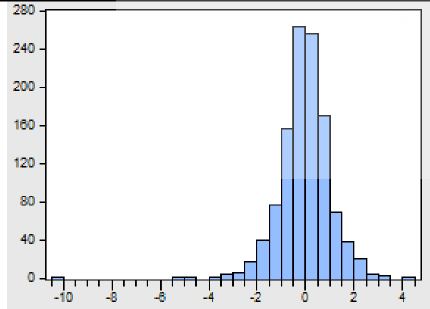
	AC	PAC	Q-Stat	Prob
1	0.011	0.011	0.1268	0.722
2	-0.004	-0.004	0.1481	0.929
3	-0.011	-0.011	0.2869	0.962
4	-0.017	-0.017	0.6100	0.962
5	-0.018	-0.018	0.9816	0.964
6	0.010	0.010	1.1003	0.982
7	0.022	0.022	1.6706	0.976
8	-0.004	-0.005	1.6906	0.989
9	0.011	0.011	1.8259	0.994
10	-0.001	-0.000	1.8262	0.998
11	-0.013	-0.012	2.0215	0.998
12	-0.003	-0.002	2.0330	0.999
13	-0.008	-0.009	2.1139	1.000
14	-0.013	-0.013	2.3169	1.000
15	0.024	0.024	2.9823	1.000
16	-0.015	-0.017	3.2435	1.000
17	-0.015	-0.015	3.5182	1.000
18	0.012	0.013	3.6900	1.000
19	-0.016	-0.016	3.9699	1.000
20	0.004	0.006	3.9909	1.000

Heteroskedasticity Test: ARCH

F-statistic	0.182813	Prob. F(10,1103)	0.9975
Obs*R-squared	1.843304	Prob. Chi-Square(10)	0.9974

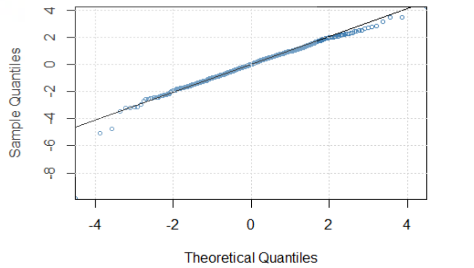
Heteroskedasticity Test: ARCH

F-statistic	0.196746	Prob. F(15,1093)	0.9996
Obs*R-squared	2.986329	Prob. Chi-Square(15)	0.9996



Series: Standardized Residuals
Sample 2 1125
Observations 1124

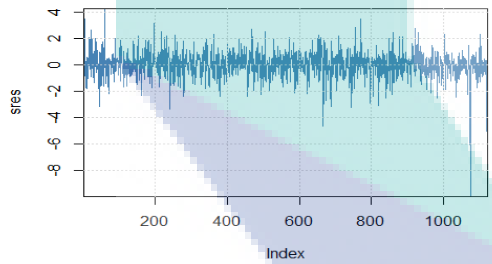
Mean	-0.020144
Median	-0.001772
Maximum	4.227669
Minimum	-10.02880
Std. Dev.	1.019383
Skewness	-1.034127
Kurtosis	12.35896
Jarque-Bera	4302.472
Probability	0.000000



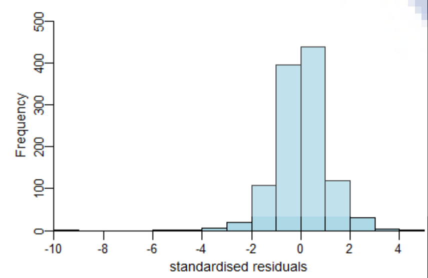
v. ARIMA (0,1,0)-GARCH (1,1) with skewed-GED distribution

```
> fit45_r<-garchFit(formula = ~ arma(0,0)+garch(1, 1), data = dldt4,cond.d
ist="sged", include.mean = FALSE,trace=F) # for ARIMA without constant
> sum45<-summary(fit45_r)
Title:GARCH Modelling

Call:garchFit(formula = ~arma(0, 0) + garch(1, 1), data = dldt4, cond.dist
= "sged", include.mean = FALSE, trace = F)

Mean and Variance Equation:data ~ arma(0, 0) + garch(1, 1)
<environment: 0x1487fa5c>[data = dldt4]

Conditional Distribution:sged

Coefficient(s):
     omega       alpha1        beta1         skew        shape
2.9207e-06   4.3095e-02   9.3697e-01   9.2895e-01   1.1754e+00

Std. Errors:based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  2.921e-06   1.212e-06    2.409  0.01598 *
alpha1 4.309e-02   1.232e-02    3.497  0.00047 ***
beta1  9.370e-01   1.588e-02   59.008  < 2e-16 ***
skew   9.289e-01   1.821e-02   50.999  < 2e-16 ***
shape  1.175e+00   6.157e-02   19.089  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 3461.27    normalized:  3.079422

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  4147.252  0
 Shapiro-Wilk Test  R    W      0.9438308 0
 Ljung-Box Test     R    Q(10)  10.60208  0.3893472
 Ljung-Box Test     R    Q(15)  19.28142  0.2012632
 Ljung-Box Test     R    Q(20)  22.47182  0.3154664
 Ljung-Box Test     R^2  Q(10)  1.52967   0.9988399
 Ljung-Box Test     R^2  Q(15)  2.694424  0.9997949
 Ljung-Box Test     R^2  Q(20)  3.746284  0.999973
 LM Arch Test       R    TR^2   1.85204   0.9996013

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.149947 -6.127596 -6.149987 -6.141501
```

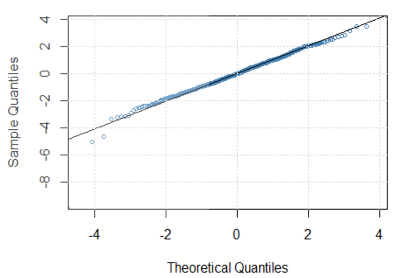
Standardised residuals




```
> dw45<-2*(1-f45$acf[2]);dw45 #use formula dw=2(1-r1)
[1] 2.064687
> ArchTest (x45, lags=10, demean = FALSE) #arch test at lag=10

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x45
Chi-squared = 1.614, df = 10, p-value = 0.9985

> ArchTest (x45, lags=15, demean = FALSE)

 ARCH LM-test; Null hypothesis: no ARCH effects

data:  x45
Chi-squared = 2.8092, df = 15, p-value = 0.9997
```




```
                    x45
nobs        1124.000000
NAs            0.000000
Minimum       -9.936151
Maximum        4.279370
1. Quartile   -0.503191
3. Quartile    0.596552
Mean           0.035785
Median         0.059051
Sum           40.222624
SE Mean        0.030452
LCL Mean      -0.023964
UCL Mean       0.095535
Variance       1.042329
Stdev          1.020945
Skewness      -0.977405
Kurtosis       9.182744
```

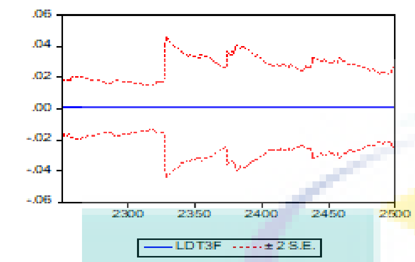


STAGE IV: BOX-JENKINS – GARCH FORECASTING

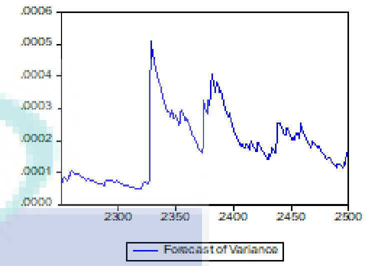
Sample 1: Refer Appendix 2

Sample 2: ARIMA (0,1,0)-GARCH (1,1) with GED innovations

i. For Stationary Data (Daily Log Return Gold Price)



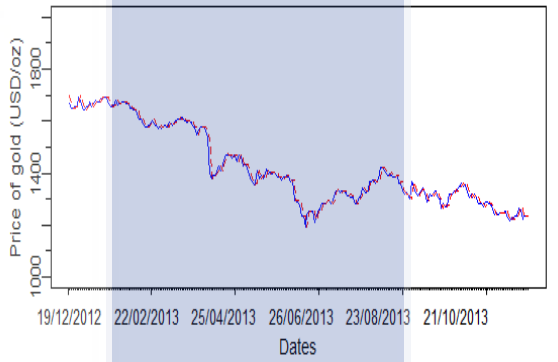
Forecast: LDT3F	
Actual: D(LDT3)	
Forecast sample: 2251 2500	
Included observations: 250	
Root Mean Squared Error	0.013897
Mean Absolute Error	0.009119
Mean Abs. Percent Error	105.7116
Theil Inequality Coefficient	0.951699
Bias Proportion	0.022338
Variance Proportion	0.977662
Covariance Proportion	0.000000




```
> mae_st_AG3=sum(error2st_AG3)/T; mae_st_
AG3
[1] 0.009118529
> rmse_st_AG3=sqrt(sum(error3st_AG3)/T);
rmse_st_AG3
[1] 0.01389693
```

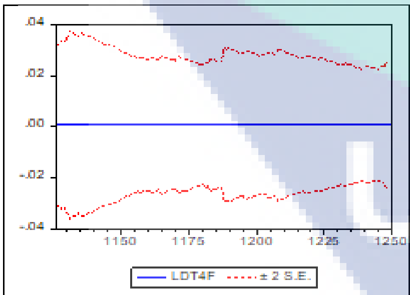
For Daily Gold Price

```
> mae_AG3=sum(error2_AG3)/T; mae_AG3
[1] 12.68691
> rmse_AG3=sqrt(sum(error3_AG3)/T); rmse_
AG3
[1] 19.21902
> mape_AG3=(100/T)*sum(error4_AG3); mape_
AG3
[1] 0.9154706
```

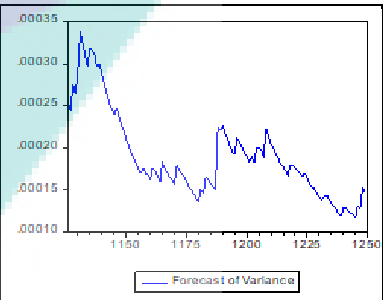


Sample 3: ARIMA (0,1,0)-GARCH (1,1) with *t* innovations

i. For Stationary Data (Daily Log Return Gold Price)



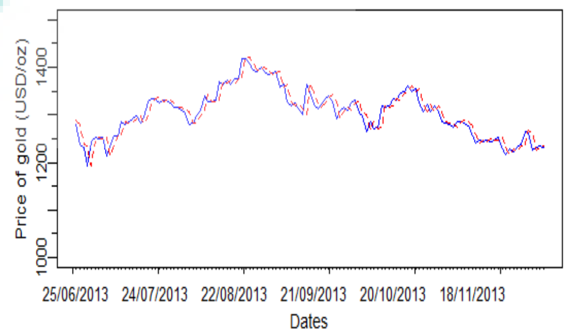
Forecast: LDT4F	
Actual: D(LDT4)	
Forecast sample: 1126 1250	
Included observations: 125	
Root Mean Squared Error	0.013808
Mean Absolute Error	0.009951
Mean Abs. Percent Error	100.3051
Theil Inequality Coefficient	0.955540
Bias Proportion	0.005518
Variance Proportion	0.994482
Covariance Proportion	-0.000000




```
> mae_st_AG4=sum(error2st_AG4)/T; mae_st_AG4
[1] 0.009951418
> rmse_st_AG4=sqrt(sum(error3st_AG4)/T); rmse
_st_AG4
[1] 0.01380593
```

For Daily Gold Price

```
> mae_AG4=sum(error2_AG4)/T; mae_AG4
[1] 12.93009
> rmse_AG4=sqrt(sum(error3_AG4)/T); rmse_AG4
[1] 17.87645
> mape_AG4=(100/T) *sum(error4_AG4); mape_AG4
[1] 0.9956225
```



APPENDIX 4
ANALYSIS OF CHAPTER 4 SECTION 4.4

A. OUT-OF-SAMPLE DATA

- 25 June -17 Dec 2013 (125 data)

i) Original data (Price in USD)

Date	Price ($)	Date	Price($)	Date	Price($)	Date	Price($)	Date	Price($)
25/6/2013	1279.00	30/7/2013	1324.15	4/9/2013	1390.00	9/10/2013	1304.00	13/11/2013	1272.50
26/6/2013	1236.25	31/7/2013	1314.50	5/9/2013	1385.00	10/10/2013	1298.50	14/11/2013	1286.00
27/6/2013	1232.75	1/8/2013	1315.00	6/9/2013	1387.00	11/10/2013	1265.50	15/11/2013	1287.25
28/6/2013	1192.00	2/8/2013	1309.25	9/9/2013	1390.00	14/10/2013	1285.50	16/11/2013	1283.50
1/7/2013	1242.75	5/8/2013	1304.75	10/9/2013	1358.25	15/10/2013	1270.50	17/11/2013	1275.75
2/7/2013	1252.50	6/8/2013	1280.50	11/9/2013	1363.75	16/10/2013	1273.50	18/11/2013	1257.00
3/7/2013	1250.00	7/8/2013	1282.50	12/9/2013	1328.00	17/10/2013	1319.25	19/11/2013	1240.00
4/7/2013	1251.75	8/8/2013	1298.25	13/9/2013	1318.50	18/10/2013	1316.50	20/11/2013	1246.25
5/7/2013	1212.75	9/8/2013	1309.00	16/9/2013	1324.00	19/10/2013	1317.50	21/11/2013	1243.00
8/7/2013	1235.25	12/8/2013	1341.00	17/9/2013	1312.25	20/10/2013	1333.00	22/11/2013	1247.50
9/7/2013	1255.50	13/8/2013	1328.50	18/9/2013	1301.00	21/10/2013	1331.25	23/11/2013	1245.00
10/7/2013	1256.00	14/8/2013	1326.50	19/9/2013	1365.50	22/10/2013	1344.75	24/11/2013	1245.50
11/7/2013	1285.00	15/8/2013	1329.75	20/9/2013	1349.25	23/10/2013	1347.75	25/11/2013	1253.00
12/7/2013	1279.75	16/8/2013	1369.25	21/9/2013	1323.00	24/10/2013	1361.00	2/12/2013	1229.50
15/7/2013	1284.75	19/8/2013	1365.00	22/9/2013	1314.25	25/10/2013	1349.25	3/12/2013	1217.25
16/7/2013	1291.50	20/8/2013	1372.50	23/9/2013	1322.75	26/10/2013	1354.75	4/12/2013	1227.50
17/7/2013	1297.25	21/8/2013	1363.00	24/9/2013	1333.00	27/10/2013	1324.00	5/12/2013	1222.50
18/7/2013	1283.25	22/8/2013	1375.50	25/9/2013	1341.00	1/11/2013	1306.75	6/12/2013	1233.00
19/7/2013	1295.75	23/8/2013	1377.50	26/9/2013	1326.50	4/11/2013	1320.50	9/12/2013	1237.00
22/7/2013	1327.00	27/8/2013	1419.25	1/10/2013	1290.75	5/11/2013	1307.25	10/12/2013	1266.25
23/7/2013	1333.50	28/8/2013	1419.50	2/10/2013	1306.25	6/11/2013	1319.00	11/12/2013	1260.75
24/7/2013	1335.00	29/8/2013	1407.75	3/10/2013	1316.00	7/11/2013	1307.25	12/12/2013	1225.25
25/7/2013	1326.00	30/8/2013	1394.75	4/10/2013	1309.75	8/11/2013	1285.50	13/12/2013	1232.00
26/7/2013	1331.00	2/9/2013	1392.25	7/10/2013	1323.50	11/11/2013	1282.50	16/12/2013	1234.75
29/7/2013	1329.75	3/9/2013	1399.50	8/10/2013	1329.50	12/11/2013	1281.25	17/12/2013	1231.75

Out-of-sample data in original scale (Price)

[1]	1279.00	1236.25	1232.75	1192.00	1242.75	1252.50	1250.00	1251.75	1212.75	1235.25
[11]	1255.50	1256.00	1285.00	1279.75	1284.75	1291.50	1297.25	1283.25	1295.75	1327.00
[21]	1333.50	1335.00	1326.00	1331.00	1329.75	1324.15	1314.50	1315.00	1309.25	1304.75
[31]	1280.50	1282.50	1298.25	1309.00	1341.00	1328.50	1326.50	1329.75	1369.25	1365.00
[41]	1372.50	1363.00	1375.50	1377.50	1419.25	1419.50	1407.75	1394.75	1392.25	1399.50
[51]	1390.00	1385.00	1387.00	1390.00	1358.25	1363.75	1328.00	1318.50	1324.00	1312.25
[61]	1301.00	1365.50	1349.25	1323.00	1314.25	1322.75	1333.00	1341.00	1326.50	1290.75
[71]	1306.25	1316.00	1309.75	1323.50	1329.50	1304.00	1298.50	1265.50	1285.50	1270.50
[81]	1273.50	1319.25	1316.50	1317.50	1333.00	1331.25	1344.75	1347.75	1361.00	1349.25
[91]	1354.75	1324.00	1306.75	1320.50	1307.25	1319.00	1307.25	1285.50	1282.50	1281.25
[101]	1272.50	1286.00	1287.25	1283.50	1275.75	1257.00	1240.00	1246.25	1243.00	1247.50
[111]	1245.00	1245.50	1253.00	1229.50	1217.25	1227.50	1222.50	1233.00	1237.00	1266.25
[121]	1260.75	1225.25	1232.00	1234.75	1231.75					

ii. Out-of-sample data in Transformed scale (log price)

[1]	7.153834	7.119838	7.117003	7.083388	7.125082	7.132897	7.130899	7.132298	7.100646	7.119029
[11]	7.135289	7.135687	7.158514	7.154420	7.158319	7.163560	7.168002	7.157151	7.166845	7.190676
[21]	7.195562	7.196687	7.189922	7.193686	7.192746	7.188526	7.181212	7.181592	7.177210	7.173767
[31]	7.155006	7.156567	7.168772	7.177019	7.201171	7.191806	7.190299	7.192746	7.222018	7.218910
[41]	7.224389	7.217443	7.226573	7.228026	7.257884	7.258060	7.249748	7.240470	7.238676	7.243870
[51]	7.237059	7.233455	7.234898	7.237059	7.213952	7.217994	7.191429	7.184250	7.188413	7.179499
[61]	7.170888	7.219276	7.207304	7.187657	7.181021	7.187468	7.195187	7.201171	7.190299	7.162979
[71]	7.174916	7.182352	7.177592	7.188035	7.192558	7.173192	7.168965	7.143223	7.158903	7.147166
[81]	7.149524	7.184819	7.182732	7.183491	7.195187	7.193874	7.203963	7.206192	7.215975	7.207304
[91]	7.211372	7.188413	7.175298	7.185766	7.175681	7.184629	7.175681	7.158903	7.156567	7.155591
[101]	7.148739	7.159292	7.160263	7.157346	7.151290	7.136483	7.122867	7.127894	7.125283	7.128897
[111]	7.126891	7.127292	7.133296	7.114363	7.104349	7.112735	7.108653	7.117206	7.120444	7.143815
[121]	7.139462	7.110900	7.116394	7.118624	7.116191					

iii. Out-of-sample data in Stationary scale (First differenced of log price)

[1]	-0.0060411370	-0.0339959186	-0.0028351578	-0.0336148775	0.0416940973	0.0078148880
[7]	-0.0019980027	0.0013990209	-0.0316520641	0.0183828706	0.0162605209	0.0003981684
[13]	0.0228266503	-0.0040939720	0.0038994005	0.0052401867	0.0044423057	-0.0108507168
[19]	0.0096937556	0.0238310773	0.0048863093	0.0011242272	-0.0067644001	0.0037636476
[25]	-0.0009395848	-0.0042202105	-0.0073143790	0.0003803004	-0.0043822115	-0.0034430026
[31]	-0.0187608249	0.0015606714	0.0122059057	0.0082462832	0.0241521174	-0.0093651181
[37]	-0.0015065916	0.0024470600	0.0292721900	-0.0031087160	0.0054794658	-0.0069457417
[43]	0.0091291487	0.0014529607	0.0298583010	0.0001761339	-0.0083120118	-0.0092774974
[49]	-0.0017940443	0.0051938865	-0.0068112828	-0.0036036075	0.0014430017	0.0021606058
[55]	-0.0231066407	0.0040411517	-0.0265642071	-0.0071793242	0.0041627307	-0.0089142364
[61]	-0.0086100216	0.0483874624	-0.0119717794	-0.0196469974	-0.0066357244	0.0064467421
[67]	0.0077191384	0.0059835631	-0.0108717097	-0.0273204498	0.0119369920	0.0074363962
[73]	-0.0047605536	0.0104434634	0.0045231890	-0.0193664682	-0.0042267114	-0.0257424511
[79]	0.0156804467	-0.0117372239	0.0023584917	0.0352943779	-0.0020866933	0.0007593015
[85]	0.0116960398	-0.0013136907	0.0100897717	0.0022284132	0.0097831883	-0.0086708412
[91]	0.0040680529	-0.0229594779	-0.0131143189	0.0104673141	-0.0100847587	0.0089481796
[97]	-0.0089481796	-0.0167779464	-0.0023364497	-0.0009751342	-0.0068526945	0.0105531564
[103]	0.0009715341	-0.0029174386	-0.0060564804	-0.0148063114	-0.0136165500	0.0050276627
[109]	-0.0026112298	0.0036137361	-0.0020060187	0.0004015258	0.0060036202	-0.0189330932
[115]	-0.0100133666	0.0083853646	-0.0040816383	0.0085522818	0.0032388692	0.0233706832
[121]	-0.0043529946	-0.0285618772	0.0054939603	0.0022296553	-0.0024325980	

B. SIMULATION DATA USING ARIMA(0,1,0)-GARCH(1,1) WITH *t* INNOVATIONS

1. 1-STEP AHEAD OF ARIMA(0,1,0)-GARCH(1,1)

a. Simulation data in original scale (price) with 95% and 80% prediction intervals

Date	Actual Price	Forecast 1-step AG	95% PI	80% PI	Date	Actual Price	Forecast 1-step AG	95% PI	80% PI
25/6/2013	1279.00	1287.62	(1241.06, 1334.18)	(1261.05, 1314.20)	10/9/2013	1358.25	1390.94	(1045.62, 1736.26)	(1193.85, 1588.03)
26/6/2013	1236.25	1279.87	(1214.02, 1345.72)	(1242.28, 1317.45)	11/9/2013	1363.75	1359.17	(1010.72, 1707.61)	(1160.30, 1558.04)
27/6/2013	1232.75	1237.09	(1156.44, 1317.74)	(1191.06, 1283.12)	12/9/2013	1328.00	1364.67	(1013.13, 1716.22)	(1164.04, 1565.31)
28/6/2013	1192.00	1233.58	(1140.46, 1326.71)	(1180.43, 1286.73)	13/9/2013	1318.50	1328.90	(974.28, 1683.51)	(1126.51, 1531.29)
1/7/2013	1242.75	1192.81	(1088.69, 1296.92)	(1133.38, 1252.23)	16/9/2013	1324.00	1319.39	(961.73, 1677.05)	(1115.27, 1523.52)
2/7/2013	1252.50	1243.59	(1129.53, 1357.65	(1178.50, 1308.69)	17/9/2013	1312.25	1324.90	(964.22, 1685.57)	(1119.05, 1530.74)
3/7/2013	1250.00	1253.35	(1130.15, 1376.54	(1183.04, 1323.66)	18/9/2013	1301.00	1313.14	(949.47, 1676.81)	(1105.58, 1520.70)
4/7/2013	1251.75	1250.85	(1119.15, 1382.55	(1175.68, 1326.01)	19/9/2013	1365.50	1301.88	(935.24, 1668.52)	(1092.63, 1511.13)
5/7/2013	1212.75	1252.60	(1112.91, 1392.29	(1172.87, 1332.32)	20/9/2013	1349.25	1366.42	(996.84, 1736.01)	(1155.49, 1577.36)
8/7/2013	1235.25	1213.57	(1066.33, 1360.82	(1129.53, 1297.61)	21/9/2013	1323.00	1350.16	(977.66, 1722.67)	(1137.56, 1562.76)
9/7/2013	1255.50	1236.09	(1081.65, 1390.52	(1147.95, 1324.23)	22/9/2013	1314.25	1323.90	(948.49, 1699.30)	(1109.64, 1538.15)
10/7/2013	1256.00	1256.35	(1095.05, 1417.65	(1164.29, 1348.41)	23/9/2013	1322.75	1315.14	(936.86, 1693.42)	(1099.24, 1531.04)
11/7/2013	1285.00	1256.85	(1088.96, 1424.74	(1161.03, 1352.67)	24/9/2013	1333.00	1323.65	(942.51, 1704.78)	(1106.12, 1541.17)
12/7/2013	1279.75	1285.87	(1111.65, 1460.09	(1186.43, 1385.30)	25/9/2013	1341.00	1333.90	(949.93, 1717.87)	(1114.76, 1553.04)
15/7/2013	1284.75	1280.62	(1100.28, 1460.95	(1177.69, 1383.54)	26/9/2013	1326.50	1341.91	(955.13, 1728.69)	(1121.16, 1562.66)
16/7/2013	1291.50	1285.62	(1099.37, 1471.87	(1179.32, 1391.92)	1/10/2013	1290.75	1327.40	(937.82, 1716.97)	(1105.06, 1549.74)
17/7/2013	1297.25	1292.37	(1100.39, 1484.36	(1182.80, 1401.95)	2/10/2013	1306.25	1291.62	(899.28, 1683.97)	(1067.70, 1515.55)
18/7/2013	1283.25	1298.13	(1100.58, 1495.68	(1185.38, 1410.88)	3/10/2013	1316.00	1307.13	(912.03, 1702.23)	(1081.64, 1532.63)
19/7/2013	1295.75	1284.12	(1081.15, 1487.08	(1168.28, 1399.96)	4/10/2013	1309.75	1316.89	(919.06, 1714.73)	(1089.83, 1543.95)
22/7/2013	1327.00	1296.63	(1088.39, 1504.86	(1177.78, 1415.47)	7/10/2013	1323.50	1310.64	(910.09, 1711.19)	(1082.03, 1539.24)
23/7/2013	1333.50	1327.90	(1114.52, 1541.28	(1206.12, 1449.68)	8/10/2013	1329.50	1324.40	(921.15, 1727.64)	(1094.25, 1554.54)
24/7/2013	1335.00	1334.40	(1116.00, 1552.80	(1209.75, 1459.05)	9/10/2013	1304.00	1330.40	(924.47, 1736.33)	(1098.72, 1562.07)
25/7/2013	1326.00	1335.90	(1112.59, 1559.21	(1208.45, 1463.35)	10/10/2013	1298.50	1304.88	(896.29, 1713.47)	(1071.69, 1538.08)
26/7/2013	1331.00	1326.90	(1098.79, 1555.01	(1196.71, 1457.09)	11/10/2013	1265.50	1299.38	(888.14, 1710.61)	(1064.67, 1534.08)
29/7/2013	1329.75	1331.90	(1099.09, 1564.72	(1199.03, 1464.78)	14/10/2013	1285.50	1266.36	(852.49, 1680.22)	(1030.15, 1502.56)
30/7/2013	1324.15	1330.65	(1093.22, 1568.08	(1195.14, 1466.16)	15/10/2013	1270.50	1286.37	(869.90, 1702.84)	(1048.68, 1524.06)
31/7/2013	1314.50	1325.05	(1083.10, 1566.99	(1186.96, 1463.13)	16/10/2013	1273.50	1271.36	(852.29, 1690.43)	(1032.18, 1510.53)
1/8/2013	1315.00	1315.39	(1069.00, 1561.78	(1174.77, 1456.01)	17/10/2013	1319.25	1274.36	(852.72, 1696.01)	(1033.71, 1515.01)
2/8/2013	1309.25	1315.89	(1065.14, 1566.64	(1172.78, 1459.00)	18/10/2013	1316.50	1320.14	(895.93, 1744.35)	(1078.03, 1562.25)
5/8/2013	1304.75	1310.14	(1055.10, 1565.17	(1164.58, 1455.69)	19/10/2013	1317.50	1317.39	(890.63, 1744.15)	(1073.83, 1560.95)
6/8/2013	1280.50	1305.63	(1046.38, 1564.88	(1157.67, 1453.60)	20/10/2013	1333.00	1318.39	(889.10, 1747.68)	(1073.38, 1563.40)
7/8/2013	1282.50	1281.37	(1017.97, 1544.77	(1131.04, 1431.70)	21/10/2013	1331.25	1333.90	(902.09, 1765.71)	(1087.46, 1580.35)
8/8/2013	1298.25	1283.37	(1015.88, 1550.85	(1130.71, 1436.03)	22/10/2013	1344.75	1332.15	(897.84, 1766.46)	(1084.28, 1580.03)
9/8/2013	1309.00	1299.13	(1027.62, 1570.64)	(1144.17, 1454.09)	23/10/2013	1347.75	1345.66	(908.86, 1782.46)	(1096.36, 1594.96)
12/8/2013	1341.00	1309.89	(1034.41, 1585.36)	(1152.67, 1467.11)	24/10/2013	1361.00	1348.66	(909.39, 1787.94)	(1097.95, 1599.37)
13/8/2013	1328.50	1341.91	(1062.53, 1621.29)	(1182.46, 1501.36)	25/10/2013	1349.25	1361.92	(920.18, 1803.66)	(1109.81, 1614.03)
14/8/2013	1326.50	1329.40	(1046.17, 1612.63)	(1167.75, 1491.05)	26/10/2013	1354.75	1350.16	(905.98, 1794.35)	(1096.65, 1603.67)
15/8/2013	1329.75	1327.40	(1040.36, 1614.43)	(1163.58, 1491.22)	27/10/2013	1324.00	1355.67	(909.05, 1802.28)	(1100.77, 1610.56)
16/8/2013	1369.25	1330.65	(1039.86, 1621.44)	(1164.69, 1496.61)	1/11/2013	1306.75	1324.90	(875.86, 1773.93)	(1068.62, 1581.18)
19/8/2013	1365.00	1370.18	(1075.69, 1664.67)	(1202.10, 1538.25)	4/11/2013	1320.50	1307.63	(856.19, 1759.08)	(1049.98, 1565.29)
20/8/2013	1372.50	1365.92	(1067.77, 1664.07)	(1195.76, 1536.09)	5/11/2013	1307.25	1321.39	(867.55, 1775.23)	(1062.37, 1580.41)
21/8/2013	1363.00	1373.43	(1071.67, 1675.19)	(1201.20, 1545.65)	6/11/2013	1319.00	1308.13	(851.91, 1764.36)	(1047.75, 1568.51)
22/8/2013	1375.50	1363.92	(1058.59, 1669.26)	(1189.66, 1538.19)	7/11/2013	1307.25	1319.89	(861.30, 1778.49)	(1058.16, 1581.63)
23/8/2013	1377.50	1376.43	(1067.57, 1685.29)	(1200.15, 1552.71)	8/11/2013	1285.50	1308.13	(847.18, 1769.08)	(1045.06, 1571.21)
27/8/2013	1419.25	1378.43	(1066.08, 1690.79)	(1200.16, 1556.70)	11/11/2013	1282.50	1286.37	(823.07, 1749.67)	(1021.95, 1550.79)
28/8/2013	1419.50	1420.21	(1104.40, 1736.02)	(1239.97, 1600.45)	12/11/2013	1281.25	1283.37	(817.74, 1749.00)	(1017.62, 1549.12)
29/8/2013	1407.75	1420.46	(1101.24, 1739.68)	(1238.27, 1602.65)	13/11/2013	1272.50	1282.12	(814.16, 1750.07)	(1015.04, 1549.19)
30/8/2013	1394.75	1408.70	(1086.10, 1731.30)	(1224.59, 1592.82)	14/11/2013	1286.00	1273.36	(803.10, 1743.62)	(1004.97, 1541.76)
2/9/2013	1392.25	1395.69	(1069.75, 1721.63)	(1209.67, 1581.72)	15/11/2013	1287.25	1286.87	(814.31, 1759.43)	(1017.16, 1556.58)
3/9/2013	1399.50	1393.19	(1063.94, 1722.44)	(1205.28, 1581.11)	16/11/2013	1283.50	1288.12	(813.27, 1762.97)	(1017.11, 1559.13)
4/9/2013	1390.00	1400.45	(1067.92, 1732.97)	(1210.66, 1590.23)	17/11/2013	1275.75	1284.37	(807.24, 1761.50)	(1012.06, 1556.68)
5/9/2013	1385.00	1390.94	(1055.17, 1726.71)	(1199.31, 1582.58)	18/11/2013	1257.00	1276.61	(797.22, 1756.01)	(1003.01, 1550.22)
6/9/2013	1387.00	1385.94	(1046.95, 1724.92)	(1192.47, 1579.41)	19/11/2013	1240.00	1257.85	(776.20, 1739.50)	(982.96, 1532.74)
9/9/2013	1390.00	1387.94	(1045.77, 1730.11)	(1192.65, 1583.22)	20/11/2013	1246.25	1240.84	(756.94, 1724.74)	(964.66, 1517.01)

Date	Actual Price	Forecast 1-step AG	95% PI	80% PI	Date	Actual Price	Forecast 1-step AG	95% PI	80% PI
21/11/2013	1243.00	1247.09	(760.96, 1733.23)	(969.64, 1524.54)	6/12/2013	1233.00	1223.33	(717.52, 1729.13)	(934.65, 1512.01)
22/11/2013	1247.50	1243.84	(755.48, 1732.20)	(965.12, 1522.56)	9/12/2013	1237.00	1233.83	(725.89, 1741.78)	(943.94, 1523.73)
23/11/2013	1245.00	1248.34	(757.77, 1738.92)	(968.36, 1528.33)	10/12/2013	1266.25	1237.84	(727.76, 1747.91)	(946.72, 1528.95)
24/11/2013	1245.50	1245.84	(753.07, 1738.62)	(964.60, 1527.09)	11/12/2013	1260.75	1267.11	(754.91, 1779.30)	(974.78, 1559.43)
25/11/2013	1253.00	1246.34	(751.37, 1741.31)	(963.85, 1528.84)	12/12/2013	1225.25	1261.60	(747.30, 1775.91)	(968.07, 1555.13)
2/12/2013	1229.50	1253.85	(756.69, 1751.00)	(970.10, 1537.59)	13/12/2013	1232.00	1226.08	(709.67, 1742.49)	(931.35, 1520.81)
3/12/2013	1217.25	1230.33	(731.00, 1729.66)	(945.35, 1515.32)	16/12/2013	1234.75	1232.83	(714.33, 1751.34)	(936.91, 1528.76)
4/12/2013	1227.50	1218.07	(716.57, 1719.57)	(931.85, 1504.29)	17/12/2013	1231.75	1235.59	(714.99, 1756.18)	(938.47, 1532.70)
5/12/2013	1222.50	1228.33	(724.67, 1731.99)	(940.88, 1515.78)					

To find Prediction Interval for 1-step ahead ARIMA-GARCH:

Residual data for 1-step AG

```
for(i in 1:125){
  resiAG4[i]<- dt4_o[i]-f_AG4[i];resiAG4[125]
}
resiAG4
```

Data	Residual	Data	Residual	Data	Residual	Data	Residual	Data	Residual
[1,]	-8.620238	[26,]	-6.499319	[51,]	-10.446491	[76,]	-26.399150	[101,]	-9.616518
[2,]	-43.614997	[27,]	-10.545531	[52,]	-5.940066	[77,]	-6.381903	[102,]	12.639400
[3,]	-4.336084	[28,]	-0.389005	[53,]	1.063315	[78,]	-33.878184	[103,]	0.380270
[4,]	-41.583717	[29,]	-6.639344	[54,]	2.061963	[79,]	19.144134	[104,]	-4.620576
[5,]	49.943842	[30,]	-5.385454	[55,]	-32.690066	[80,]	-15.869393	[105,]	-8.618039
[6,]	8.909520	[31,]	-25.132412	[56,]	4.581407	[81,]	2.140752	[106,]	-19.612798
[7,]	-3.347074	[32,]	1.133989	[57,]	-36.672313	[82,]	44.888724	[107,]	-17.850118
[8,]	0.904617	[33,]	14.882637	[58,]	-10.398135	[83,]	-3.642217	[108,]	5.411379
[9,]	-39.846567	[34,]	9.871985	[59,]	4.608290	[84,]	0.109642	[109,]	-4.092847
[10,]	21.679809	[35,]	31.114715	[60,]	-12.645429	[85,]	14.608967	[110,]	3.659350
[11,]	19.414592	[36,]	-13.406927	[61,]	-12.137483	[86,]	-2.651516	[111,]	-3.343693
[12,]	-0.349103	[37,]	-2.898473	[62,]	63.620125	[87,]	12.599668	[112,]	-0.342002
[13,]	28.150559	[38,]	2.352879	[63,]	-17.173497	[88,]	2.090537	[113,]	6.657660
[14,]	-6.119054	[39,]	38.600681	[64,]	-27.162506	[89,]	12.338508	[114,]	-24.347412
[15,]	4.134497	[40,]	-5.176033	[65,]	-9.644754	[90,]	-12.670454	[115,]	-13.081519
[16,]	5.881115	[41,]	6.576841	[66,]	7.611164	[91,]	4.587494	[116,]	9.426766
[17,]	4.876549	[42,]	-10.428230	[67,]	9.355415	[92,]	-31.666226	[117,]	-5.830167
[18,]	-14.877338	[43,]	11.578194	[68,]	7.098484	[93,]	-18.145429	[118,]	9.673216
[19,]	11.632130	[44,]	1.069741	[69,]	-15.406927	[94,]	12.866236	[119,]	3.166114
[20,]	30.373676	[45,]	40.818388	[70,]	-36.647121	[95,]	-14.143063	[120,]	28.413409
[21,]	5.602542	[46,]	-0.709848	[71,]	14.627057	[96,]	10.865898	[121,]	-6.356374
[22,]	0.598145	[47,]	-12.710017	[72,]	8.866574	[97,]	-12.642049	[122,]	-36.352654
[23,]	-9.902870	[48,]	-13.952070	[73,]	-7.140020	[98,]	-22.634102	[123,]	5.921355
[24,]	4.103217	[49,]	-3.443279	[74,]	12.864208	[99,]	-3.869393	[124,]	1.916790
[25,]	-2.150164	[50,]	6.308413	[75,]	5.104908	[100,]	-2.117363	[125,]	-3.835069

```
> basicStats(f_AG4)
                  f_AG4
nobs         125.000000
NAs            0.000000
Minimum     1192.806158
Maximum     1420.460017
1. Quartile 1271.359248
3. Quartile 1332.150333
Mean        1305.565566
Median      1308.134102
Sum       163195.695733
SE Mean        4.425373
LCL Mean    1296.806512
UCL Mean    1314.324619
Variance    2447.991208
Stdev         49.477179
Skewness       0.120733
Kurtosis      -0.560318
```

```
> basicStats(resiAG4) #to get variance of the error for 1-step ahead
                resiAG4
nobs         125.000000
NAs            0.000000
Minimum      -43.614997
Maximum       63.620124
1. Quartile  -10.446491
3. Quartile    6.657660
Mean          -1.322366
Median        -0.709848
Sum         -165.295733
SE Mean        1.600954
LCL Mean      -4.491103
UCL Mean       1.846371
Variance     320.381830
Stdev         17.899213
Skewness       0.416493
Kurtosis       1.500949
```

```
> v1<-qt(c(.025, .975), df=4.81);v1 #t dist with alpha 0.025 and v=4.8
1, for PI 95%
[1] -2.601425  2.601425
> v2<-qt(c(.1, .9), df=4.81); v2#t dist with alpha 0.1 and v=4.81, for
 PI 80%
[1] -1.484687  1.484687
for(i in 1:125){
  h[i]=i;h[1]
  lo95_AG4[i]<-f_AG4[i]-(2.6014*(sqrt(h[i]*320.3818)))#lower limit 95%
  hi95_AG4[i]<-f_AG4[i]+(2.6014*(sqrt(h[i]*320.3818)))#upper limit 95%
  lo80_AG4[i]<-f_AG4[i]-(1.4847*(sqrt(h[i]*320.3818)))#lower limit 80%
  hi80_AG4[i]<-f_AG4[i]+(1.4847*(sqrt(h[i]*320.3818)))#upper limit 80%
}
cbind(dt4_o,f_AG4,lo95_AG4,hi95_AG4,lo80_AG4,hi80_AG4)
```

b. Simulation data in transformation scale (log) - based on EViews results

[1]	7.160551	7.154510	7.120514	7.117679	7.084064	7.125758	7.133573	7.131575	7.132974	7.101322
[11]	7.119705	7.135965	7.136363	7.159190	7.155096	7.158996	7.164236	7.168678	7.157827	7.167521
[21]	7.191352	7.196238	7.197363	7.190598	7.194362	7.193422	7.189202	7.181888	7.182268	7.177886
[31]	7.174443	7.155682	7.157243	7.169449	7.177695	7.201847	7.192482	7.190975	7.193422	7.222695
[41]	7.219586	7.225065	7.218120	7.227249	7.228702	7.258560	7.258736	7.250424	7.241147	7.239352
[51]	7.244546	7.237735	7.234131	7.235574	7.237735	7.214628	7.218670	7.192105	7.184926	7.189089
[61]	7.180175	7.171565	7.219952	7.207980	7.188333	7.181698	7.188144	7.195863	7.201847	7.190975
[71]	7.163655	7.175592	7.183028	7.178268	7.188711	7.193234	7.173868	7.169641	7.143899	7.159579
[81]	7.147842	7.150200	7.185495	7.183408	7.184167	7.195863	7.194550	7.204639	7.206868	7.216651
[91]	7.207980	7.212048	7.189089	7.175974	7.186442	7.176357	7.185305	7.176357	7.159579	7.157243
[101]	7.156268	7.149415	7.159968	7.160940	7.158022	7.151966	7.137159	7.123543	7.128570	7.125959
[111]	7.129573	7.127567	7.127968	7.133972	7.115039	7.105026	7.113411	7.109329	7.117882	7.121120
[121]	7.144491	7.140138	7.111576	7.117070	7.119300					

Simulation data in stationary scale (First differenced of log data) - based on R (simulation for 1 series only)

```
library("fGarch", lib.loc="~/R/win-library/3.4")
set.seed(1234)
spec = garchSpec(model = list(mu=0.0007,omega = 2.5e-6, alpha = 3.45e-
2,
                                beta = 9.474e-1, shape=4.81),
cond.dist="std") #simulation for ARIMA(0,1,0)-GARCH(1,1)-t dist
f1=garchSim(spec, n = 125);f1 #results of the 125 simulation data
```

[1]	0.0032336690	0.0035946250	-0.0096160530	0.0021458690	-0.0118923500	-0.0019453350
[7]	-0.0117594500	-0.0052013150	0.0185603600	0.0009740198	-0.0062786830	0.0229451100
[13]	-0.0044352470	-0.0107953100	0.0155319800	0.0096796750	-0.0100745300	0.0050352500
[19]	0.0015485060	0.0017492980	0.0117372000	-0.0030484830	0.0026762660	0.0016542430
[25]	0.0107005100	0.0009504282	0.0083335200	0.0049469740	0.0128437100	0.0348630800
[31]	0.0063654970	0.0202228200	-0.0023715650	-0.0144468700	0.0117994200	0.0087944170
[37]	0.0040368000	0.0346885400	-0.0000290915	0.0066798420	0.0006408244	0.0154800400
[43]	0.0078802900	0.0137405700	0.0044007740	0.0034056120	-0.0031219260	-0.0136936100
[49]	-0.0080342930	-0.0465339800	-0.0351341300	0.0058209920	0.0009878060	0.0190192600
[55]	-0.0093911090	0.0019784160	-0.0165331400	-0.0013902020	-0.0085223650	0.0114143700
[61]	0.0046950500	-0.0089345740	-0.0120049900	-0.0024123590	-0.0101578500	0.0189279500
[67]	-0.0137561300	0.0065717330	0.0111003600	-0.0033840930	0.0207243500	0.0113094300
[73]	-0.0028587170	-0.0036971970	-0.0072829290	0.0048327740	-0.0096956680	0.0024382980
[79]	0.0021914480	0.0074380220	-0.0164284900	-0.0045731370	-0.0139514400	0.0017935680
[85]	-0.0382821500	-0.0069438830	0.0156552400	-0.0035594270	-0.0085826680	-0.0075571530
[91]	-0.0100078500	-0.0000245724	-0.0051379020	-0.0038235180	0.0125958500	0.0006047664
[97]	-0.0018435180	0.0061130170	-0.0119203300	0.0157608900	0.0004433674	0.0022603040
[103]	-0.0009655506	0.0080645190	-0.0110209400	0.0035320880	0.0166671700	-0.0165803400
[109]	-0.0139785000	0.0055407800	-0.0222916800	-0.0030069800	-0.0009050866	-0.0204717200
[115]	0.0092287740	0.0107621900	0.0056033040	-0.0070149470	0.0204072100	0.0028432990
[121]	-0.0007357040	-0.0059651230	-0.0243788200	-0.0071469930	-0.0165564800	

2. *n*-step ahead of ARIMA(0,1,0)-GARCH(1,1) – in original scale (price)

Date	Actual Price	Forecast *n*-step AG	95% PI	80% PI	Date	Actual Price	Forecast *n*-step AG	95% PI	80% PI
25/6/2013	1279.00	1287.62	(1116.44, 1458.80)	(1189.92, 1385.32)	6/9/2013	1387.00	1333.69	(1162.52, 1504.87)	(1236.00, 1431.39)
26/6/2013	1236.25	1288.49	(1117.31, 1459.67)	(1190.80, 1386.19)	9/9/2013	1390.00	1334.60	(1163.42, 1505.77)	(1236.90, 1432.29)
27/6/2013	1232.75	1289.36	(1118.19, 1460.54)	(1191.67, 1387.06)	10/9/2013	1358.25	1335.50	(1164.32, 1506.68)	(1237.80, 1433.19)
28/6/2013	1192.00	1290.23	(1119.06, 1461.41)	**(1192.54, 1387.93)**	11/9/2013	1363.75	1336.40	(1165.22, 1507.58)	(1238.71, 1434.10)
1/7/2013	1242.75	1291.11	(1119.93, 1462.28)	(1193.41, 1388.80)	12/9/2013	1328.00	1337.31	(1166.13, 1508.48)	(1239.61, 1435.00)
2/7/2013	1252.50	1291.98	(1120.80, 1463.16)	(1194.28, 1389.68)	13/9/2013	1318.50	1338.21	(1167.03, 1509.39)	(1240.51, 1435.91)
3/7/2013	1250.00	1292.85	(1121.68, 1464.03)	(1195.16, 1390.55)	16/9/2013	1324.00	1339.11	(1167.94, 1510.29)	(1241.42, 1436.81)
4/7/2013	1251.75	1293.73	(1122.55, 1464.91)	(1196.03, 1391.42)	17/9/2013	1312.25	1340.02	(1168.84, 1511.20)	(1242.32, 1437.72)
5/7/2013	1212.75	1294.60	(1123.43, 1465.78)	(1196.91, 1392.30)	18/9/2013	1301.00	1340.93	(1169.75, 1512.10)	(1243.23, 1438.62)
8/7/2013	1235.25	1295.48	(1124.30, 1466.66)	(1197.78, 1393.18)	19/9/2013	1365.50	1341.83	(1170.66, 1513.01)	(1244.14, 1439.53)
9/7/2013	1255.50	1296.36	(1125.18, 1467.53)	(1198.66, 1394.05)	20/9/2013	1349.25	1342.74	(1171.56, 1513.92)	(1245.04, 1440.44)
10/7/2013	1256.00	1297.23	(1126.06, 1468.41)	(1199.54, 1394.93)	21/9/2013	1323.00	1343.65	(1172.47, 1514.83)	(1245.95, 1441.35)
11/7/2013	1285.00	1298.11	(1126.93, 1469.29)	(1200.41, 1395.81)	22/9/2013	1314.25	1344.56	(1173.38, 1515.73)	(1246.86, 1442.25)
12/7/2013	1279.75	1298.99	(1127.81, 1470.16)	(1201.29, 1396.68)	23/9/2013	1322.75	1345.47	(1174.29, 1516.64)	(1247.77, 1443.16)
15/7/2013	1284.75	1299.87	(1128.69, 1471.04)	(1202.17, 1397.56)	24/9/2013	1333.00	1346.38	(1175.20, 1517.55)	(1248.68, 1444.07)
16/7/2013	1291.50	1300.75	(1129.57, 1471.92)	(1203.05, 1398.44)	25/9/2013	1341.00	1347.29	(1176.11, 1518.46)	(1249.59, 1444.98)
17/7/2013	1297.25	1301.62	(1130.45, 1472.80)	(1203.93, 1399.32)	26/9/2013	1326.50	1348.20	(1177.02, 1519.38)	(1250.50, 1445.89)
18/7/2013	1283.25	1302.51	(1131.33, 1473.68)	(1204.81, 1400.20)	1/10/2013	1290.75	1349.11	(1177.93, 1520.29)	(1251.41, 1446.81)
19/7/2013	1295.75	1303.39	(1132.21, 1474.56)	(1205.69, 1401.08)	2/10/2013	1306.25	1350.02	(1178.85, 1521.20)	(1252.33, 1447.72)
22/7/2013	1327.00	1304.27	(1133.09, 1475.44)	(1206.57, 1401.96)	3/10/2013	1316.00	1350.94	(1179.76, 1522.11)	(1253.24, 1448.63)
23/7/2013	1333.50	1305.15	(1133.97, 1476.33)	(1207.45, 1402.85)	4/10/2013	1309.75	1351.85	(1180.67, 1523.03)	(1254.15, 1449.55)
24/7/2013	1335.00	1306.03	(1134.86, 1477.21)	(1208.34, 1403.73)	7/10/2013	1323.50	1352.76	(1181.59, 1523.94)	(1255.07, 1450.46)
25/7/2013	1326.00	1306.92	(1135.74, 1478.09)	(1209.22, 1404.61)	8/10/2013	1329.50	1353.68	(1182.50, 1524.86)	(1255.98, 1451.37)
26/7/2013	1331.00	1307.80	(1136.62, 1478.98)	(1210.10, 1405.50)	9/10/2013	1304.00	1354.59	(1183.42, 1525.77)	(1256.90, 1452.29)
29/7/2013	1329.75	1308.68	(1137.51, 1479.86)	(1210.99, 1406.38)	10/10/2013	1298.50	1355.51	(1184.33, 1526.69)	(1257.81, 1453.21)
30/7/2013	1324.15	1309.57	(1138.39, 1480.75)	(1211.87, 1407.27)	11/10/2013	1265.50	1356.43	(1185.25, 1527.60)	(1258.73, 1454.12)
31/7/2013	1314.50	1310.45	(1139.28, 1481.63)	(1212.76, 1408.15)	14/10/2013	1285.50	1357.34	(1186.17, 1528.52)	(1259.65, 1455.04)
1/8/2013	1315.00	1311.34	(1140.16, 1482.52)	(1213.64, 1409.04)	15/10/2013	1270.50	1358.26	(1187.09, 1529.44)	(1260.57, 1455.96)
2/8/2013	1309.25	1312.23	(1141.05, 1483.40)	(1214.53, 1409.92)	16/10/2013	1273.50	1359.18	(1188.00, 1530.36)	(1261.48, 1456.88)
5/8/2013	1304.75	1313.12	(1141.94, 1484.29)	(1215.42, 1410.81)	17/10/2013	1319.25	1360.10	(1188.92, 1531.28)	(1262.40, 1457.80)
6/8/2013	1280.50	1314.00	(1142.83, 1485.18)	(1216.31, 1411.70)	18/10/2013	1316.50	1361.02	(1189.84, 1532.20)	(1263.32, 1458.72)
7/8/2013	1282.50	1314.89	(1143.71, 1486.07)	(1217.20, 1412.59)	19/10/2013	1317.50	1361.94	(1190.76, 1533.12)	(1264.24, 1459.64)
8/8/2013	1298.25	1315.78	(1144.60, 1486.96)	(1218.09, 1413.48)	20/10/2013	1333.00	1362.86	(1191.68, 1534.04)	(1265.17, 1460.56)
9/8/2013	1309.00	1316.67	(1145.49, 1487.85)	(1218.98, 1414.37)	21/10/2013	1331.25	1363.78	(1192.61, 1534.96)	(1266.09, 1461.48)
12/8/2013	1341.00	1317.56	(1146.38, 1488.74)	(1219.87, 1415.26)	22/10/2013	1344.75	1364.71	(1193.53, 1535.88)	(1267.01, 1462.40)
13/8/2013	1328.50	1318.45	(1147.28, 1489.63)	(1220.76, 1416.15)	23/10/2013	1347.75	1365.63	(1194.45, 1536.81)	(1267.93, 1463.32)
14/8/2013	1326.50	1319.34	(1148.17, 1490.52)	(1221.65, 1417.04)	24/10/2013	1361.00	1366.55	(1195.38, 1537.73)	(1268.86, 1464.25)
15/8/2013	1329.75	1320.24	(1149.06, 1491.41)	(1222.54, 1417.93)	25/10/2013	1349.25	1367.48	(1196.30, 1538.65)	(1269.78, 1465.17)
16/8/2013	1369.25	1321.13	(1149.95, 1492.31)	(1223.43, 1418.83)	26/10/2013	1354.75	1368.40	(1197.22, 1539.58)	(1270.71, 1466.10)
19/8/2013	1365.00	1322.02	(1150.85, 1493.20)	(1224.33, 1419.72)	27/10/2013	1324.00	1369.33	(1198.15, 1540.50)	(1271.63, 1467.02)
20/8/2013	1372.50	1322.92	(1151.74, 1494.09)	(1225.22, 1420.61)	1/11/2013	1306.75	1370.25	(1199.08, 1541.43)	(1272.56, 1467.95)
21/8/2013	1363.00	1323.81	(1152.63, 1494.99)	(1226.12, 1421.51)	4/11/2013	1320.50	1371.18	(1200.00, 1542.36)	(1273.48, 1468.88)
22/8/2013	1375.50	1324.71	(1153.53, 1495.88)	(1227.01, 1422.40)	5/11/2013	1307.25	1372.11	(1200.93, 1543.28)	(1274.41, 1469.80)
23/8/2013	1377.50	1325.60	(1154.43, 1496.78))	(1227.91, 1423.30)	6/11/2013	1319.00	1373.03	(1201.86, 1544.21)	(1275.34, 1470.73)
27/8/2013	1419.25	1326.50	(1155.32, 1497.68)	(1228.80, 1424.20)	7/11/2013	1307.25	1373.96	(1202.79, 1545.14)	(1276.27, 1471.66)
28/8/2013	1419.50	1327.40	(1156.22, 1498.57)	(1229.70, 1425.09)	8/11/2013	1285.50	1374.89	(1203.72, 1546.07)	(1277.20, 1472.59)
29/8/2013	1407.75	1328.29	(1157.12, 1499.47)	(1230.60, 1425.99)	11/11/2013	1282.50	1375.82	(1204.65, 1547.00)	(1278.13, 1473.52)
30/8/2013	1394.75	1329.19	(1158.02, 1500.37)	(1231.50, 1426.89)	12/11/2013	1281.25	1376.75	(1205.58, 1547.93)	(1279.06, 1474.45)
2/9/2013	1392.25	1330.09	(1158.91, 1501.27)	(1232.40, 1427.79)	13/11/2013	1272.50	1377.68	(1206.51, 1548.86)	**(1279.99, 1475.38)**
3/9/2013	1399.50	1330.99	(1159.81, 1502.17)	(1233.30, 1428.69)	14/11/2013	1286.00	1378.62	(1207.44, 1549.79)	(1280.92, 1476.31)
4/9/2013	1390.00	1331.89	(1160.71, 1503.07)	(1234.20, 1429.59)	15/11/2013	1287.25	1379.55	(1208.37, 1550.73)	(1281.85, 1477.24)
5/9/2013	1385.00	1332.79	(1161.62, 1503.97)	(1235.10, 1430.49)	16/11/2013	1283.50	1380.48	(1209.30, 1551.66)	(1282.79, 1478.18)

Date	Actual Price	Forecast *n*-step AG	95% PI	80% PI	Date	Actual Price	Forecast *n*-step AG	95% PI	80% PI
17/11/2013	1275.75	1381.41	(1210.24, 1552.59)	**(1283.72, 1479.11)**	4/12/2013	1227.50	1391.73	(1220.55, 1562.90)	**(1294.03, 1489.42)**
18/11/2013	1257.00	1382.35	(1211.17, 1553.53)	**(1284.65, 1480.05)**	5/12/2013	1222.50	1392.67	(1221.49, 1563.84)	**(1294.97, 1490.36)**
19/11/2013	1240.00	1383.28	(1212.11, 1554.46)	**(1285.59, 1480.98)**	6/12/2013	1233.00	1393.61	(1222.43, 1564.79)	**(1295.91, 1491.31)**
20/11/2013	1246.25	1384.22	(1213.04, 1555.40)	**(1286.52, 1481.92)**	9/12/2013	1237.00	1394.55	(1223.38, 1565.73)	**(1296.86, 1492.25)**
21/11/2013	1243.00	1385.16	(1213.98, 1556.33)	**(1287.46, 1482.85)**	10/12/2013	1266.25	1395.50	(1224.32, 1566.67)	**(1297.80, 1493.19)**
22/11/2013	1247.50	1386.09	(1214.92, 1557.27)	**(1288.40, 1483.79)**	11/12/2013	1260.75	1396.44	(1225.26, 1567.62)	**(1298.74, 1494.14)**
23/11/2013	1245.00	1387.03	(1215.85, 1558.21)	**(1289.33, 1484.73)**	12/12/2013	1225.25	1397.38	**(1226.21, 1568.56)**	**(1299.69, 1495.08)**
24/11/2013	1245.50	1387.97	(1216.79, 1559.14)	**(1290.27, 1485.66)**	13/12/2013	1232.00	1398.33	(1227.15, 1569.51)	**(1300.63, 1496.02)**
25/11/2013	1253.00	1388.91	(1217.73, 1560.08)	**(1291.21, 1486.60)**	16/12/2013	1234.75	1399.27	(1228.10, 1570.45)	**(1301.58, 1496.97)**
2/12/2013	1229.50	1389.85	(1218.67, 1561.02)	**(1292.15, 1487.54)**	17/12/2013	1231.75	1400.22	(1229.04, 1571.40)	**(1302.52, 1497.92)**
3/12/2013	1217.25	1390.79	**(1219.61, 1561.96)**	**(1293.09, 1488.48)**					

Note:

1. Actual data outside 80% prediction interval: 23/125 (18.4%), specifically starting at 104th data
2. Actual data outside 95% prediction interval: 2/125 (1.6%)

To find Prediction Interval for *n*-step ahead ARIMA-GARCH:

Residual data for 125-step AG

```
for(i in 1:125){
  resiAG4_n[i]<- dt4_o[i]-f_AG4_n[i];resiAG4_n[125]
}
resiAG4_n
```

Data	Residual	Data	Residual	Data	Residual	Data	Residual	Data	Residual
[1,]	-8.620238	[26,]	14.581507	[51,]	58.109129	[76,]	-50.593745	[101,]	-105.183600
[2,]	-52.241065	[27,]	4.045837	[52,]	52.208363	[77,]	-57.009865	[102,]	-92.615336
[3,]	-56.612480	[28,]	3.659569	[53,]	53.306987	[78,]	-90.926605	[103,]	-92.297703
[4,]	-98.234484	[29,]	-2.977300	[54,]	55.405002	[79,]	-71.843965	[104,]	-96.980700
[5,]	-48.357079	[30,]	-8.364768	[55,]	22.752406	[80,]	-87.761946	[105,]	-105.664328
[6,]	-39.480263	[31,]	-33.502837	[56,]	27.349201	[81,]	-85.680548	[106,]	-125.348587
[7,]	-42.854038	[32,]	-32.391506	[57,]	-9.304615	[82,]	-40.849770	[107,]	-143.283479
[8,]	-41.978403	[33,]	-17.530776	[58,]	-19.709043	[83,]	-44.519614	[108,]	-137.969004
[9,]	-81.853360	[34,]	-7.670647	[59,]	-15.114083	[84,]	-44.440081	[109,]	-142.155160
[10,]	-60.228908	[35,]	23.438880	[60,]	-27.769734	[85,]	-29.861170	[110,]	-138.591950
[11,]	-40.855050	[36,]	10.047805	[61,]	-39.925998	[86,]	-32.532881	[111,]	-142.029374
[12,]	-41.231784	[37,]	7.156127	[62,]	23.667124	[87,]	-19.955216	[112,]	-142.467431
[13,]	-13.109110	[38,]	9.513846	[63,]	6.509634	[88,]	-17.878175	[113,]	-135.906123
[14,]	-19.237029	[39,]	48.120961	[64,]	-20.648470	[89,]	-5.551758	[114,]	-160.345449
[15,]	-15.115542	[40,]	42.977472	[65,]	-30.307188	[90,]	-18.225966	[115,]	-173.535411
[16,]	-9.244650	[41,]	49.583379	[66,]	-22.716521	[91,]	-13.650799	[116,]	-164.226009
[17,]	-4.374353	[42,]	39.188682	[67,]	-13.376469	[92,]	-45.326256	[117,]	-170.167243
[18,]	-19.254650	[43,]	50.793379	[68,]	-6.287033	[93,]	-63.502340	[118,]	-160.609112
[19,]	-7.635542	[44,]	51.897472	[69,]	-21.698211	[94,]	-50.679051	[119,]	-157.551619
[20,]	22.732970	[45,]	92.750958	[70,]	-58.360007	[95,]	-64.856389	[120,]	-129.244763
[21,]	28.350886	[46,]	92.103838	[71,]	-43.772419	[96,]	-54.034354	[121,]	-135.688546
[22,]	28.968205	[47,]	79.456111	[72,]	-34.935449	[97,]	-66.712946	[122,]	-172.132966
[23,]	19.084927	[48,]	65.557778	[73,]	-42.099096	[98,]	-89.392166	[123,]	-166.328026
[24,]	23.201051	[49,]	62.158836	[74,]	-29.263360	[99,]	-93.322015	[124,]	-164.523724
[25,]	21.066578	[50,]	68.509287	[75,]	-24.178243	[100,]	-95.502493	[125,]	-168.470062

```
> basicStats(f_AG4_n)
                  f_AG4_n
nobs          125.000000
NAs             0.000000
Minimum      1287.620238
Maximum      1400.220062
1. Quartile  1314.891506
3. Quartile  1371.179051
Mean         1343.139947
Median       1342.740366
Sum        167892.493435
SE Mean         2.942299
LCL Mean     1337.316313
UCL Mean     1348.963582
Variance     1082.140728
Stdev          32.895907
Skewness        0.028917
Kurtosis       -1.227913
```

```
> basicStats(resiAG4_n)

               resiAG4_n
nobs          125.000000
NAs             0.000000
Minimum      -173.535411
Maximum        92.750958
1. Quartile   -85.680548
3. Quartile     4.045837
Mean          -38.896747
Median        -30.307188
Sum         -4862.093435
SE Mean         5.885497
LCL Mean      -50.545795
UCL Mean      -27.247700
Variance     4329.884925
Stdev          65.801861
Skewness       -0.337425
Kurtosis       -0.568069
```

```
> v1<-qt(c(.025, .975), df=4.81);v1 #t dist with alpha 0.025 and v=4.8
1, for PI 95%
[1] -2.601425  2.601425
> v2<-qt(c(.1, .9), df=4.81); v2#t dist with alpha 0.1 and v=4.81, for
 PI 80%
[1] -1.484687  1.484687
for(i in 1:125){
        h[i]=i;h[1]
        lo95_AG4_n[i]<-f_AG4_n[i]-(2.6014*65.8019)#lower limit 95%
        hi95_AG4_n[i]<-f_AG4_n[i]+(2.6014*65.8019)#upper limit 95%
        lo80_AG4_n[i]<-f_AG4_n[i]-(1.4847*65.8019)#lower limit 80%
        hi80_AG4_n[i]<-f_AG4_n[i]+(1.4847*65.8019)#upper limit 80%
      }
cbind(dt4_o,f_AG4_n,lo95_AG4_n,hi95_AG4_n,lo80_AG4_n,hi80_AG4_n)
```

C. ERRORS CALCULATION

1(a) 1-step ahead error of ARIMA(0,1,0)-GARCH(1,1) in original scale

Data	Error	Abs. error	error ^2	Abs (error/y)	Data	Error	Abs. error	error ^2	Abs (error/y)
[1,]	-8.620238	8.620238	74.308510	0.006740	[64,]	-27.162506	27.162506	737.801700	0.020531
[2,]	-43.614997	43.614997	1902.268000	0.035280	[65,]	-9.644754	9.644754	93.021270	0.007339
[3,]	-4.336084	4.336084	18.801630	0.003517	[66,]	7.611164	7.611164	57.929810	0.005754
[4,]	-41.583717	41.583717	1729.206000	0.034886	[67,]	9.355415	9.355415	87.523790	0.007018
[5,]	49.943842	49.943842	2494.387000	0.040188	[68,]	7.098484	7.098484	50.388470	0.005293
[6,]	8.909520	8.909520	79.379550	0.007113	[69,]	-15.406927	15.406927	237.373400	0.011615
[7,]	-3.347074	3.347074	11.202900	0.002678	[70,]	-36.647121	36.647121	1343.011000	0.028392
[8,]	0.904617	0.904617	0.818332	0.000723	[71,]	14.627057	14.627057	213.950800	0.011198
[9,]	-39.846567	39.846567	1587.749000	0.032856	[72,]	8.866574	8.866574	78.616140	0.006738
[10,]	21.679809	21.679809	470.014100	0.017551	[73,]	-7.140020	7.140020	50.979880	0.005451
[11,]	19.414592	19.414592	376.926400	0.015464	[74,]	12.864208	12.864208	165.487800	0.009720
[12,]	-0.349103	0.349103	0.121873	0.000278	[75,]	5.104908	5.104908	26.060080	0.003840
[13,]	28.150559	28.150559	792.454000	0.021907	[76,]	-26.399150	26.399150	696.915100	0.020245
[14,]	-6.119054	6.119054	37.442820	0.004781	[77,]	-6.381903	6.381903	40.728690	0.004915
[15,]	4.134497	4.134497	17.094070	0.003218	[78,]	-33.878184	33.878184	1147.731000	0.026771
[16,]	5.881115	5.881115	34.587510	0.004554	[79,]	19.144134	19.144134	366.497900	0.014892
[17,]	4.876549	4.876549	23.780730	0.003759	[80,]	-15.869393	15.869393	251.837600	0.012491
[18,]	-14.877338	14.877338	221.335200	0.011593	[81,]	2.140752	2.140752	4.582820	0.001681
[19,]	11.632130	11.632130	135.306500	0.008977	[82,]	44.888724	44.888724	2014.998000	0.034026
[20,]	30.373676	30.373676	922.560200	0.022889	[83,]	-3.642217	3.642217	13.265750	0.002767
[21,]	5.602542	5.602542	31.388470	0.004201	[84,]	0.109642	0.109642	0.012021	0.000083
[22,]	0.598145	0.598145	0.357777	0.000448	[85,]	14.608967	14.608967	213.421900	0.010959
[23,]	-9.902870	9.902870	98.066820	0.007468	[86,]	-2.651516	2.651516	7.030539	0.001992
[24,]	4.103217	4.103217	16.836390	0.003083	[87,]	12.599668	12.599668	158.751600	0.009370
[25,]	-2.150164	2.150164	4.623203	0.001617	[88,]	2.090537	2.090537	4.370345	0.001551
[26,]	-6.499319	6.499319	42.241150	0.004908	[89,]	12.338508	12.338508	152.238800	0.009066
[27,]	-10.545531	10.545531	111.208200	0.008022	[90,]	-12.670454	12.670454	160.540400	0.009391
[28,]	-0.389005	0.389005	0.151325	0.000296	[91,]	4.587494	4.587494	21.045100	0.003386
[29,]	-6.639344	6.639344	44.080880	0.005071	[92,]	-31.666226	31.666226	1002.750000	0.023917
[30,]	-5.385454	5.385454	29.003120	0.004128	[93,]	-18.145429	18.145429	329.256600	0.013886
[31,]	-25.132412	25.132412	631.638100	0.019627	[94,]	12.866236	12.866236	165.540000	0.009743
[32,]	1.133989	1.133989	1.285931	0.000884	[95,]	-14.143063	14.143063	200.026200	0.010819
[33,]	14.882637	14.882637	221.492900	0.011464	[96,]	10.865898	10.865898	118.067700	0.008238
[34,]	9.871985	9.871985	97.456080	0.007542	[97,]	-12.642049	12.642049	159.821400	0.009671
[35,]	31.114715	31.114715	968.125500	0.023203	[98,]	-22.634102	22.634102	512.302600	0.017607
[36,]	-13.406927	13.406927	179.745700	0.010092	[99,]	-3.869393	3.869393	14.972200	0.003017
[37,]	-2.898473	2.898473	8.401146	0.002185	[100,]	-2.117363	2.117363	4.483227	0.001653
[38,]	2.352879	2.352879	5.536038	0.001769	[101,]	-9.616518	9.616518	92.477420	0.007557
[39,]	38.600681	38.600681	1490.013000	0.028191	[102,]	12.639400	12.639400	159.754400	0.009828
[40,]	-5.176033	5.176033	26.791320	0.003792	[103,]	0.380270	0.380270	0.144605	0.000295
[41,]	6.576841	6.576841	43.254840	0.004792	[104,]	-4.620576	4.620576	21.349720	0.003600
[42,]	-10.428230	10.428230	108.748000	0.007651	[105,]	-8.618039	8.618039	74.270600	0.006755
[43,]	11.578194	11.578194	134.054600	0.008417	[106,]	-19.612798	19.612798	384.661900	0.015603
[44,]	1.069741	1.069741	1.144345	0.000777	[107,]	-17.850118	17.850118	318.626700	0.014395
[45,]	40.818388	40.818388	1666.141000	0.028761	[108,]	5.411379	5.411379	29.283030	0.004342
[46,]	-0.709848	0.709848	0.503884	0.000500	[109,]	-4.092847	4.092847	16.751390	0.003293
[47,]	-12.710017	12.710017	161.544500	0.009029	[110,]	3.659350	3.659350	13.390840	0.002933
[48,]	-13.952070	13.952070	194.660300	0.010003	[111,]	-3.343693	3.343693	11.180280	0.002686
[49,]	-3.443279	3.443279	11.856170	0.002473	[112,]	-0.342002	0.342002	0.116965	0.000275
[50,]	6.308413	6.308413	39.796070	0.004508	[113,]	6.657660	6.657660	44.324430	0.005313
[51,]	-10.446491	10.446491	109.129200	0.007515	[114,]	-24.347412	24.347412	592.796500	0.019803
[52,]	-5.940066	5.940066	35.284380	0.004289	[115,]	-13.081519	13.081519	171.126100	0.010747
[53,]	1.063315	1.063315	1.130638	0.000767	[116,]	9.426766	9.426766	88.863910	0.007680
[54,]	2.061963	2.061963	4.251692	0.001483	[117,]	-5.830167	5.830167	33.990840	0.004769
[55,]	-32.690066	32.690066	1068.640000	0.024068	[118,]	9.673216	9.673216	93.571100	0.007845
[56,]	4.581407	4.581407	20.989290	0.003359	[119,]	3.166114	3.166114	10.024280	0.002560
[57,]	-36.672313	36.672313	1344.859000	0.027615	[120,]	28.413409	28.413409	807.321800	0.022439
[58,]	-10.398135	10.398135	108.121200	0.007886	[121,]	-6.356374	6.356374	40.403490	0.005042
[59,]	4.608290	4.608290	21.236330	0.003481	[122,]	-36.352654	36.352654	1321.515000	0.029670
[60,]	-12.645429	12.645429	159.906900	0.009636	[123,]	5.921355	5.921355	35.062450	0.004806
[61,]	-12.137483	12.137483	147.318500	0.009329	[124,]	1.916790	1.916790	3.674085	0.001552
[62,]	63.620125	63.620125	4047.520000	0.046591	[125,]	-3.835069	3.835069	14.707750	0.003114
[63,]	-17.173497	17.173497	294.929000	0.012728					

```
> mae_AG4=sum(error2_AG4)/T; mae_AG4  [1] 12.93009
> rmse_AG4=sqrt(sum(error3_AG4)/T); rmse_AG4 [1] 17.87645
> mape_AG4=(100/T)*sum(error4_AG4); mape_AG4 [1] 0.9956225
```

1(b) 1-step ahead error of ARIMA(0,1,0)-GARCH(1,1) in transformed (log) scale

Data	Error	Abs. error	error ^2	Abs (error/y*)	Data	Error	Abs. error	error ^2	Abs (error/y*)
[1,]	-0.006717	0.006717	0.000045	0.000939	[64,]	-0.020323	0.020323	0.000413	0.002827
[2,]	-0.034672	0.034672	0.001202	0.004870	[65,]	-0.007312	0.007312	0.000053	0.001018
[3,]	-0.003511	0.003511	0.000012	0.000493	[66,]	0.005771	0.005771	0.000033	0.000803
[4,]	-0.034291	0.034291	0.001176	0.004841	[67,]	0.007043	0.007043	0.000050	0.000979
[5,]	0.041018	0.041018	0.001682	0.005757	[68,]	0.005307	0.005307	0.000028	0.000737
[6,]	0.007139	0.007139	0.000051	0.001001	[69,]	-0.011548	0.011548	0.000133	0.001606
[7,]	-0.002674	0.002674	0.000007	0.000375	[70,]	-0.027997	0.027997	0.000784	0.003909
[8,]	0.000723	0.000723	0.000001	0.000101	[71,]	0.011261	0.011261	0.000127	0.001569
[9,]	-0.032328	0.032328	0.001045	0.004553	[72,]	0.006760	0.006760	0.000046	0.000941
[10,]	0.017707	0.017707	0.000314	0.002487	[73,]	-0.005437	0.005437	0.000030	0.000757
[11,]	0.015584	0.015584	0.000243	0.002184	[74,]	0.009767	0.009767	0.000095	0.001359
[12,]	-0.000278	0.000278	0.000000	0.000039	[75,]	0.003847	0.003847	0.000015	0.000535
[13,]	0.022151	0.022151	0.000491	0.003094	[76,]	-0.020043	0.020043	0.000402	0.002794
[14,]	-0.004770	0.004770	0.000023	0.000667	[77,]	-0.004903	0.004903	0.000024	0.000684
[15,]	0.003223	0.003223	0.000010	0.000450	[78,]	-0.026419	0.026419	0.000698	0.003698
[16,]	0.004564	0.004564	0.000021	0.000637	[79,]	0.015004	0.015004	0.000225	0.002096
[17,]	0.003766	0.003766	0.000014	0.000525	[80,]	-0.012413	0.012413	0.000154	0.001737
[18,]	-0.011527	0.011527	0.000133	0.001611	[81,]	0.001682	0.001682	0.000003	0.000235
[19,]	0.009018	0.009018	0.000081	0.001258	[82,]	0.034618	0.034618	0.001198	0.004818
[20,]	0.023155	0.023155	0.000536	0.003220	[83,]	-0.002763	0.002763	0.000008	0.000385
[21,]	0.004210	0.004210	0.000018	0.000585	[84,]	0.000083	0.000083	0.000000	0.000012
[22,]	0.000448	0.000448	0.000000	0.000062	[85,]	0.011020	0.011020	0.000121	0.001532
[23,]	-0.007440	0.007440	0.000055	0.001035	[86,]	-0.001990	0.001990	0.000004	0.000277
[24,]	0.003088	0.003088	0.000010	0.000429	[87,]	0.009414	0.009414	0.000089	0.001307
[25,]	-0.001616	0.001616	0.000003	0.000225	[88,]	0.001552	0.001552	0.000002	0.000215
[26,]	-0.004896	0.004896	0.000024	0.000681	[89,]	0.009107	0.009107	0.000083	0.001262
[27,]	-0.007990	0.007990	0.000064	0.001113	[90,]	-0.009347	0.009347	0.000087	0.001297
[28,]	-0.000296	0.000296	0.000000	0.000041	[91,]	0.003392	0.003392	0.000012	0.000470
[29,]	-0.005058	0.005058	0.000026	0.000705	[92,]	-0.023636	0.023636	0.000559	0.003288
[30,]	-0.004119	0.004119	0.000017	0.000574	[93,]	-0.013790	0.013790	0.000190	0.001922
[31,]	-0.019437	0.019437	0.000378	0.002717	[94,]	0.009791	0.009791	0.000096	0.001363
[32,]	0.000885	0.000885	0.000001	0.000124	[95,]	-0.010761	0.010761	0.000116	0.001500
[33,]	0.011530	0.011530	0.000133	0.001608	[96,]	0.008272	0.008272	0.000068	0.001151
[34,]	0.007570	0.007570	0.000057	0.001055	[97,]	-0.009624	0.009624	0.000093	0.001341
[35,]	0.023476	0.023476	0.000551	0.003260	[98,]	-0.017454	0.017454	0.000305	0.002438
[36,]	-0.010041	0.010041	0.000101	0.001396	[99,]	-0.003013	0.003013	0.000009	0.000421
[37,]	-0.002183	0.002183	0.000005	0.000304	[100,]	-0.001651	0.001651	0.000003	0.000231
[38,]	0.001771	0.001771	0.000003	0.000246	[101,]	-0.007529	0.007529	0.000057	0.001053
[39,]	0.028596	0.028596	0.000818	0.003960	[102,]	0.009877	0.009877	0.000098	0.001380
[40,]	-0.003785	0.003785	0.000014	0.000524	[103,]	0.000295	0.000295	0.000000	0.000041
[41,]	0.004803	0.004803	0.000023	0.000665	[104,]	-0.003594	0.003594	0.000013	0.000502
[42,]	-0.007622	0.007622	0.000058	0.001056	[105,]	-0.006733	0.006733	0.000045	0.000941
[43,]	0.008453	0.008453	0.000071	0.001170	[106,]	-0.015482	0.015482	0.000240	0.002169
[44,]	0.000777	0.000777	0.000001	0.000107	[107,]	-0.014293	0.014293	0.000204	0.002007
[45,]	0.029182	0.029182	0.000852	0.004021	[108,]	0.004352	0.004352	0.000019	0.000611
[46,]	-0.000500	0.000500	0.000000	0.000069	[109,]	-0.003287	0.003287	0.000011	0.000461
[47,]	-0.008988	0.008988	0.000081	0.001240	[110,]	0.002938	0.002938	0.000009	0.000412
[48,]	-0.009954	0.009954	0.000099	0.001375	[111,]	-0.002682	0.002682	0.000007	0.000376
[49,]	-0.002470	0.002470	0.000006	0.000341	[112,]	-0.000275	0.000275	0.000000	0.000039
[50,]	0.004518	0.004518	0.000020	0.000624	[113,]	0.005328	0.005328	0.000028	0.000747
[51,]	-0.007487	0.007487	0.000056	0.001035	[114,]	-0.019609	0.019609	0.000385	0.002756
[52,]	-0.004280	0.004280	0.000018	0.000592	[115,]	-0.010689	0.010689	0.000114	0.001505
[53,]	0.000767	0.000767	0.000001	0.000106	[116,]	0.007709	0.007709	0.000059	0.001084
[54,]	0.001485	0.001485	0.000002	0.000205	[117,]	-0.004758	0.004758	0.000023	0.000669
[55,]	-0.023783	0.023783	0.000566	0.003297	[118,]	0.007876	0.007876	0.000062	0.001107
[56,]	0.003365	0.003365	0.000011	0.000466	[119,]	0.002563	0.002563	0.000007	0.000360
[57,]	-0.027240	0.027240	0.000742	0.003788	[120,]	0.022695	0.022695	0.000515	0.003177
[58,]	-0.007855	0.007855	0.000062	0.001093	[121,]	-0.005029	0.005029	0.000025	0.000704
[59,]	0.003487	0.003487	0.000012	0.000485	[122,]	-0.029238	0.029238	0.000855	0.004112
[60,]	-0.009590	0.009590	0.000092	0.001336	[123,]	0.004818	0.004818	0.000023	0.000677
[61,]	-0.009286	0.009286	0.000086	0.001295	[124,]	0.001554	0.001554	0.000002	0.000218
[62,]	0.047711	0.047711	0.002276	0.006609	[125,]	-0.003109	0.003109	0.000010	0.000437
[63,]	-0.012648	0.012648	0.000160	0.001755					

```
> mae_t_AG4=sum(error2t_AG4)/T; mae_t_AG4 [1] 0.009951418
> rmse_t_AG4=sqrt(sum(error3t_AG4)/T); rmse_t_AG4 [1] 0.01380593
> mape_t_AG4=(100/T)*sum(error4t_AG4); mape_t_AG4 [1] 0.1388268
```

1(c) 1-step ahead error of ARIMA(0,1,0)-GARCH(1,1) in stationary scale

Data	Error	Absolute error	error ^2	Data	Error	Absolute error	error ^2
[1,]	-0.006717	0.006717	0.000045	[64,]	-0.020323	0.020323	0.000413
[2,]	-0.034672	0.034672	0.001202	[65,]	-0.007312	0.007312	0.000053
[3,]	-0.003511	0.003511	0.000012	[66,]	0.005771	0.005771	0.000033
[4,]	-0.034291	0.034291	0.001176	[67,]	0.007043	0.007043	0.000050
[5,]	0.041018	0.041018	0.001682	[68,]	0.005307	0.005307	0.000028
[6,]	0.007139	0.007139	0.000051	[69,]	-0.011548	0.011548	0.000133
[7,]	-0.002674	0.002674	0.000007	[70,]	-0.027997	0.027997	0.000784
[8,]	0.000723	0.000723	0.000001	[71,]	0.011261	0.011261	0.000127
[9,]	-0.032328	0.032328	0.001045	[72,]	0.006760	0.006760	0.000046
[10,]	0.017707	0.017707	0.000314	[73,]	-0.005437	0.005437	0.000030
[11,]	0.015584	0.015584	0.000243	[74,]	0.009767	0.009767	0.000095
[12,]	-0.000278	0.000278	0.000000	[75,]	0.003847	0.003847	0.000015
[13,]	0.022151	0.022151	0.000491	[76,]	-0.020043	0.020043	0.000402
[14,]	-0.004770	0.004770	0.000023	[77,]	-0.004903	0.004903	0.000024
[15,]	0.003223	0.003223	0.000010	[78,]	-0.026419	0.026419	0.000698
[16,]	0.004564	0.004564	0.000021	[79,]	0.015004	0.015004	0.000225
[17,]	0.003766	0.003766	0.000014	[80,]	-0.012413	0.012413	0.000154
[18,]	-0.011527	0.011527	0.000133	[81,]	0.001682	0.001682	0.000003
[19,]	0.009018	0.009018	0.000081	[82,]	0.034618	0.034618	0.001198
[20,]	0.023155	0.023155	0.000536	[83,]	-0.002763	0.002763	0.000008
[21,]	0.004210	0.004210	0.000018	[84,]	0.000083	0.000083	0.000000
[22,]	0.000448	0.000448	0.000000	[85,]	0.011020	0.011020	0.000121
[23,]	-0.007440	0.007440	0.000055	[86,]	-0.001990	0.001990	0.000004
[24,]	0.003088	0.003088	0.000010	[87,]	0.009414	0.009414	0.000089
[25,]	-0.001616	0.001616	0.000003	[88,]	0.001552	0.001552	0.000002
[26,]	-0.004896	0.004896	0.000024	[89,]	0.009107	0.009107	0.000083
[27,]	-0.007990	0.007990	0.000064	[90,]	-0.009347	0.009347	0.000087
[28,]	-0.000296	0.000296	0.000000	[91,]	0.003392	0.003392	0.000012
[29,]	-0.005058	0.005058	0.000026	[92,]	-0.023636	0.023636	0.000559
[30,]	-0.004119	0.004119	0.000017	[93,]	-0.013790	0.013790	0.000190
[31,]	-0.019437	0.019437	0.000378	[94,]	0.009791	0.009791	0.000096
[32,]	0.000885	0.000885	0.000001	[95,]	-0.010761	0.010761	0.000116
[33,]	0.011530	0.011530	0.000133	[96,]	0.008272	0.008272	0.000068
[34,]	0.007570	0.007570	0.000057	[97,]	-0.009624	0.009624	0.000093
[35,]	0.023476	0.023476	0.000551	[98,]	-0.017454	0.017454	0.000305
[36,]	-0.010041	0.010041	0.000101	[99,]	-0.003013	0.003013	0.000009
[37,]	-0.002183	0.002183	0.000005	[100,]	-0.001651	0.001651	0.000003
[38,]	0.001771	0.001771	0.000003	[101,]	-0.007529	0.007529	0.000057
[39,]	0.028596	0.028596	0.000818	[102,]	0.009877	0.009877	0.000098
[40,]	-0.003785	0.003785	0.000014	[103,]	0.000295	0.000295	0.000000
[41,]	0.004803	0.004803	0.000023	[104,]	-0.003594	0.003594	0.000013
[42,]	-0.007622	0.007622	0.000058	[105,]	-0.006733	0.006733	0.000045
[43,]	0.008453	0.008453	0.000071	[106,]	-0.015482	0.015482	0.000240
[44,]	0.000777	0.000777	0.000001	[107,]	-0.014293	0.014293	0.000204
[45,]	0.029182	0.029182	0.000852	[108,]	0.004352	0.004352	0.000019
[46,]	-0.000500	0.000500	0.000000	[109,]	-0.003287	0.003287	0.000011
[47,]	-0.008988	0.008988	0.000081	[110,]	0.002938	0.002938	0.000009
[48,]	-0.009954	0.009954	0.000099	[111,]	-0.002682	0.002682	0.000007
[49,]	-0.002470	0.002470	0.000006	[112,]	-0.000275	0.000275	0.000000
[50,]	0.004518	0.004518	0.000020	[113,]	0.005328	0.005328	0.000028
[51,]	-0.007487	0.007487	0.000056	[114,]	-0.019609	0.019609	0.000385
[52,]	-0.004280	0.004280	0.000018	[115,]	-0.010689	0.010689	0.000114
[53,]	0.000767	0.000767	0.000001	[116,]	0.007709	0.007709	0.000059
[54,]	0.001485	0.001485	0.000002	[117,]	-0.004758	0.004758	0.000023
[55,]	-0.023783	0.023783	0.000566	[118,]	0.007876	0.007876	0.000062
[56,]	0.003365	0.003365	0.000011	[119,]	0.002563	0.002563	0.000007
[57,]	-0.027240	0.027240	0.000742	[120,]	0.022695	0.022695	0.000515
[58,]	-0.007855	0.007855	0.000062	[121,]	-0.005029	0.005029	0.000025
[59,]	0.003487	0.003487	0.000012	[122,]	-0.029238	0.029238	0.000855
[60,]	-0.009590	0.009590	0.000092	[123,]	0.004818	0.004818	0.000023
[61,]	-0.009286	0.009286	0.000086	[124,]	0.001554	0.001554	0.000002
[62,]	0.047711	0.047711	0.002276	[125,]	-0.003109	0.003109	0.000010
[63,]	-0.012648	0.012648	0.000160				

```
> mae_st_AG4=sum(error2st_AG4)/T; mae_st_AG4 [1] 0.009951418
> rmse_st_AG4=sqrt(sum(error3st_AG4)/T); rmse_st_AG4 [1] 0.01380593
```

2. *n*-step ahead forecasting performance for ARIMA(0,1,0)-GARCH(1,1) in original scale

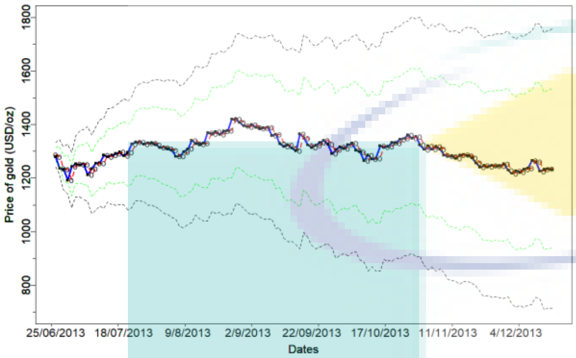
Data	Error	Abs. error	error ^2	Abs (error/y*)	Data	Error	Abs. error	error ^2	Abs (error/y*)
[1,]	-8.620238	8.620238	74.308505	0.006740	[64,]	-20.648470	20.648470	426.359307	0.015607
[2,]	-52.241065	52.241065	2729.128842	0.042258	[65,]	-30.307188	30.307188	918.525645	0.023060
[3,]	-56.612480	56.612480	3204.972868	0.045924	[66,]	-22.716521	22.716521	516.040340	0.017174
[4,]	-98.234484	98.234484	9650.013847	0.082411	[67,]	-13.376469	13.376469	178.929936	0.010035
[5,]	-48.357079	48.357079	2338.407048	0.038911	[68,]	-6.287033	6.287033	39.526778	0.004688
[6,]	-39.480263	39.480263	1558.691153	0.031521	[69,]	-21.698211	21.698211	470.812371	0.016357
[7,]	-42.854038	42.854038	1836.468543	0.034283	[70,]	-58.360007	58.360007	3405.890434	0.045214
[8,]	-41.978403	41.978403	1762.186317	0.033536	[71,]	-43.772419	43.772419	1916.024706	0.033510
[9,]	-81.853360	81.853360	6699.972503	0.067494	[72,]	-34.935449	34.935449	1220.485591	0.026547
[10,]	-60.228908	60.228908	3627.521404	0.048758	[73,]	-42.099096	42.099096	1772.333852	0.032143
[11,]	-40.855050	40.855050	1669.135083	0.032541	[74,]	-29.263360	29.263360	856.344220	0.022111
[12,]	-41.231784	41.231784	1700.059983	0.032828	[75,]	-24.178243	24.178243	584.587430	0.018186
[13,]	-13.109110	13.109110	171.848765	0.010202	[76,]	-50.593745	50.593745	2559.727019	0.038799
[14,]	-19.237029	19.237029	370.063298	0.015032	[77,]	-57.009865	57.009865	3250.124694	0.043904
[15,]	-15.115542	15.115542	228.479622	0.011765	[78,]	-90.926605	90.926605	8267.647497	0.071850
[16,]	-9.244650	9.244650	85.463557	0.007158	[79,]	-71.843965	71.843965	5161.555346	0.055888
[17,]	-4.374353	4.374353	19.134964	0.003372	[80,]	-87.761946	87.761946	7702.159226	0.069077
[18,]	-19.254650	19.254650	370.741538	0.015005	[81,]	-85.680548	85.680548	7341.156336	0.067280
[19,]	-7.635542	7.635542	58.301497	0.005893	[82,]	-40.849770	40.849770	1668.703740	0.030964
[20,]	22.732970	22.732970	516.787932	0.017131	[83,]	-44.519614	44.519614	1981.996069	0.033817
[21,]	28.350886	28.350886	803.772749	0.021261	[84,]	-44.440081	44.440081	1974.920804	0.033731
[22,]	28.968205	28.968205	839.156887	0.021699	[85,]	-29.861170	29.861170	891.689446	0.022401
[23,]	19.084927	19.084927	364.234425	0.014393	[86,]	-32.532881	32.532881	1058.388365	0.024438
[24,]	23.201051	23.201051	538.288769	0.017431	[87,]	-19.955216	19.955216	398.210646	0.014839
[25,]	21.066578	21.066578	443.800729	0.015843	[88,]	-17.878175	17.878175	319.629139	0.013265
[26,]	14.581507	14.581507	212.620340	0.011012	[89,]	-5.551758	5.551758	30.822015	0.004079
[27,]	4.045837	4.045837	16.368798	0.003078	[90,]	-18.225966	18.225966	332.185846	0.013508
[28,]	3.659569	3.659569	13.392442	0.002783	[91,]	-13.650799	13.650799	186.344302	0.010076
[29,]	-2.977300	2.977300	8.864316	0.002274	[92,]	-45.326256	45.326256	2054.469525	0.034234
[30,]	-8.364768	8.364768	69.969342	0.006411	[93,]	-63.502340	63.502340	4032.547205	0.048596
[31,]	-33.502837	33.502837	1122.440066	0.026164	[94,]	-50.679051	50.679051	2568.366251	0.038379
[32,]	-32.391506	32.391506	1049.209657	0.025257	[95,]	-64.856389	64.856389	4206.351220	0.049613
[33,]	-17.530776	17.530776	307.328096	0.013503	[96,]	-54.034354	54.034354	2919.711380	0.040966
[34,]	-7.670647	7.670647	58.838829	0.005860	[97,]	-66.712946	66.712946	4450.617200	0.051033
[35,]	23.438880	23.438880	549.381077	0.017479	[98,]	-89.392166	89.392166	7990.959410	0.069539
[36,]	10.047805	10.047805	100.958379	0.007563	[99,]	-93.322015	93.322015	8708.998561	0.072766
[37,]	7.156127	7.156127	51.210154	0.005395	[100,]	-95.502493	95.502493	9120.726141	0.074539
[38,]	9.513846	9.513846	90.513257	0.007155	[101,]	-105.183600	105.183600	11063.589666	0.082659
[39,]	48.120961	48.120961	2315.626851	0.035144	[102,]	-92.615336	92.615336	8577.600409	0.072018
[40,]	42.977472	42.977472	1847.063091	0.031485	[103,]	-92.297703	92.297703	8518.865911	0.071701
[41,]	49.583379	49.583379	2458.511457	0.036126	[104,]	-96.980700	96.980700	9405.256161	0.075560
[42,]	39.188682	39.188682	1535.752806	0.028752	[105,]	-105.664328	105.664328	11164.950150	0.082825
[43,]	50.793379	50.793379	2579.967390	0.036927	[106,]	-125.348587	125.348587	15712.268363	0.099720
[44,]	51.897472	51.897472	2693.347606	0.037675	[107,]	-143.283479	143.283479	20530.155477	0.115551
[45,]	92.750958	92.750958	8602.740278	0.065352	[108,]	-137.969004	137.969004	19035.445988	0.110707
[46,]	92.103838	92.103838	8483.117046	0.064885	[109,]	-142.155160	142.155160	20208.089451	0.114365
[47,]	79.456111	79.456111	6313.273654	0.056442	[110,]	-138.591950	138.591950	19207.728610	0.111096
[48,]	65.557778	65.557778	4297.822225	0.047003	[111,]	-142.029374	142.029374	20172.342983	0.114080
[49,]	62.158836	62.158836	3863.720885	0.044646	[112,]	-142.467431	142.467431	20296.968870	0.114386
[50,]	68.509287	68.509287	4693.522424	0.048953	[113,]	-135.906123	135.906123	18470.474228	0.108465
[51,]	58.109129	58.109129	3376.670906	0.041805	[114,]	-160.345449	160.345449	25710.663130	0.130415
[52,]	52.208363	52.208363	2725.713154	0.037696	[115,]	-173.535411	173.535411	30114.539037	0.142563
[53,]	53.306987	53.306987	2841.634820	0.038433	[116,]	-164.226009	164.226009	26970.181944	0.133789
[54,]	55.405002	55.405002	3069.714207	0.039860	[117,]	-170.167243	170.167243	28956.890477	0.139196
[55,]	22.752406	22.752406	517.671998	0.016751	[118,]	-160.609112	160.609112	25795.286928	0.130259
[56,]	27.349201	27.349201	747.978804	0.020054	[119,]	-157.551619	157.551619	24822.512655	0.127366
[57,]	-9.304615	9.304615	86.575866	0.007006	[120,]	-129.244763	129.244763	16704.208743	0.102069
[58,]	-19.709043	19.709043	388.446383	0.014948	[121,]	-135.688546	135.688546	18411.381473	0.107625
[59,]	-15.114083	15.114083	228.435507	0.011415	[122,]	-172.132966	172.132966	29629.758155	0.140488
[60,]	-27.769734	27.769734	771.158131	0.021162	[123,]	-166.328026	166.328026	27665.012267	0.135007
[61,]	-39.925998	39.925998	1594.085323	0.030689	[124,]	-164.523724	164.523724	27068.055902	0.133245
[62,]	23.667124	23.667124	560.132767	0.017332	[125,]	-168.470062	168.470062	28382.161753	0.136773
[63,]	6.509634	6.509634	42.375330	0.004825					

```
> mae_AG4_n=sum(error2_AG4_n)/T; mae_AG4_n [1] 59.02881
> rmse_AG4_n=sqrt(sum(error3_AG4_n)/T); rmse_AG4_n [1] 76.21157
> mape_AG4_n=(100/T)*sum(error4_AG4_n); mape_AG4_n [1] 4.613464
```

D. SUMMARY OF 1-STEP AHEAD AND MULTISTEP AHEAD FORECASTING RESULTS OF ARIMA-GARCH (AG)

1-step ahead forecast AG

```
> mae_AG4=sum(error2_AG4)/T;[1] 12.93009
> rmse_AG4=sqrt(sum(error3_AG4)/T;
17.87645
> mape_AG4=(100/T)*sum(error4_AG4);
0.9956225
```



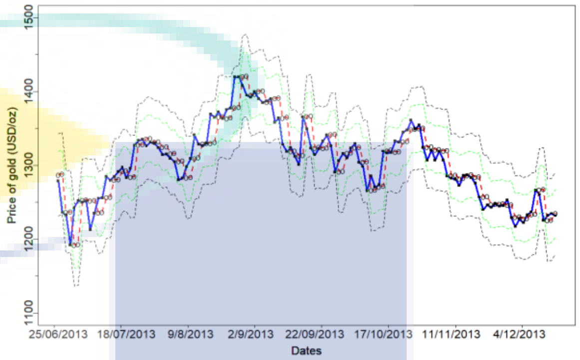

```
> outAG_1_lo80<-dt4_o[dt4_o<lo80_AG4; 1236.25
> outAG_1_hi80<-dt4_o[dt4_o>hi80_AG4;
numeric(0)
> outAG_1_lo95<-dt4_o[dt4_o<lo95_AG4;
numeric(0)
> outAG_1_hi95<-dt4_o[dt4_o>hi95_AG4;
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 1/125 = 0.8%   95% PI: 0%

2-step ahead forecast AG

```
> mae_AG4_2=sum(error2_AG4_2)/T;[1] 15.79385
> rmse_AG4_2=sqrt(sum(error3_AG4_2)/T);
[1] 21.32971
> mape_AG4_2=(100/T)*sum(error4_AG4_2);
[1] 1.213224
```



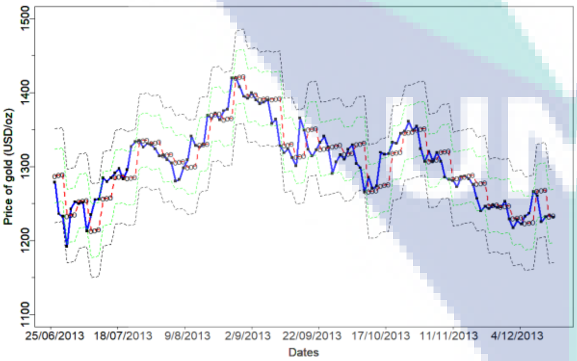

```
> outAG_2_lo80<-dt4_o[dt4_o<lo80_AG4_2;
1236.25 1192.00 1212.75 1358.25 1328.00 1318
.50 1323.00 1290.75 1265.50 1285.50 1225.25
> outAG_2_hi80<-dt4_o[dt4_o>hi80_AG4_2;
[1] 1242.75 1252.50 1327.00 1369.25 1365.00
1419.25 1419.50 1365.50 1319.25
> outAG_2_lo95<-dt4_o[dt4_o<lo95_AG4_2;
numeric(0)
> outAG_2_hi95<-dt4_o[dt4_o>hi95_AG4_2;
[1] 1252.5
```

% Forecast data NOT within prediction interval

80% PI: 20/125 = 16%   95% PI: 1/125 = 0.8%

3-step ahead forecast AG

```
> mae_AG4_3=sum(error2_AG4_3)/T; [1] 18.29534
> rmse_AG4_3=sqrt(sum(error3_AG4_3)/T);
24.44724
> mape_AG4_3=(100/T)*sum(error4_AG4_3);
1.409793
```



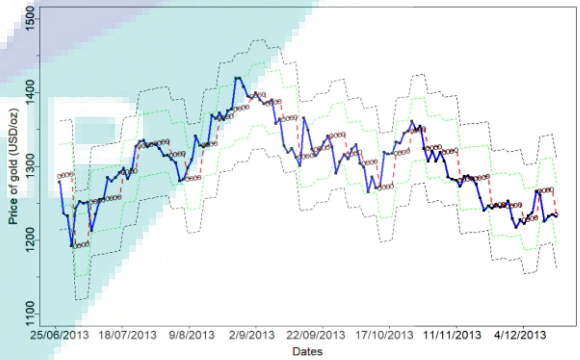

```
> outAG_3_lo80<-dt4_o[dt4_o<lo80_AG4_3];
[1] 1236.25 1232.75 1192.00 1212.75 1328.00 1
314.25 1290.75 1265.50 1306.75 1282.50 [11] 1
240.00 1225.25 1232.00
> outAG_3_hi80<-dt4_o[dt4_o>hi80_AG4_3];
[1] 1255.50 1256.00 1327.00 1333.50 1341.00 1
369.25 1419.25 1365.50 1319.25 1316.50 [11] 1
317.50 1266.25
>outAG_3_lo95<-dt4_o[dt4_o<lo95_AG4_3];
1328.0 1265.5
> outAG_3_hi95<-dt4_o[dt4_o>hi95_AG4_3];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 25/125 = 20%   95% PI: 2/125 = 1.6%

4-step ahead forecast AG

```
> mae_AG4_4=sum(error2_AG4_4)/T; 21.60963
> rmse_AG4_4=sqrt(sum(error3_AG4_4)/T);
28.36626
> mape_AG4_4=(100/T)*sum(error4_AG4_4);
1.67163
```



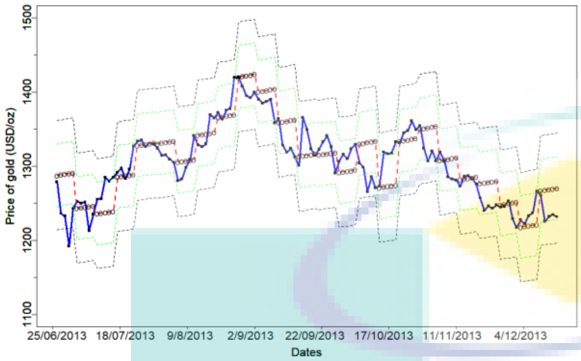

```
> outAG_4_lo80<-dt4_o[dt4_o<lo80_AG4_4];
[1] 1236.25 1232.75 1192.00 1318.50 1324.00
1312.25 1290.75 1240.00 1225.25
> outAG_4_hi80<-dt4_o[dt4_o>hi80_AG4_4];
 [1] 1242.75 1252.50 1250.00 1251.75 1341.00
 1328.50 1365.50 1319.25 1316.50 1317.50
> outAG_4_lo95<-dt4_o[dt4_o<lo95_AG4_4];
[1] 1192
> outAG_4_hi95<-dt4_o[dt4_o>hi95_AG4_4];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 20/125 = 16%   95% PI: 1/125 = 0.8%

5-step ahead forecast AG

```
> mae_AG4_5=sum(error2_AG4_5)/T; [1] 22.83944
> rmse_AG4_5=sqrt(sum(error3_AG4_5)/T);
[1] 28.93036
> mape_AG4_5=(100/T)*sum(error4_AG4_5);
[1] 1.764663
```



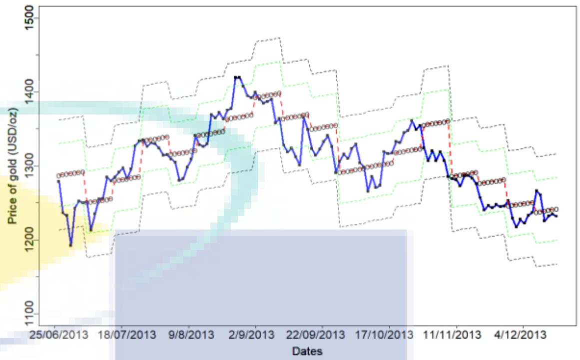

```
> outAG_5_lo80<-dt4_o[dt4_o<lo80_AG4_5];
[1] 1236.25 1232.75 1192.00 1242.75 1358.25 1318.50 1312.25 1265.50 1285.50 1270.50 [11] 1306.75 1307.25 1225.25
> outAG_5_hi80<-dt4_o[dt4_o>hi80_AG4_5];
[1] 1285.00 1284.75 1419.25 1365.50 1319.25 1316.50 1317.50 1333.00 1266.25
> outAG_5_lo95<-dt4_o[dt4_o<lo95_AG4_5];
[1] 1192
> outAG_5_hi95<-dt4_o[dt4_o>hi95_AG4_5];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 22/125 = 17.6% 95% PI: 1/125 = 0.8%

7-step ahead forecast AG

```
> mae_AG4_7=sum(error2_AG4_7)/T;[1] 24.59806
> rmse_AG4_7=sqrt(sum(error3_AG4_7)/T);
[1] 30.12332
> mape_AG4_7=(100/T)*sum(error4_AG4_7);
[1] 1.894061
```



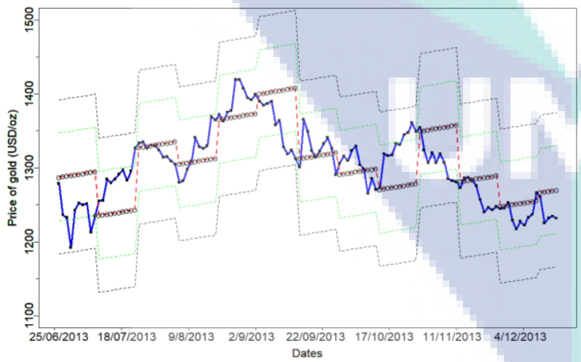

```
> outAG_7_lo80<-dt4_o[dt4_o<lo80_AG4_7];
[1] 1236.25 1232.75 1192.00 1242.75 1250.00 1318.50 1324.00 1312.25 1301.00 1290.75 1306.75 1307.25 1307.25 1285.50
> outAG_7_hi80<-dt4_o[dt4_o>hi80_AG4_7];
[1] 1333.50 1419.25 1419.50
>outAG_7_lo95<-dt4_o[dt4_o<lo95_AG4_7];
1192.0 1285.5
> outAG_7_hi95<-dt4_o[dt4_o>hi95_AG4_7];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 17/125 = 13.6% 95% PI: 2/125 = 1.6%

10-step ahead forecast AG

```
> mae_AG4_10=sum(error2_AG4_10)/T; 32.28698
> rmse_AG4_10=sqrt(sum(error3_AG4_10)/T);
40.19701
> mape_AG4_10=(100/T)*sum(error4_AG4_10);
2.485947
```



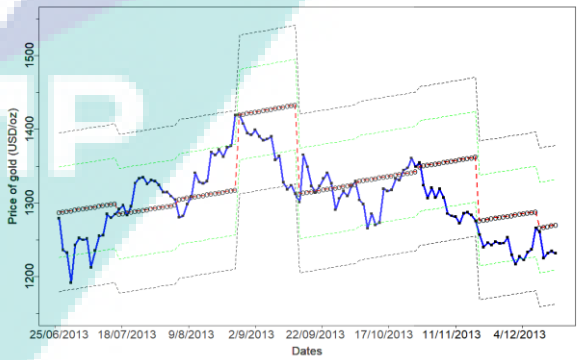

```
> outAG_10_lo80<-dt4_o[dt4_o<lo80_AG4_10];
 [1] 1192.00 1212.75 1235.25 1328.00 1318.50 1324.00 1312.25 1285.50 1282.50 1281.25
> outAG_10_hi80<-dt4_o[dt4_o>hi80_AG4_10];
[1] 1327.00 1344.75 1347.75 1361.00 1349.25
> outAG_10_lo95<-dt4_o[dt4_o<lo95_AG4_10];
numeric(0)
> outAG_10_hi95<-dt4_o[dt4_o>hi95_AG4_10];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 15/125 = 12% 95% PI: 0%

15-step ahead forecast AG

```
> mae_AG4_15=sum(error2_AG4_15)/T; 37.65506
> rmse_AG4_15=sqrt(sum(error3_AG4_15)/T);
46.20913
> mape_AG4_15=(100/T)*sum(error4_AG4_15);
2.906795
```



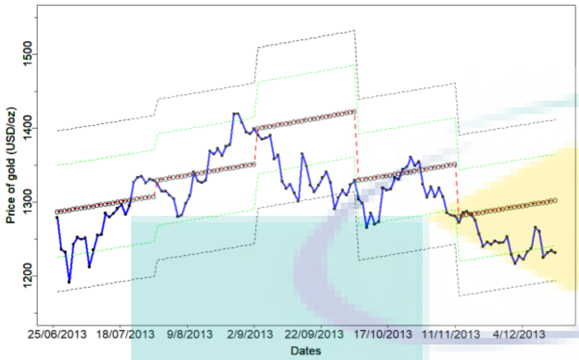

```
> outAG_15_lo80<-dt4_o[dt4_o<lo80_AG4_15];
1192.00 1212.75 1358.25 1363.75 1328.00 1318.50 1324.00 1312.25 1265.50 1270.50 1285.50 1282.50 1281.25 1272.50 1286.00 1287.25 1283.50 1275.75 1217.25 1222.50
> outAG_15_hi80<-dt4_o[dt4_o>hi80_AG4_15];
 1419.25
> outAG_15_lo95<-dt4_o[dt4_o<lo95_AG4_15];
[1] 1318.50 1324.00 1312.25
> outAG_15_hi95<-dt4_o[dt4_o>hi95_AG4_15];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 21/125 = 16.8% 95% PI: 3/125 =2.4%

25-step ahead forecast AG

```
> mae_AG4_25=sum(error2_AG4_25)/T;[1] 43.7949
> rmse_AG4_25=sqrt(sum(error3_AG4_25)/T);
53.01155
> mape_AG4_25=(100/T)*sum(error4_AG4_25);
[1] 3.38398
```



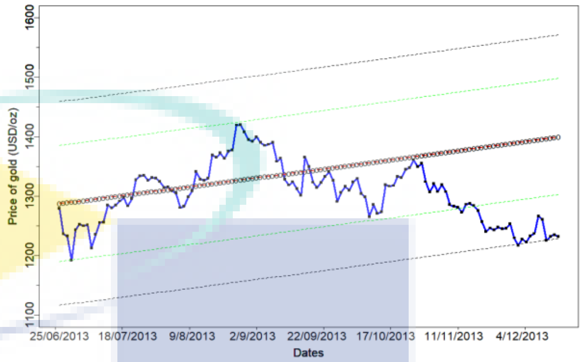

```
>outAG_25_lo80<-dt4_o[dt4_o<lo80_AG4_25];
1192.00 1212.75 1328.00 1318.50 1324.00 1312.
25 1301.00 1349.25 1323.00 1314.25 1322.75 13
33.00 1341.00 1326.50 1290.75 1306.25 1316.00
 1309.75 1323.50 1329.50 1265.50 1270.50 1285
.50 1282.50 1281.25 1229.50 1217.25 1227.50 1
222.50 1233.00 1225.25 1232.00 1234.75 1231.7
5
> outAG_25_hi80<-dt4_o[dt4_o>hi80_AG4_25];
[1] 1419.25 1419.50
> outAG_25_lo95<-dt4_o[dt4_o<lo95_AG4_25];
[1] 1301.00 1290.75 1306.25 1309.75
> outAG_25_hi95<-dt4_o[dt4_o>hi95_AG4_25];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 36/125 = 28.8%   95% PI: 4/125 =3.2%

125-step ahead forecast AG

```
> mae_AG4_n=sum(error2_AG4_n)/T; 59.02881
> rmse_AG4_n=sqrt(sum(error3_AG4_n)/T);
[1] 76.21157
> mape_AG4_n=(100/T)*sum(error4_AG4_n);
[1] 4.613464
```




```
> outAG_n_lo80<-dt4_o[dt4_o<lo80_AG4_n];
1192.00 1272.50 1275.75 1257.00 1240.00 1246
.25 1243.00 1247.50 1245.00 1245.50 1253.00
1229.50 1217.25 1227.50 1222.50 1233.00 1237
.00 1266.25 1260.75 1225.25 1232.00 1234.75
1231.75
> outAG_n_hi80<-dt4_o[dt4_o>hi80_AG4_n];
numeric(0)
> outAG_n_lo95<-dt4_o[dt4_o<lo95_AG4_n];
1217.25 1225.25
> outAG_n_hi95<-dt4_o[dt4_o>hi95_AG4_n];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 23/125 = 18.4%   95% PI: 2/125 =1.6%

To find Prediction Intervals for 7-step ahead ARIMA-GARCH:

	f_AG4_7		resiAG4_7
nobs	125.000000	nobs	125.000000
NAs	0.000000	NAs	0.000000
Minimum	1237.836592	Minimum	-98.234485
Maximum	1398.854507	Maximum	53.482710
1. Quartile	1284.083377	1. Quartile	-31.666227
3. Quartile	1352.903734	3. Quartile	12.655764
Mean	1314.086016	Mean	-9.842816
Median	1315.889344	Median	-10.076676
Sum	164260.751941	Sum	-1230.351941
SE Mean	3.915441	SE Mean	2.556670
LCL Mean	1306.336261	LCL Mean	-14.903182
UCL Mean	1321.835770	UCL Mean	-4.782449
Variance	1916.334742	Variance	817.070178
Stdev	43.775961	Stdev	28.584439
Skewness	0.049188	Skewness	-0.116312
Kurtosis	-1.023547	Kurtosis	-0.201849

```
v1<-qt(c(.025,.975), df=4.81);v1 #t dist, alpha 0.025, v=4.81, PI 95%
v2<-qt(c(.1, .9), df=4.81); v2#t dist, alpha 0.1, v=4.81, for PI 80%
T<- 125
lo95_AG4_7=matrix(0,T,1); lo95_AG4_7 #lower limit PI 95%
hi95_AG4_7=matrix(0,T,1); hi95_AG4_7 # upper limit PI 95%
lo80_AG4_7=matrix(0,T,1); lo80_AG4_7 #lower limit PI 80%
hi80_AG4_7=matrix(0,T,1); hi80_AG4_7 #upper limit PI 80%
for(i in 1:125){
  lo95_AG4_7[i]<-f_AG4_7[i]-(2.6014*28.5844)#lower limit 95%
  hi95_AG4_7[i]<-f_AG4_7[i]+(2.6014*28.5844)#upper limit 95%
  lo80_AG4_7[i]<-f_AG4_7[i]-(1.4847*28.5844)#lower limit 80%
  hi80_AG4_7[i]<-f_AG4_7[i]+(1.4847*28.5844)#upper limit 80%
}
cbind(dt4_o,f_AG4_7,lo95_AG4_7,hi95_AG4_7,lo80_AG4_7,hi80_AG4_7)
```

Simulation Price Data of 7-step ahead using ARIMA(0,1,0)-GARCH(1,1) with *t* innovations

```
> cbind(dt4_o,f_AG4_7,lo95_AG4_7,hi95_AG4_7,lo80_AG4_7,hi80_AG4_7)
         [,1]     [,2]     [,3]     [,4]     [,5]     [,6]
 [1,] 1279.00 1287.620 1213.261 1361.980 1245.181 1330.059
 [2,] 1236.25 1288.491 1214.132 1362.851 1246.052 1330.930
 [3,] 1232.75 1289.362 1215.003 1363.722 1246.923 1331.802
 [4,] 1192.00 1290.234 1215.875 1364.594 1247.795 1332.674
 [5,] 1242.75 1291.107 1216.748 1365.467 1248.668 1333.546
 [6,] 1252.50 1291.980 1217.621 1366.340 1249.541 1334.420
 [7,] 1250.00 1292.854 1218.495 1367.213 1250.415 1335.293
 [8,] 1251.75 1250.845 1176.486 1325.205 1208.406 1293.285
 [9,] 1212.75 1251.691 1177.332 1326.051 1209.252 1294.131
[10,] 1235.25 1252.538 1178.178 1326.897 1210.099 1294.977
[11,] 1255.50 1253.385 1179.026 1327.744 1210.946 1295.824
[12,] 1256.00 1254.233 1179.873 1328.592 1211.793 1296.672
[13,] 1285.00 1255.081 1180.721 1329.440 1212.642 1297.520
[14,] 1279.75 1255.930 1181.570 1330.289 1213.490 1298.369
[15,] 1284.75 1280.616 1206.256 1354.975 1238.176 1323.055
[16,] 1291.50 1281.482 1207.122 1355.841 1239.042 1323.921
[17,] 1297.25 1282.348 1207.989 1356.708 1239.909 1324.788
[18,] 1283.25 1283.216 1208.856 1357.575 1240.776 1325.655
[19,] 1295.75 1284.083 1209.724 1358.443 1241.644 1326.523
[20,] 1327.00 1284.952 1210.592 1359.311 1242.513 1327.391
[21,] 1333.50 1285.821 1211.461 1360.180 1243.382 1328.260
[22,] 1335.00 1334.402 1260.042 1408.761 1291.963 1376.841
[23,] 1326.00 1335.304 1260.945 1409.664 1292.865 1377.744
[24,] 1331.00 1336.207 1261.848 1410.567 1293.768 1378.647
[25,] 1329.75 1337.111 1262.752 1411.471 1294.672 1379.550
[26,] 1324.15 1338.015 1263.656 1412.375 1295.576 1380.455
[27,] 1314.50 1338.920 1264.561 1413.280 1296.481 1381.360
[28,] 1315.00 1339.826 1265.466 1414.185 1297.387 1382.265
[29,] 1309.25 1315.889 1241.530 1390.249 1273.450 1358.329
[30,] 1304.75 1316.779 1242.420 1391.139 1274.340 1359.219
[31,] 1280.50 1317.670 1243.310 1392.029 1275.231 1360.109
[32,] 1282.50 1318.561 1244.202 1392.920 1276.122 1361.000
[33,] 1298.25 1319.453 1245.093 1393.812 1277.013 1361.892
[34,] 1309.00 1320.345 1245.986 1394.705 1277.906 1362.784
[35,] 1341.00 1321.238 1246.879 1395.598 1278.799 1363.677
[36,] 1328.50 1341.907 1267.547 1416.266 1299.468 1384.346
[37,] 1326.50 1342.814 1268.455 1417.174 1300.375 1385.254
[38,] 1329.75 1343.723 1269.363 1418.082 1301.283 1386.162
[39,] 1369.25 1344.631 1270.272 1418.991 1302.192 1387.071
[40,] 1365.00 1345.541 1271.181 1419.900 1303.102 1387.980
[41,] 1372.50 1346.451 1272.091 1420.810 1304.012 1388.890
[42,] 1363.00 1347.361 1273.002 1421.721 1304.922 1389.801
[43,] 1375.50 1363.922 1289.562 1438.281 1321.483 1406.361
[44,] 1377.50 1364.844 1290.485 1439.204 1322.405 1407.283
[45,] 1419.25 1365.767 1291.408 1440.127 1323.328 1408.207
[46,] 1419.50 1366.691 1292.332 1441.050 1324.252 1409.130
[47,] 1407.75 1367.615 1293.256 1441.975 1325.176 1410.055
[48,] 1394.75 1368.540 1294.181 1442.900 1326.101 1410.979
[49,] 1392.25 1369.466 1295.106 1443.825 1327.026 1411.905
[50,] 1399.50 1393.192 1318.832 1467.551 1350.752 1435.631
[51,] 1390.00 1394.134 1319.774 1468.493 1351.695 1436.573
[52,] 1385.00 1395.077 1320.717 1469.436 1352.637 1437.516
[53,] 1387.00 1396.020 1321.661 1470.380 1353.581 1438.459
[54,] 1390.00 1396.964 1322.605 1471.324 1354.525 1439.404
[55,] 1358.25 1397.909 1323.550 1472.269 1355.470 1440.348
[56,] 1363.75 1398.855 1324.495 1473.214 1356.415 1441.294
[57,] 1328.00 1364.672 1290.313 1439.032 1322.233 1407.112
[58,] 1318.50 1365.595 1291.236 1439.955 1323.156 1408.035
[59,] 1324.00 1366.519 1292.159 1440.878 1324.080 1408.958
[60,] 1312.25 1367.443 1293.084 1441.802 1325.004 1409.882
[61,] 1301.00 1368.368 1294.008 1442.727 1325.929 1410.807
[62,] 1365.50 1369.293 1294.934 1443.653 1326.854 1411.733
[63,] 1349.25 1370.219 1295.860 1444.579 1327.780 1412.659
[64,] 1323.00 1350.163 1275.803 1424.522 1307.723 1392.602
[65,] 1314.25 1351.076 1276.716 1425.435 1308.636 1393.515
[66,] 1322.75 1351.989 1277.630 1426.349 1309.550 1394.429
[67,] 1333.00 1352.904 1278.544 1427.263 1310.464 1395.343
```

[68,]	1341.00	1353.819	1279.459	1428.178	1311.379	1396.258
[69,]	1326.50	1354.734	1280.375	1429.094	1312.295	1397.174
[70,]	1290.75	1355.651	1281.291	1430.010	1313.211	1398.090
[71,]	1306.25	1291.623	1217.263	1365.982	1249.184	1334.062
[72,]	1316.00	1292.496	1218.137	1366.856	1250.057	1334.936
[73,]	1309.75	1293.371	1219.011	1367.730	1250.931	1335.810
[74,]	1323.50	1294.245	1219.886	1368.605	1251.806	1336.685
[75,]	1329.50	1295.121	1220.761	1369.480	1252.681	1337.560
[76,]	1304.00	1295.997	1221.637	1370.356	1253.557	1338.436
[77,]	1298.50	1296.873	1222.514	1371.232	1254.434	1339.312
[78,]	1265.50	1299.378	1225.019	1373.738	1256.939	1341.817
[79,]	1285.50	1300.257	1225.898	1374.616	1257.818	1342.696
[80,]	1270.50	1301.136	1226.777	1375.496	1258.697	1343.576
[81,]	1273.50	1302.016	1227.657	1376.376	1259.577	1344.456
[82,]	1319.25	1302.897	1228.537	1377.256	1260.458	1345.336
[83,]	1316.50	1303.778	1229.419	1378.137	1261.339	1346.217
[84,]	1317.50	1304.660	1230.300	1379.019	1262.221	1347.099
[85,]	1333.00	1318.391	1244.032	1392.750	1275.952	1360.830
[86,]	1331.25	1319.283	1244.923	1393.642	1276.843	1361.722
[87,]	1344.75	1320.175	1245.815	1394.534	1277.736	1362.614
[88,]	1347.75	1321.068	1246.708	1395.427	1278.628	1363.507
[89,]	1361.00	1321.961	1247.602	1396.321	1279.522	1364.400
[90,]	1349.25	1322.855	1248.496	1397.215	1280.416	1365.295
[91,]	1354.75	1323.750	1249.390	1398.109	1281.311	1366.189
[92,]	1324.00	1355.666	1281.307	1430.026	1313.227	1398.105
[93,]	1306.75	1356.583	1282.224	1430.943	1314.144	1399.022
[94,]	1320.50	1357.501	1283.141	1431.860	1315.061	1399.940
[95,]	1307.25	1358.419	1284.059	1432.778	1315.979	1400.858
[96,]	1319.00	1359.337	1284.978	1433.697	1316.898	1401.777
[97,]	1307.25	1360.257	1285.897	1434.616	1317.817	1402.696
[98,]	1285.50	1361.177	1286.817	1435.536	1318.737	1403.616
[99,]	1282.50	1286.369	1212.010	1360.729	1243.930	1328.809
[100,]	1281.25	1287.239	1212.880	1361.599	1244.800	1329.679
[101,]	1272.50	1288.110	1213.750	1362.469	1245.671	1330.549
[102,]	1286.00	1288.981	1214.622	1363.341	1246.542	1331.420
[103,]	1287.25	1289.853	1215.493	1364.212	1247.414	1332.292
[104,]	1283.50	1290.725	1216.366	1365.085	1248.286	1333.164
[105,]	1275.75	1291.598	1217.239	1365.958	1249.159	1334.037
[106,]	1257.00	1276.613	1202.253	1350.972	1234.174	1319.052
[107,]	1240.00	1277.476	1203.117	1351.836	1235.037	1319.915
[108,]	1246.25	1278.340	1203.981	1352.700	1235.901	1320.779
[109,]	1243.00	1279.205	1204.845	1353.564	1236.765	1321.644
[110,]	1247.50	1280.070	1205.710	1354.429	1237.631	1322.509
[111,]	1245.00	1280.936	1206.576	1355.295	1238.496	1323.375
[112,]	1245.50	1281.802	1207.442	1356.161	1239.363	1324.241
[113,]	1253.00	1246.342	1171.983	1320.702	1203.903	1288.782
[114,]	1229.50	1247.185	1172.826	1321.545	1204.746	1289.625
[115,]	1217.25	1248.029	1173.669	1322.388	1205.589	1290.468
[116,]	1227.50	1248.873	1174.513	1323.232	1206.434	1291.312
[117,]	1222.50	1249.717	1175.358	1324.077	1207.278	1292.157
[118,]	1233.00	1250.563	1176.203	1324.922	1208.123	1293.002
[119,]	1237.00	1251.408	1177.049	1325.768	1208.969	1293.848
[120,]	1266.25	1237.837	1163.477	1312.196	1195.397	1280.276
[121,]	1260.75	1238.674	1164.314	1313.033	1196.234	1281.113
[122,]	1225.25	1239.511	1165.152	1313.871	1197.072	1281.951
[123,]	1232.00	1240.350	1165.990	1314.709	1197.911	1282.789
[124,]	1234.75	1241.189	1166.829	1315.548	1198.749	1283.628
[125,]	1231.75	1242.028	1167.669	1316.388	1199.589	1284.467

APPENDIX 5
ANALYSIS OF CHAPTER 4 SECTION 4.5

OUT-OF-SAMPLE DATA

- 25 June -17 Dec 2013 (125 data)

A) ARIMA-EGARCH Model

*** Note: ARIMA(0,1,0) - EGARCH(1,1) with normal distribution is significant**

(i) ARIMA-EGARCH(1,1) normal distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 57 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(2) + C(3)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(4)
*RESID(-1)/@SQRT(GARCH(-1)) + C(5)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000474	0.000345	1.373557	0.1696
Variance Equation				
C(2)	-0.464552	0.057634	-8.060352	0.0000
C(3)	0.141116	0.017591	8.022216	0.0000
C(4)	-0.049245	0.007964	-6.183262	0.0000
C(5)	0.959373	0.006277	152.8311	0.0000

R-squared	-0.000073	Mean dependent var	0.000370
Adjusted R-squared	-0.000073	S.D. dependent var	0.012243
S.E. of regression	0.012243	Akaike info criterion	-6.045677
Sum squared resid	0.168331	Schwarz criterion	-6.023326
Log likelihood	3402.671	Hannan-Quinn criter.	-6.037230
Durbin-Watson stat	2.033763		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 31 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	-0.938337	0.017025	-55.11536	0.0000
MA(1)	0.959228	0.015541	61.72315	0.0000
Variance Equation				
C(3)	-5.124438	0.591494	-8.663548	0.0000
C(4)	0.148458	0.043108	3.443848	0.0006
C(5)	-0.260152	0.026567	-9.792468	0.0000
C(6)	0.434178	0.066421	6.536807	0.0000

R-squared	0.014300	Mean dependent var	0.000376
Adjusted R-squared	0.013420	S.D. dependent var	0.012246
S.E. of regression	0.012164	Akaike info criterion	-6.028036
Sum squared resid	0.165866	Schwarz criterion	-6.001195
Log likelihood	3390.742	Hannan-Quinn criter.	-6.017892
Durbin-Watson stat	2.041491		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 88 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000482	0.000340	1.419698	0.1557
AR(1)	-0.016809	0.036579	-0.459528	0.6459
Variance Equation				
C(3)	-0.472306	0.058369	-8.091672	0.0000
C(4)	0.142020	0.017613	8.063163	0.0000
C(5)	-0.048669	0.007932	-6.135647	0.0000
C(6)	0.958584	0.006383	150.1726	0.0000

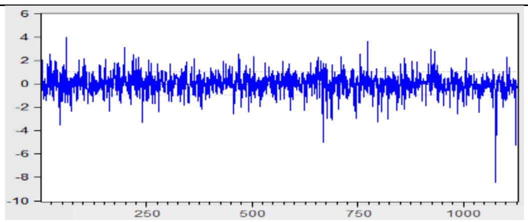
R-squared	0.000220	Mean dependent var	0.000376
Adjusted R-squared	-0.000672	S.D. dependent var	0.012246
S.E. of regression	0.012251	Akaike info criterion	-6.044598
Sum squared resid	0.168235	Schwarz criterion	-6.017757
Log likelihood	3400.042	Hannan-Quinn criter.	-6.034454
F-statistic	0.049274	Durbin-Watson stat	1.987972
Prob(F-statistic)	0.998528		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 74 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000485	0.000339	1.431705	0.1522
MA(1)	-0.017221	0.036531	-0.471406	0.6374
Variance Equation				
C(3)	-0.470911	0.058805	-8.008008	0.0000
C(4)	0.142895	0.017653	8.094777	0.0000
C(5)	-0.048897	0.007951	-6.149685	0.0000
C(6)	0.958786	0.006428	149.1649	0.0000

R-squared	0.000233	Mean dependent var	0.000370
Adjusted R-squared	-0.000658	S.D. dependent var	0.012243
S.E. of regression	0.012247	Akaike info criterion	-6.044146
Sum squared resid	0.168280	Schwarz criterion	-6.017324
Log likelihood	3402.810	Hannan-Quinn criter.	-6.034010
F-statistic	0.052400	Durbin-Watson stat	2.000908
Prob(F-statistic)	0.998292		

ARIMA(0,1,0)-EGARCH(1,1) with normal distribution



Correlogram of Standardized Residuals

Date: 01/18/19 Time: 16:12
Sample: 2 1125
Included observations: 1124

	AC	PAC	Q-Stat	Prob
1	-0.001	-0.001	0.0015	0.969
2	-0.031	-0.031	1.0814	0.582
3	-0.020	-0.020	1.5352	0.674
4	0.006	0.005	1.5812	0.812
5	0.048	0.047	4.1580	0.527
6	-0.040	-0.040	5.9336	0.431
7	-0.018	-0.015	6.2854	0.507
8	-0.010	-0.011	6.4008	0.602
9	0.069	0.067	11.881	0.220
10	-0.004	-0.007	11.903	0.292
11	-0.016	-0.008	12.184	0.350
12	-0.044	-0.042	14.366	0.278
13	0.022	0.021	14.922	0.312
14	0.022	0.012	15.470	0.347
15	-0.073	-0.068	21.524	0.121
16	0.020	0.025	22.003	0.143
17	-0.002	-0.001	22.006	0.184
18	0.041	0.030	23.942	0.157
19	-0.005	-0.005	23.968	0.197
20	0.024	0.035	24.604	0.217

Correlogram of Standardized Residuals Squared

Date: 01/18/19 Time: 16:13
Sample: 2 1125
Included observations: 1124

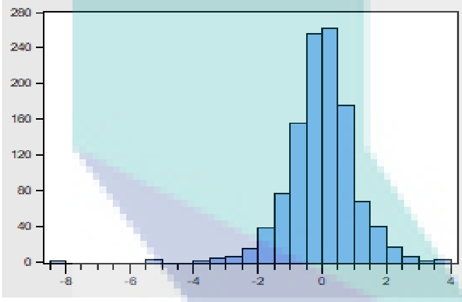
	AC	PAC	Q-Stat	Prob
1	-0.008	-0.008	0.0758	0.783
2	-0.008	-0.008	0.1402	0.932
3	-0.013	-0.013	0.3308	0.954
4	-0.019	-0.019	0.7276	0.948
5	-0.024	-0.024	1.3697	0.928
6	0.012	0.011	1.5395	0.957
7	0.014	0.013	1.7511	0.972
8	-0.007	-0.008	1.8111	0.986
9	0.024	0.024	2.4736	0.982
10	0.003	0.004	2.4849	0.991
11	-0.009	-0.007	2.5672	0.995
12	-0.001	-0.001	2.5691	0.998
13	-0.007	-0.007	2.6265	0.999
14	-0.011	-0.010	2.7709	0.999
15	0.047	0.047	5.3342	0.989
16	-0.017	-0.018	5.6820	0.991
17	-0.017	-0.017	6.0226	0.993
18	0.020	0.020	6.4919	0.994
19	-0.016	-0.015	6.7763	0.995
20	0.019	0.021	7.1697	0.996

Heteroskedasticity Test: ARCH

F-statistic	0.262579	Prob. F(10,1103)	0.9888
Obs*R-squared	2.645680	Prob. Chi-Square(10)	0.9886

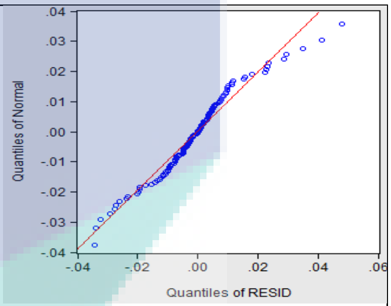
Heteroskedasticity Test: ARCH

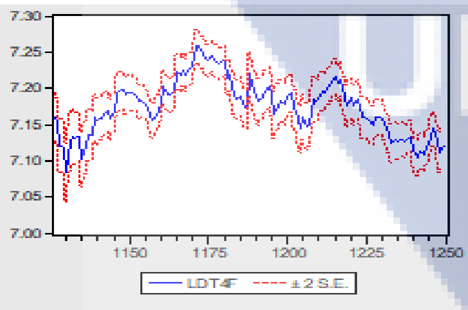
F-statistic	0.365984	Prob. F(15,1093)	0.9869
Obs*R-squared	5.542282	Prob. Chi-Square(15)	0.9865



Series: Standardized Residuals
Sample 2 1125
Observations 1124

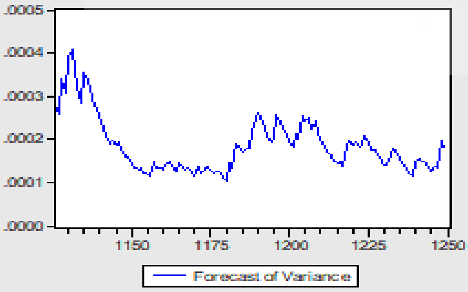
Mean	-0.010411
Median	0.013316
Maximum	3.995083
Minimum	-8.440496
Std. Dev.	1.000667
Skewness	-0.786646
Kurtosis	8.993679
Jarque-Bera	1798.373
Probability	0.000000



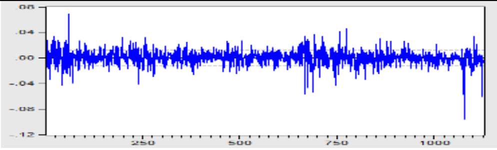


Forecast: LDT4F
Actual: LDT4
Forecast sample: 1126 1250
Included observations: 125

Root Mean Squared Error	0.013792
Mean Absolute Error	0.009940
Mean Abs. Percent Error	0.138667
Theil Inequality Coefficient	0.000961
Bias Proportion	0.003567
Variance Proportion	0.000542
Covariance Proportion	0.995891



ARIMA(1,1,1)-EGARCH(1,1) with normal distribution



Correlogram of Standardized Residuals

Date: 01/14/19 Time: 11:48
Sample: 3 1125
Included observations: 1123
Q-statistic probabilities adjusted for 2 ARMA term(s)

Lag	AC	PAC	Q-Stat	Prob
1	-0.012	-0.012	0.1701	
2	-0.031	-0.031	1.2523	
3	-0.049	-0.050	3.9756	0.046
4	0.030	0.028	5.0128	0.082
5	0.055	0.053	8.4465	0.038
6	-0.045	-0.044	10.714	0.030
7	-0.034	-0.029	11.993	0.035
8	-0.020	-0.019	12.454	0.053
9	0.069	0.060	17.871	0.013
10	-0.013	-0.016	18.059	0.021
11	-0.021	-0.014	18.577	0.029
12	-0.036	-0.029	20.043	0.029
13	0.025	0.018	20.744	0.036
14	0.007	-0.004	20.799	0.053
15	-0.079	-0.076	27.971	0.009
16	0.044	0.051	30.172	0.007
17	0.010	0.010	30.276	0.011
18	0.041	0.026	32.202	0.009
19	-0.017	-0.008	32.533	0.013
20	0.039	0.050	34.273	0.012

Correlogram of Standardized Residuals Squared

Date: 01/14/19 Time: 11:50
Sample: 3 1125
Included observations: 1123
Q-statistic probabilities adjusted for 2 ARMA term(s)

Lag	AC	PAC	Q-Stat	Prob
1	-0.026	-0.026	0.7652	
2	0.036	0.035	2.2092	
3	0.033	0.035	3.4524	0.063
4	0.018	0.019	3.8356	0.147
5	-0.004	-0.005	3.8519	0.278
6	0.084	0.082	11.893	0.018
7	0.024	0.028	12.534	0.028
8	0.029	0.024	13.464	0.036
9	0.123	0.118	30.491	0.000
10	0.066	0.069	35.472	0.000
11	0.042	0.039	37.495	0.000
12	0.036	0.022	38.975	0.000
13	0.036	0.027	40.459	0.000
14	0.019	0.014	40.874	0.000
15	0.089	0.069	49.841	0.000
16	0.008	-0.004	49.911	0.000
17	0.014	-0.006	50.135	0.000
18	0.072	0.048	56.093	0.000
19	0.023	0.003	56.709	0.000
20	0.093	0.077	66.663	0.000

Heteroskedasticity Test: ARCH

F-statistic	3.777080	Prob. F(10,1102)	0.0001
Obs*R-squared	36.88364	Prob. Chi-Square(10)	0.0001

Heteroskedasticity Test: ARCH

F-statistic	3.079598	Prob. F(15,1092)	0.0001
Obs*R-squared	44.96854	Prob. Chi-Square(15)	0.0001

(ii) ARIMA-EGARCH(1,1) with *t* distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 30 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(2) + C(3)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(4)
*RESID(-1)/@SQRT(GARCH(-1)) + C(5)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000720	0.000296	2.436393	0.0148
Variance Equation				
C(2)	-0.204890	0.079297	-2.583830	0.0098
C(3)	0.095768	0.026251	3.648089	0.0003
C(4)	0.012598	0.016699	0.754429	0.4506
C(5)	0.984903	0.007741	127.2334	0.0000
T-DIST. DOF	4.801590	0.633334	7.581453	0.0000

R-squared	-0.000820	Mean dependent var	0.000370
Adjusted R-squared	-0.000820	S.D. dependent var	0.012243
S.E. of regression	0.012248	Akaike info criterion	-6.163370
Sum squared resid	0.168457	Schwarz criterion	-6.136548
Log likelihood	3469.814	Hannan-Quinn criter.	-6.153234
Durbin-Watson stat	2.032245		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 66 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(4) + C(5)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(6)
*RESID(-1)/@SQRT(GARCH(-1)) + C(7)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000717	0.000299	2.401037	0.0163
AR(1)	-0.938630	0.028993	-32.37454	0.0000
MA(1)	0.952866	0.026971	35.32923	0.0000
Variance Equation				
C(4)	-0.203667	0.084401	-2.413088	0.0158
C(5)	0.097763	0.026703	3.661194	0.0003
C(6)	0.014522	0.016926	0.857989	0.3909
C(7)	0.985209	0.008231	119.6975	0.0000
T-DIST. DOF	4.889857	0.646991	7.557839	0.0000

R-squared	0.013974	Mean dependent var	0.000376
Adjusted R-squared	0.012214	S.D. dependent var	0.012246
S.E. of regression	0.012171	Akaike info criterion	-6.166189
Sum squared resid	0.165920	Schwarz criterion	-6.130401
Log likelihood	3470.315	Hannan-Quinn criter.	-6.152664
F-statistic	2.267578	Durbin-Watson stat	2.029464
Prob(F-statistic)	0.027040		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 33 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000737	0.000287	2.564731	0.0103
AR(1)	-0.032808	0.028495	-1.151381	0.2496
Variance Equation				
C(3)	-0.212868	0.078407	-2.714914	0.0066
C(4)	0.096164	0.026360	3.648105	0.0003
C(5)	0.012372	0.016492	0.750165	0.4532
C(6)	0.984012	0.007700	127.7913	0.0000
T-DIST. DOF	4.705181	0.621280	7.573361	0.0000

R-squared	-0.000872	Mean dependent var	0.000376
Adjusted R-squared	-0.001765	S.D. dependent var	0.012246
S.E. of regression	0.012257	Akaike info criterion	-6.163355
Sum squared resid	0.168419	Schwarz criterion	-6.132041
Log likelihood	3467.724	Hannan-Quinn criter.	-6.151521
Durbin-Watson stat	1.955874		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 124 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000735	0.000286	2.573189	0.0101
MA(1)	-0.035874	0.028384	-1.263886	0.2063
Variance Equation				
C(3)	-0.203658	0.079115	-2.574220	0.0100
C(4)	0.095987	0.026283	3.652046	0.0003
C(5)	0.013234	0.016205	0.816640	0.4141
C(6)	0.985014	0.007740	127.2583	0.0000
T-DIST. DOF	4.720234	0.622533	7.582301	0.0000

R-squared	-0.000884	Mean dependent var	0.000370
Adjusted R-squared	-0.001776	S.D. dependent var	0.012243
S.E. of regression	0.012254	Akaike info criterion	-6.163088
Sum squared resid	0.168468	Schwarz criterion	-6.131796
Log likelihood	3470.656	Hannan-Quinn criter.	-6.151263
Durbin-Watson stat	1.963687		

(iii) ARIMA-EGARCH(1,1) with GED distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 18 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(2) + C(3)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(4)
*RESID(-1)/@SQRT(GARCH(-1)) + C(5)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000593	0.000279	2.129633	0.0332
Variance Equation				
C(2)	-16.45243	0.625892	-26.28636	0.0000
C(3)	-0.021974	0.035876	-0.612492	0.5402
C(4)	-0.072755	0.037274	-1.951885	0.0510
C(5)	-0.860435	0.068600	-12.54283	0.0000
GED PARAMETER	1.097400	0.049938	21.97521	0.0000

R-squared	-0.000333	Mean dependent var	0.000370
Adjusted R-squared	-0.000333	S.D. dependent var	0.012243
S.E. of regression	0.012245	Akaike info criterion	-6.107523
Sum squared resid	0.168375	Schwarz criterion	-6.080701
Log likelihood	3438.428	Hannan-Quinn criter.	-6.097387
Durbin-Watson stat	2.033234		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 69 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(4) + C(5)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(6)
*RESID(-1)/@SQRT(GARCH(-1)) + C(7)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000667	0.000278	2.398111	0.0165
AR(1)	-0.934372	0.030877	-30.26139	0.0000
MA(1)	0.946522	0.029486	32.10041	0.0000
Variance Equation				
C(4)	-0.266319	0.094729	-2.811370	0.0049
C(5)	0.107591	0.030091	3.575498	0.0003
C(6)	-0.005579	0.016469	-0.338753	0.7348
C(7)	0.979084	0.009310	105.1648	0.0000
GED PARAMETER	1.174127	0.051790	22.67113	0.0000

R-squared	0.013716	Mean dependent var	0.000376
Adjusted R-squared	0.011955	S.D. dependent var	0.012246
S.E. of regression	0.012173	Akaike info criterion	-6.154015
Sum squared resid	0.165964	Schwarz criterion	-6.118227
Log likelihood	3463.480	Hannan-Quinn criter.	-6.140490
F-statistic	2.225151	Durbin-Watson stat	2.027011
Prob(F-statistic)	0.030064		

Inverted AR Roots	-.93
Inverted MA Roots	-.95

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 43 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000702	0.000257	2.733771	0.0063
AR(1)	-0.055968	0.026453	-2.115740	0.0344
Variance Equation				
C(3)	-0.269352	0.091092	-2.956922	0.0031
C(4)	0.104077	0.029937	3.476549	0.0005
C(5)	-0.004317	0.015768	-0.273788	0.7842
C(6)	0.978438	0.008994	108.7899	0.0000
GED PARAMETER	1.144016	0.050102	22.83391	0.0000

R-squared	-0.001995	Mean dependent var	0.000376
Adjusted R-squared	-0.002889	S.D. dependent var	0.012246
S.E. of regression	0.012264	Akaike info criterion	-6.154562
Sum squared resid	0.168608	Schwarz criterion	-6.123248
Log likelihood	3462.787	Hannan-Quinn criter.	-6.142728
Durbin-Watson stat	1.912249		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 19 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000742	0.000253	2.933656	0.0033
MA(1)	-0.069394	0.021639	-3.206845	0.0013
Variance Equation				
C(3)	-16.03562	0.815190	-19.67101	0.0000
C(4)	-0.030832	0.044538	-0.692268	0.4888
C(5)	-0.100043	0.044340	-2.256266	0.0241
C(6)	-0.813955	0.089689	-9.075344	0.0000
GED PARAMETER	1.069379	0.049058	21.79838	0.0000

R-squared	-0.003023	Mean dependent var	0.000370
Adjusted R-squared	-0.003917	S.D. dependent var	0.012243
S.E. of regression	0.012267	Akaike info criterion	-6.114629
Sum squared resid	0.168828	Schwarz criterion	-6.083337
Log likelihood	3443.422	Hannan-Quinn criter.	-6.102804
Durbin-Watson stat	1.899774		

B) ARIMA-APARCH Model

```
> m1=garchFit(~1+aparch(1,1), data=dldt4, trace=F)#aparch(1,1)
> summary (m1)
Title:GARCH Modelling
Call:garchFit(formula = ~1 + aparch(1, 1), data = dldt4, trace = F)
Mean and Variance Equation:data ~ 1 + aparch(1, 1)
<environment: 0x0000000004f88f90>[data = dldt4]
Conditional Distribution:norm
Coefficient(s):
       mu      omega     alpha1     gamma1      beta1      delta
0.0004012  0.0065080  0.0694739  0.5535739  0.9058669  0.4201452

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     4.012e-04   7.754e-05    5.174 2.29e-07 ***
omega  6.508e-03   1.821e-03    3.574 0.000352 ***
alpha1 6.947e-02   1.751e-02    3.968 7.25e-05 ***
gamma1 5.536e-01   1.814e-01    3.051 0.002280 **
beta1  9.059e-01   1.608e-02   56.345  < 2e-16 ***
delta  4.201e-01   2.972e-01    1.414 0.157505
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Log Likelihood: -1699.917    normalized:  -1.512382
Standardised Residuals Tests:
                              Statistic  p-Value
 Jarque-Bera Test   R    Chi^2  40015205   0
 Shapiro-Wilk Test  R    W      0.02701801 0
 Ljung-Box Test     R    Q(10)  102.544    0
 Ljung-Box Test     R    Q(15)  102.5723   4.218847e-15
 Ljung-Box Test     R    Q(20)  102.5903   4.319878e-13
 Ljung-Box Test     R^2  Q(10)  12.19241   0.2723854
 Ljung-Box Test     R^2  Q(15)  12.19241   0.6644103
 Ljung-Box Test     R^2  Q(20)  12.19241   0.909294
 LM Arch Test       R    TR^2   24.17223   0.01927104
Information Criterion Statistics:
     AIC      BIC      SIC     HQIC
3.035440 3.062262 3.035383 3.045576

> m3=garchFit(~1+aparch(1,1), data=dldt4, trace=F, cond.dist="std")#aparch(1,1) with t
> summary (m3)
Title:GARCH Modelling
Call:garchFit(formula = ~1 + aparch(1, 1), data = dldt4, cond.dist = "std", trace = F)
Mean and Variance Equation:data ~ 1 + aparch(1, 1)
<environment: 0x0000000017028298>[data = dldt4]
Conditional Distribution:std
Coefficient(s):
mu          omega        alpha1        gamma1         beta1         delta         shape
7.0276e-04    4.2577e-05    5.2499e-02  -1.4137e-01    9.4462e-01    1.3289e+00    4.8214e+00

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      7.028e-04   2.957e-04    2.377 0.017468 *
omega   4.258e-05   2.082e-05    2.045 0.040814 *
alpha1  5.250e-02   1.416e-02    3.708 0.000209 ***
gamma1 -1.414e-01   1.608e-01   -0.879 0.379301
beta1   9.446e-01   1.297e-02   72.851  < 2e-16 ***
delta   1.329e+00   4.888e-01    2.718 0.006559 **
shape   4.821e+00   6.928e-01    6.959 3.42e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:3449.734    normalized:  3.069159
Standardised Residuals Tests:
                               Statistic p-Value
 Jarque-Bera Test   R    Chi^2  221063    0
 Shapiro-Wilk Test  R    W      0.7925914 0
 Ljung-Box Test     R    Q(10)  15.57377  0.1125012
 Ljung-Box Test     R    Q(15)  23.15934  0.08081493
 Ljung-Box Test     R    Q(20)  29.36809  0.08076042
 Ljung-Box Test     R^2  Q(10)  4.374218  0.9288905
 Ljung-Box Test     R^2  Q(15)  4.855187  0.9932831
 Ljung-Box Test     R^2  Q(20)  5.405159  0.999495
 LM Arch Test       R    TR^2   6.25924   0.9024539
Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.125862 -6.094570 -6.125939 -6.114036

> m4=garchFit(~1+aparch(1,1), data=dldt4, trace=F, cond.dist="ged")#aparch(1,1) with ged
> summary (m4)
Title:GARCH Modelling
Call:garchFit(formula = ~1 + aparch(1, 1), data = dldt4, cond.dist = "ged", trace = F)
Mean and Variance Equation:data ~ 1 + aparch(1, 1)
<environment: 0x0000000009bc5b48>[data = dldt4]
Conditional Distribution:ged

Coefficient(s):
mu         omega       alpha1       gamma1        beta1        delta     shape
6.2446e-04  9.7087e-05  5.8887e-02  3.2423e-02  9.3436e-01  1.2132e+00  1.1686e+00
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     6.245e-04   2.344e-04    2.664 0.007715 **
omega  9.709e-05   4.459e-05    2.177 0.029459 *
alpha1 5.889e-02   1.663e-02    3.541 0.000398 ***
gamma1 3.242e-02   1.912e-01    0.170 0.865364
beta1  9.344e-01   1.654e-02   56.491  < 2e-16 ***
delta  1.213e+00   6.083e-01    1.995 0.046097 *
shape  1.169e+00   6.107e-02   19.134  < 2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Log Likelihood:3409.638    normalized:  3.033486

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  483061.4  0
 Shapiro-Wilk Test  R    W      0.7451382 0
 Ljung-Box Test     R    Q(10)  25.60213  0.004313801
 Ljung-Box Test     R    Q(15)  33.46694  0.004043101
 Ljung-Box Test     R    Q(20)  40.05189  0.004920489
 Ljung-Box Test     R^2  Q(10)  9.696887  0.4674767
 Ljung-Box Test     R^2  Q(15)  9.937081  0.8236789
 Ljung-Box Test     R^2  Q(20)  10.26936  0.9630205
 LM Arch Test       R    TR^2   6.340672  0.8979447

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.054516 -6.023223 -6.054593 -6.042690
```

C) ARIMA-TGARCH Model (or APARCH model with delta=2)

*** Note: only ARIMA(0,1,0)-TGARCH(1,1) with normal distribution is significant**

(i) ARIMA-TGARCH(1,1) with normal distribution, delta=2 (or TGARCH)

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 44 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*(ABS(RESID(-1)) - C(4)*RESID(-1)) + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000481	0.000354	1.358389	0.1743
Variance Equation				
C(2)	4.48E-06	6.50E-07	6.886939	0.0000
C(3)	0.051849	0.008001	6.480186	0.0000
C(4)	0.146288	0.049887	2.932392	0.0034
C(5)	0.917330	0.007643	120.0269	0.0000

R-squared	-0.000082	Mean dependent var	0.000370
Adjusted R-squared	-0.000082	S.D. dependent var	0.012243
S.E. of regression	0.012243	Akaike info criterion	-6.041285
Sum squared resid	0.168333	Schwarz criterion	-6.018934
Log likelihood	3400.202	Hannan-Quinn criter.	-6.032838
Durbin-Watson stat	2.033745		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 96 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*(ABS(RESID(-1)) - C(5)*RESID(-1)) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	-0.940932	0.024043	-39.13599	0.0000
MA(1)	0.959275	0.022035	43.53378	0.0000
Variance Equation				
C(3)	4.71E-06	7.05E-07	6.680098	0.0000
C(4)	0.053582	0.008259	6.487311	0.0000
C(5)	0.174926	0.049996	3.498785	0.0005
C(6)	0.914699	0.007984	114.5648	0.0000

R-squared	0.014290	Mean dependent var	0.000376
Adjusted R-squared	0.013411	S.D. dependent var	0.012246
S.E. of regression	0.012164	Akaike info criterion	-6.046336
Sum squared resid	0.165867	Schwarz criterion	-6.019495
Log likelihood	3401.018	Hannan-Quinn criter.	-6.036192
Durbin-Watson stat	2.036310		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 78 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*(ABS(RESID(-1)) - C(5)*RESID(-1)) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000479	0.000351	1.364068	0.1725
AR(1)	-0.014398	0.037553	-0.383409	0.7014
Variance Equation				
C(3)	4.62E-06	6.80E-07	6.796666	0.0000
C(4)	0.051780	0.008044	6.437330	0.0000
C(5)	0.153642	0.050352	3.051341	0.0023
C(6)	0.915904	0.007625	120.1146	0.0000

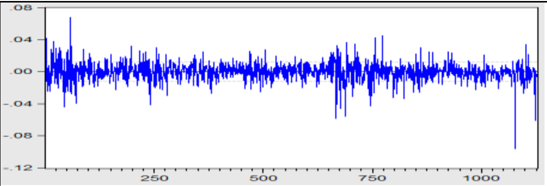
R-squared	0.000216	Mean dependent var	0.000376
Adjusted R-squared	-0.000676	S.D. dependent var	0.012246
S.E. of regression	0.012251	Akaike info criterion	-6.040035
Sum squared resid	0.168236	Schwarz criterion	-6.013194
Log likelihood	3397.480	Hannan-Quinn criter.	-6.029891
F-statistic	0.048500	Durbin-Watson stat	1.992571
Prob(F-statistic)	0.998583		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 70 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*(ABS(RESID(-1)) - C(5)*RESID(-1)) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000485	0.000350	1.383899	0.1664
MA(1)	-0.015199	0.037429	-0.406072	0.6847
Variance Equation				
C(3)	4.49E-06	6.70E-07	6.700076	0.0000
C(4)	0.052240	0.008013	6.519300	0.0000
C(5)	0.143494	0.049621	2.891831	0.0038
C(6)	0.916926	0.007615	120.4169	0.0000

R-squared	0.000224	Mean dependent var	0.000370
Adjusted R-squared	-0.000667	S.D. dependent var	0.012243
S.E. of regression	0.012247	Akaike info criterion	-6.039687
Sum squared resid	0.168281	Schwarz criterion	-6.012865
Log likelihood	3400.304	Hannan-Quinn criter.	-6.029550
F-statistic	0.050284	Durbin-Watson stat	2.004761
Prob(F-statistic)	0.998454		

ARIMA(0,1,0)-TGARCH(1,1) with normal distribution



Correlogram of Standardized Residuals

Date: 01/18/19 Time: 11:44
Sample: 2 1125
Included observations: 1124

	AC	PAC	Q-Stat	Prob
1	0.003	0.003	0.0112	0.916
2	-0.033	-0.033	1.2329	0.540
3	-0.022	-0.022	1.7649	0.623
4	0.004	0.003	1.7861	0.775
5	0.047	0.045	4.2663	0.512
6	-0.041	-0.042	6.2046	0.401
7	-0.015	-0.012	6.4559	0.488
8	-0.010	-0.010	6.5589	0.585
9	0.069	0.066	11.904	0.219
10	-0.001	-0.004	11.905	0.291
11	-0.016	-0.008	12.200	0.349
12	-0.047	-0.045	14.715	0.257
13	0.021	0.021	15.240	0.293
14	0.024	0.013	15.886	0.320
15	-0.072	-0.068	21.789	0.113
16	0.020	0.025	22.240	0.136
17	-0.002	-0.001	22.244	0.176
18	0.041	0.030	24.159	0.150
19	-0.005	-0.006	24.187	0.189
20	0.020	0.031	24.627	0.216

Correlogram of Standardized Residuals Squared

Date: 01/18/19 Time: 11:45
Sample: 2 1125
Included observations: 1124

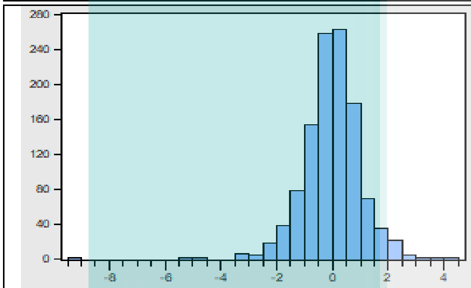
	AC	PAC	Q-Stat	Prob
1	-0.002	-0.002	0.0027	0.958
2	-0.008	-0.008	0.0739	0.964
3	-0.013	-0.013	0.2709	0.965
4	-0.018	-0.018	0.6360	0.959
5	-0.023	-0.023	1.2358	0.941
6	0.010	0.009	1.3388	0.969
7	0.016	0.015	1.6395	0.977
8	-0.007	-0.008	1.7022	0.989
9	0.012	0.012	1.8667	0.993
10	0.003	0.003	1.8738	0.997
11	-0.012	-0.011	2.0362	0.998
12	-0.003	-0.002	2.0444	0.999
13	-0.011	-0.011	2.1838	1.000
14	-0.014	-0.013	2.3940	1.000
15	0.035	0.034	3.7900	0.998
16	-0.020	-0.021	4.2338	0.998
17	-0.018	-0.018	4.5971	0.999
18	0.011	0.010	4.7245	0.999
19	-0.017	-0.017	5.0576	0.999
20	0.008	0.010	5.1358	1.000

Heteroskedasticity Test: ARCH

F-statistic	0.194797	Prob. F(10,1103)	0.9967
Obs*R-squared	1.963932	Prob. Chi-Square(10)	0.9966

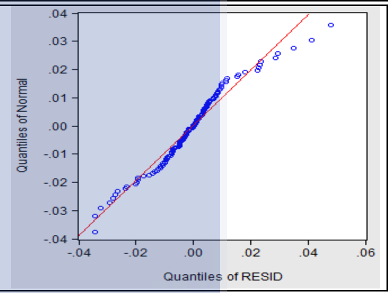
Heteroskedasticity Test: ARCH

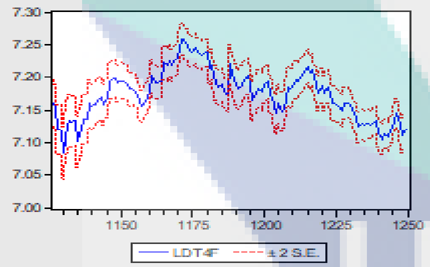
F-statistic	0.257710	Prob. F(15,1093)	0.998
Obs*R-squared	3.908417	Prob. Chi-Square(15)	0.998



Series: Standardized Residuals
Sample 2 1125
Observations 1124

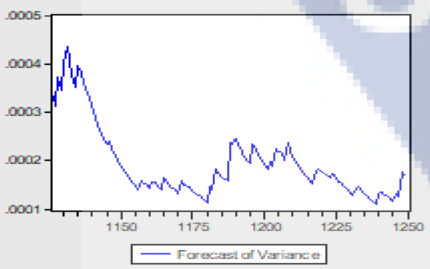
Mean	-0.008223
Median	0.012475
Maximum	4.145550
Minimum	-9.027717
Std. Dev.	1.000767
Skewness	-0.879610
Kurtosis	10.26843
Jarque-Bera	2619.154
Probability	0.000000





Forecast: LDT4F
Actual: LDT4
Forecast sample: 1126 1250
Included observations: 125

Root Mean Squared Error	0.013793
Mean Absolute Error	0.009940
Mean Abs. Percent Error	0.138673
Theil Inequality Coefficient	0.000961
Bias Proportion	0.003622
Variance Proportion	0.000542
Covariance Proportion	0.995836



(ii) ARIMA-TGARCH(1,1) with *t* distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 20 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*(ABS(RESID(-1)) - C(4)*RESID(-1)) + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000694	0.000296	2.344021	0.0191
Variance Equation				
C(2)	2.28E-06	1.07E-06	2.128181	0.0333
C(3)	0.034915	0.011428	3.055210	0.0022
C(4)	-0.106291	0.129729	-0.819329	0.4126
C(5)	0.948873	0.015438	61.46181	0.0000
T-DIST. DOF	4.815781	0.634447	7.590514	0.0000

R-squared	-0.000702	Mean dependent var	0.000370
Adjusted R-squared	-0.000702	S.D. dependent var	0.012243
S.E. of regression	0.012247	Akaike info criterion	-6.162823
Sum squared resid	0.168437	Schwarz criterion	-6.136001
Log likelihood	3469.506	Hannan-Quinn criter.	-6.152686
Durbin-Watson stat	2.032484		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 141 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*(ABS(RESID(-1)) - C(6)*RESID(-1)) + C(7)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000694	0.000266	2.613125	0.0090
AR(1)	0.548619	0.333671	1.644191	0.1001
MA(1)	-0.595382	0.320567	-1.857280	0.0633
Variance Equation				
C(4)	2.30E-06	1.06E-06	2.173929	0.0297
C(5)	0.034844	0.011413	3.052948	0.0023
C(6)	-0.121062	0.124033	-0.976041	0.3290
C(7)	0.948949	0.015118	62.77143	0.0000
T-DIST. DOF	4.705219	0.617875	7.615164	0.0000

R-squared	0.002168	Mean dependent var	0.000376
Adjusted R-squared	0.000386	S.D. dependent var	0.012246
S.E. of regression	0.012244	Akaike info criterion	-6.163663
Sum squared resid	0.167907	Schwarz criterion	-6.127875
Log likelihood	3468.897	Hannan-Quinn criter.	-6.150138
F-statistic	0.347628	Durbin-Watson stat	1.931378
Prob(F-statistic)	0.931826		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 18 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*(ABS(RESID(-1)) - C(5)*RESID(-1)) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000714	0.000287	2.488003	0.0128
AR(1)	-0.035318	0.029176	-1.210542	0.2261
Variance Equation				
C(3)	2.48E-06	1.09E-06	2.273560	0.0230
C(4)	0.035191	0.011530	3.052072	0.0023
C(5)	-0.095446	0.130071	-0.733796	0.4631
C(6)	0.947256	0.015267	62.04596	0.0000
T-DIST. DOF	4.690657	0.618815	7.580068	0.0000

R-squared	-0.000845	Mean dependent var	0.000376
Adjusted R-squared	-0.001738	S.D. dependent var	0.012246
S.E. of regression	0.012257	Akaike info criterion	-6.162862
Sum squared resid	0.168414	Schwarz criterion	-6.131547
Log likelihood	3467.447	Hannan-Quinn criter.	-6.151027
Durbin-Watson stat	1.951326		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 18 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*(ABS(RESID(-1)) - C(5)*RESID(-1)) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000709	0.000285	2.484401	0.0130
MA(1)	-0.038810	0.029054	-1.335777	0.1816
Variance Equation				
C(3)	2.29E-06	1.08E-06	2.113164	0.0346
C(4)	0.034991	0.011482	3.047473	0.0023
C(5)	-0.112779	0.127260	-0.886212	0.3755
C(6)	0.949003	0.015485	61.28482	0.0000
T-DIST. DOF	4.718501	0.621501	7.592100	0.0000

R-squared	-0.000854	Mean dependent var	0.000370
Adjusted R-squared	-0.001746	S.D. dependent var	0.012243
S.E. of regression	0.012253	Akaike info criterion	-6.162634
Sum squared resid	0.168463	Schwarz criterion	-6.131342
Log likelihood	3470.400	Hannan-Quinn criter.	-6.150809
Durbin-Watson stat	1.958352		

(iii) ARIMA-TGARCH(1,1) with GED distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 25 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*(ABS(RESID(-1)) - C(4)*RESID(-1)) + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000621	0.000275	2.256333	0.0240
Variance Equation				
C(2)	2.72E-06	1.06E-06	2.564739	0.0103
C(3)	0.037159	0.012612	2.946406	0.0032
C(4)	-0.013833	0.120114	-0.115169	0.9083
C(5)	0.942763	0.016371	57.58660	0.0000
GED PARAMETER	1.167927	0.047497	24.58942	0.0000

R-squared	-0.000422	Mean dependent var	0.000370
Adjusted R-squared	-0.000422	S.D. dependent var	0.012243
S.E. of regression	0.012245	Akaike info criterion	-6.151450
Sum squared resid	0.168390	Schwarz criterion	-6.124628
Log likelihood	3463.115	Hannan-Quinn criter.	-6.141314
Durbin-Watson stat	2.033052		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 22 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*(ABS(RESID(-1)) - C(6)*RESID(-1)) + C(7)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000714	0.000236	3.029696	0.0024
AR(1)	0.498249	0.248625	2.004017	0.0451
MA(1)	-0.561725	0.236350	-2.376666	0.0175
Variance Equation				
C(4)	2.70E-06	1.07E-06	2.539238	0.0111
C(5)	0.036106	0.012564	2.873627	0.0041
C(6)	-0.037420	0.118147	-0.316722	0.7515
C(7)	0.943985	0.016281	57.98156	0.0000
GED PARAMETER	1.135260	0.046953	24.17843	0.0000

R-squared	0.001167	Mean dependent var	0.000376
Adjusted R-squared	-0.000616	S.D. dependent var	0.012246
S.E. of regression	0.012250	Akaike info criterion	-6.156649
Sum squared resid	0.168076	Schwarz criterion	-6.120861
Log likelihood	3464.959	Hannan-Quinn criter.	-6.143124
F-statistic	0.187003	Durbin-Watson stat	1.898075
Prob(F-statistic)	0.988138		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 17 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*(ABS(RESID(-1)) - C(5)*RESID(-1)) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000691	0.000257	2.694189	0.0071
AR(1)	-0.057976	0.026507	-2.187188	0.0287
	Variance Equation			
C(3)	2.87E-06	1.09E-06	2.635851	0.0084
C(4)	0.036548	0.012508	2.921976	0.0035
C(5)	-0.013929	0.121068	-0.115053	0.9084
C(6)	0.942052	0.016213	58.10367	0.0000
GED PARAMETER	1.142719	0.046815	24.40900	0.0000

R-squared	-0.002103	Mean dependent var	0.000376
Adjusted R-squared	-0.002997	S.D. dependent var	0.012246
S.E. of regression	0.012265	Akaike info criterion	-6.154655
Sum squared resid	0.168626	Schwarz criterion	-6.123340
Log likelihood	3462.839	Hannan-Quinn criter.	-6.142820
Durbin-Watson stat	1.908552		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 21 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*(ABS(RESID(-1)) - C(5)*RESID(-1)) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000707	0.000254	2.780117	0.0054
MA(1)	-0.061891	0.026280	-2.355005	0.0185
	Variance Equation			
C(3)	2.72E-06	1.09E-06	2.495171	0.0126
C(4)	0.036691	0.012582	2.916233	0.0035
C(5)	-0.030417	0.119029	-0.255546	0.7983
C(6)	0.943405	0.016563	56.95943	0.0000
GED PARAMETER	1.141577	0.047176	24.19828	0.0000

R-squared	-0.002188	Mean dependent var	0.000370
Adjusted R-squared	-0.003081	S.D. dependent var	0.012243
S.E. of regression	0.012262	Akaike info criterion	-6.154842
Sum squared resid	0.168687	Schwarz criterion	-6.123550
Log likelihood	3466.021	Hannan-Quinn criter.	-6.143016
Durbin-Watson stat	1.914418		

D) ARIMA-GARCH-M Model

(i) ARIMA-GARCH-M(1,1) with normal distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 58 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-3.481288	6.186807	-0.562695	0.5736
C	0.001007	0.000885	1.138945	0.2547
	Variance Equation			
C	4.08E-06	4.94E-07	8.262945	0.0000
RESID(-1)^2	0.048471	0.007987	6.068642	0.0000
GARCH(-1)	0.924675	0.007928	116.6294	0.0000

R-squared	-0.000364	Mean dependent var	0.000370
Adjusted R-squared	-0.001256	S.D. dependent var	0.012243
S.E. of regression	0.012250	Akaike info criterion	-6.039244
Sum squared resid	0.168380	Schwarz criterion	-6.016892
Log likelihood	3399.055	Hannan-Quinn criter.	-6.030797
Durbin-Watson stat	2.035602		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 12 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	3.253207	2.417042	1.345945	0.1783
AR(1)	-0.915480	0.026392	-34.68719	0.0000
MA(1)	0.934346	0.020650	45.24648	0.0000
	Variance Equation			
C	4.35E-06	5.38E-07	8.086023	0.0000
RESID(-1)^2	0.049767	0.008273	6.015792	0.0000
GARCH(-1)	0.921899	0.008481	108.6990	0.0000

R-squared	0.013488	Mean dependent var	0.000376
Adjusted R-squared	0.011726	S.D. dependent var	0.012246
S.E. of regression	0.012174	Akaike info criterion	-6.044027
Sum squared resid	0.166002	Schwarz criterion	-6.017186
Log likelihood	3399.721	Hannan-Quinn criter.	-6.033883
Durbin-Watson stat	2.038567		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 79 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-2.933323	6.197345	-0.473319	0.6360
C	0.000940	0.000879	1.069136	0.2850
AR(1)	-0.016471	0.037551	-0.438629	0.6609
	Variance Equation			
C	4.22E-06	5.27E-07	8.006269	0.0000
RESID(-1)^2	0.048812	0.008099	6.027023	0.0000
GARCH(-1)	0.923139	0.007921	116.5456	0.0000

R-squared	-0.000252	Mean dependent var	0.000376
Adjusted R-squared	-0.002038	S.D. dependent var	0.012246
S.E. of regression	0.012259	Akaike info criterion	-6.037681
Sum squared resid	0.168314	Schwarz criterion	-6.010840
Log likelihood	3396.158	Hannan-Quinn criter.	-6.027537
Durbin-Watson stat	1.989838		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 62 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-3.713021	6.266136	-0.592554	0.5535
C	0.001040	0.000888	1.171027	0.2416
MA(1)	-0.017657	0.037413	-0.471936	0.6370
	Variance Equation			
C	4.09E-06	5.20E-07	7.866394	0.0000
RESID(-1)^2	0.048875	0.008007	6.104035	0.0000
GARCH(-1)	0.924267	0.007864	117.5388	0.0000

R-squared	-0.000059	Mean dependent var	0.000370
Adjusted R-squared	-0.001843	S.D. dependent var	0.012243
S.E. of regression	0.012254	Akaike info criterion	-6.037705
Sum squared resid	0.168329	Schwarz criterion	-6.010883
Log likelihood	3399.190	Hannan-Quinn criter.	-6.027569
Durbin-Watson stat	2.002125		

(ii) ARIMA-GARCH-M(1,1) with *t* distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 16 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-3.188526	5.420968	-0.588184	0.5564
C	0.001072	0.000737	1.453737	0.1460
Variance Equation				
C	2.46E-06	1.11E-06	2.210787	0.0271
RESID(-1)^2	0.034084	0.011200	3.043149	0.0023
GARCH(-1)	0.948022	0.015641	60.61030	0.0000
T-DIST. DOF	4.803791	0.634248	7.574000	0.0000

R-squared	-0.000475	Mean dependent var	0.000370
Adjusted R-squared	-0.001366	S.D. dependent var	0.012243
S.E. of regression	0.012251	Akaike info criterion	-6.162585
Sum squared resid	0.168399	Schwarz criterion	-6.135763
Log likelihood	3469.373	Hannan-Quinn criter.	-6.152448
Durbin-Watson stat	2.034561		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 56 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
GARCH = C(5) + C(6)*RESID(-1)^2 + C(7)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-3.221013	5.691207	-0.565963	0.5714
C	0.001068	0.000765	1.396256	0.1626
AR(1)	-0.597331	0.213148	-2.802424	0.0051
MA(1)	0.592321	0.214075	2.766889	0.0057
Variance Equation				
C	2.45E-06	1.21E-06	2.023503	0.0430
RESID(-1)^2	0.036111	0.011780	3.065375	0.0022
GARCH(-1)	0.946600	0.016924	55.93341	0.0000
T-DIST. DOF	4.841377	0.640383	7.560132	0.0000

R-squared	0.007791	Mean dependent var	0.000376
Adjusted R-squared	0.005131	S.D. dependent var	0.012246
S.E. of regression	0.012215	Akaike info criterion	-6.161206
Sum squared resid	0.166961	Schwarz criterion	-6.125418
Log likelihood	3467.517	Hannan-Quinn criter.	-6.147681
F-statistic	1.255300	Durbin-Watson stat	2.010592
Prob(F-statistic)	0.269419		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 18 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-2.335238	5.226131	-0.446839	0.6550
C	0.000994	0.000716	1.388994	0.1648
AR(1)	-0.034567	0.029452	-1.173677	0.2405
Variance Equation				
C	2.66E-06	1.13E-06	2.357768	0.0184
RESID(-1)^2	0.034309	0.011236	3.053517	0.0023
GARCH(-1)	0.946437	0.015409	61.42276	0.0000
T-DIST. DOF	4.682086	0.618609	7.568731	0.0000

R-squared	-0.000766	Mean dependent var	0.000376
Adjusted R-squared	-0.002553	S.D. dependent var	0.012246
S.E. of regression	0.012262	Akaike info criterion	-6.162613
Sum squared resid	0.168401	Schwarz criterion	-6.131298
Log likelihood	3467.307	Hannan-Quinn criter.	-6.150778
Durbin-Watson stat	1.954116		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Student's t distribution
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 18 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-3.264531	5.427857	-0.601440	0.5475
C	0.001103	0.000738	1.494115	0.1351
MA(1)	-0.037779	0.029408	-1.284623	0.1989
Variance Equation				
C	2.50E-06	1.13E-06	2.204589	0.0275
RESID(-1)^2	0.034059	0.011229	3.033245	0.0024
GARCH(-1)	0.948027	0.015718	60.31508	0.0000
T-DIST. DOF	4.707630	0.621930	7.569386	0.0000

R-squared	-0.000493	Mean dependent var	0.000370
Adjusted R-squared	-0.002278	S.D. dependent var	0.012243
S.E. of regression	0.012257	Akaike info criterion	-6.162322
Sum squared resid	0.168402	Schwarz criterion	-6.131030
Log likelihood	3470.225	Hannan-Quinn criter.	-6.150496
Durbin-Watson stat	1.962640		

(iii) ARIMA-GARCH-M(1,1) with GED distribution

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 22 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-1.064972	5.012894	-0.212447	0.8318
C	0.000745	0.000677	1.100524	0.2711
Variance Equation				
C	2.74E-06	1.06E-06	2.576434	0.0100
RESID(-1)^2	0.036986	0.012510	2.956581	0.0031
GARCH(-1)	0.942668	0.016314	57.78427	0.0000
GED PARAMETER	1.168742	0.044438	26.30067	0.0000

R-squared	-0.000283	Mean dependent var	0.000370
Adjusted R-squared	-0.001175	S.D. dependent var	0.012243
S.E. of regression	0.012250	Akaike info criterion	-6.151494
Sum squared resid	0.168367	Schwarz criterion	-6.124672
Log likelihood	3463.140	Hannan-Quinn criter.	-6.141358
Durbin-Watson stat	2.033929		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 25 iterations
MA Backcast: 2
Presample variance: backcast (parameter = 0.7)
GARCH = C(5) + C(6)*RESID(-1)^2 + C(7)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-2.066900	4.880579	-0.423495	0.6719
C	0.000963	0.000659	1.459985	0.1443
AR(1)	0.501099	0.253714	1.975060	0.0483
MA(1)	-0.563394	0.241584	-2.332087	0.0197
Variance Equation				
C	2.78E-06	1.08E-06	2.575803	0.0100
RESID(-1)^2	0.035757	0.012339	2.897989	0.0038
GARCH(-1)	0.943494	0.016157	58.39472	0.0000
GED PARAMETER	1.137221	0.043639	26.05968	0.0000

R-squared	0.001674	Mean dependent var	0.000376
Adjusted R-squared	-0.001002	S.D. dependent var	0.012246
S.E. of regression	0.012253	Akaike info criterion	-6.156705
Sum squared resid	0.167990	Schwarz criterion	-6.120917
Log likelihood	3464.990	Hannan-Quinn criter.	-6.143180
F-statistic	0.268100	Durbin-Watson stat	1.902445
Prob(F-statistic)	0.966217		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 3 1125
Included observations: 1123 after adjustments
Convergence achieved after 25 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-0.538249	4.793793	-0.112280	0.9106
C	0.000748	0.000651	1.149485	0.2504
AR(1)	-0.057980	0.026562	-2.182839	0.0290
Variance Equation				
C	2.90E-06	1.10E-06	2.646057	0.0081
RESID(-1)^2	0.036461	0.012449	2.928862	0.0034
GARCH(-1)	0.941803	0.016246	57.97257	0.0000
GED PARAMETER	1.142503	0.043881	26.03611	0.0000

R-squared	-0.002005	Mean dependent var	0.000376
Adjusted R-squared	-0.003794	S.D. dependent var	0.012246
S.E. of regression	0.012270	Akaike info criterion	-6.154656
Sum squared resid	0.168609	Schwarz criterion	-6.123341
Log likelihood	3462.839	Hannan-Quinn criter.	-6.142821
Durbin-Watson stat	1.909049		

Dependent Variable: D(LDT4)
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Sample (adjusted): 2 1125
Included observations: 1124 after adjustments
Convergence achieved after 38 iterations
MA Backcast: 1
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

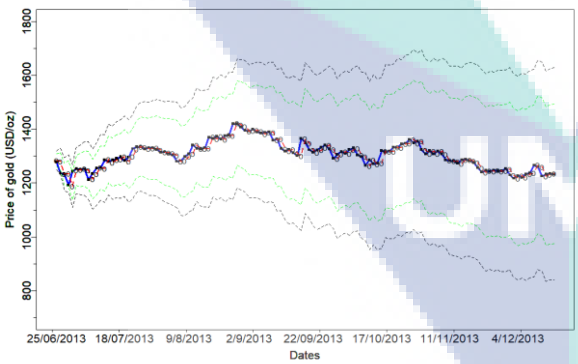
Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-0.211792	4.932258	-0.042940	0.9657
C	0.000727	0.000667	1.089835	0.2758
MA(1)	-0.061397	0.026374	-2.327975	0.0199
Variance Equation				
C	2.78E-06	1.11E-06	2.516359	0.0119
RESID(-1)^2	0.036667	0.012530	2.926362	0.0034
GARCH(-1)	0.942787	0.016599	56.79761	0.0000
GED PARAMETER	1.141629	0.044076	25.90141	0.0000

R-squared	-0.002071	Mean dependent var	0.000370
Adjusted R-squared	-0.003859	S.D. dependent var	0.012243
S.E. of regression	0.012266	Akaike info criterion	-6.154801
Sum squared resid	0.168668	Schwarz criterion	-6.123509
Log likelihood	3465.998	Hannan-Quinn criter.	-6.142976
Durbin-Watson stat	1.915630		

1-STEP AND MULTISTEP AHEAD FORECASTING OF ARIMA-EGARCH (AEG)

1-step ahead forecast AEG

```
> mae_AEG4=sum(error2_AEG4)/T; [1] 12.90242
> rmse_AEG4=sqrt(sum(error3_AEG4)/T); rmse_A
EG4 [1] 17.91294
> mape_AEG4=(100/T)*sum(error4_AEG4); mape_A
EG4 [1] 0.99303
```

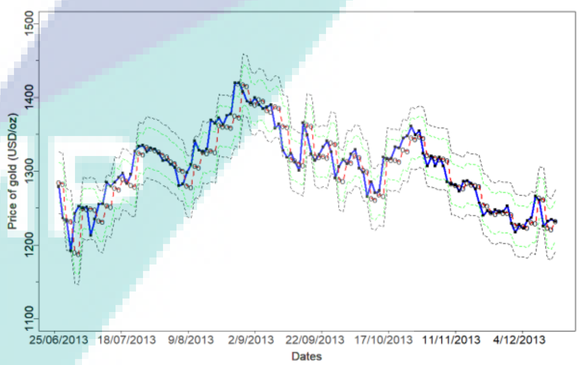



```
> outAEG_1_lo80<-dt4_o[dt4_o<lo80_AEG4];[1]
1236.25
> outAEG_1_hi80<-dt4_o[dt4_o>hi80_AEG4];[1]
1242.75
> outAEG_1_lo95<-dt4_o[dt4_o<lo95_AEG4];nume
ric(0)
> outAEG_1_hi95<-dt4_o[dt4_o>hi95_AEG4];nume
ric(0)
```

% Forecast data NOT within prediction interval
80% PI: 1/125 = 0.8% 95% PI: 0/125 = 0%

2-step ahead forecast AEG

```
> mae_AEG4_2=sum(error2_AEG4_2)/T; 15.64307
> rmse_AEG4_2=sqrt(sum(error3_AEG4_2)/T);
[1] 21.4087
> mape_AEG4_2=(100/T)*sum(error4_AEG4_2);
[1] 1.198982
```

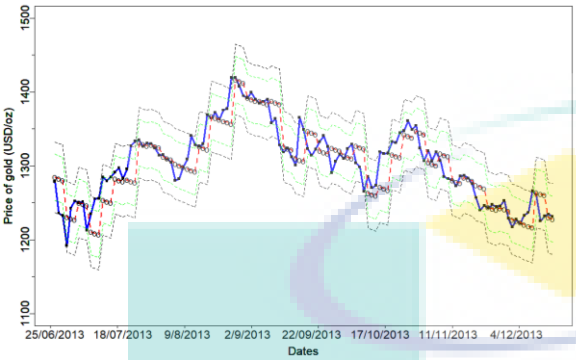



```
> outAEG_2_lo80<-dt4_o[dt4_o<lo80_AEG4_2];
1236.25 1192.00 1212.75 1358.25 1328.00 1318
.50 1323.00 1290.75 1265.50 1285.50 1225.25
> outAEG_2_hi80<-dt4_o[dt4_o>hi80_AEG4_2];
 [1] 1242.75 1252.50 1285.00 1279.75 1327.00
 1309.00 1341.00 1369.25 1365.00 1419.25 141
9.50 1365.50 1316.00 1319.25 1266.25
> outAEG_2_lo95<-dt4_o[dt4_o<lo95_AEG4_2];
[1] 1236.25 1290.75
> outAEG_2_hi95<-dt4_o[dt4_o>hi95_AEG4_2];
[1] 1242.75 1252.50 1327.00 1419.25 1419.50
1365.50 1319.25
```

% Forecast data NOT within prediction interval
80% PI: 23/125 = 18.4% 95% PI: 9/125 = 7.2%

3-step ahead forecast AEG

```
> mae_AEG4_3=sum(error2_AEG4_3)/T; 18.18917
> rmse_AEG4_3=sqrt(sum(error3_AEG4_3)/T);
[1] 24.48962
> mape_AEG4_3=(100/T)*sum(error4_AEG4_3);
[1] 1.398816
```



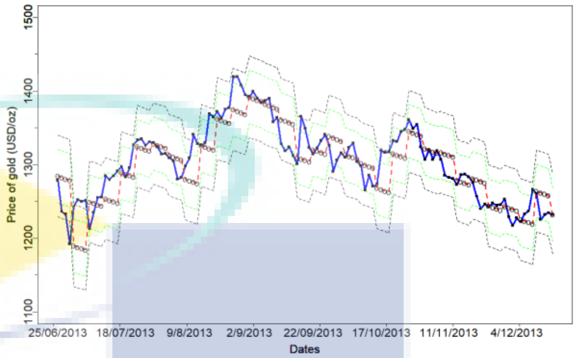

```
> outAEG_3_lo80<-dt4_o[dt4_o<lo80_AEG4_3];
 [1] 1236.25 1232.75 1192.00 1212.75 1328.00
1290.75 1265.50 1306.75 1240.00 1225.25
> outAEG_3_hi80<-dt4_o[dt4_o>hi80_AEG4_3];
 [1] 1255.50 1256.00 1284.75 1327.00 1333.50
1341.00 1328.50 1369.25 1419.25 1365.50 1349.
25 1319.25 1316.50 1317.50 1344.75 1266.25
> outAEG_3_lo95<-dt4_o[dt4_o<lo95_AEG4_3];
[1] 1232.75 1328.00 1265.50
> outAEG_3_hi95<-dt4_o[dt4_o>hi95_AEG4_3];
[1] 1256.00 1333.50 1419.25 1365.50 1317.50
1266.25
```

% Forecast data NOT within prediction interval

80% PI: 26/125 = 20.8% 95% PI: 9/125 = 7.2%

4-step ahead forecast AEG

```
> mae_AEG4_4=sum(error2_AEG4_4)/T; 21.82219
> rmse_AEG4_4=sqrt(sum(error3_AEG4_4)/T);
[1] 28.48176
> mape_AEG4_4=(100/T)*sum(error4_AEG4_4);
[1] 1.681157
```



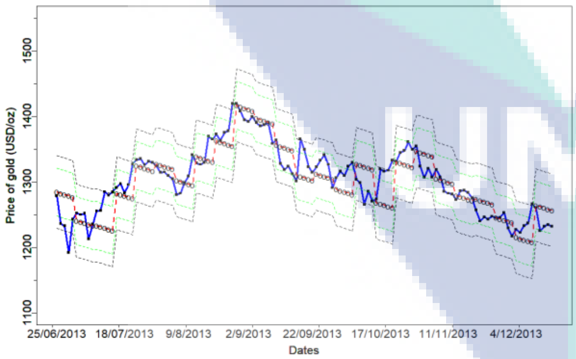

```
> outAEG_4_lo80<-dt4_o[dt4_o<lo80_AEG4_4];
[1] 1236.25 1232.75 1192.00 1212.75 1318.50
1312.25 1290.75 1240.00 1225.25
> outAEG_4_hi80<-dt4_o[dt4_o>hi80_AEG4_4];
 [1] 1242.75 1252.50 1250.00 1251.75 1291.50
 1327.00 1341.00 1328.50 1369.25 1365.00 141
9.25 1419.50 1407.75 1365.50 1349.25 1319.25
 1316.50 1317.50 1347.75 1266.25
> outAEG_4_lo95<-dt4_o[dt4_o<lo95_AEG4_4];
[1] 1192
> outAEG_4_hi95<-dt4_o[dt4_o>hi95_AEG4_4];
 [1] 1252.50 1250.00 1251.75 1341.00 1365.50
```

% Forecast data NOT within prediction interval

80% PI: 29/125 = 23.2% 95% PI: 6/125 = 4.8%

5-step ahead forecast AEG

```
> mae_AEG4_5=sum(error2_AEG4_5)/T; 22.46578
> rmse_AEG4_5=sqrt(sum(error3_AEG4_5)/T);
[1] 28.35659
> mape_AEG4_5=(100/T)*sum(error4_AEG4_5);
[1] 1.729642
```



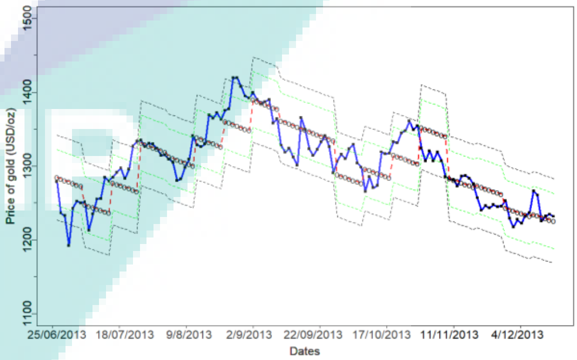

```
> outAEG_5_lo80<-dt4_o[dt4_o<lo80_AEG4_5];
[1] 1236.25 1232.75 1192.00 1265.50 1270.50
1225.25
> outAEG_5_hi80<-dt4_o[dt4_o>hi80_AEG4_5];
 [1] 1285.00 1279.75 1284.75 1327.00 1341.00
1369.25 1419.25 1365.50 1349.25 1323.50 1329.
50 1319.25 1316.50 1317.50 1333.00 1266.25
> outAEG_5_lo95<-dt4_o[dt4_o<lo95_AEG4_5];
[1] 1192.0 1265.5
> outAEG_5_hi95<-dt4_o[dt4_o>hi95_AEG4_5];
[1] 1284.75 1419.25 1365.50 1333.00 1266.25
```

% Forecast data NOT within prediction interval

80% PI: 22/125 = 17.6% 95% PI: 7/125 = 5.6%

7-step ahead forecast AEG

```
> mae_AEG4_7=sum(error2_AEG4_7)/T; 23.35416
> rmse_AEG4_7=sqrt(sum(error3_AEG4_7)/T);
[1] 29.25207
> mape_AEG4_7=(100/T)*sum(error4_AEG4_7);
[1] 1.785301
```



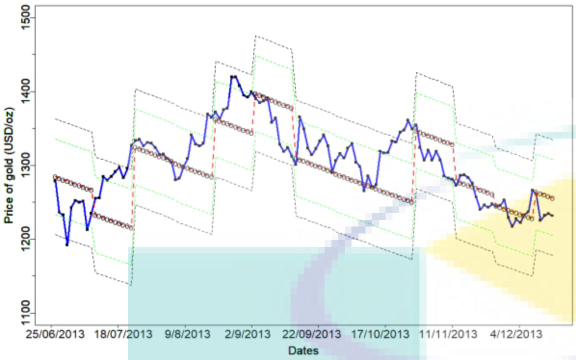

```
> outAEG_7_lo80<-dt4_o[dt4_o<lo80_AEG4_7];
 [1] 1236.25 1232.75 1192.00 1318.50 1312.25
 1301.00 1290.75 1306.75 1307.25 1285.50
> outAEG_7_hi80<-dt4_o[dt4_o>hi80_AEG4_7];
 [1] 1285.00 1279.75 1327.00 1333.50 1341.00
 1372.50 1419.25 1419.50 1407.75 1394.75 139
2.25 1323.50 1329.50 1347.75 1361.00 1349.25
 1354.75
> outAEG_7_lo95<-dt4_o[dt4_o<lo95_AEG4_7];
[1] 1192
> outAEG_7_hi95<-dt4_o[dt4_o>hi95_AEG4_7];
[1] 1327.00 1333.50 1419.25 1419.50
```

% Forecast data NOT within prediction interval

80% PI: 27/125 = 21.6% 95% PI: 5/125 = 4%

10-step ahead forecast AEG

```
> mae_AEG4_10=sum(error2_AEG4_10)/T; 32.01497
> rmse_AEG4_10=sqrt(sum(error3_AEG4_10)/T);
[1] 41.2452
> mape_AEG4_10=(100/T)*sum(error4_AEG4_10);
[1] 2.440347
```



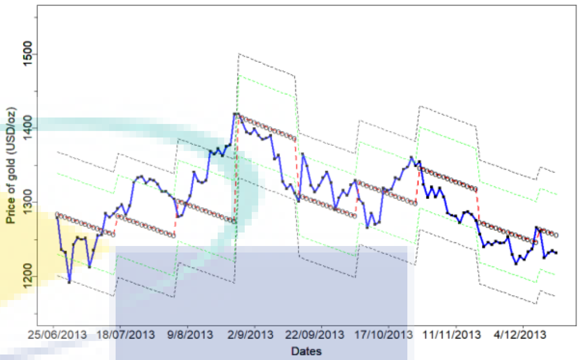

```
> outAEG_10_lo80<-dt4_o[dt4_o<lo80_AEG4_10];
[1] 1192.00 1212.75 1328.00 1318.50 1324.00
1312.25
> outAEG_10_hi80<-dt4_o[dt4_o>hi80_AEG4_10];
 [1] 1285.00 1279.75 1284.75 1291.50 1297.25
1283.25 1295.75 1327.00 1369.25 1365.00
1419.25 1419.50 1407.75 1399.50 1365.50
1319.25 1316.50 1317.50 1333.00 1331.25
1344.75 1347.75 1361.00 1349.25
> outAEG_10_lo95<-dt4_o[dt4_o<lo95_AEG4_10];
[1] 1192
> outAEG_10_hi95<-dt4_o[dt4_o>hi95_AEG4_10];o
utAEG_10_hi95
[1] 1327.00 1369.25 1365.00 1344.75 1347.75 1
361.00 1349.25
```

% Forecast data NOT within prediction interval

80% PI: 30/125 = 24%   95% PI: 8/125 = 6.4%

15-step ahead forecast AEG

```
> mae_AEG4_15=sum(error2_AEG4_15)/T;35.03482
> rmse_AEG4_15=sqrt(sum(error3_AEG4_15)/T);
[1] 42.50387
> mape_AEG4_15=(100/T)*sum(error4_AEG4_15);
[1] 2.674686
```



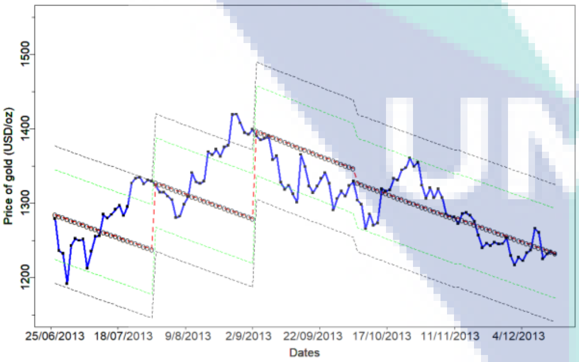

```
> outAEG_15_lo80<-dt4_o[dt4_o<lo80_AEG4_15];
[1] 1192.00 1212.75 1328.00 1318.50 1324.00
1312.25 1265.50
> outAEG_15_hi80<-dt4_o[dt4_o>hi80_AEG4_15];
 [1] 1333.50 1335.00 1326.00 1331.00 1329.75
 1324.15 1315.00 1369.25 1365.00 1372.50 136
3.00 1375.50 1377.50 1419.25 1365.50 1361.00
> outAEG_15_lo95<-dt4_o[dt4_o<lo95_AEG4_15];
[1] 1192
> outAEG_15_hi95<-dt4_o[dt4_o>hi95_AEG4_15];
[1] 1372.50 1375.50 1377.50 1419.25
```

% Forecast data NOT within prediction interval

80% PI: 23/125 = 18.4%   95% PI: 5/125 =4%

25-step ahead forecast AEG

```
> mae_AEG4_25=sum(error2_AEG4_25)/T; 35.96708
> rmse_AEG4_25=sqrt(sum(error3_AEG4_25)/T);
[1] 47.13697
> mape_AEG4_25=(100/T)*sum(error4_AEG4_25);
[1] 2.721687
```



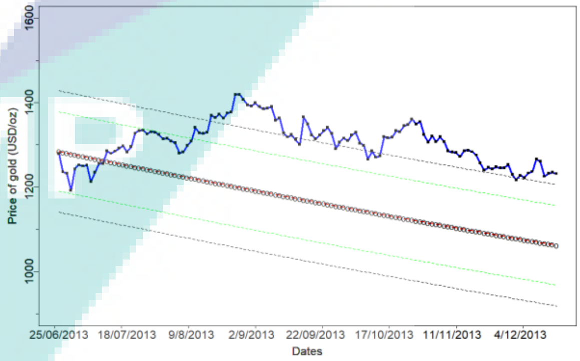

```
> outAEG_25_lo80<-dt4_o[dt4_o<lo80_AEG4_25];
[1] 1192.00 1318.50 1312.25 1301.00 1290.75
> outAEG_25_hi80<-dt4_o[dt4_o>hi80_AEG4_25];
 [1] 1327.00 1333.50 1335.00 1326.00 1331.00
1329.75 1369.25 1365.00 1372.50 1363.00
1375.50 1377.50 1419.25 1419.50 1407.75 1394.
75 1392.25 1399.50
> outAEG_25_lo95<-dt4_o[dt4_o<lo95_AEG4_25];
numeric(0)
> outAEG_25_hi95<-dt4_o[dt4_o>hi95_AEG4_25];
[1] 1419.25 1419.50 1407.75 1394.75 1392.25
1399.50
```

% Forecast data NOT within prediction interval

80% PI: 23/125 = 18.4%   95% PI: 6/125 =4.8%

125-step ahead forecast AEG

```
> mae_AEG4_n=sum(error2_AEG4_n)/T;  140.4588
> rmse_AEG4_n=sqrt(sum(error3_AEG4_n)/T);
[1] 152.7912
> mape_AEG4_n=(100/T)*sum(error4_AEG4_n);
[1] 10.7086
```




```
> outAEG_n_lo80<-dt4_o[dt4_o<lo80_AEG4_n];
numeric(0)
> outAEG_n_hi80<-dt4_o[dt4_o>hi80_AEG4_n];
1341.00 1328.50 1326.50 1329.75 1369.25 1365
.00 1372.50 1363.00 1375.50 1377.50 1419.25
1419.50 1407.75 1394.75 1392.25 1399.50 1390
.00 1385.00 1387.00 1390.00 1358.25 1363.75
1328.00 1318.50 1324.00 1312.25 1301.00 1365
.50 1349.25 1323.00 1314.25 1322.75 1333.00
1341.00 1326.50 1290.75 1306.25 1316.00 1309
.75 1323.50 1329.50 1304.00 1298.50 1265.50
1285.50 1270.50 1273.50 1319.25 1316.50 1317
.50 1333.00 1331.25 1344.75 1347.75 1361.00
1349.25 1354.75 1324.00 1306.75 1320.50 1307
.25 1319.00 1307.25 1285.50 1282.50 1281.25
```

	1272.50 1286.00 1287.25 1283.50 1275.75 1257.00 1240.00 1246.25 1243.00 1247.50 1245 .00 1245.50 1253.00 1229.50 1217.25 1227.50 1222.50 1233.00 1237.00 1266.25 1260.75 1225 .25 1232.00 1234.75 1231.75 > outAEG_n_lo95<-dt4_o[dt4_o<lo95_AEG4_n]; numeric(0) > outAEG_n_hi95<-dt4_o[dt4_o>hi95_AEG4_n]; 1369.25 1365.00 1372.50 1363.00 1375.50 1377 .50 1419.25 1419.50 1407.75 1394.75 1392.25 1399.50 1390.00 1385.00 1387.00 1390.00 1358 .25 1363.75 1328.00 1324.00 1365.50 1349.25 1323.00 1314.25 1322.75 1333.00 1341.00 1326 .50 1306.25 1316.00 1309.75 1323.50 1329.50 1304.00 1298.50 1285.50 1319.25 1316.50 1317 .50 1333.00 1331.25 1344.75 1347.75 1361.00 1349.25 1354.75 1324.00 1306.75 1320.50 1307 .25 1319.00 1307.25 1285.50 1282.50 1281.25 1272.50 1286.00 1287.25 1283.50 1275.75 1257 .00 1240.00 1246.25 1243.00 1247.50 1245.00 1245.50 1253.00 1229.50 1227.50 1222.50 1233 .00 1237.00 1266.25 1260.75 1225.25 1232.00 1234.75 1231.75 % Forecast data NOT within prediction interval 80% PI: 91/125 = 72.8% 95% PI: 79/125 =63.2%

To find Prediction Intervals for 7-step ahead ARIMA-EGARCH:

```
for(i in 1:125){
  lo95_AEG4_7[i]<-f_AEG4_7[i]-(1.96*29.3230)#lower limit 95%
  hi95_AEG4_7[i]<-f_AEG4_7[i]+(1.96*29.3230)#upper limit 95%
  lo80_AEG4_7[i]<-f_AEG4_7[i]-(1.2816*29.3230)#lower limit 80%
  hi80_AEG4_7[i]<-f_AEG4_7[i]+(1.2816*29.3230)#upper limit 80%
}
cbind(dt4_o,f_AEG4_7,lo95_AEG4_7,hi95_AEG4_7,lo80_AEG4_7,hi80_AEG4_7)
```

	f_AEG4_7		resiAEG4_7
nobs	125.000000	nobs	125.000000
NAs	0.000000	NAs	0.000000
Minimum	1225.671358	Minimum	-86.881786
Maximum	1390.116771	Maximum	67.413050
1. Quartile	1273.012216	1. Quartile	-20.307467
3. Quartile	1340.286224	3. Quartile	21.530490
Mean	1302.593530	Mean	1.649670
Median	1300.960609	Median	0.834798
Sum	162824.191245	Sum	206.208755
SE Mean	3.884357	SE Mean	2.622733
LCL Mean	1294.905300	LCL Mean	-3.541452
UCL Mean	1310.281760	UCL Mean	6.840793
Variance	1886.028291	Variance	859.840786
Stdev	43.428427	Stdev	29.323042
Skewness	0.059025	Skewness	0.045667
Kurtosis	-1.019764	Kurtosis	-0.271759

Simulation Price Data of 7-step ahead using ARIMA(0,1,0)-EGARCH(1,1) with normal

```
> cbind(dt4_o,f_AEG4_7,lo95_AEG4_7,hi95_AEG4_7,lo80_AEG4_7,hi80_AEG4_7)
          [,1]     [,2]     [,3]     [,4]     [,5]     [,6]
  [1,] 1279.00 1284.778 1227.305 1342.251 1247.198 1322.359
  [2,] 1236.25 1282.810 1225.337 1340.283 1245.230 1320.390
  [3,] 1232.75 1280.844 1223.371 1338.317 1243.264 1318.425
  [4,] 1192.00 1278.882 1221.409 1336.355 1241.301 1316.462
  [5,] 1242.75 1276.922 1219.449 1334.395 1239.342 1314.503
  [6,] 1252.50 1274.966 1217.493 1332.439 1237.385 1312.546
  [7,] 1250.00 1273.012 1215.539 1330.485 1235.432 1310.593
  [8,] 1251.75 1248.085 1190.612 1305.558 1210.504 1285.665
  [9,] 1212.75 1246.172 1188.699 1303.645 1208.592 1283.753
 [10,] 1235.25 1244.263 1186.790 1301.736 1206.683 1281.843
 [11,] 1255.50 1242.357 1184.883 1299.830 1204.776 1279.937
 [12,] 1256.00 1240.453 1182.980 1297.926 1202.873 1278.033
 [13,] 1285.00 1238.552 1181.079 1296.025 1200.972 1276.133
 [14,] 1279.75 1236.655 1179.181 1294.128 1199.074 1274.235
 [15,] 1284.75 1277.789 1220.316 1335.262 1240.209 1315.370
 [16,] 1291.50 1275.831 1218.358 1333.304 1238.251 1313.412
 [17,] 1297.25 1273.876 1216.403 1331.350 1236.296 1311.457
```

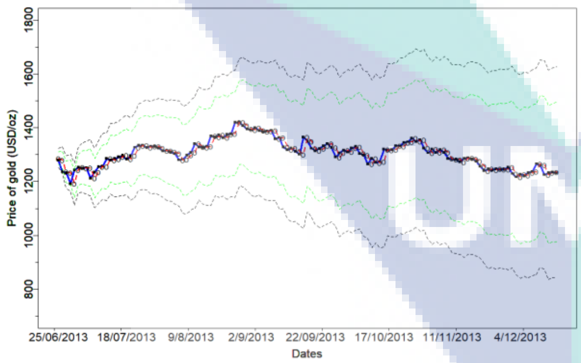
[18,]	1283.25	1271.925	1214.452	1329.398	1234.344	1309.505
[19,]	1295.75	1269.976	1212.503	1327.449	1232.395	1307.556
[20,]	1327.00	1268.030	1210.557	1325.503	1230.449	1305.610
[21,]	1333.50	1266.087	1208.614	1323.560	1228.507	1303.667
[22,]	1335.00	1331.457	1273.984	1388.930	1293.876	1369.037
[23,]	1326.00	1329.417	1271.944	1386.890	1291.836	1366.997
[24,]	1331.00	1327.380	1269.907	1384.853	1289.799	1364.960
[25,]	1329.75	1325.346	1267.873	1382.819	1287.766	1362.926
[26,]	1324.15	1323.315	1265.842	1380.788	1285.735	1360.896
[27,]	1314.50	1321.288	1263.815	1378.761	1283.707	1358.868
[28,]	1315.00	1319.263	1261.790	1376.736	1281.683	1356.843
[29,]	1309.25	1312.985	1255.512	1370.458	1275.405	1350.565
[30,]	1304.75	1310.973	1253.500	1368.446	1273.393	1348.554
[31,]	1280.50	1308.965	1251.492	1366.438	1271.384	1346.545
[32,]	1282.50	1306.959	1249.486	1364.432	1269.379	1344.539
[33,]	1298.25	1304.956	1247.483	1362.430	1267.376	1342.537
[34,]	1309.00	1302.957	1245.484	1360.430	1265.377	1340.537
[35,]	1341.00	1300.961	1243.488	1358.434	1263.380	1338.541
[36,]	1328.50	1338.945	1281.472	1396.418	1301.365	1376.526
[37,]	1326.50	1336.894	1279.421	1394.367	1299.313	1374.474
[38,]	1329.75	1334.845	1277.372	1392.318	1297.265	1372.426
[39,]	1369.25	1332.800	1275.327	1390.273	1295.220	1370.380
[40,]	1365.00	1330.758	1273.285	1388.231	1293.178	1368.338
[41,]	1372.50	1328.719	1271.246	1386.192	1291.139	1366.299
[42,]	1363.00	1326.683	1269.210	1384.156	1289.103	1364.263
[43,]	1375.50	1360.912	1303.439	1418.385	1323.331	1398.492
[44,]	1377.50	1358.826	1301.353	1416.299	1321.246	1396.407
[45,]	1419.25	1356.744	1299.271	1414.217	1319.164	1394.325
[46,]	1419.50	1354.666	1297.192	1412.139	1317.085	1392.246
[47,]	1407.75	1352.590	1295.117	1410.063	1315.010	1390.170
[48,]	1394.75	1350.517	1293.044	1407.991	1312.937	1388.098
[49,]	1392.25	1348.448	1290.975	1405.921	1310.868	1386.029
[50,]	1399.50	1390.117	1332.644	1447.590	1352.536	1427.697
[51,]	1390.00	1387.987	1330.514	1445.460	1350.406	1425.567
[52,]	1385.00	1385.860	1328.387	1443.333	1348.280	1423.440
[53,]	1387.00	1383.737	1326.264	1441.210	1346.156	1421.317
[54,]	1390.00	1381.616	1324.143	1439.090	1344.036	1419.197
[55,]	1358.25	1379.500	1322.026	1436.973	1341.919	1417.080
[56,]	1363.75	1377.386	1319.913	1434.859	1339.806	1414.966
[57,]	1328.00	1361.660	1304.187	1419.134	1324.080	1399.241
[58,]	1318.50	1359.574	1302.101	1417.047	1321.994	1397.154
[59,]	1324.00	1357.491	1300.018	1414.964	1319.911	1395.071
[60,]	1312.25	1355.411	1297.938	1412.884	1317.831	1392.991
[61,]	1301.00	1353.334	1295.861	1410.807	1315.754	1390.915
[62,]	1365.50	1351.261	1293.787	1408.734	1313.680	1388.841
[63,]	1349.25	1349.190	1291.717	1406.663	1311.610	1386.770
[64,]	1323.00	1347.183	1289.710	1404.656	1309.602	1384.763
[65,]	1314.25	1345.118	1287.645	1402.592	1307.538	1382.699
[66,]	1322.75	1343.057	1285.584	1400.531	1305.477	1380.638
[67,]	1333.00	1341.000	1283.527	1398.473	1303.419	1378.580
[68,]	1341.00	1338.945	1281.472	1396.418	1301.365	1376.525
[69,]	1326.50	1336.893	1279.420	1394.366	1299.313	1374.474
[70,]	1290.75	1334.845	1277.372	1392.318	1297.265	1372.425
[71,]	1306.25	1288.772	1231.299	1346.245	1251.192	1326.353
[72,]	1316.00	1286.798	1229.325	1344.271	1249.217	1324.378
[73,]	1309.75	1284.826	1227.353	1342.299	1247.246	1322.406
[74,]	1323.50	1282.857	1225.384	1340.330	1245.277	1320.438
[75,]	1329.50	1280.892	1223.419	1338.365	1243.311	1318.472
[76,]	1304.00	1278.929	1221.456	1336.402	1241.349	1316.509
[77,]	1298.50	1276.970	1219.496	1334.443	1239.389	1314.550
[78,]	1265.50	1296.510	1239.037	1353.983	1258.930	1334.091
[79,]	1285.50	1294.524	1237.051	1351.997	1256.944	1332.104
[80,]	1270.50	1292.540	1235.067	1350.013	1254.960	1330.121
[81,]	1273.50	1290.560	1233.087	1348.033	1252.980	1328.140
[82,]	1319.25	1288.583	1231.109	1346.056	1251.002	1326.163
[83,]	1316.50	1286.608	1229.135	1344.081	1249.028	1324.188
[84,]	1317.50	1284.637	1227.164	1342.110	1247.056	1322.217
[85,]	1333.00	1315.481	1258.008	1372.954	1277.901	1353.062
[86,]	1331.25	1313.466	1255.993	1370.939	1275.885	1351.046
[87,]	1344.75	1311.453	1253.980	1368.926	1273.873	1349.034
[88,]	1347.75	1309.444	1251.971	1366.917	1271.863	1347.024
[89,]	1361.00	1307.437	1249.964	1364.910	1269.857	1345.018
[90,]	1349.25	1305.434	1247.961	1362.907	1267.854	1343.014
[91,]	1354.75	1303.434	1245.961	1360.907	1265.854	1341.014
[92,]	1324.00	1352.674	1295.201	1410.147	1315.094	1390.255
[93,]	1306.75	1350.602	1293.129	1408.075	1313.021	1388.182
[94,]	1320.50	1348.532	1291.059	1406.005	1310.952	1386.113

```
 [95,] 1307.25 1346.466 1288.993 1403.939 1308.886 1384.046
 [96,] 1319.00 1344.403 1286.930 1401.876 1306.823 1381.983
 [97,] 1307.25 1342.343 1284.870 1399.816 1304.763 1379.923
 [98,] 1285.50 1340.286 1282.813 1397.759 1302.706 1377.867
 [99,] 1282.50 1283.530 1226.057 1341.003 1245.950 1321.111
[100,] 1281.25 1281.564 1224.091 1339.037 1243.983 1319.144
[101,] 1272.50 1279.600 1222.127 1337.073 1242.020 1317.180
[102,] 1286.00 1277.639 1220.166 1335.113 1240.059 1315.220
[103,] 1287.25 1275.682 1218.209 1333.155 1238.101 1313.262
[104,] 1283.50 1273.727 1216.254 1331.200 1236.147 1311.308
[105,] 1275.75 1271.776 1214.302 1329.249 1234.195 1309.356
[106,] 1257.00 1273.795 1216.322 1331.268 1236.215 1311.376
[107,] 1240.00 1271.844 1214.370 1329.317 1234.263 1309.424
[108,] 1246.25 1269.895 1212.422 1327.368 1232.314 1307.475
[109,] 1243.00 1267.949 1210.476 1325.422 1230.369 1305.529
[110,] 1247.50 1266.006 1208.533 1323.479 1228.426 1303.587
[111,] 1245.00 1264.066 1206.593 1321.540 1226.486 1301.647
[112,] 1245.50 1262.130 1204.657 1319.603 1224.549 1299.710
[113,] 1253.00 1243.592 1186.119 1301.065 1206.011 1281.172
[114,] 1229.50 1241.686 1184.213 1299.159 1204.106 1279.267
[115,] 1217.25 1239.784 1182.311 1297.257 1202.203 1277.364
[116,] 1227.50 1237.884 1180.411 1295.357 1200.304 1275.464
[117,] 1222.50 1235.987 1178.514 1293.460 1198.407 1273.568
[118,] 1233.00 1234.094 1176.620 1291.567 1196.513 1271.674
[119,] 1237.00 1232.203 1174.730 1289.676 1194.622 1269.783
[120,] 1266.25 1235.105 1177.632 1292.578 1197.524 1272.685
[121,] 1260.75 1233.212 1175.739 1290.685 1195.632 1270.793
[122,] 1225.25 1231.323 1173.850 1288.796 1193.742 1268.903
[123,] 1232.00 1229.436 1171.963 1286.909 1191.856 1267.016
[124,] 1234.75 1227.552 1170.079 1285.025 1189.972 1265.133
[125,] 1231.75 1225.671 1168.198 1283.144 1188.091 1263.252
```

1-STEP AND MULTISTEP AHEAD FORECASTING OF ARIMA-TGARCH (ATG)

1-step ahead forecast ATG

```
> mae_ATG4=sum(error2_ATG4)/T; 12.91445
> rmse_ATG4=sqrt(sum(error3_ATG4)/T);
[1] 17.85786
> mape_ATG4=(100/T)*sum(error4_ATG4);
[1] 0.9943338
```



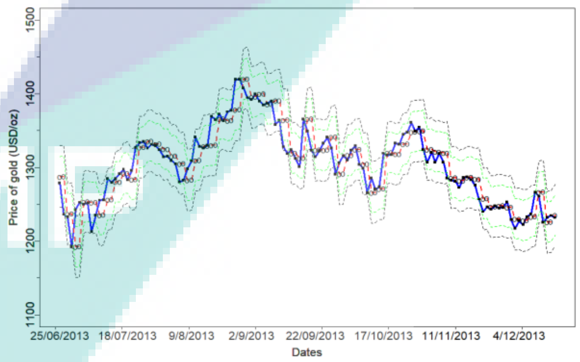

```
> outATG_1_lo80<-dt4_o[dt4_o<lo80_ATG4];
[1] 1236.25
> outATG_1_hi80<-dt4_o[dt4_o>hi80_ATG4];
numeric(0)
> outATG_1_lo95<-dt4_o[dt4_o<lo95_ATG4];
numeric(0)
> outATG_1_hi95<-dt4_o[dt4_o>hi95_ATG4];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 1/125 = 0.8%   95% PI: 0/125 = 0%

2-step ahead forecast ATG

```
> mae_ATG4_2=sum(error2_ATG4_2)/T; 15.73296
> rmse_ATG4_2=sqrt(sum(error3_ATG4_2)/T);
[1] 21.29756
> mape_ATG4_2=(100/T)*sum(error4_ATG4_2);
[1] 1.208305
```



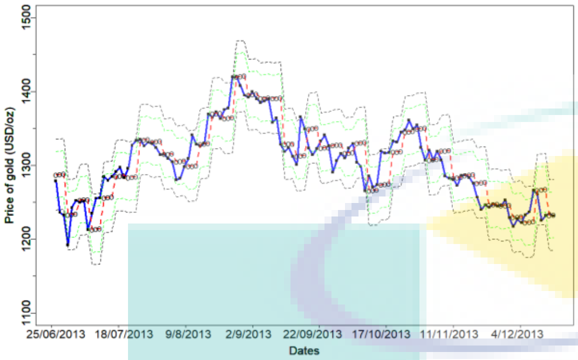

```
> outATG_2_lo80<-dt4_o[dt4_o<lo80_ATG4_2];
 [1] 1236.25 1192.00 1212.75 1358.25 1363.75
 1328.00 1318.50 1323.00 1290.75 1265.50 128
5.50 1257.00 1225.25
> outATG_2_hi80<-dt4_o[dt4_o>hi80_ATG4_2];
 [1] 1242.75 1252.50 1285.00 1327.00 1341.00
 1369.25 1365.00 1419.25 1419.50 1365.50 131
9.25 1266.25
> outATG_2_lo95<-dt4_o[dt4_o<lo95_ATG4_2];
[1] 1236.25 1192.00 1318.50 1323.00 1290.75
1225.25
> outATG_2_hi95<-dt4_o[dt4_o>hi95_ATG4_2];
[1] 1242.75 1252.50 1327.00 1365.50 1319.25
```

% Forecast data NOT within prediction interval

80% PI: 25/125 = 20%   95% PI: 11/125 = 8.8%

3-step ahead forecast ATG

```
> mae_ATG4_3=sum(error2_ATG4_3)/T; 18.20806
> rmse_ATG4_3=sqrt(sum(error3_ATG4_3)/T);
[1] 24.38725
> mape_ATG4_3=(100/T)*sum(error4_ATG4_3);
[1] 1.40284
```



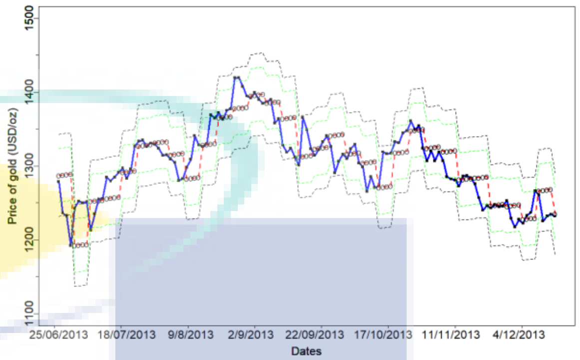

```
> outATG_3_lo80<-dt4_o[dt4_o<lo80_ATG4_3];
1236.25 1232.75 1192.00 1212.75 1358.25 1328.
00 1314.25 1290.75 1298.50 1265.50 1306.75 12
85.50 1282.50 1240.00 1246.25 1225.25 1232.00
> outATG_3_hi80<-dt4_o[dt4_o>hi80_ATG4_3];
 [1] 1255.50 1256.00 1327.00 1333.50 1341.00
1369.25 1419.25 1365.50 1349.25 1319.25 1316.
50 1317.50 1266.25
> outATG_3_lo95<-dt4_o[dt4_o<lo95_ATG4_3];
[1] 1236.25 1232.75 1328.00 1265.50
> outATG_3_hi95<-dt4_o[dt4_o>hi95_ATG4_3];
[1] 1333.50 1419.25 1365.50
```

% Forecast data NOT within prediction interval

80% PI: 30/125 = 24% 95% PI: 7/125 = 5.6%

4-step ahead forecast ATG

```
> mae_ATG4_4=sum(error2_ATG4_4)/T; 21.55504
> rmse_ATG4_4=sqrt(sum(error3_ATG4_4)/T);
[1] 28.28821
> mape_ATG4_4=(100/T)*sum(error4_ATG4_4);
[1] 1.66681
```



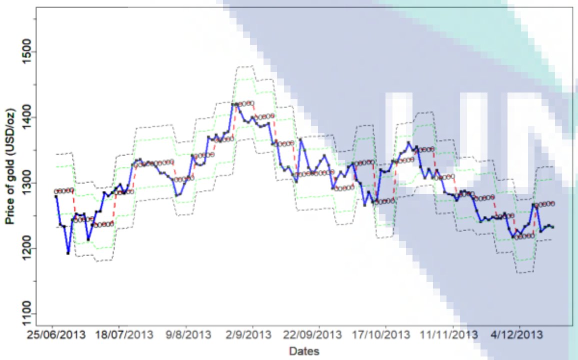

```
>outATG_4_lo80<-dt4_o[dt4_o<lo80_ATG4_4];123
6.25 1232.75 1192.00 1212.75 1280.50 1328.00
 1318.50 1324.00 1312.25 1290.75 1306.25 126
5.50 1282.50 1281.25 1240.00 1246.25 1225.25
> outATG_4_hi80<-dt4_o[dt4_o>hi80_ATG4_4];
 [1] 1242.75 1252.50 1250.00 1251.75 1341.00
 1328.50 1369.25 1419.25 1419.50 1365.50 131
9.25 1316.50 1317.50 1266.25
> outATG_4_lo95<-dt4_o[dt4_o<lo95_ATG4_4];
[1] 1232.75 1192.00
> outATG_4_hi95<-dt4_o[dt4_o>hi95_ATG4_4];
[1] 1252.50 1250.00 1251.75 1341.00
```

% Forecast data NOT within prediction interval

80% PI: 31/125 = 24.8% 95% PI: 6/125 = 4.8%

5-step ahead forecast ATG

```
> mae_ATG4_5=sum(error2_ATG4_5)/T; 22.71886
> rmse_ATG4_5=sqrt(sum(error3_ATG4_5)/T);
[1] 28.75122
> mape_ATG4_5=(100/T)*sum(error4_ATG4_5);
[1] 1.754874
```



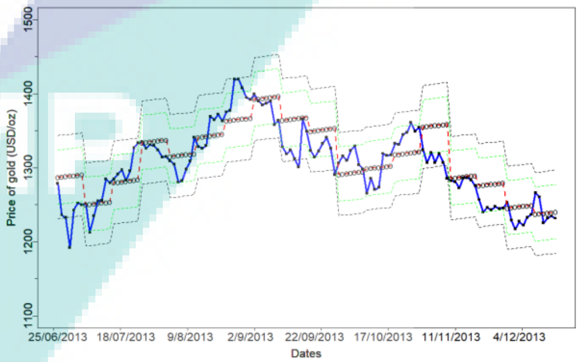

```
> outATG_5_lo80<-dt4_o[dt4_o<lo80_ATG4_5];
 [1] 1236.25 1232.75 1192.00 1242.75 1358.25
1318.50 1324.00 1312.25 1265.50 1285.50 1270.
50 1306.75 1307.25 1240.00 1225.25 1231.75
> outATG_5_hi80<-dt4_o[dt4_o>hi80_ATG4_5];
 1285.00 1279.75 1284.75 1327.00 1419.25 1365
.50 1319.25 1316.50 1317.50 1333.00 1266.25
> outATG_5_lo95<-dt4_o[dt4_o<lo95_ATG4_5];
[1] 1232.75 1192.00 1265.50 1270.50
> outATG_5_hi95<-dt4_o[dt4_o>hi95_ATG4_5];
[1] 1333
```

% Forecast data NOT within prediction interval

80% PI: 27/125 = 21.6% 95% PI: 5/125 = 4%

7-step ahead forecast ATG

```
> mae_ATG4_7=sum(error2_ATG4_7)/T; 24.34648
> rmse_ATG4_7=sqrt(sum(error3_ATG4_7)/T);
[1] 29.82328
> mape_ATG4_7=(100/T)*sum(error4_ATG4_7);
[1] 1.873586
```



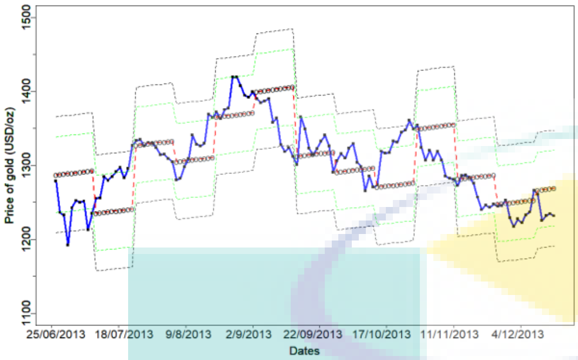

```
> outATG_7_lo80<-dt4_o[dt4_o<lo80_ATG4_7];
1236.25 1232.75 1192.00 1242.75 1252.50 1250
.00 1212.75 1358.25 1318.50 1324.00 1312.25
[12] 1301.00 1290.75 1306.75 1307.25 1319.00
 1307.25 1285.50 1240.00
> outATG_7_hi80<-dt4_o[dt4_o>hi80_ATG4_7];
[1] 1327.00 1333.50 1419.25 1419.50 1407.75
1361.00
> outATG_7_lo95<-dt4_o[dt4_o<lo95_ATG4_7];
[1] 1192.00 1301.00 1290.75 1285.50
> outATG_7_hi95<-dt4_o[dt4_o>hi95_ATG4_7];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 25/125 = 20% 95% PI: 4/125 = 3.2%

10-step ahead forecast ATG

```
> mae_ATG4_10=sum(error2_ATG4_10)/T; 32.14324
> rmse_ATG4_10=sqrt(sum(error3_ATG4_10)/T);
[1] 39.9784
> mape_ATG4_10=(100/T)*sum(error4_ATG4_10);
[1] 2.472696
```



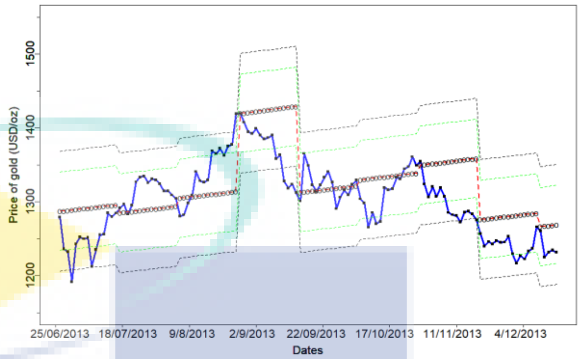

```
> outATG_10_lo80<-dt4_o[dt4_o<lo80_ATG4_10];
1236.25 1232.75 1192.00 1212.75 1235.25 1328.
00 1318.50 1324.00 1312.25 1285.50 1282.50 12
81.25
> outATG_10_hi80<-dt4_o[dt4_o>hi80_ATG4_10];
 [1] 1291.50 1297.25 1295.75 1327.00 1369.25
1365.00 1365.50 1333.00 1331.25 1344.75 1347.
75 1361.00 1349.25
> outATG_10_lo95<-dt4_o[dt4_o<lo95_ATG4_10];
[1] 1192.00 1212.75 1318.50 1324.00 1312.25
> outATG_10_hi95<-dt4_o[dt4_o>hi95_ATG4_10];
[1] 1327 1361
```

% Forecast data NOT within prediction interval

80% PI: 25/125 = 20% 95% PI: 6/125 = 4.8%

15-step ahead forecast ATG

```
> mae_ATG4_15=sum(error2_ATG4_15)/T;37.06404
> rmse_ATG4_15=sqrt(sum(error3_ATG4_15)/T);
[1] 45.28961
> mape_ATG4_15=(100/T)*sum(error4_ATG4_15);
[1] 2.858706
```



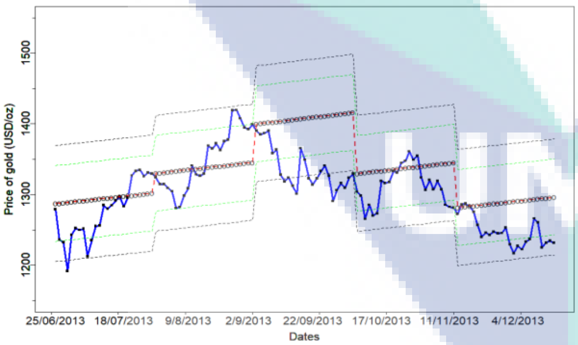

```
> outATG_15_lo80<-dt4_o[dt4_o<lo80_ATG4_15];
1232.75 1192.00 1212.75 1235.25 1358.25 1363
.75 1328.00 1318.50 1324.00 1312.25 1265.50
1270.50 1273.50 1285.50 1282.50 1281.25 1272
.50 1286.00 1287.25 1283.50 1275.75 1217.25
1227.50 1222.50
> outATG_15_hi80<-dt4_o[dt4_o>hi80_ATG4_15];
1369.25 1365.00 1372.50 1375.50 1377.50 1419
.25
> outATG_15_lo95<-dt4_o[dt4_o<lo95_ATG4_15];
1192.00 1328.00 1318.50 1324.00 1312.25 1272
.50 1275.75
> outATG_15_hi95<-dt4_o[dt4_o>hi95_ATG4_15];
1419.25
```

% Forecast data NOT within prediction interval

80% PI: 30/125 =24 % 95% PI: 8/125 =6.4%

25-step ahead forecast ATG

```
> mae_ATG4_25=sum(error2_ATG4_25)/T; 42.28412
> rmse_ATG4_25=sqrt(sum(error3_ATG4_25)/T);
[1] 51.03541
> mape_ATG4_25=(100/T)*sum(error4_ATG4_25);
[1] 3.262471
```



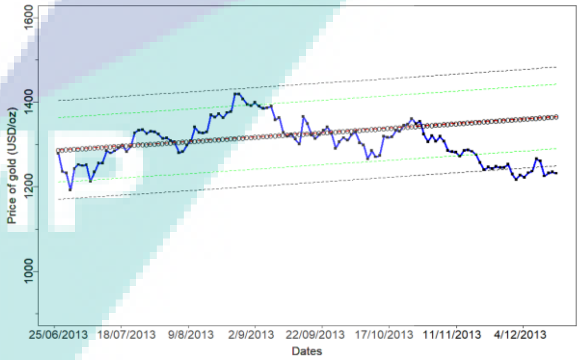

```
> outATG_25_lo80<-dt4_o[dt4_o<lo80_ATG4_25];
1232.75 1192.00 1212.75 1235.25 1328.00 1318.50
1324.00 1312.25 1301.00 1349.25 1323.00 1314.25
1322.75 1333.00 1341.00 1326.50 1290.751306.25 1
316.00 1309.75 1323.50 1329.50 1265.50 1270.50 1
273.50 1285.50 1282.50 1281.25 1229.50 1217.25 1
227.50 1222.50 1233.00 1237.00 1225.25 1232.00 1
234.75 1231.75
> outATG_25_hi80<-dt4_o[dt4_o>hi80_ATG4_25];
[1] 1419.25 1419.50 1407.75 1399.50
> outATG_25_lo95<-dt4_o[dt4_o<lo95_ATG4_25];
1192.00 1318.50 1312.25 1301.00 1323.00 1314.25
1322.75 1326.50 1290.75 1306.25 1316.00 1309.75
1323.50 1329.50
> outATG_25_hi95<-dt4_o[dt4_o>hi95_ATG4_25];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 42/125 = 33.6% 95% PI: 14/125 =11.2%

125-step ahead forecast ATG

```
> mae_ATG4_n=sum(error2_ATG4_n)/T; 49.04749
> rmse_ATG4_n=sqrt(sum(error3_ATG4_n)/T);
[1] 63.36834
> mape_ATG4_n=(100/T)*sum(error4_ATG4_n);
[1] 3.817525
```




```
> outATG_n_lo80<-dt4_o[dt4_o<lo80_ATG4_n];
 [1] 1192.00 1212.75 1272.50 1275.75 1257.00
 1240.00 1246.25 1243.00 1247.50 1245.00 124
5.50 1253.00 1229.50 1217.25 1227.50 1222.50
 1233.00 1237.00 1266.25 1260.75 1225.25 123
2.00 1234.75 1231.75
> outATG_n_hi80<-dt4_o[dt4_o>hi80_ATG4_n];
[1] 1419.25 1419.50 1407.75 1394.75 1399.50
> outATG_n_lo95<-dt4_o[dt4_o<lo95_ATG4_n];
 [1] 1229.50 1217.25 1227.50 1222.50 1233.00
 1237.00 1225.25 1232.00 1234.75 1231.75
> outATG_n_hi95<-dt4_o[dt4_o>hi95_ATG4_n];
numeric(0)
```

% Forecast data NOT within prediction interval

80% PI: 29/125 = 23.2% 95% PI: 10/125 =8%

```
                  f_ATG4_7                       resiATG4_7
nobs             125.000000      nobs            125.000000
NAs                0.000000      NAs               0.000000
Minimum         1237.594661      Minimum         -97.226094
Maximum         1396.941822      Maximum          54.283356
1. Quartile     1282.829020      1. Quartile     -31.319397
3. Quartile     1351.846363      3. Quartile      13.189219
Mean            1313.064209      Mean             -8.821009
Median          1315.632158      Median           -8.822258
Sum           164133.026167      Sum           -1102.626167
SE Mean            3.909900      SE Mean           2.558380
LCL Mean        1305.325422      LCL Mean        -13.884759
UCL Mean        1320.802996      UCL Mean         -3.757259
Variance        1910.914661      Variance        818.163288
Stdev             43.714010      Stdev            28.603554
Skewness           0.049821      Skewness         -0.101724
Kurtosis          -1.025866      Kurtosis         -0.207700
```

To find Prediction Interval for 7-step ahead ARIMA-TGARCH:

```
for(i in 1:125){
  lo95_ATG4_7[i]<-f_ATG4_7[i]-(1.96*28.6036)#lower limit 95%
  hi95_ATG4_7[i]<-f_ATG4_7[i]+(1.96*28.6036)#upper limit 95%
  lo80_ATG4_7[i]<-f_ATG4_7[i]-(1.2816*28.6036)#lower limit 80%
  hi80_ATG4_7[i]<-f_ATG4_7[i]+(1.2816*28.6036)#upper limit 80%
}
cbind(dt4_o,f_ATG4_7,lo95_ATG4_7,hi95_ATG4_7,lo80_ATG4_7,hi80_ATG4_7)
```

Simulation Price Data of 7-step ahead using ARIMA(0,1,0)-TGARCH(1,1) with normal

```
> cbind(dt4_o,f_ATG4_7,lo95_ATG4_7,hi95_ATG4_7,lo80_ATG4_7,hi80_ATG4_7)
         [,1]     [,2]     [,3]     [,4]     [,5]     [,6]
 [1,] 1279.00 1287.369 1231.306 1343.432 1250.710 1324.027
 [2,] 1236.25 1287.987 1231.924 1344.051 1251.329 1324.646
 [3,] 1232.75 1288.607 1232.544 1344.670 1251.948 1325.265
 [4,] 1192.00 1289.226 1233.163 1345.289 1252.568 1325.884
 [5,] 1242.75 1289.846 1233.783 1345.909 1253.187 1326.504
 [6,] 1252.50 1290.466 1234.403 1346.529 1253.808 1327.124
 [7,] 1250.00 1291.086 1235.023 1347.149 1254.428 1327.745
 [8,] 1251.75 1250.601 1194.538 1306.664 1213.943 1287.259
 [9,] 1212.75 1251.202 1195.139 1307.265 1214.544 1287.860
[10,] 1235.25 1251.804 1195.741 1307.867 1215.145 1288.462
[11,] 1255.50 1252.405 1196.342 1308.468 1215.747 1289.064
[12,] 1256.00 1253.007 1196.944 1309.070 1216.349 1289.666
[13,] 1285.00 1253.610 1197.547 1309.673 1216.951 1290.268
[14,] 1279.75 1254.212 1198.149 1310.276 1217.554 1290.871
[15,] 1284.75 1280.365 1224.302 1336.428 1243.707 1317.024
[16,] 1291.50 1280.981 1224.918 1337.044 1244.322 1317.639
[17,] 1297.25 1281.597 1225.533 1337.660 1244.938 1318.255
[18,] 1283.25 1282.213 1226.150 1338.276 1245.554 1318.871
[19,] 1295.75 1282.829 1226.766 1338.892 1246.171 1319.487
[20,] 1327.00 1283.446 1227.383 1339.509 1246.787 1320.104
[21,] 1333.50 1284.063 1228.000 1340.126 1247.404 1320.721
[22,] 1335.00 1334.141 1278.078 1390.204 1297.483 1370.799
[23,] 1326.00 1334.782 1278.719 1390.845 1298.124 1371.441
[24,] 1331.00 1335.424 1279.361 1391.487 1298.766 1372.082
[25,] 1329.75 1336.066 1280.003 1392.129 1299.408 1372.724
[26,] 1324.15 1336.708 1280.645 1392.771 1300.050 1373.367
[27,] 1314.50 1337.351 1281.288 1393.414 1300.693 1374.009
[28,] 1315.00 1337.994 1281.931 1394.057 1301.335 1374.652
[29,] 1309.25 1315.632 1259.569 1371.695 1278.974 1352.291
[30,] 1304.75 1316.265 1260.202 1372.328 1279.606 1352.923
[31,] 1280.50 1316.897 1260.834 1372.960 1280.239 1353.556
[32,] 1282.50 1317.530 1261.467 1373.594 1280.872 1354.189
[33,] 1298.25 1318.164 1262.101 1374.227 1281.505 1354.822
[34,] 1309.00 1318.798 1262.734 1374.861 1282.139 1355.456
[35,] 1341.00 1319.431 1263.368 1375.495 1282.773 1356.090
[36,] 1328.50 1341.645 1285.582 1397.708 1304.986 1378.303
[37,] 1326.50 1342.290 1286.227 1398.353 1305.631 1378.948
[38,] 1329.75 1342.935 1286.872 1398.998 1306.277 1379.593
[39,] 1369.25 1343.580 1287.517 1399.644 1306.922 1380.239
[40,] 1365.00 1344.226 1288.163 1400.289 1307.568 1380.885
[41,] 1372.50 1344.873 1288.810 1400.936 1308.214 1381.531
[42,] 1363.00 1345.519 1289.456 1401.582 1308.861 1382.177
[43,] 1375.50 1363.655 1307.592 1419.718 1326.997 1400.314
[44,] 1377.50 1364.311 1308.248 1420.374 1327.652 1400.969
[45,] 1419.25 1364.967 1308.904 1421.030 1328.308 1401.625
[46,] 1419.50 1365.623 1309.560 1421.686 1328.964 1402.281
[47,] 1407.75 1366.279 1310.216 1422.342 1329.621 1402.938
[48,] 1394.75 1366.936 1310.873 1422.999 1330.278 1403.594
```

[49,]	1392.25	1367.593	1311.530	1423.656	1330.935	1404.252
[50,]	1399.50	1392.919	1336.856	1448.982	1356.261	1429.578
[51,]	1390.00	1393.589	1337.526	1449.652	1356.931	1430.247
[52,]	1385.00	1394.259	1338.196	1450.322	1357.600	1430.917
[53,]	1387.00	1394.929	1338.866	1450.992	1358.271	1431.587
[54,]	1390.00	1395.600	1339.537	1451.663	1358.941	1432.258
[55,]	1358.25	1396.271	1340.208	1452.334	1359.612	1432.929
[56,]	1363.75	1396.942	1340.879	1453.005	1360.283	1433.600
[57,]	1328.00	1364.406	1308.343	1420.469	1327.747	1401.064
[58,]	1318.50	1365.062	1308.998	1421.125	1328.403	1401.720
[59,]	1324.00	1365.718	1309.655	1421.781	1329.059	1402.376
[60,]	1312.25	1366.374	1310.311	1422.437	1329.716	1403.033
[61,]	1301.00	1367.031	1310.968	1423.094	1330.373	1403.689
[62,]	1365.50	1367.688	1311.625	1423.751	1331.030	1404.347
[63,]	1349.25	1368.346	1312.283	1424.409	1331.687	1405.004
[64,]	1323.00	1349.899	1293.836	1405.962	1313.240	1386.557
[65,]	1314.25	1350.548	1294.485	1406.611	1313.889	1387.206
[66,]	1322.75	1351.197	1295.134	1407.260	1314.538	1387.855
[67,]	1333.00	1351.846	1295.783	1407.909	1315.188	1388.505
[68,]	1341.00	1352.496	1296.433	1408.559	1315.838	1389.155
[69,]	1326.50	1353.146	1297.083	1409.209	1316.488	1389.805
[70,]	1290.75	1353.797	1297.734	1409.860	1317.139	1390.455
[71,]	1306.25	1291.371	1235.307	1347.434	1254.712	1328.029
[72,]	1316.00	1291.991	1235.928	1348.054	1255.333	1328.650
[73,]	1309.75	1292.612	1236.549	1348.675	1255.954	1329.271
[74,]	1323.50	1293.234	1237.171	1349.297	1256.575	1329.892
[75,]	1329.50	1293.855	1237.792	1349.919	1257.197	1330.514
[76,]	1304.00	1294.477	1238.414	1350.541	1257.819	1331.136
[77,]	1298.50	1295.100	1239.037	1351.163	1258.441	1331.758
[78,]	1265.50	1299.124	1243.061	1355.187	1262.466	1335.783
[79,]	1285.50	1299.749	1243.686	1355.812	1263.090	1336.407
[80,]	1270.50	1300.374	1244.311	1356.437	1263.715	1337.032
[81,]	1273.50	1300.999	1244.936	1357.062	1264.340	1337.657
[82,]	1319.25	1301.624	1245.561	1357.687	1264.966	1338.283
[83,]	1316.50	1302.250	1246.187	1358.313	1265.591	1338.908
[84,]	1317.50	1302.876	1246.813	1358.939	1266.218	1339.534
[85,]	1333.00	1318.133	1262.070	1374.196	1281.475	1354.792
[86,]	1331.25	1318.767	1262.704	1374.830	1282.109	1355.425
[87,]	1344.75	1319.401	1263.338	1375.464	1282.743	1356.059
[88,]	1347.75	1320.035	1263.972	1376.098	1283.377	1356.694
[89,]	1361.00	1320.670	1264.607	1376.733	1284.011	1357.328
[90,]	1349.25	1321.305	1265.242	1377.368	1284.646	1357.963
[91,]	1354.75	1321.940	1265.877	1378.003	1285.282	1358.598
[92,]	1324.00	1355.401	1299.338	1411.464	1318.743	1392.060
[93,]	1306.75	1356.053	1299.990	1412.116	1319.394	1392.711
[94,]	1320.50	1356.705	1300.642	1412.768	1320.046	1393.363
[95,]	1307.25	1357.357	1301.294	1413.420	1320.699	1394.015
[96,]	1319.00	1358.009	1301.946	1414.073	1321.351	1394.668
[97,]	1307.25	1358.662	1302.599	1414.725	1322.004	1395.321
[98,]	1285.50	1359.315	1303.252	1415.379	1322.657	1395.974
[99,]	1282.50	1286.118	1230.055	1342.181	1249.460	1322.776
[100,]	1281.25	1286.736	1230.673	1342.799	1250.078	1323.395
[101,]	1272.50	1287.355	1231.292	1343.418	1250.696	1324.013
[102,]	1286.00	1287.974	1231.911	1344.037	1251.315	1324.632
[103,]	1287.25	1288.593	1232.530	1344.656	1251.934	1325.251
[104,]	1283.50	1289.212	1233.149	1345.275	1252.554	1325.871
[105,]	1275.75	1289.832	1233.769	1345.895	1253.174	1326.490
[106,]	1257.00	1276.363	1220.300	1332.426	1239.705	1313.022
[107,]	1240.00	1276.977	1220.914	1333.040	1240.318	1313.635
[108,]	1246.25	1277.591	1221.528	1333.654	1240.932	1314.249
[109,]	1243.00	1278.205	1222.142	1334.268	1241.547	1314.863
[110,]	1247.50	1278.819	1222.756	1334.882	1242.161	1315.478
[111,]	1245.00	1279.434	1223.371	1335.497	1242.776	1316.093
[112,]	1245.50	1280.049	1223.986	1336.112	1243.391	1316.708
[113,]	1253.00	1246.099	1190.036	1302.162	1209.440	1282.757
[114,]	1229.50	1246.698	1190.635	1302.761	1210.039	1283.356
[115,]	1217.25	1247.297	1191.234	1303.360	1210.639	1283.955
[116,]	1227.50	1247.897	1191.834	1303.960	1211.238	1284.555
[117,]	1222.50	1248.497	1192.434	1304.560	1211.838	1285.155
[118,]	1233.00	1249.097	1193.034	1305.160	1212.438	1285.755
[119,]	1237.00	1249.697	1193.634	1305.760	1213.039	1286.356
[120,]	1266.25	1237.595	1181.532	1293.658	1200.936	1274.253
[121,]	1260.75	1238.190	1182.127	1294.253	1201.531	1274.848
[122,]	1225.25	1238.785	1182.722	1294.848	1202.126	1275.443
[123,]	1232.00	1239.380	1183.317	1295.443	1202.722	1276.039
[124,]	1234.75	1239.976	1183.913	1296.039	1203.318	1276.635
[125,]	1231.75	1240.572	1184.509	1296.635	1203.914	1277.231