University of Al-Qadisiyah

College of Administration & Economics

Statistics Department

Using Machine Learning Models with Time-Varying to Estimate GARCH Models

A thesis submitted to the council of the college of Administration & Economics\ University of Al-Qadisiyah as partial fulfillment of the requirements for the degree of master in Statistics

By

Sura Hassan Sami

Supervised by

Prof. Dr. Tahir R. Dikheel

1442 A.H. **2021 A.D.**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

الرَّحْمَنُ ﴿1﴾ عَلَّمَ الْقُرْآنَ ﴿2﴾ خَلَقَ الْإِنْسَانَ ﴿3﴾ عَلَّمَهُ الْبَيَانَ

﴿4﴾ الشَّمْسُ وَالْقَمَرُ بِحُسْبَانٍ ﴿5﴾ وَالنَّجْمُ وَالشَّجَرُ يَسْجُدَانِ ﴿6﴾

وَالسَّمَاءَ رَفَعَهَا وَوَضَعَ الْمِيزَانَ ﴿7﴾ أَلَّا تَطْغَوْا فِي الْمِيزَانِ ﴿8﴾

وَأَقِيمُوا الْوَزْنَ بِالْقِسْطِ وَلَا تُخْسِرُوا الْمِيزَانَ ﴿9﴾

صدق الله العلي العظيم

سورة الرحمن المباركة (الايات 1-9)

Dedication

To my Allah, may my work benefit me on the day of doom

To the last of the prophets Mohammed (peace be upon him)

To the righteous servants of Allah

TO my supervisor and teacher

To my illustrious teachers

To my paradise on the earth (my mother)

To my support and my destiny (my father)

To the love of my life (my husband)

To my family (my brothers and my sisters)

To Iraqi martyrs and to all whom helped me to complete this thesis

Thanks and acknowledge

Alhamdulillah, I praise and thank Allah for His greatness and for giving me the strength and courage to complete this thesis.

I also extend my thanks and appreciation to my professors who enlightened me through science and knowledge unitll I became what I am now. I do not forget the scientific seminar committee which played a big role in drawing the general lines of this thesis.

I would also like to thank Mr. .Chairman and the members of the discussion committee who will have the greatest impact in evaluating this thesis to show its final form.

I have to thank my parents for their love and support throughout my life. Thank you both for giving me strength to reach for the stars and chase my dreams.

I would also like to thank my dear husband and my support in this world, which his wish was to complete my studies, now I am fulfilling this dream, and my brothers and my sisters deserve my wholehearted thanks as well

And to all those who have helped me and forget to mention it, I extend to all my thanks

And thank Allah the lord of the whole universe

- Abstract

Support Vector Regression (SVR) model is an important part of the machine learning techniques that it developed to solve the regression problems. In the high dimensions space, SVR model is used to determine the optimal hyperplane for modifying the non- linear patterns in regression. In this thesis, the SVR model developed by combining it with penalized methods such as adaptive LASSO (ALASSO) and Elastic net (EN) to reduce the dimensions of the model and improve its performance where we refer to our proposed methods by ALASSO-SVR and EN – SVR. However, the penalized methods are compared with the machine learning models (random forests and regression trees) and LASSO to know the high accuracy predictive performance of these models. Financial time series are characterized by the presence of fluctuations that occur randomly during different time periods, this is disagreement with penalized methods and the machine learning models which is assume the constants of variance. Moreover, we use the General Autoregressive Conditional Variance (GARCH) models with these models to estimate the conditional variance and the parameters of the ALASSO- SVR, EN – SVR methods and the machine learning models. Then the predictive performance of these models are carried out by iterative procedure where the parameters estimated by GARCH model applied to make the prediction of one-step-ahead by recursive estimation. However, these parameters are updated by the new information. We use the economical variables and lagged variables to predict the monthly exchange rate returns of IQD/USD. Our results of the simulation and real data display that the suggested methods are better than the machine learning models, particular the EN – SVR method is capable of best predicting monthly exchange rate returns and also improving the predictive performance.

Table of Contents

Content tables

Content figures

Chapter one

1.1 Introduction

Financial forecasting plays an important role for companies and investors, especially those who invest their money in financial markets. Consider the extreme difficulties in predicting share prices, even if they are dependent on time, forecasting the exchange rates of the dollar against other currencies is less difficult, because changes in these prices may be somewhat stable if the economical conditions are stable. For instance, a country such as Iraq is subject to political and economical turmoil that greatly affects the prediction of exchange rates from one period to another (Hassan 2016). Although, this type of data is considered as time series in the statistical literature, but the difficulty lies in its modeling.

The financial time series suffers from several problems such as heteroscedasticity, noise and leptokurtosis (Cont, 2001). Homoscedasticity is one of the several assumptions of regression where the variance of the random errors is constant and this is sometimes inconsistent with real data. The Homoscedasticity assumption is often violated in the financial time series, due to the rapid occurrence of volatility associated with time, where in some periods they are more volatile than the others and lead to clustering in specific periods which is so called heteroscedasticity. Noise is one of the important concepts in the analysis of the time series and the prediction, especially in the financial data, where it refers to the missing information in the series that affects its behavior and this leads to many problems such as overfitting and under fitting. Leptokurtosis distributions are statistical distributions that used to describe heavier tails or a higher probability when kurtosis more than three of the normal distribution.

In these circumstances, the financial market is unpredictable and any improvement that occurs during this period is argument. Unfortunately, the huge efforts that have paid in the literature were unable to predict exchange rates due to volatility that occurs during different time periods and does not indicate the reasons leading to change in currency rates. This is reflected by the assumptions of the

random walk and efficient market hypothesis, where random change in the time series is unpredictable (Garg 2012). The emergence of these fluctuations that are associated with the financial time series, which made the use of regular time series models not feasible. So interest began to study other models that simulate this type of data. Moreover, Engle (1982) proposed Autoregressive Conditional Homoscedasticity (ARCH) which is the first model in the family of the ARCH models. It represents a special case of the Generalized Autoregressive Conditional Homoscedasticity (GARCH) models, before ARCH model was introduced, no accurate prediction method has been found. It is a prediction model which is far affected by the values of the squared residuals for previous period that can be defined as (Engle, 1982);

$$\varepsilon_t = \sigma_t * e_t \,, \qquad (1.1)$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^{s} \alpha_i \varepsilon_{t-i}^2 \,, \qquad (1.2)$$

where a_0 is the intercept value, $a_1, \ldots\ldots, \alpha_i$ is the parameter of the ARCH model, $a_0, \alpha_i > 0$, $i = 1,2, \ldots\ldots, s$, s is the order of model, e_t is the series of the identify independent distribution, ε_t is the residuals series and σ_t^2 is the conditional variance.

Researchers have noted that when performing practical applications using the ARCH models that expansion in the values of s may result in negative values for α and this contradicts with assumption of the model that the values of $\alpha \geq 0$. Moreover, Bollerslev (1986) proposed a general model is called GARCH models from order (s, m) that it required a lot of parameters to describe the homoscedasticity process accuracy in the time series. The GARCH models have been studied and applied widely in studies and research over the past years by applying stages of time series. It depends on the variance of the previous time and the squared residuals of the previous period. It is a more globally in scientific fields,

so in order to predict the volatility in the future depending on the variance of the previous time and its volatility can be defined as (Bollerslev, 1986):

$$\sigma_t^2 = a_0 + \sum_{i=1}^{s} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2 \ , \qquad (1.3)$$

where a_0 is the intercept value, α_i, β_j are the parameters of the model, $a_0, \alpha_i, \beta_j > 0$, $i = 1,2,..,s$, $j = 1,2,..m$ and s, m represent the orders of model.

This thesis is divided into five chapters, the first chapter includes an introduction, the literature review and thesis objectives. The second chapter consists of the theoretical side, where it explained ARMA model, ARCH and GARCH models and its tests, the machine learning models, LASSO, Elastic net, adaptive LASSO, variable selection and selection of penalty parameter. The third chapter displays the simulation side. The fourth chapter shows the applied side. The fifth chapter contains the conclusions and the recommendations.

1.2 literature review

In this section, we show the studies conducted by researchers and economists that are related to the subject of our research.

Fernando and et al (2003) focused on estimation the parameter of a GARCH model by using SVM method to predict the volatility of stock market returns and it compared with the maximum likelihood method (MLE). They found, Support Vector Machine (SVM) estimator has a higher predicting ability than MLE estimator methods.

Ince and Rafalis (2006) suggested a two stages predicting model which combines the parametric technique such as autoregressive integrated moving average (ARIMA), vector autoregressive (VAR) and co-integration techniques and non-parametric technique such as support vector regression (SVR) and artificial neural networks (ANN). They showed that input section is very important and then SVR technique outperforms the ANN for both input section methods.

Haydee (2008) introduced an applied study on the Philippine stock market. The study included prediction of volatility in the rate of inflation using the GARCH model with three types of statistical distributions of error, including the normal distribution, student t with fixed df distribution and general error distribution with fixed parameter.

Chen and et al (2009) suggested a performance comparison SVM-GARCH with simple moving average, standard GARCH, ANN-GARCH and Exponential GARCH (EGARCH) models to predict the volatility of the daily British Pound (GBP) exchange rates and New York Stock Exchange (NYSE) composite index by using two evaluation measures and robust Diebold–Mariano tests. Simulation study and real data pointed out that SVM-GARCH models outperformed other models in most situations for volatility predicting.

Hang and Shin (2010) introduced a technique to estimate GARCH models by using kernel machine learning for forecasting the conditional volatility of stock market returns, although GARCH models usually estimated by the maximum likelihood methods (MLE). They concluded that when estimating conditional volatility of financial return, GARCH model can be estimated by the kernel machine learning and that the kernel machine learning has a higher forecasting ability than MLE.

Alamili (2011) suggested a model for predicting exchange rate returns for the euro against the US dollar by using the machine learning models such ANN and SVM. He concluded in his thesis that the SVM model offers some advantages that exceed the advantages of ANN in financial forecasting.

Garg (2012) proposed using machine learning model as Regression trees (RTs), Random forests (RFs), SVM and LASSO to predict exchange rate returns with GARCH models extend these models and compare the predictive performance of the machine learning models with the autoregressive process. He showed that the

predictive performance of SVR with GARCH-extended is better than other models to predict exchange rate returns.

Okasha (2014) presented a research to predict the movement of the Palestinian Stock Exchange Market by using the machine learning models such as SVM and ANN and compare them with the ARIMA time series model. He showed that the SVM model is the most reliable and accurate model in prediction.

Rofael and Hosni (2015) presented forecasting and estimation for the daily exchange rates in Egypt by using ARCH model, stochastic volatility and time-varying parameter. They found that the exchange rate returns in Egypt suffer from volatility and that there is a risk mismatch between the stock market volatility and exchange rate market volatility

Mohsin (2018) introduced research to study the Pakistan daily exchange rate against the US dollar by using two types of symmetric and asymmetric GARCH models. The result of the study showed that fluctuations are the best for the Pakistan exchange rate and that it is more appropriate for the GARCH family models.

Chen and Du (2020) introduced a study to predict the Bitcoin price against the Fed funds rate, Euro/GBP rate, the USD/GBP rate and the West Texas Intermediate price by using GARCH model and SVM and decision trees. The simulation study showed there are positive and negative relationships among the Bitcoin price with the Fed funds rate, Euro/GBP rate, the USD/GBP rate and the West Texas Intermediate price and the machine learning methods more suitable than traditional statistics methods to predict the Bitcoin price.

Through literature review, it is noticed that SVR method did not employ the variable selection in a single algorithmic framework. Therefore, our suggestion employs the ALASSO and EN methods to reduce dimensions and combine them

with SVR to take advantage of the residuals of them to reduce fluctuations and obtain accurate results.

1.3 Thesis objectives

This thesis aims to suggest two methods: ALASSO and EN by combining them with SVR model and GARCH models to deal with the presence of fluctuations, in order to get the fit model for analyzing the time series and obtaining high predictions accuracy. Our methods will refer to as (ALASSO-SVR) and (EN-SVR). The simulation and real data that represent the monthly returns of Iraqi dinar (IQD) against (USD) are used to compare our proposed methods with familiar previous methods such as Regression trees, Random forests, LASSO by the root of mean squared error (RMSE) criterion.

Chapter two

The theoretical part

2.1 Time series

The researchers focused on the subject of time series because of its importance in studying the behavior of various phenomena over specific time periods through analyzing and interpreting these phenomena. It has covered a wide range of economy, medicine, environment and other fields. The time series y_t are defined as a stochastic process of a set of observations collected over time and in many time series effects appear to be monthly, weekly, or daily at certain periods of the year as a result of many economic, natural and seasonal factors etc. Therefore, the time series is affected by several changes and these changes may affect by secular trend of the time series in the long and short term (Box, Jenkins, Reinsel &Ljung, 2016). Depending on the time series data, a prediction process for the future is made.

Time series analysis is an important and vital issue that is indispensable in the planning and decision-making process. It consists of two variables, the first variable is explanatory that represents the time variable and the second is the response variable, which is the value of the phenomenon being studied as follows:

$$y = f(t) \ . \qquad (2.1)$$

There are two types of time series, discrete and continuous (Brockwell and Davis, 2002). One of the goals of time series analysis is to obtain an accurate description, build models that explain its behavior and it uses the results to predict the behavior of the series in the future.

The construction of the appropriate model in the time series begins with the stage of diagnosis the model by using one or more statistical tests, then the stage of selecting the order of the model and the stage of estimating parameters of this model. The estimation common methods are least square method (LS), maximum likelihood method (ML) and moment method (MOM). Then checking the model

using one or more statistical tests, if this model is appropriate for estimation, the stage of forecasting is implement.

The stationarity plays an important role in time series analysis, because it is considered a natural alternative the hypothesis of observations, where it is independent and identically distributed (iid) in standard statistics (Francq & Zakoian, 2010). The non-stationary of time series either in the mean , where it means that the mean is depending on time that it can be covered into stationary by using differences or the non-stationary in variance that can be converted into stationary by using transformation. There are two types of stationary, the weak stationary and the strictly stationary. Time series is called the weak stationary if the mean and variance are not depending on time and the covariance depends only on the lag l (Nason, 2006):

$$cov(y_t, y_{t+l}) = \gamma_l \quad . \qquad (2.2)$$

The time series is said to be the strictly stationary if all statistical properties do not change with time. More formally, it stochastic process that equals 1 for each integer $h \geq 1$, for any time $t_1, \ldots\ldots, t_h$ and the joint statistical distribution of $y_{t_1}, y_{t_2}, \ldots . y_{t_l}$ is constant over time. This means the joint statistical distribution for l is the same as the joint distribution of $y_{t_1+l}, y_{t_2+l}, \ldots . y_{t_l+l}$ for h (see Nason 2006, Box and Jenkins 1976).

2.2 ARMA model

Yule (1926) studied the linear models of time series, particularly Autoregressive model (AR) of order p (Yule, 1926), whilst walker (1931) was introduced the general model of AR (Walker 1931). Slutsky (1937) was studied and put the general formula of moving average model (MA) of order z (Slutsky, 1937). Then these models were developed into mixed model by wold (1938) with an operation series into three directions in the procedure of estimation, which it called the

Autoregressive Moving Average model (ARMA) (Wold, 1938). It refers to this model by ARMA (p, z) where p and z are the order of the model. This model requires stationary data.

ARMA model is stochastic procedure that represents the correlation type between the time series and itself. However, it represents the continuity pattern of phenomena. It is the most common model of time series where it used in many sciences such as physical system, economics, medicine and other fields, where can be written as (Sampson, 2001):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_z \varepsilon_{t-z} \quad , \qquad (2.3)$$

where

ϕ_0 is the constant, $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots \theta_z$ are the coefficient of the model and $\varepsilon_t \sim$ (i.i.d). The process y_t is not stationary when θ and $|\phi| = \pm 1$ (Brockwell and Davis, 2002). In case of p=1 and z=1, we get the first-order of the model which also called ARMA process that can be written as:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad . \qquad (2.4)$$

The fit model and its order are determined by the autocorrelation function (ACF) and the partial autocorrelation function (PACF). When the autocorrelations decay exponentially to zero, it means that the model is an AR model. So the order model of AR is selected by a number of PACF that it significantly different from zero. We selected the MA model if the PACF decay to zero exponentially, whilst the order of this model determined by the number of ACF are significantly different from zero. The fit model is ARMA model if the ACF and PACF decay exponentially to zero, where it determined this model by AR & MA. However, it used to know the stability of the time series. The time series is non stationary, if the ACF function is decay very slowly with increasing delay periods l , so we need to take differences (Box and Jenkins 1976, Box and Jenkins 2008).

2.3 Autoregressive Distributed Lag (ADL)

The time series models as AR, Moving Average (MA) and ARMA models used to predict the time series. AR model is one of the most commonly model of time series that uses a linear combination of past values of the target to make forecasts. Moreover, MA is a common approach for modeling univariate time series, so the aim of the iterative estimation methods reduced to the loss function (Chatfield, 1995).

Consider the series y_t , $AR\ (p)$ model can be written as (Yaseen, 2019):

$$y_t = a_0 + \sum_{l=1}^{p} \beta_l L^l y_t + \varepsilon_t \ , \qquad (2.5)$$

where a_0 is the intercept, β_l is the regression coefficient for l^{th}, where $l = 0,1, \dots p,$ ε_t represents the white noise and L is the lag operator that is $L^0 y_t = y_t$, $L^1 y_t = y_{t-1}$. Sometimes the time series suffers from some problems that affect the accuracy of prediction as the inability of the past series y_{t-1} to describe the present series y_t. Fortunately, Pesaran &Shin (1998) suggested the Autoregressive Distributed Lag **(**ADL) model to overcome this problem by adding more explanatory variables $x_{j,t-l}\ (j = 1, \dots, q \ and \ l = 0, \dots., p_j)$ (Pesaran &Shin1998). The ADL model has the dependent lag variables and lagged independent variables and can be written as:

$$\mathrm{y}_t = \alpha + \sum_{l=0}^{p_0} \beta_{0,l} L^l y_t + \sum_{l=0}^{p_1} \beta_{1,l} L^l x_{1,t} + \dots. + \sum_{l=0}^{p_q} \beta_{q,l} L^l x_{q,t} + \varepsilon_t \ , \qquad (2.6)$$

where q and p_j represent the order of ADL model.

when α =0, the model can be written as (Yaseen 2019):

$$\mathrm{y}_t = \sum_{j=0}^{q} \sum_{l=0}^{p_j} \beta_{j,l} L^l x_{j,t} + \varepsilon_t \ . \qquad (2.7)$$

The formula of this model can be written as matrix (X) as follows:

$$Y = X\beta + \varepsilon \ , \qquad (2.8)$$

where

$$X = \begin{bmatrix} x_{0,\max(p_0,p_1,..,p_q)} & x_{0,(\max(p_0,p_1,..,p_q)-1)} & \dots & x_{0,(\max(p_0,p_1,..,p_q)-p_0)} \\ x_{0,(\max(p_0,p_1,..,p_q)+1)} & x_{0,\max(p_0,p_1,..,p_q)} & \cdots & x_{0,(\max(p_0,p_1,..,p_q)-p_0+1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{0,n^*-1} & x_{0,n^*-2} & \cdots & x_{0,n^*-p_0} \\ x_{1,\max(p_0,p_1,..,p_q)} & x_{1,(\max(p_0,p_1,..,p_q)-1)} & \dots & x_{1,(\max(p_0,p_1,..,p_q)-p_0)} \cdots \\ x_{1,(\max(p_0,p_1,..,p_q)+1)} & x_{1,\max(p_0,p_1,..,p_q)} & \cdots & x_{1,(\max(p_0,p_1,..,p_q)-p_0+1)} \cdots \\ \vdots & \vdots & \ddots & \vdots \quad \ddots \\ x_{1,n^*-1} & x_{1,n^*-2} & \dots & x_{1,n^*-2} \quad \dots \\ x_{q,\max(p_0,p_1,..,p_q)} & x_{q,(\max(p_0,p_1,..,p_q)-1)} & \dots & x_{q,(\max(p_0,p_1,..,p_q)-p_0)} \\ x_{q,(\max(p_0,p_1,..,p_q)+1)} & x_{q,(\max(p_0,p_1,..,p_q))} & \dots & x_{q,(\max(p_0,p_1,..,p_q)-p_0+1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{q,n^*-1} & x_{q,n^*-2} & \cdots & x_{q,n^*-p_q} \end{bmatrix}_{(n^* \times q)} . \qquad (2.9)$$

The dependent lag variables can be in matrix form:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n^*} \end{bmatrix}_{(n^* \times 1)} . \qquad (2.10)$$

The white noise is:

$$\hat{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{n^*} \end{bmatrix}_{(n^* \times 1)} , \qquad (2.11)$$

$$\beta = \begin{bmatrix} \beta_{0,0} \\ \beta_{0,1} \\ \vdots \\ \beta_{0,p_0} \\ \beta_{1,0} \\ \beta_{1,1} \\ \vdots \\ \beta_{0,p_1} \\ \vdots \\ \beta_{q,0} \\ \beta_{q,1} \\ \vdots \\ \beta_{q,p_q} \end{bmatrix}_{(q\times 1)} , \qquad (2.12)$$

where $n^* = n - \max(p_j)$. There are several assumptions such as $\beta_{0,0} = 0$ and $x_{0,t} = y_t$. As well as, the equation (2.7) must be satisfies the following assumptions: (Park &Sakaori, 2013):

1- $E\left(\varepsilon_t | y_{t-1}, y_{t-2}, \dots., x_{1,t}, x_{1,t-1}, \dots., x_{p,t-1}, x_{p,t-2}, \dots.\right) = 0$
2- Stationary between the variables (y_t) and $(x_{1,t}, \dots., x_{p,t})$
3- The correlation coefficient between $(y_t, x_{1,t}, \dots., x_{p,t})$ and $(y_{t-1}, y_{t-2}, \dots., x_{p,t-l})$ decrease with l increase.

2.4 ARCH/GARCH models

In financial return, volatility or variance is important of the asset returns for risk management. Since the volatility is implicit in return time series, so it needs to use one of the volatility estimation technique. Fortunately, economists have increased interest in the level of these volatility and their instability over time. Economists suggested models that the improvement of prediction accuracy of such volatility such as ARCH and GARCH. They are the linear models that used to study of the time-varying volatility for financial returns and it was modeled. The GARCH model is good to capture some the properties of time series such as the insignificant

ACF of the returns, the leptokurtosis distribution of the returns, the fluctuations clustering and so on (Takaishi, 2018). They exist in many types of the GARCH model, the first type is an EGARCH model presented by Daniel B. Nelson (1991) to capture the conditional variance but also ensure this variance is always positive, Integrated GARCH process (IGARCH) has to use when β is large that means the effect of the shocks is long. Threshold GARCH process (TGARCH) developed by Lawrence R. Glosten, Ravi Jagannathan and Davids E. Runkle (1993) to capture the positive and negative shocks. The difference between TGARCH and EGARCH models is, if the effect of the negative shocks are larger than the positive shocks will lead to the leverage effect, where the effect of the EGARCH model is greater than TGARCH model and the estimated of kurtosis of the TGARCH model is lager than EGARCH model (Kirchgässner and Wolters, 2007).

However, the stationarity property of GARCH models is $\alpha + \beta < 1$ and the considerations to calculate the unconditional variance of GARCH (s, m) is:

$$V[\varepsilon] = E[\varepsilon^2] = \frac{\alpha_0}{1 - \sum_{i=1}^{s} \alpha_i - \sum_{j=1}^{m} \beta_j} \ , \qquad (2.13)$$

and the kurtosis is :

$$K[\varepsilon] = 3 + \frac{6 \sum_{i=1}^{s} \alpha_i^2}{1 - \sum_{j=1}^{m} \beta_j - 2 \sum_{i=1}^{s} \alpha_i \sum_{j=1}^{m} \beta_j - 3 \sum_{i=1}^{s} \alpha_i^2} \ , \qquad (2.14)$$

where s, m represent the orders of GARCH model and α_i^2, β_j are the parameter of the model.

There are several stages of the construction of the ARCH and GARCH models, where can be mentioned as follows:

2.4.1 Identification

This stage is important in identification the model based on the data available by plotting the original series, making sure of the time series is stationary and

plotting of the basic characteristics of the time series such as autocorrelation function (ACF) and partial autocorrelation function (PACF). Moreover, to identify whether the time series follows the GARCH model, several tests are used with 5% the significant level:

2.4.1.1 Lagrange multiplier test

Engle (1982) proposed the Lagrange multiplier (LM) test to detect of conditional heteroscedasticity and also is called score statistic (Engle, 1982). It used for testing the null hypothesis as follows:

$$H_0: \alpha_i = 0 \qquad for\ i = 1,2, \ldots., s$$

$$H_1: \alpha_i \neq 0$$

It has a suitable form and can be expressed as n times the determination coefficient where it can be the test statistic as follows (Box, Jenkins, Ljung and Reinsel, 2016):

$$LM\ test = nR^2\ , \qquad (2.15)$$

where n represents the sample size, s represents the order of ARCH model and R^2 is the coefficient of determination. LM test depends on the χ^2 distribution with s degree of freedom under the null hypothesis of no ARCH effect. When the test statistic (LM) is large than χ^2 distribution with s degree of freedom, so the alternative hypothesis H_1 is to accept that means the ARCH effect is present.

2.4.1.2 Jarque-Bera test

Jarque and Bera (1980) introduced the normality test that can be applied directly on the time series itself or on its differences to test the normality of residuals. This test is based on the skewness and kurtosis that depends on the third and fourth moment. The test statistics can be written as follows (Kirchgässner and Wolters, 2007):

H_0: The residuals follow the normal distribution.

H_a: The residuals do not follow the normal distribution.

$$JB = \frac{n(s+m)}{6}\left(S^2 + {}^1\!/_4\,(K-3)^2\right), \qquad (2.16)$$

where $(s+m)$ represents the number of estimated parameters in the model, S represents the skewness and K represent the kurtosis are:

$$S = \frac{1}{n}\frac{\sum_{t=1}^{n}(y_t-\mu)^3}{\sqrt{\gamma(0)^3}},$$

$$k = \frac{1}{n}\frac{\sum_{t=1}^{n}(y_t-\mu)^4}{\sqrt{\gamma(0)^4}},$$

where $\gamma(0) = \frac{1}{n}\sum_{t=1}^{n}(y_t-\mu)^2$.

When $k = 3$ and $S = 0$ that means the distribution of residuals is normality. This statistic follows the Chi-squared distribution with a degree of freedom equal to two as $\chi^2_{1-\alpha(2)}$. When $JB \geq \chi^2_{1-\alpha(2)}$ or $p\_value \leq \alpha$ the null hypothesis H_0 is reject (Thadewald and Büning, 2004), so the residuals are not normality distribution.

2.4.1.3 Ljung-Box test

It is one of the statistical tests that are used to test the random time series errors to check the suitability of the time series model by calculating ACF for the residuals of a set of the lag. Ljung and Box (1978) was proposed this test to check the effect of the GARCH models, where the test statistics can be written as follows (Francq & Zakoian, 2010):

$H_0: \rho_1 = \rho_2 = \rho_3 = \cdots = \rho_k = \cdots \rho_r = 0 \qquad for\ k = 1,2,3,\ldots,r$

$H_a: \rho_k \neq 0$

$$Q_{(r)} = n(n+2)\sum_{k=1}^{r}\frac{\widehat{\rho_k^2}}{n-k} \sim \chi^2_{(r-c)}, \qquad (2.17)$$

where r represents the lags number of ACF, c represents the number of the estimated parameters and $\widehat{\rho_k^2}$ represents the ACF coefficients for the residuals of the series.

The test statistic $Q_{(r)}$ compares with the tabular value of the Chi-square test $\chi^2_{(r-c)}$ with a degree of freedom $(r-c)$ at a significant level $1-\alpha$. If it was $Q_{(r)} < \chi^2_{(r-c)}$ or $\_value \geq \alpha$, that means the null hypothesis H_0 is accepted. Therefore, there is no effect of the GARCH models, while $Q_{(r)} > \chi^2_{(r-c)}$ or $p\_value \leq \alpha$ that means the effect of the GARCH models is presented (Brockwell and Davis, 2002).

2.4.2 The Order Model Selection

In statistical modeling, the most important goal is to choose the appropriate model from a set of suggested models to characterize the basic data .If it chooses an order higher than the actual order of the model this leads to increase the number of its parameters. Thus this leads to increase the variance and loss of accuracy of the model. On the other hand, if it chooses a lower order than actual order this leads to inconsistency of the model parameters. Several criteria were established to choose the appropriate model by comparing the available models and choosing their order such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Hannan-Quinn (HQ) criterion (Francq & Zakoian, 2010).

2.4.2.1 Akaike Information Criterion

Akaike (1974) proposed AIC. It does not mean anything by itself, but rather it useful when compared to the values of AIC for other models with the same of datasets, where it selects the model that gives the lowest value of criterion. The general form of the AIC criterion can be written as follows (Tizro, Ghashghaie, Georgiou and Voudouris, 2014):

$$AIC(h) = nln\left(\widehat{\sigma_e^2}\right) + 2(h)\,, \qquad (2.18)$$

$$\widehat{\sigma_e^2} = \frac{1}{n-h}\sum_{t=1}^{h}(y_t - \hat{y}_t)^2 ,$$

where $\widehat{\sigma_e^2}$ is the variance of the residuals of the model and h represents the number of estimated parameters.

2.4.2.2 Bayesian Information Criterion

Akaike (1979 and 1978) adjusted the AIC criterion, the new criterion was called BIC which can be written as follows (Sampson, 2001):

$$BIC(h) = ln\left(\widehat{\sigma_e^2}\right) + \frac{\ln(n)(h)}{n} \quad . \qquad (2.19)$$

2.4.2.3 Hannan – Quinn criterion

Hannan and Quinn (1979) suggested a new criterion to determine the order of the model called Hannan - Quinn (HQ) which can be written as follow (Maïnassara and Kokonendji, 2016):

$$HQ\ (h) = ln\widehat{\sigma_e^2} + 2(h)\ c\ \ln(\ln n)/n \ , \quad (2.20)$$

where $c\ > 1$ is constant .

2.4.3 The Estimation of GARCH model

After identifying the appropriate model and ensuring that the time series follows the GARCH models, the parameters of this model that treats the heterogeneity of variance are estimated using several common methods such as Maximum Likelihood Method (MLE). It is one of the estimation methods that used by (Engel 1982) to estimate the ARCH model and then Bollerselve (1986) used to estimate the GARCH models as following (Bollerselve, 1986):

$$f(\varepsilon_t\,|F_{t-1}) = \frac{1}{\sqrt{2\pi{\sigma_t}^2}}\exp\left(-\frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}\right), \qquad \varepsilon_t \sim N(0,\sigma_t^2) \qquad (2.21)$$

Let $\theta = (\alpha_0, \alpha_1, \ldots\ldots, \alpha_s, \beta_1, \ldots\ldots, \beta_m)$,

Let $\acute{Z}_t = (1, \varepsilon_{t-1}, \ldots\ldots, \varepsilon_{t-s}, \sigma_{t-1}^2, \ldots\ldots, \sigma_{t-m}^2)$,

The log likelihood function for a sample of n observations:

$$L(\theta) = n^{-1} \sum_{t=1}^{n} I_t(\theta) , \qquad (2.22)$$

where

$$I_t(\theta) = -\frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\left(\frac{\varepsilon_t^2}{\sigma_t^2}\right) , \qquad (2.23)$$

So the partial derivative of (I_t) as follows:

$$\frac{\partial I_t}{\partial \theta} = \frac{1}{2\sigma_t^2}\frac{\partial L}{\partial \theta}\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right) , \qquad (2.24)$$

$$\frac{\partial^2 I_t}{\partial \theta \partial \acute{\theta}} = \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right)\frac{\partial}{\partial \acute{\theta}}\left[\frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\right] - \frac{1}{(\sigma_t^2)^2}\left(\frac{1}{2} - \frac{\varepsilon_t^2}{\sigma_t^2}\right)\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \acute{\theta}} , \qquad (2.25)$$

as for the asymptotically normal is :

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, I_{\acute{\theta}\theta}^{-1})$$

$$I_{\acute{\theta}\theta} = -E\left[\frac{\partial^2 I_t}{\partial \theta \partial \acute{\theta}}\right] ,$$

Hence, the information matrix that is negative expectation of the Hessian average for all observations,

$$I_{\acute{\theta}\theta} = -\frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial^2 I_t}{\partial \theta \partial \acute{\theta}}\right] ,$$

so

$$I_{\acute{\theta}\theta} = -\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{1}{\sigma_t^2} Z_t \acute{Z}_t\right) , \qquad (2.26)$$

To obtain the maximum likelihood estimations by using an iterative methods such as Newton Raphson, let θ_j is the parameter estimates after j^{th} iterative and θ_{j+1} computed as follows:

$$\theta_{j+1} = \theta_j + I_{\acute{\theta}\theta}^{-1}(\theta_j)\frac{\partial L}{\partial \theta}(\theta_j) \ . \qquad (2.27)$$

In the same way, the GARCH (1, 1) model is estimated, where information vector θ is as follows:

$$\theta = (\alpha_0, \alpha_1, \beta_1) \ ,$$

The conditional log-likelihood function with n observations for parameters $(\alpha_0, \alpha_1, \beta_1)$ are:

$$L = (\alpha_0, \alpha_1, \beta_1) = \sum_{t=1}^{n} I_t \ \ ,$$

$$I_t = -\frac{1}{2}\ln(\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2) - \frac{1}{2}\frac{\varepsilon_t^2}{\alpha_0+\alpha_1\varepsilon_{t-1}^2+\beta_1\sigma_{t-1}^2} \ ,$$

$$\frac{\partial I_t}{\partial \alpha_0} = \frac{1}{2(\alpha_0+\alpha_1\varepsilon_{t-1}^2)}\left(\frac{\varepsilon_t^2}{\alpha_0+\alpha_1\varepsilon_{t-1}^2+\beta_1\sigma_{t-1}^2} - 1\right), \qquad (2.28)$$

$$\frac{\partial I_t}{\partial \alpha_1} = \frac{1}{2(\alpha_0+\alpha_1\varepsilon_{t-1}^2)}\varepsilon_{t-1}^2\left(\frac{\varepsilon_t^2}{\alpha_0+\alpha_1\varepsilon_{t-1}^2+\beta_1\sigma_{t-1}^2} - 1\right), \qquad (2.29)$$

$$\frac{\partial I_t}{\partial \beta_1} = \frac{1}{2(\alpha_0+\alpha_1\varepsilon_{t-1}^2)}\sigma_{t-1}^4\left(\frac{\varepsilon_t^2}{\alpha_0+\alpha_1\varepsilon_{t-1}^2+\beta_1\sigma_{t-1}^2} - 1\right), \qquad (2.30)$$

$$\frac{\partial I_t}{\partial \theta} = \sum_t \frac{1}{2\sigma_t^2} Z_t\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right) \ \ ,$$

$$I_{\acute{\theta}\theta} = -E\left[\frac{\partial^2 I_t}{\partial\theta\partial\acute{\theta}}\right] = \begin{bmatrix} I_{\alpha_0\alpha_0} & I_{\alpha_0\alpha_1} & I_{\alpha_0\beta_1} \\ I_{\alpha_1\alpha_0} & I_{\alpha_1\alpha_1} & I_{\alpha_1\beta_1} \\ I_{\beta_1\alpha_0} & I_{\beta_1\alpha_1} & I_{\beta_1\beta_1} \end{bmatrix},$$

Element of Hessian matrix is:

$$\frac{\partial^2 I_t}{\partial {\alpha_0}^2} = -\frac{1}{2\sigma_t^2}\left(\frac{2\varepsilon_t^2}{\sigma_t^2} - 1\right),$$

$$\frac{\partial^2 I_t}{\partial {\alpha_1}^2} = -\frac{1}{2\sigma_t^2}\ \varepsilon_{t-1}^4 \left(\frac{2\varepsilon_t^2}{\sigma_t^2} - 1\right),$$

$$\frac{\partial^2 I_t}{\partial {\beta_1}^2} = -\frac{1}{2\sigma_t^2}\ \sigma_{t-1}^4 \left(\frac{2\varepsilon_t^2}{\sigma_t^2} - 1\right),$$

$$\frac{\partial^2 I_t}{\partial \alpha_0 \alpha_1} = -\frac{1}{2\sigma_t^2}\ \varepsilon_{t-1}^2 \left(\frac{2\varepsilon_t^2}{\sigma_t^2} - 1\right), \qquad (2.31)$$

$$\frac{\partial^2 I_t}{\partial \alpha_0 \beta_1} = -\frac{1}{2\sigma_t^2}\ \beta_{t-1}^2 \left(\frac{2\varepsilon_t^2}{\sigma_t^2} - 1\right), \qquad (2.32)$$

$$\frac{\partial^2 I_t}{\partial \alpha_1 \beta_1} = -\frac{1}{2\sigma_t^2}\ \varepsilon_{t-1}^2 \sigma_{t-1}^2 \left(\frac{2\varepsilon_t^2}{\sigma_t^2} - 1\right), \qquad (2.33)$$

the element of information matrix is :

$$I_{\alpha_0 \alpha_0} = \frac{1}{2n}\ \sum_t \frac{1}{\sigma_t^4}, \qquad (2.34)$$

$$I_{\alpha_0 \alpha_1} = \frac{1}{2n}\ \sum_t \frac{\varepsilon_{t-1}^2}{\sigma_t^4}, \qquad (2.35)$$

$$I_{\alpha_0 \beta_1} = \frac{1}{2n}\ \sum_t \frac{\beta_{t-1}^2}{\sigma_t^4}, \qquad (2.36)$$

$$I_{\alpha_1 \alpha_1} = \frac{1}{2n}\ \sum_t \frac{\varepsilon_{t-1}^4}{\sigma_t^4}, \qquad (2.37)$$

$$I_{\alpha_1 \beta_1} = \frac{1}{2n}\ \sum_t \frac{\sigma_{t-1}^2 \varepsilon_{t-1}^2}{\sigma_t^4}, \qquad (2.38)$$

$$I_{\beta_1 \beta_1} = \frac{1}{2n}\ \sum_t \frac{\sigma_{t-1}^4}{\sigma_t^4}, \qquad (2.39)$$

$$I_{\alpha_1 \alpha_1} = \frac{1}{2n}\ \sum_{i=1}^{n} \frac{\varepsilon_{t-1}^4}{\sigma_t^2}\ , \qquad (2.40)$$

then substituting them in the formulas (2.27) to obtain the estimations .

2.4.4 Diagnostic Checking Model

After estimating the parameters of the model and before using the model to calculate forecasts of the future, it must be tested to ensure the suitability or validity

of the model to represent the time series studied through two methods to check the accuracy of the model:

The first method is by plot the values of the ACF of the standard residual series and plot the ACF values of the standard square residual series according to the following formula (Shumway and stoffer, 2011):

$$\widetilde{\varepsilon_t} = \frac{\widehat{\varepsilon_t}}{\widehat{\sigma_t}} \quad , \qquad (2.41)$$

where $\widetilde{\varepsilon_t}$ represents the series of standardized residuals ,$\widehat{\varepsilon_t}$ represents the residuals series and $\widehat{\sigma_t}$ represents the conditional standard deviation for residuals.

The residuals formula is calculated as $\widehat{\varepsilon_t} = y_t - \widehat{y_t}$ for the studied models, where y_t represent the time series and $\widehat{y_t}$ is the predictive value of y_t. For conditional standard deviation series, it is calculated from the square root of the variance equation for the studied models after estimated of the parameters. These plot can be clear whether if the two series are normality distributed or not. If the model fits well, it means that the standardized residuals are randomly distributed with mean zero and variance one (Shumway and stoffer, 2011).

The second method is by using the Ljung Box test of the residuals that was previously used in the identification section, it used to show the suitability and then forecasting for the GARCH models. This means that the model is sufficient to predict (Tsay, 2002).

2.5 Variables selection

In regression, the variables selection is one of the important techniques which we needs to be insight of variable importance, where the analysis is parried out on the whole sample (Garg 2012). It used to reduce the problem of bias when the number of independent variables is large, which is an important issue in the statistics. It is well known that the independent variables can affect the dependent

variable, and this effect may be minimal or not affect the dependent variable. On the other hand, variable selection is to find good set of predictors that can improve the predictive ability of model and interpretability, so statistician's interest is to use different methods such as subset selection and stepwise selection. Although they are very useful practically, but these selection processes disregard the random errors transmitted in the variable selection stage (Fan &Li 2001).

The predictive ability of the improved model should not be much affected by small and large changes in the data (Hesterberg et al., 2008). Unfortunately, its theoretical properties are difficult to understand. Furthermore, there are several disadvantages in selection of the best subset such as instability (Fan &Li 2001). There are several criteria used the variable selection methods, most of methods depend on the Sum of Squares Errors (SSE) , it can be written as follows:

$$SSE = \sum_{t=1}^{n} (y_t - \hat{y_t})^2 \qquad ,t = 1,2,\dots.,n \qquad (2.42)$$

equivalent to

$$\|y - \hat{y}\|^2$$

where $\hat{y}$ is the predictive value of the dependent variable at time t.

2.6 Regression trees (RT)

Linear regression analysis has been widely used in many research fields. It is the linear relation between one response variable with one or more than one independent variables. The target of regression analysis is to get accuracy prediction model that can be interpreted the underlying phenomena. However, the performance linear regression will be very poor when there is an interaction between the independent variables or nonlinear relationships. Regression trees method has been proposed in the statistical literature by Breiman and et al (1984)

to overcome the complicate problem. It is a variable selection procedure that is employed the idea of decision trees in which consists of internal and terminal nodes where each node contains a statement that is either true or false. It is grown as a binary tree where each node can be split to sub-branch nodes and so on. It used a classification and regression trees (CART) algorithm (Ekici, 2014). CART is one of the machine learning algorithms that is characterized by flexibility and ease of interpretation, but does not avoid overfitting, which causes weakness in prediction.

Moreover, the algorithm of RT start with a root node splits into child nodes and then these child node splits into sub-branches on the right and left under conditions that must be achieved and then access to the terminal nodes are called leaves where making decision or prediction. The prediction at terminal nodes a is computed by (Garg, 2012):

$$m_a = \frac{1}{n_a} \sum_{i \in a} y_i \ , \qquad (2.43)$$

where $n_a$is the number of the observations in terminal nodes, y_i is the response variable. To build a function at the end of the regression tree, it must have a constant m_i as follow:

$$g(x) = \sum_{i=1}^{a} m_a \, I(x \in R_a) \, , \qquad (2.44)$$

where $I(.)$ represents the indicator function. The goal of decision trees is to get the best split and reduce the sum of the squared errors for a tree 'T' that can be calculated as:

$$MSE = \frac{1}{n} \sum_{i \in a} (y_i - m_i)^2 \, . \qquad (2.45)$$

The rpart package is widely used in R library to estimate and predict of the RT, where the growth of the tree needs several the tuning parameters such as cp and minsplit that can be defined as follows (Breiman 1984):

1. **cp**: is the threshold parameter that minimizes deviance when a split is made.

2. **minsplit**: represents the minimum observations for the node that it will the split.

2.7 Random Forests (RF)

The random forests method was introduced by Breiman (2001). It is considered as a one of machine learning techniques. It is more familiar than regression trees due to its using for large sample sizes with large number of variables, so it is a variable selection method (Breiman, 2001). Moreover, predictive performance of RF is poor in case of few variables, so it is difficult to determine the variables to be inserted into the model (Garg, 2012). RF method combines many regression trees in one algorithmic form that similar to the CART algorithm in building trees.

The idea of RF algorithm relies on resampling technique, firstly is to sampling bootstrap sample at random with replacement and then select the predictor variables at random without replacement to determine the root node as mentioned in regression trees. The number of predictor variables that selected at previous step should be less than the total number of it. Similar to this procedure have remaining variables to create the second node and so on. Resampling another dataset with replacement and repeat the same procedure to build the second regression trees and so on (Ekici, 2014). The objective function should be determined to do deciding which node must be split. After creating B regression trees, predictor of the random forest is:

$$\widehat{f_{rf}^{B}} = 1/B \sum_{b=1}^{B} T_b(X) \quad , \qquad (2.46)$$

If the trees are grown adaptive, it will be reduced the variance and correlation among the trees and in addition the average is useful to remove bias. The advantage of the random forests uses of multiple trees to reduce the risk of over fitting, it's relatively robust to outlier data, it uses the internal estimates to give useful result of error, correlation and strength and it's simple and easy (Breiman 2001). In the R package, the random forest () is carried out the estimations and predictions with RF, where the tuning parameters of this package that can be defined as follows (Breiman, 2001):

1. **ntree**: represents the number of trees will be constructed to grow the random forests, the default of ntree is 500 .

2. **mtry**: represents the number of predictors that must be employed at each node. In regression, the default of mtry is p/3 where p is the number of predictors.

2.8 Support Vector Regression (SVR)

Vapnik et al. (1992) proposed Support Vector Regression (SVR) for classification and regression. It is considered as a one of an important supervised machine learning method. It is a will-known algorithm with characters' simplicity which parry over fit to the data. It is one of the prediction methods that work out the machine learning to be high predictive accuracy. SVR deals with two different data and its task of separating classes in feature space. The basic idea of SVR gets the optimal hyper plane that it selected from set of the hyper plane (Evgeniou and Pontil, 2014).

SVR algorithm is carried out by hypothesis dimensional space that was closest to the points that the coordinates are support vector so that the distance between the hyper plane edge and hypothesis space margin are called marginal and have to be equals to both sides (Ashfaq and et. al 2013). It uses the kernel function to

implicitly set input data in a high-dimensional space. The kernel function is used when data cannot be separated linearly. The best way to choose the appropriate function to get the best solution to the problem is through trial and error for different kernel functions and its examples such as linear, nonlinear, polynomial, Gaussian kernel, Radial Basis Function (RBF), sigmoid etc (Garg, 2012). Recently, SVR has been applied in regression analysis by employ the Ɛ-insensitive loss function that can be written by (Awad and Khanna, 2015);

$$L_{\varepsilon}\big(y, f(x,w)\big) = \begin{cases} 0 & |y - f(x,w)| \leq \varepsilon; \\ |y - f(x,w)| & othereise\ , \end{cases} \qquad (2.47)$$

$$f(x,w) = w^T x + b \qquad (2.48)$$

where w is the weight vector and b is the bias. By using the above function, the errors that less than the threshold ε, will be ignored.

The optimal of hyper plane has maximized the margin. In spite of the difficulty of interpretation and reduced vision about the variables that contribute significantly during the prediction, but it is good model to predict the returns. However, it uses the Structural Risk Minimization (SRM) to reduce maximum risk while it uses the Empirical Risk Minimization (ERM) minimizing error in training data. This well make the SVR avoid the overfitting easily (Awad and Khanna, 2015).

The R package named e1071 parry out SVR where the following tuning parameters:

1. **kernel**: determine the kernel type to be employed such as linear, radial, polynomial, sigmoid.

2. **epsilon**: represents the epsilon in the loss function of Ɛ-insensitive, where the default value of eps is 0.1.

2.9 The Least Absolute Shrinkage and Selection Operator (LASSO)

consider the linear regression model:

$$y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots + X_P\beta_P + \varepsilon \quad , \qquad (2.49)$$

where $X_1, X_2, \ldots\ldots X_P$ are the independent variables, β_0 is the intercept value, $\beta_1, \beta_2, \ldots, \beta_P$ are the regression coefficients, y is the response variable and ε is the white noise. However, this model similar to the time series model that it shown in equation (2.6). It is well known that the Ordinary Least Squared (OLS) is the best linear unbiased estimator, when its assumptions are met. Unfortunately, in real world these assumptions may be violated such as the collinearity between two or more than two variables are considered, where the assumption is assumed the independently among the explanatory variables. The OLS results will be poorly when a multicollinearity is present. So the regularization techniques were introduced to improve the OLS (Tibshirani, 1996). For example, ridge regression was proposed by Hoerl and Kennard (1970) to overcome this problem. The aim of the ridge regression is to minimize the MSE that depends on the penalty function which l_2 – norm ($\sum_{j=1}^{p}\left|\beta_j\right|^2$) a part by shrinking the coefficient and giving the unimportant variables a value close to zero. Moreover, the performance prediction of the ridge regression is better than the OLS due to MSE of the ridge regression smaller than MSE of the OLS method (Kibria and Banik, 2016). The ridge regression always keeps all the independent variable (important and unimportant). The ridge estimator can be defined as follows: -

$$\widehat{\beta_{ridge}} = \arg min_{\beta \in R^p} \left\{ \|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} \left\|\beta_j\right\|_2^2 \right\}, \qquad (2.50)$$

where λ is the threshold.

Thus, it is not variable selections. In 1990's, the penalized methods were proposed to solve the ridge regression problem. Tibshirani (1996) suggested the Least

Absolute Shrinkage and Selection Operator (LASSO) method that used in high dimensional data. The aim of the LASSO carry out shrinkage and variable selection to estimate regression coefficient by depending on penalty function. LASSO performs coefficient shrinkage in general and important variables selection in particular depend on the sum of absolute value$\sum_{i=1}^{k} |\beta_i|$. It forces some coefficient to be zero. The LASSO coefficients are estimated by minimizing $\sum_{j=1}^{k} |\beta_j|$. The LASSO estimator can be defined as follows (Rajaratnam et al., 2019):

$$\widehat{\beta_{LASSO}} = \arg min_{\beta \in R^p} \left\{ \|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \quad . \qquad (2.51)$$

General LASSO path is inconsistent where the probability mass is greater than zero, even if the whole path contains the true parameter value or not, it cannot be achieved by using prediction accuracy as the selection criterion (leng et al. 2006) and do not have oracle properties.

2.10 Adaptive LASSO (ALASSO)

Zou (2006) showed that the lasso estimator may be inconsistent and then it is not satisfied the unbiasedness property due to the positive probability mass at 0. Moreover, he proposed adaptive LASSO that selects the exact nonzero coefficients with probability converging to 1 by depending on weights. The estimators of these coefficients are asymptotically normal with the same means and covariance. ALASSO can be solved by the same efficient algorithm for solving the LASSO. The ALASSO estimator can be written as follows (Zou, 2006):

$$\widehat{\beta_{Ad\ LASSO}} = \arg min_{\beta \in R^p} \left\{ \|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j| \right\} , \qquad (2.52)$$

where w_i is a known weight vector.

2.11 Elastic net Method

Zou and Hastie (2005) proposed Elastic Net (EN) method. EN is one of the variable selection methods and new regularization. It combined the penalty function of the LASSO and the penalty function of the ridge, the EN estimator can be defined by the following (Hastie and Zou, 2005):

$$\widehat{\beta_{net}} = \arg min_{\beta \in R^p} \{\|Y - X\beta\|_2^2 + \lambda(\alpha \|\beta_i\|_2^2 + (1-\alpha)\|\beta_i\|) \} , \qquad (2.53)$$

where $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$ and $\lambda = \lambda_1 + \lambda_2$

When $\alpha = 1$ the elastic net equivalent the ridge regression.

When $\alpha = 0$ the elastic net equivalent the LASSO.

The elastic net algorithm obtained shrinkage and variable selection of the coefficient simultaneously, elastic net method exhibits a grouping effect where strongly correlated between the predictors. Thus, the Predictive performance of the elastic net better than LASSO. The results of the elastic net has a smaller MSE than the LASSO (Matthew and Yahaya, 2015).

2.12 Selection of penalty parameter

The selection of penalty parameter is important step in the penalization technique such as LASSO, ALASSO and EN models. The penalty parameter is controlling the amount of shrinkage of estimates and yields variable selection, sometimes it is called threshold λ (Hastie et al. 2001) and tuning parameter.

If the penalty of parameter equal to zero, the estimates of the regularization techniques similar to the estimates of OLS, whilst the threshold is large enough $\lambda \to \infty$, it leads to exaggerated results in term of the variable selection and shrinkage, so all the coefficients forced to be zero. This is disagreement with the small value of the penalty of parameter, where it affects the amount of shrinkage of the coefficient and variable selection. It means some non-useful variables are

appeared and the amount of shrinkage is minimizing (Sartori 2011). There are several criteria used to select the penalty parameter.

2.12.1 Cross -Validation Criterion

One of the most used criteria by statisticians to select the penalty of parameter is CV criterion. It is used to assess validity and strength of predictive by estimating the Predictive Mean Square Errors of models ($PMSE$).

K-fold cross validation is common technique for assessing predictive models through estimating the $PMSE$, where a data sets $D = (x_t, y_t)$ divided into K of set, including the training sets and test sets. K-fold cross validation is frequency procedure where each time is removed the part of D to obtain the model $f^{-k}(x, \lambda)$ and estimate $PMSE$ as follows:

$$CV(\lambda) = \frac{1}{n}\sum_{k=1}^{K}\sum_{t\in C_k}\left\|y_t - f^{-k}(x_t, \lambda)\right\|^2 \qquad (2.54)$$

where C_k is an indicator observation in the each part of K , λ is the threshold and select the tuning parameter that makes at its the lower value of $CV(\lambda)$ (Tibshirani ,R .J ., & Tibshirani, R. 2009).

2.12.2 Generalized Cross - Validation criterion

The GCV criterion is also an important for selecting the threshold, where can be defined the formula as follows: (Fan & Li 2001)

$$\mathrm{GCV}(\lambda) = \frac{\left\|Y - X\beta_\lambda\right\|^2}{n\left(1 - \frac{df(\lambda)}{n}\right)^2} \qquad (2.55)$$

where $df(\lambda)$ is the degree of freedom. The threshold which has the least value of $\mathrm{GCV}(\lambda)$ criterion is selected.

2.12.3 Bayesian information criterion (BIC)

BIC criterion is widely used to choose the tuning parameter λ because of its ease in the computational. It is also called Schwarz criterion that can be calculated as follows: (Alfons et al 2013)

$$\mathrm{BIC}(\lambda) = \log(\widehat{\sigma^2}) + df(\lambda)\frac{\log(n)}{n} \qquad (2.56)$$

The penalty of parameter λ that it has the least value of $\mathrm{BIC}(\lambda)$ criterion is selected.

2. 13 Root Mean Square Errors (RMSE)

The standard statistical criterion of this thesis is the Root of Mean Square Errors (RMSE) that it used to measure the predictive performance of the methods by the errors. The random error of n samples size is calculated as:

$$\varepsilon_t = y_t - \widehat{y_t} \qquad (2.57)$$

Also we suppose the sample set of error is unbiased. Therefore the RMSE is computed for the data set, where can be written as: (Chai & Draxler, 2014)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2} \qquad (2.58)$$

2. 14 ALASSO-SVR and EN-SVR Algorithms

The algorithm of the regularized methods (ALASSO -SVR and EN-SVR) can be written as follows:

1. Split the original data set into two data sets, training (D_{tr}) and testing (D_{ts}) sets respectively.

2. Determine the important variables of D_{tr} set in Equ (2.36) by using the ALASSO -SVR and EN-SVR methods.

3. Using these variables in the SVR function to estimate the model:

$$\widehat{Y_t}^{(c)} = f\left(\Theta, \mathrm{X_t}^{(c)}\right) \qquad (2.59)$$

where $\mathrm{X_t}^{(c)}$ is the matrix of the important variables that selected in step (2) and $\widehat{Y_t}^{(c)}$ is the time series represents that affected by the important variables. Then the residual ε_t of this model is computed where $\varepsilon_t = y_t - \widehat{y_t}$.

4. Estimate the time varying variance of GARCH (1, 1) by using ε_t from step 3 as follows (Garg, 2012):

$$\sigma_t^2 = a_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \qquad (2.60)$$

and then the standard devision of conditional variance $\sqrt{\sigma_t^2}$ is estimated, where $t = 1,2, \ldots, n$.

5. Implement the transformation of Y_t by using $\sqrt{\sigma_t^2}$ from step 4 to reduce the fluctuations clustering effect on $\widehat{Y_t}^{(c)}$ as follows:

$$Y_t^* = \widehat{Y_t}^{(c)} \Big/ \sqrt{\sigma_t^2} \qquad (2.61)$$

6. Using Y_t^* to obtain the adjusted returns and repeat the steps 1-5 depending on the new returns, then the RMSE is computed for them to obtain accurate results.

Chapter three

The simulation part

3.1 Introduction

In this chapter, we carried out the simulation study to compare the performance of the methods that are mentioned in the theoretical part by using R programing. Each simulated dataset is divided into (30%) training and (70%) testing sets**.** First, we generate three models that are GARCH (1, 0), GARCH (1, 1) and GARCH (1, 2) with different sample size. Random error term is generated from standard normal distribution and attached with each model.

Model_1 $\sigma_t^2 = 0.05 + (0.6362)\varepsilon_{t-1}^2$ (3.1)

Model_2 $\sigma_t^2 = 0.006 + (0.3877)\varepsilon_{t-1}^2 + (0.5752)\sigma_{t-1}^2$ (3.2)

Model_3 $\sigma_t^2 = 0.1 + (0.2)\varepsilon_{t-1}^2 + (0.3)\sigma_{t-1}^2 + (0.1)\,\sigma_{t-2}^2$ (3.3)

Each of these GRACH models is employed to generate y_t in the ADL model with fixed order (5, 3, 3, 5, 5) as shown in equation (3.1):

$$y_t = \alpha + \sum_{l=0}^{5} \beta_{(0,l)} y_{(t-l)} + \sum_{l=0}^{3} \beta_{(1,l)}\, x_{(1,t-l)} + \ldots. + \sum_{l=0}^{5} \beta_{(5,l)}\, x_{(5,t-l)} + \varepsilon_t \quad , \qquad (3.4)$$

where $(x_1, x_{2,\ldots,} x_5)$ are generated from standard normal distribution. We used the RMSE criterion to select the best method, where the method that has the least value of RMSE criterion is the best.

3.2 Simulation steps

The steps are carried out of the experiment as the following:

1. Generate the time series

In order to generate the data that are stable and clean, we used the model which is described in the equation (3.4) with default parameters are:

Table (3.1) The default parameter values

cases	Standard devision (σ)	n	True Parameters
Sparse case	1, 3, 6	150, 200, 250	$\beta = (\{\beta_{0,1}, \dots, \beta_{4,4}\} = 7, \{\beta_{4,5}, \dots, \beta_{5,5}\} = 0)$
Very sparse case	1, 3, 6	150, 200, 250	$\beta = (\{\beta_{0,1}, \dots, \beta_{3,1}\} = 5, \{\beta_{3,2}, \dots, \beta_{5,5}\} = 0)$

2. Generating the error term

The error term was generated by the normal distribution with zero mean and standard devision ($\sigma = 1,3,6$)

3. Computation the response variable

The response variable was computed by the following:

$$Y = X\beta + \varepsilon$$

4. Replication of the experiment 1000 times to obtain the results is stable

5. Computation the RMSE criterion

The RMSE criterion is computed and then comparable among results.

3.3 The results

In this section, we compare between predictive performance of the machine learning models and regularization methods. After completing the implementation of the program, the following results were obtained depending on different parameter of GARCH models.

In this experiment, we choose the optimal set of penalty parameter $(\lambda_1, \lambda_2, \lambda)$ by k-fold cross validation, where $k = 3,3,7$ respectively.

3.3.1 The result of sparse case

When the data is generated by model as shown in equation (3.4)with the default values of parameters as follows:

Table(3.2). The values of parameters (sparse case)

y_{t-1}	$x_{1,t-1}$	$x_{2,t-1}$	$x_{3,t-1}$	$x_{4,t-1}$	$x_{5,t-1}$
$\{\beta_{0,1}, \dots, \beta_{0,5}\} = 7$	$\{\beta_{1,0}, \dots, \beta_{1,3}\} = 7$	$\{\beta_{2,0}, \dots, \beta_{2,3}\} = 7$	$\{\beta_{3,0}, \dots, \beta_{3,3,}\} = 7$	$\{\beta_{4,0} = 5, \beta_{4,1}, \dots, \beta_{4,4}\} = 7$, $\beta_{4,4} = 0$	$\{\beta_{5,0}, \dots, \beta_{5,5}\} = 0$

3.3.1.1 GARCH (1, 0) models

After carried out of the program, the results will be obtained depending on the standard devision and parameters of the GARCH (1, 0) models $(a_0 = 0.05, a_1 = 0.6362$, as shown in the following table:

Table(3.3). The RMSE of methods when the model is GARCH (1, 0) over all 1000 simulation dataset (Sparse case).

Methods used	RMSE								
	Parameters	$\sigma = 1$			$\sigma = 3$			$\sigma = 6$		
		n=150	n=200	n=250	n=150	n=200	n=250	n=150	n=200	n=250
RT	-	1.3343	1.3582	1.3634	1.3442	1.3584	1.3616	1.3437	1.3546	1.3693
RF	-	1.1932	1.2353	1.2341	1.1861	1.2043	1.2277	1.2025	1.2221	1.2513
LASSO	-	1.1797	1.1344	1.0971	1.2338	1.1638	1.1279	1.2301	1.1610	1.1303
ALASS0-SVR	kernel=linear	1.0248	1.0588	1.0670	1.0262	1.1638	1.0766	1.0361	1.0370	1.0835
	kernel=radial	0.9328	0.9536	0.9764	0.9300	1.0306	1.0747	1.0276	1.0425	1.0672
	kernel=polynomial	0.8920	0.9127	0.9411	0.9133	1.0422	1.0712	1.0326	1.0475	1.0714
	kernel=sigmoid	0.9005	0.9258	0.9511	0.9047	1.0306	1.0674	1.0393	1.0408	1.0776
EN-SVR	kernel=linear	0.9985	1.0030	0.9880	0.9934	0.9975	0.99218	1.0018	0.9955	1.0010
	kernel=radial	0.9093	0.9059	0.9048	0.9122	0.9867	0.9963	0.9991	0.9983	**0.9917**
	kernel=polynomial	0.8704	0.8675	**0.8671**	0.8901	0.9998	0.9967	1.0050	0.9989	0.9932
	kernel=sigmoid	0.8820	0.8795	0.8778	**0.8874**	0.9867	0.9907	1.0103	0.9946	0.9963

In sparse case, table (3.3) shows the results obtained when the model is GARCH (1, 0) where it can be illustrated as follows:

1. When assuming $\sigma = 1$ with $n = 150, 200, 250$, there is a preference for suggested methods based on the value of the RMSE with different kernel functions, particularly EN- SVR method has a smallest value of the RMSE (0.8671) with

polynomial kernel and $n = 250$, then LASSO, RF and RT methods. However, we note the results of the methods sometimes convergent with increase sample size.

2. When assuming $\sigma = 3$ with the same sizes as the previous samples, we note that the proposed methods are better than another methods with different kernel functions and with increase the sample sizes, we note that the results are non-stationary. However, EN- SVR method is better than the ALASSO- SVR method, where it has the smallest value of the RMSE (0.8874) with sigmoid kernel and $n = 150$.

3. When assuming $\sigma = 6$ with the same sizes as the previous samples, we note that the results of the methods sometimes are stationary, where our proposed methods outperformed the others methods, in particular EN- SVR method is very good compare to the ALASSO- SVR method because it has the smallest value of RMSE (0.9917) with radial kernel and $n = 250$. We conclusion the proposed methods better than other methods with large sample sizes.

3.3.1.2 GARCH (1, 1) models

After carried out of the program, the results will be obtained depending on the standard devision and parameter of the GARCH (1, 1) models ($a_0 = 0.006$, $a_1 = 0.3877$, $\beta_1 = 0.5752$) , as shown in the following table:

Table(3.4). The RMSE of method when the model is GACH (1, 1) over all 1000 simulation dataset (Sparse case).

Methods used	RMSE									
	Parameters	$\sigma = 1$			$\sigma = 3$			$\sigma = 6$		
		n=150	n=200	n=250	n=150	n=200	n=250	n=150	n=200	n=250
RT	-	1.2511	1.2585	1.2572	1.2495	1.2608	1.2625	1.2467	1.2689	1.2601
RF	-	1.0889	1.0997	1.1105	1.0792	1.1034	1.1100	1.0926	1.0998	1.1079
LASSO	-	2.9920	3.7178	3.2470	2.3288	1.8124	1.6076	1.6533	1.3688	1.2339
ALASS0-SVR	kernel=linear	0.9337	0.9482	0.9005	0.9270	0.9517	0.9809	0.9331	0.9509	0.9785
	kernel=radial	0.9331	0.9506	0.9694	0.9422	0.9450	0.9739	0.9433	0.9463	0.95863
	kernel=polynomial	0.9317	0.9430	0.9696	0.9284	0.9362	0.9789	0.9416	0.9510	0.95864
	kernel=sigmoid	0.9287	0.9504	0.9718	0.9406	0.9343	0.9793	0.9314	0.9508	0.9581
EN-SVR	kernel=linear	0.9140	0.9076	0.8954	0.9073	0.9075	0.9033	0.9101	0.9186	0.9085
	kernel=radial	0.9112	0.9058	0.8985	0.9230	0.9055	0.9030	0.9155	0.9052	0.88161
	kernel=polynomial	0.9072	0.9011	0.8938	0.9094	0.8936	0.9021	0.9146	0.9125	0.88162
	kernel=sigmoid	0.9155	0.9092	0.9052	0.91676	0.8970	0.9017	0.9173	0.9041	0.8817

In sparse case, table (3.4) displays the results obtained when the model is GARCH (1, 1) where it can be illustrated as follows:

1. When assuming $\sigma = 1$ with $n = 150, 200, 250$, we note that the results of our proposed methods are very good compare to another methods based on the value of RMSE with different kernel functions, particularly EN- SVR method has a smallest value of the RMSE (0.8938) with polynomial kernel and $n = 250$, then RF, RT and LASSO methods. However, we note the results of the methods sometimes convergent with increase sample size.

2. When assuming $\sigma = 3$ with the same sizes as the previous samples, we note that the proposed methods are better than another methods with different kernel functions and with increase the sample sizes, we note that the results sometimes non-stationary. However, EN- SVR method is better than the ALASSO- SVR method, where it has the smallest value of the RMSE (0.8936) with sigmoid kernel and $n = 200$.

3. When assuming $\sigma = 6$ with the same sizes as the previous samples, we note that the results of the methods sometimes are stationary, where our proposed methods

outperformed the others methods, in particular EN- SVR method is very good compare to the ALASSO- SVR method because it has the smallest value of RMSE (0.8816) with radial kernel and $n = 250$ where the value of radial kernel close to the value of the polynomial kernel. We conclusion the suggested methods are very good with large sample sizes.

3.3.1.3 GARCH (1, 2) models

After carried out of the program, the results will be obtained depending on the standard devision and parameter of GARCH (1, 2) models ($a_0 = 0.5$, $a_1 = 0.2$, $\beta_1 = 0.3, \beta_2 = 0.1$) , as shown in the following table:

Table(3.5). The RMSE of method when the model is GARCH (1, 2) over all 1000 simulation dataset. (Sparse case)

Methods used	RMSE									
	Parameters	$\sigma = 1$			$\sigma = 3$			$\sigma = 6$		
		n=150	n=200	n=250	n=150	n=200	n=250	n=150	n=200	n=250
RT	-	1.0618	1.5138	1.7798	1.1191	1.4094	1.7764	1.1258	1.5424	1.8076
RF	-	1.0666	1.0904	1.1114	1.0777	1.1089	1.1041	1.0836	1.0939	1.1079
LASSO	-	3.6736	3.3208	3.0922	1.8708	1.5958	1.4959	1.4687	1.2740	1.2607
ALASS0-SVR	kernel=linear	0.9400	0.9340	0.9344	0.9366	0.9387	0.9467	0.9389	0.9344	0.9432
	kernel=radial	0.9293	0.9340	0.9418	0.9419	0.9283	0.9346	0.9343	0.9409	0.9399
	kernel=polynomial	0.9373	0.9356	0.9416	0.9367	0.9427	0.9385	0.9385	0.9401	0.9464
	kernel=sigmoid	0.9422	0.9343	0.9321	0.9373	0.9389	0.9399	0.9466	0.9362	0.9432
EN-SVR	kernel=linear	0.9204	0.9097	**0.9056**	0.9221	0.9134	0.9172	0.9219	**0.9094**	0.9146
	kernel=radial	0.9108	0.9115	0.9159	0.9259	**0.9038**	0.9052	0.9153	0.9154	0.9132
	kernel=polynomial	0.9187	0.9135	0.9146	0.9165	0.9205	0.9093	0.9191	0.9168	0.9144
	kernel=sigmoid	0.9226	0.9122	0.9066	0.9206	0.9160	0.9139	0.9317	0.9123	0.9155

In sparse case, table (3.5) exhibits the results obtained when the model is GARCH (1, 2) where it can be illustrated as follows:

1. When assuming $\sigma = 1$ with $n = 150, 200, 250$, we note that the results of our proposed methods are very good compare to another methods based on the value of RMSE with different kernel functions, particularly EN- SVR method has a smallest value of the RMSE (0.9056) with linear kernel and $n = 250$, then RF, RT

and LASSO methods. However, we note the results of the methods sometimes convergent with increase sample size.

2. When assuming $\sigma = 3$ with the same sizes as the previous samples, we note that the proposed methods are better than another methods with different kernel functions and with increase the sample sizes, we note that the results sometimes non-stationary. However, EN- SVR method is better than the ALASSO- SVR method, where it has the smallest value of the RMSE (0.9038) with radial kernel and $n = 200$.

3. When assuming $\sigma = 6$ with the same sizes as the previous samples, we note that the results of the methods sometimes are stationary, where our proposed methods outperformed the others methods, in particular EN- SVR method is very good compare to the ALASSO- SVR method because it has the smallest value of RMSE (0.9094) with linear kernel and $n = 200$. We conclusion the proposed methods outperformed with large sample sizes.

3.3.2 The result of very Sparse

When the data is generated by model as shown in equation (3.4) with the default values of parameters as follows:

Table (3.6) The values of parameters (Very Sparse case)

y_{t-1}	$x_{1,t-1}$	$x_{2,t-1}$	$x_{3,t-1}$	$x_{4,t-1}$	$x_{5,t-1}$
$\{\beta_{0,1}, \dots, \beta_{0,5}\} = 5$	$\{\beta_{1,0}, \dots, \beta_{1,3}\} = 5$	$\{\beta_{2,0}, \dots, \beta_{2,3}\} = 5$	$\{\beta_{3,0}, \dots, \beta_{3,3,}\} = 5$	$\beta_{4,0} = 5, \{\beta_{4,1}, \dots, \beta_{4,5}\} = 0$	$\{\beta_{5,0}, \dots, \beta_{5,5}\} = 0$

3.3.2.1 GARCH (1, 0) models

After carried out of the program, the results will be obtained depending on the standard devision and parameter of GARCH (1, 0) models as $(a_0 = 0.05\ , a_1 = 0.6362)$ according to the table as follows:

Table(3.7). The RMSE of method when the model is GARCH (1, 0) over all 1000 simulation dataset (Very Sparse case).

Methods used	RMSE									
	Parameters	$\sigma = 1$			$\sigma = 3$			$\sigma = 6$		
		n=150	n=200	n=250	n=150	n=200	n=250	n=150	n=200	n=250
RT	-	1.2933	1.3182	1.3207	1.3315	1.3451	1.3259	1.3414	1.3382	1.3466
RF	-	1.1509	1.1626	1.1958	1.2390	1.1868	1.1530	1.1825	1.2137	1.2340
LASSO	-	1.1988	1.1388	1.1100	1.18880	1.1419	1.1081	1.1898	1.1415	1.1096
ALASS0-SVR	kernel=linear	1.0090	1.0329	1.0243	1.0705	1.0775	0.9808	1.0111	0.9997	1.0042
	kernel=radial	1.0213	1.0407	1.0335	1.0348	1.0363	1.0101	1.0084	1.0102	1.0025
	kernel=polynomial	1.0400	1.0325	1.0248	1.0541	0.9874	1.1082	1.0055	1.0054	1.0090
	kernel=sigmoid	1.0210	1.0872	1.0141	1.0562	1.0637	1.0250	1.0041	1.0044	1.0118
EN-SVR	kernel=linear	**0.9942**	1.0153	1.0043	1.0510	1.0361	**0.9200**	1.0072	0.9963	**0.9925**
	kernel=radial	1.0091	1.0167	1.0061	1.0216	1.0071	0.9408	1.0045	1.0027	0.9952
	kernel=polynomial	1.0149	1.0192	1.0048	1.0422	0.9355	1.0667	1.0018	1.0025	1.0011
	kernel=sigmoid	1.0083	1.0685	0.9943	1.0201	1.0238	0.9681	1.0012	0.9976	1.0039

In very sparse case, table (3.7) exhibits the results obtained when the model is GARCH (1, 0) where it can be illustrated as follows:

1. When assuming $\sigma = 1$ with $n = 150, 200, 250$, we note that the results of our proposed methods are very good compare to another methods based on the value of RMSE with different kernel functions, particularly EN- SVR method has a smallest value of the RMSE (0.9942) with linear kernel and $n = 150$, then LASSO, RF and RT methods. However, we note the results of the methods are different with increase sample size.

2. When assuming $\sigma = 3$ with the same sizes as the previous samples, we note that the proposed methods are better than another methods with different kernel functions and with increase the sample sizes, we note that the results sometimes stationary. However, EN- SVR method is better than the ALASSO- SVR method, where it has the smallest value of the RMSE (0.9200) with linear kernel and $n = 250$.

3. When assuming $\sigma = 6$ with the same sizes as the previous samples, we note that the results of the methods sometimes are non-stationary, where our proposed

methods outperformed the others methods, in particular EN- SVR method is very good compare to the ALASSO- SVR method because it has the smallest value of RMSE (0.9925) with linear kernel and $n = 250$. We conclusion the proposed methods outperformed with large sample sizes.

3.3.2.2 GARCH (1, 1) models

After carried out of the program, the results will be obtained depending on the standard devision and parameter of GARCH (1, 1) models ($a_0 = 0.006$, $a_1 = 0.3877$, $\beta_1 = 0.5752$) according to the table as follows:

Table(3.8). The RMSE of method when model is GARCH (1, 1) over all 1000 simulation dataset (Very Sparse case).

Methods used	Parameters	RMSE								
		$\sigma = 1$			$\sigma = 3$			$\sigma = 6$		
		n=150	n=200	n=250	n=150	n=200	n=250	n=150	n=200	n=250
RT	-	1.3264	1.3440	1.3442	1.3288	1.3377	1.3439	1.3253	1.3419	1.3487
RF	-	1.1972	1.2151	1.2229	1.19962	1.2144	1.2383	1.1832	1.2113	1.2343
LASSO	-	3.1536	2.1369	1.7204	1.4378	1.2834	1.2354	1.2595	1.1672	1.2207
ALASS0-SVR	kernel=linear	1.0034	1.0069	0.9997	1.0110	1.0019	1.0025	1.0019	1.0113	1.0053
	kernel=radial	1.0005	1.0106	1.0115	0.9971	1.0008	1.0064	0.9967	1.0028	1.0075
	kernel=polynomial	1.0089	1.0008	1.0065	1.0119	1.0151	1.0069	1.0091	1.0038	1.0061
	kernel=sigmoid	1.0088	1.0058	1.0123	1.0136	1.0012	1.0117	1.0104	0.9953	1.0038
EN-SVR	kernel=linear	0.9992	0.9976	0.9881	1.0063	0.9988	0.9897	1.0012	1.0080	0.9967
	kernel= radial	0.9978	1.0039	1.0033	0.9937	**0.9934**	0.9969	0.9925	0.9954	0.9992
	kernel=polynomial	1.0016	0.9929	**0.9923**	1.0066	1.0095	0.9981	1.0050	0.9993	0.9962
	kernel=sigmoid	1.0024	1.0003	0.9992	1.0117	0.9946	1.0011	1.0078	**0.9884**	0.9929

In very sparse case, table (3.8) displays the results obtained when the model is GARCH (1, 1) where it can be illustrated as follows:

1. When assuming $\sigma = 1$ with $n = 150, 200, 250$, we note that the results of our proposed methods are very good compare to another methods based on the value of RMSE with different kernel functions, particularly EN- SVR method has a smallest value of the RMSE (0.9923) with polynomial kernel and $n = 250$, then

RF, RT and LASSO methods. However, we note the results of the methods sometimes convergent with increase sample size.

2. When assuming $\sigma = 3$ with the same sizes as the previous samples, we note that the proposed methods are better than another methods with different kernel functions and with increase the sample sizes, we note that the results sometimes non-stationary. However, EN- SVR method is better than the ALASSO- SVR method, where it has the smallest value of the RMSE (0.9934) with radial kernel and $n = 200$.

3. When assuming $\sigma = 6$ with the same sizes as the previous samples, we note that the results of the methods sometimes are stationary, where our proposed methods outperformed the others methods, in particular EN- SVR method is very good compare to the ALASSO- SVR method because it has the smallest value of RMSE (0.9884) with sigmoid kernel and $n = 200$. We conclusion the suggested methods are very good with large sample sizes.

3.3.2.3 GARCH (1, 2) models

After carried out of the program, the results will be obtained depending on the standard devision and parameter of GARCH (1, 2) models ($a_0 = 0.5$, $a_1 = 0.2$, $\beta_1 = 0.3, \beta_2 = 0.1$) according to the table as follows:

Table(3.9). The RMSE of method when the model is GARCH (1, 2) over all 1000 simulation dataset (Very Sparse case).

Methods used	Parameters	RMSE								
		$\sigma = 1$			$\sigma = 3$			$\sigma = 6$		
		n=150	n=200	n=250	n=150	n=200	n=250	n=150	n=200	n=250
RT	-	1.2403	1.2534	1.2520	1.2513	1.2635	1.2616	1.2533	1.2879	1.2739
RF	-	1.0785	1.1113	1.1147	1.1135	1.1183	1.1209	1.1087	1.1248	1.1399
LASSO	-	2.4599	1.9148	1.6436	1.3523	1.2579	1.1864	1.2105	1.16377	1.1363
ALASS0-SVR	kernel=linear	0.9189	0.9225	0.9094	0.9372	0.9220	0.9144	0.9468	0.9390	0.9345
	kernel=radial	0.9238	0.9177	0.9102	0.9479	0.9261	0.9150	0.9494	0.9315	0.9251
	kernel=polynomial	0.9292	0.9178	0.9131	0.9293	0.9210	0.9161	0.9405	0.9316	0.9266
	kernel=sigmoid	0.9324	0.9175	0.9136	0.9269	0.9218	0.9164	0.9380	0.9313	0.9267
EN-SVR	kernel=linear	0.91938	0.9206	0.9075	0.9383	0.9240	**0.9129**	0.9481	0.9403	0.9351
	kernel=radial	0.92293	0.9187	**0.9066**	0.9458	0.9235	0.9146	0.9499	0.9318	**0.9250**
	kernel=polynomial	0.9304	0.9191	0.9076	0.9299	0.9225	0.9163	0.9421	0.9328	0.9251
	kernel=sigmoid	0.9351	0.9181	0.9119	0.9298	0.9212	0.9141	0.9377	0.9310	0.9310

In very sparse case, table (3.9) shows the results obtained when the model is GARCH (1, 1) where it can be illustrated as follows:

1. When assuming $\sigma = 1$ with $n = 150, 200, 250$, we note that the results of our proposed methods are very good compare to another methods based on the value of RMSE with different kernel functions, particularly EN- SVR method has a smallest value of the RMSE (0.9066) with radial kernel and $n = 250$, then RF, RT and LASSO methods. However, we note the results of the methods sometimes convergent with increase sample size.

2. When assuming $\sigma = 3$ with the same sizes as the previous samples, we note that the proposed methods are better than another methods with different kernel functions and with increase the sample sizes, we note that the results sometimes non-stationary. However, EN- SVR method is better than the ALASSO- SVR method, where it has the smallest value of the RMSE (0.9129) with linear kernel and $n = 250$.

3. When assuming $\sigma = 6$ with the same sizes as the previous samples, we note that the results of the methods sometimes are stationary, where our proposed methods outperformed the others methods, in particular EN- SVR method is very good compare to the ALASSO- SVR method because it has the smallest value of RMSE (0.9250) with sigmoid kernel and $n = 250$. We conclusion the suggested methods are better than other methods with increase sample sizes.

Chapter four

Applied part

4.1 Introduction

The exchange rate is important concept in economic and financial field where the exchange process shows when it excnanges between the local currency (IQD) and foreign currency (USD). The local currency is used for buying and selling process in the internal, whilst the foreign currency is used for making relationship or financial trade between the companies in the country and the companies abroad. Therefore these companies need to market to exchange the currency in order to obtain the foreign currency of exporting country. Moreover, each person travels abroad needs to exchange the currency with the currencies of countries that he travels, especially the tourists (Abu-Ahmad, 2002).

There are two types of exchange, the first type is the cash exchange and the second type is the exchange on credit or forward exchange. The cash exchange means that the value is paid directly or sometime period up to 48 hours from the moment implement the contract. In contrast, the second type means that the value is paid after a certain period from the date implement the contract, where the price of the exchange rate is calculated depending on the prevailing exchange rate from moment of conducting the contract. That means, the price of exchange rate at the moment of implementing the contract equals to the price of the prevailing exchange rate at the moment of sign the contract (Sami, 2007). Companies use forward exchange processes in order to avoid the risks in abroad trade which creates from the potential expected fluctuations in the exchange rates of currency. The exchange rate is continuously changing during the day that based on the supply and demand of the currencies. Moreover, there are two types of prices, the purchase price which is represents an exchange process the number units of the local currency versus for the purchase one unit of foreign currency by the bank. The selling price which represents an exchange process the number units of the local currency versus for the selling one unit of foreign currency by the bank, where the

selling price is always greater than the purchase price and the margins of the bank is the difference between them (Cordenm, 1994, 39).

In 1990's, economic conditions of Iraq is effecting on all economic indicators, especially the exchange rate of the Iraqi dinar against the US dollar. Moreover, the value of Iraqi dinar is started to decline and continued because several reason, external or internal where external reasons are included economic sanctions that is imposed on Iraq in 1991.This is led to the loss of the foreign exchange resources of Iraq as the dollar, consequently oil exports stopped. On the other hand, internal reasons included several condition that led to the decrease the Iraqi dinar value against the dollar as a result of an increase in the amount of money supplied through unlimited monetary issuance, high inflation rates, and failures of monetary policy and decrease the gross domestic product (Atto, 2002). In the year 2004 the authorities issued law No. 56 of the Central Bank for improvement and stability of the exchange rate of the Iraqi dinar, for instance organizing a public sale for dollars. However, it achieved the full autonomy in formulating and carry out the monetary policy by indirect methods. Therefore, this factors contributed to issue the new Iraqi currency exchange for the old currency where it gave the central bank the full control on the export process and reduces processes of counterfeiting which was instilling a negative effect on the value of the dinar (Annual Iraqi economic reports of Central Bank for the year 2004).

4.2 Describe the variables

Here, we will use the exchange rate (IQD/USD) dataset to know the performance of methods. These dataset was collected from Central Bank of Iraq, where it represents the monthly returns in Iraq from Jan, 2005 to May, 2020. We select this period because of political and economic instability. Moreover, these period results of high volatility and low volatility of monthly returns series are given as follows,

$$y_t = \ln \frac{p_t}{p_{t-1}} \quad , \qquad (4.1)$$

where y_t represent the monthly returns at time t and p_t denote the exchange rate at time t.

For the dataset on monthly returns, the whole sample of size 184.The dataset is divided into (70%) training and (30%) testing set. Moreover, first 128 data are training set and the last 56 data are testing. The dataset consist of one dependent variable exchange rate and 13 independent variables. Table (4.1) shows the description of variables.

Table (4.1) Description of variables for monthly returns

x_1	Expenditures (exp)
x_2	General price index of Iraq stock (Gprice)
x_3	Gross foreign assets of CBI (GCBI)
x_4	Inflation (Inf)
x_5	Interest rate less than 1 year (interrat.less)
x_6	Interest rate more than 1 year (interrate.more)
x_7	Based money
x_8	M1 milion IQD
x_9	M2 board million
x_{10}	Market value
x_{11}	Number of companies (no.companies)
x_{12}	Revenues
x_{13}	Surplus&deficit

4.3 Characteristics of the time series

The monthly series for the returns of $(\mathrm{IQD/USD})$ is depicted in Figure (4.1). It displays that the returns series is mean-stationary, but non-stationary in variance.

Moreover, it shows phenomena of volatility clustering with period small and large volatility

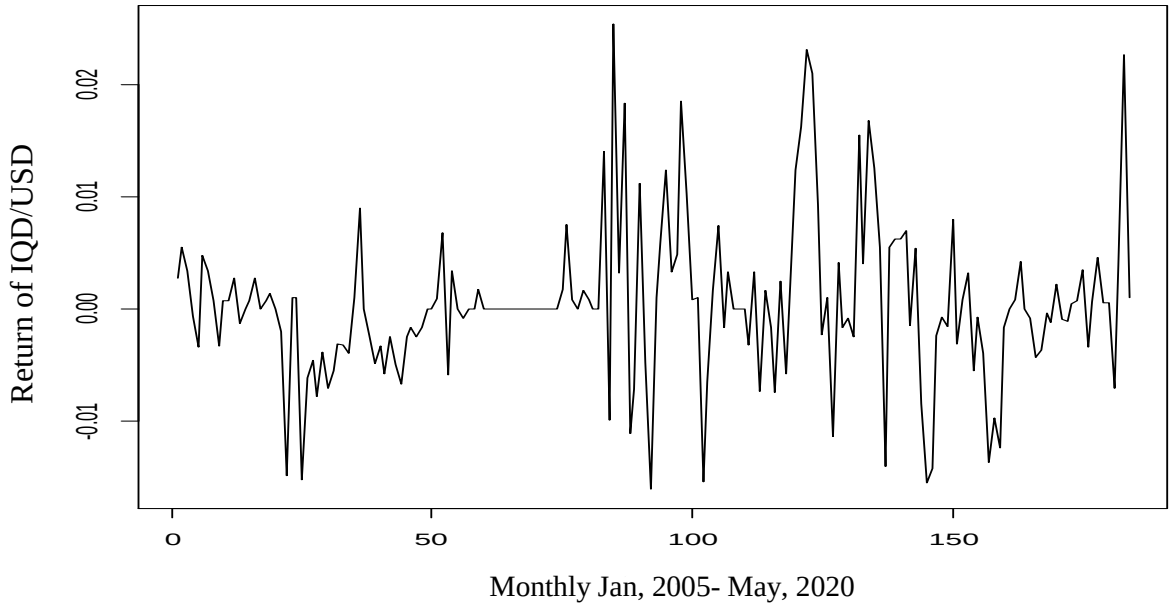

Figure (4.1). The monthly returns of IQD/USD index: Jan, 2005- May, 2020

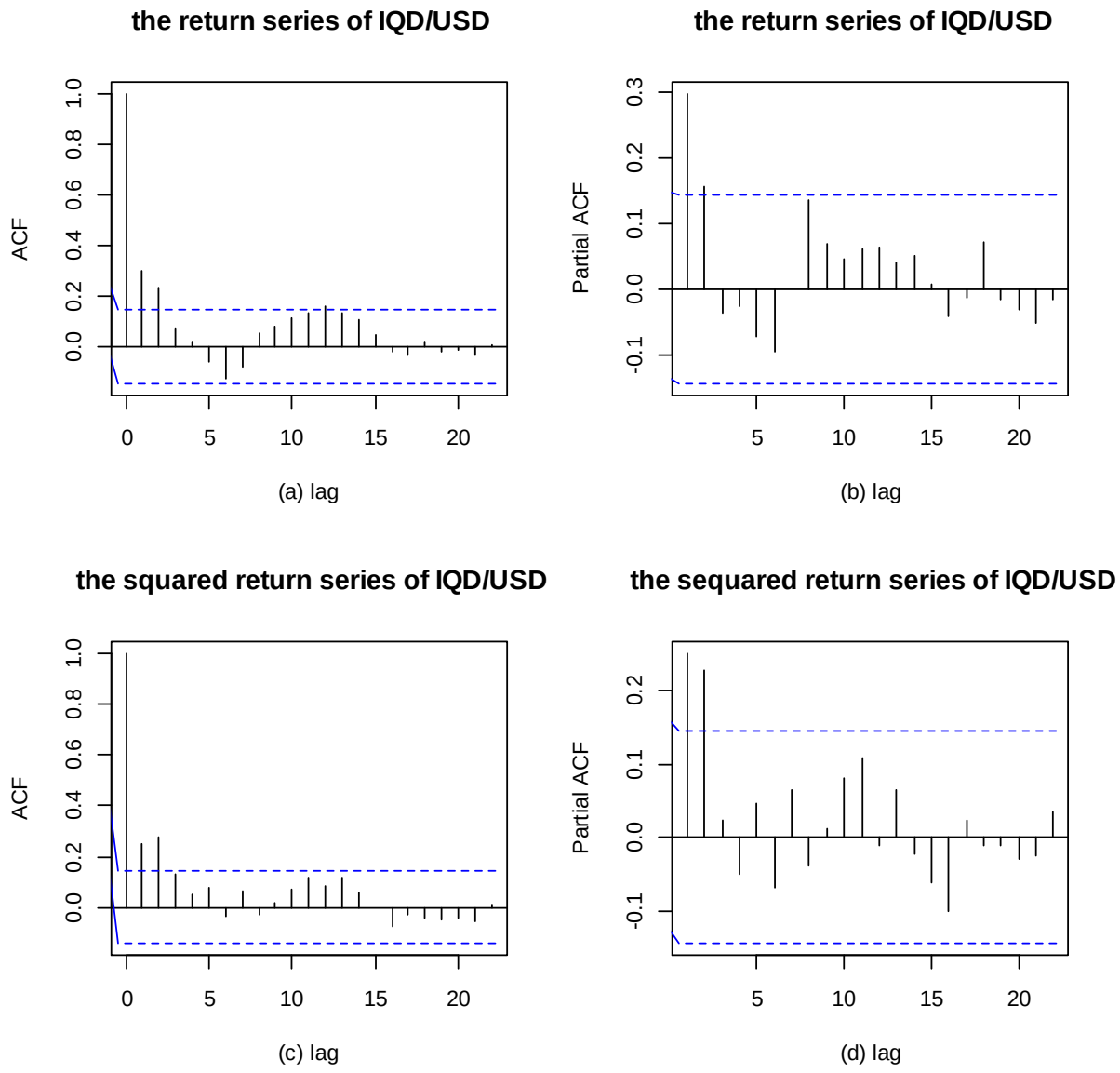

Figure (4.2). The ACF and PACF of the monthly returns.

Figure (4.2) shows that the ACF and the PACF of the return series and squared return series. In Figure (4.2) (a) noted that the lag 1, 2 and lag 12 are not significant correlation of the monthly return series, either remaining lags within the boundary. It means only the lag 1, 2 and lag 12 are affected by the clustering volatility phenomena. We note in (b) that lag 1 and lag 2 is only significant correlation of the return series. In (c) lag 1 and 2 is only significant correlation of the squared

return series, whereas in (d) almost the lagged within the boundary, it means there is little or no correlation of the squared return series.

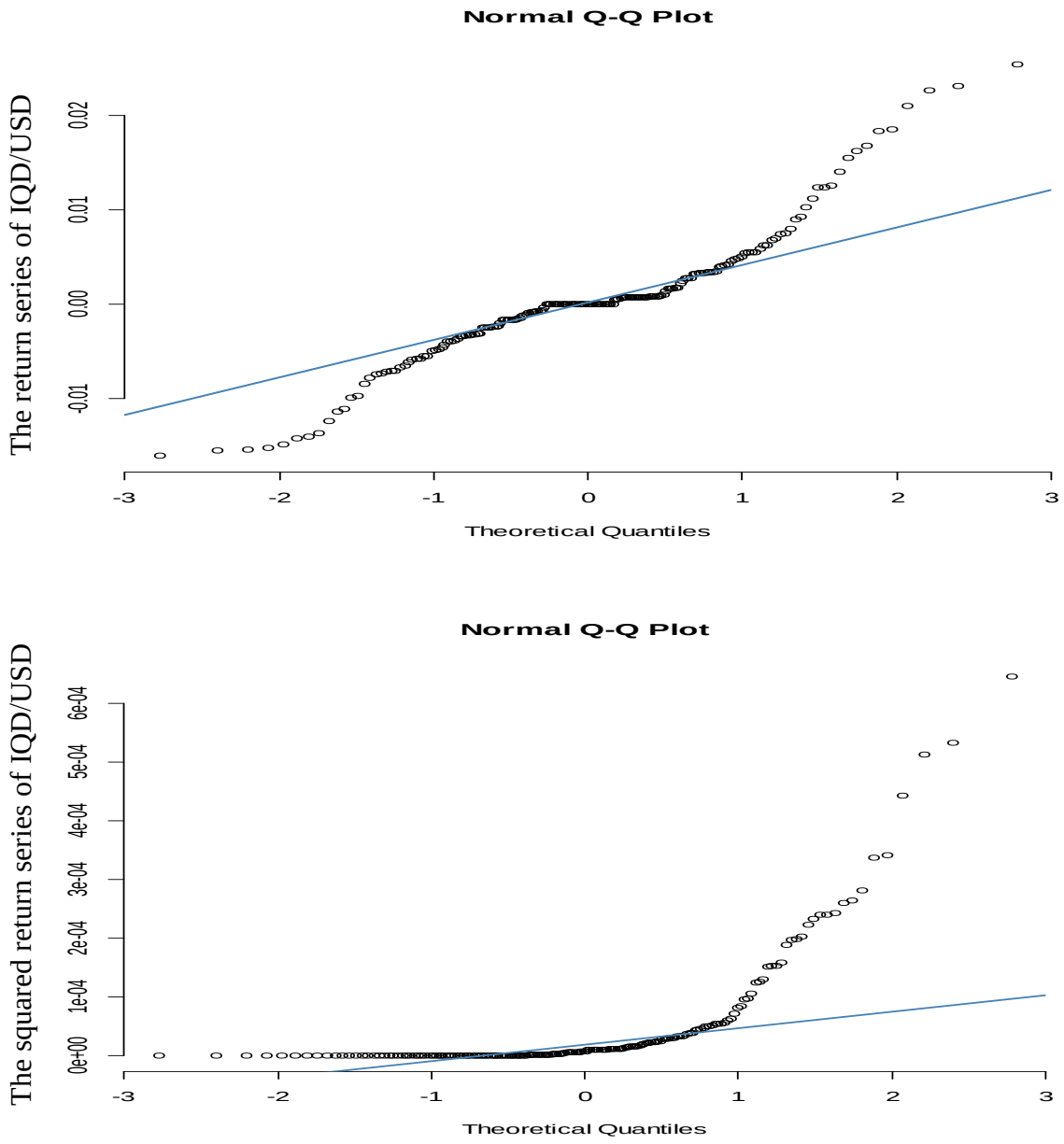

Figure (4.3). The normal Q-Q plot of the return series and squared return series.

Figure (4.3) shows that the normal Q-Q plot of the monthly returns series and the squared monthly returns series, where we note that the assumption of normality has violated due to some unusual observations.

Table (4.2). Description statistics for the returns of IQD/USD.

Return	IQD/USD
Mean	0.00038
Stander Devision	0.0069
Median	0
Maximum	0.0254
Minimum	-0.0161
Skewness	0.6813
Kurtosis	5.1843

Table (4.2) appears that the mean of monthly return series is close to zero, so the monthly returns series is characterized by asymmetric. The monthly returns of IQD/USD have positive skewness coefficient. We can see the distribution of the monthly return series is non-normal because of the kurtosis greater than 3.

4.4 Identification the appropriate model

There is more than one test to detect the presence of the GARCH effect in the time series, as it was explained in the theoretical side. The following three criteria were relied on.

4.4.1 Lagrange-Multiplier test

We use the Lagrange-Multiplier test to test the presence of heteroscedasticity as it mentioned in the theoretical side. The null hypothesis of the Lagrange Multiplier test states that there is no effect of the GARCH models. The statistic test value is 29.9117 at the significant level (5%) and the p-value of the test is 0.000005. Subsequently, the null hypothesis is rejected that means the GARCH effect is present.

4.4.2 Jarque-Bera test

Depending on the Jarque-Bera test to know the GARCH effect on the series according to the test formula after extracting the skewness and kurtosis values of the series that was mentioned in the theoretical side. Therefore, the statistic test value is 50.816 at the significant level (5%) and the p-value of the test is 0. In this case, we would reject the null hypothesis that the residuals is normally distributed and the data follows the GARCH models.

4.4.3 Ljung-Box test

In order to check the autocorrelation of residuals, used the Ljung-Box test that was explained in the theoretical side. The null hypothesis states that there is no autocorrelation among residues, so the statistic test value is 16.096 at the significant level (5%) and the p-value of the test is 0. Moreover, the null hypothesis is rejected that means the autocorrelation of the return series is present at the studied lagged. So the GARCH effect is present.

4.5 Criteria for determining the order of the model

Criterion values to determine the order of model was calculated depending on what was explained in the theoretical side. The criteria AIC, BIC and HQ are used to ensure the fit of the GARCH (1, 1) model, as shown in the table (4.4).

Table (4.4). The value of AIC, BIC and HQ to determine the order of the GARCH models at different orders.

The model	AIC	BIC	HQ
GARCH (1,0)	-7.1990	-7.1466	-7.1778
GARCH (1,1)	**-7.3243**	**-7.2545**	**-7.2960**
GARCH (1,2)	-7.3107	-7.2234	-7.2753
GARCH (1,3)	-7.2985	-7.1937	-7.2560
GARCH (2,0)	-7.2449	-7.1750	-7.2166
GARCH (2,1)	-7.3145	-7.2271	-7.2791
GARCH (2,2)	-7.3036	-7.1988	-7.2611

GARCH (1, 1) model is better than other models, additionally it has stationarity property of GARCH models that it mentioned in the theoretical side is $\alpha + \beta < 1$

4.6 Estimation of the parameters of the model

After carried out the diagnostic stage and sure that the GARCH effect is present of the return series, the estimation stage of the parameters is obtained. The estimation of parameters for the monthly return series is implemented depending on R program and based on the MLE method. The coefficients value are shown as the following table:

Table (4.3). The estimated parameters values of GARCH (1, 1) models

Parameters	Estimate	p-value
α_0	0.00006	0.8694
α_1	0.3910	0.0019
β_1	0.5722	1.08e-09

Table (4.3) shows that the estimated parameters values of GARCH (1, 1) models are 0.00006, 0.3910 and 0.5722 respectively and the $p - value$ of $\alpha_0, \alpha_1, \beta_1$ are 0.8694, 0.0019 and 1.08e-09 respectively at significant level 5%. The $p - value$

of α_1, β_1 are less than 5 %, whilst α_0 is large than 5%. It means that the parameters is significant.

4.7 Check the accuracy of the model

In this section, Ljung-Box test is used to make sure that the used model is the fit model, where criterion value is calculated in equation (2.19). Moreover, the p-value of the residuals is 0.0430 at the significant level (5%) where it is less than 5%. It means that the autocorrelation of the residuals is significant, while p-value of the squared residuals is 0.9940 at the significant level (5%) where it is greater than 5%. So the null hypothesis is not rejected that means the autocorrelation of squared residuals is not significant and the GARCH effect is not present. This indicate that the GARCH (1, 1) model is a good and fit to represent volatility of data.

4.8 The results

After ensuring that the GARCH (1, 1) model is the best model to represent the data of time series, we carried out the prediction of the monthly returns by the previous methods which is mentioned in the theoretical part. In the tabular results, there are the column specific for the parameter to choose of the tuning parameter for the corresponding model. The default parameters of a regression tree are minsplit=20 and cp=0.01 whilst for the random forests that using default ntree=500 with the mtry parameter, suggested by Breiman (2001). The SVR is estimated by using nu-regression with nu=0.5 (default) with type of kernel function such as: linear, radial, sigmoid and polynomial and the performance of each of methods is evaluated by the MSE.

Table (4.5). Predictive performance for monthly return of IQD/US

Method used	Parameter	RMSE
RT	-	1.3229
RF	-	1.1122
LASSO	-	1.2459
ALASSO - SVR	kernel=linear	0.4265
	kernel=radial	0.4040
	kernel=polynomial	0.4098
	kernel=sigmoid	0.4061
EN –SVR	kernel=linear	0.2608
	kernel=radial	0.2642
	kernel=polynomial	0.2841
	kernel=sigmoid	**0.2578**

Table (4.5) shows that the result of the suggested methods is better than another methods with different kernel functions, particularly the EN –SVR methods with sigmoid kernel perform have the potential of forecasting monthly returns cover some variation comparatively with ALASSO – SVR methods. Therefore, this indicates a strong correlation between the independent variables. We observe the results of the suggested methods are converging with different kernel functions.

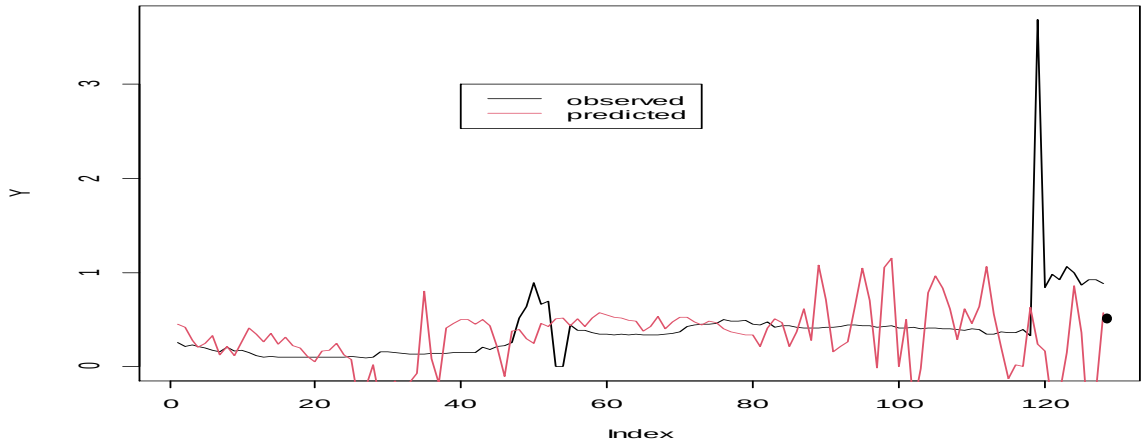

Figure (4.4). The predicted and observed monthly returns in red and blue curves using the best model.

Figure (4.4) displays that the observed and predicted monthly return for IQD/USD using the best model, where it observed the predicted values by using the Elastic net – SVR is the best. In spite of the volatility of the monthly return is very high, but Elastic net – SVR was able to improve the most variance and take it into consideration by modeling it repeatedly.

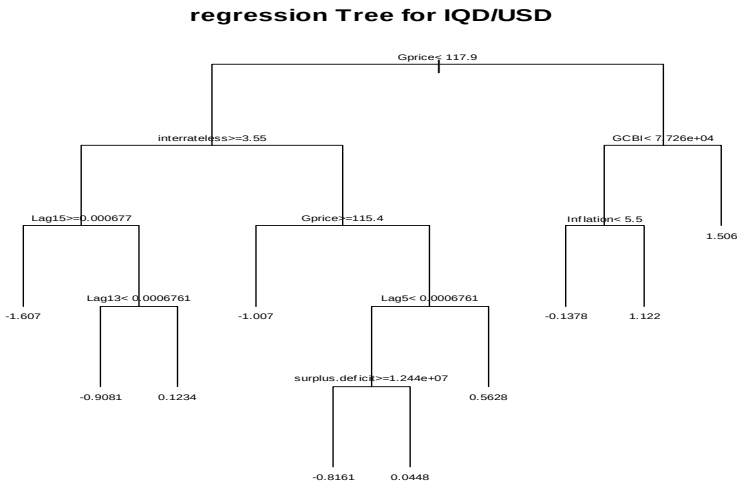

Figure (4.5).he regression tree of monthly returns (IQD/USD)

Figure (4.5) shows the important variables that contributed to grow the regression tree, where the final values in the figure represents the average of all response variables within cell.

Chapter five

Conclusion

Recommendation

5.1 Conclusion

In this thesis, we analyze the performance of the regularized methods and the machine learning models that it widely used in the statistics fields for predicting return series of IQD/USD. We propose to combine the regularized methods (adaptive LASSO and elastic net) with SVR model to reduce dimensions model and get powerful results. In the simulation part, several cases are used to implement our proposed methods. In real data, we analyze the monthly frequency of the exchange rates returns and then the GARCH models is applied to capture the clustering fluctuations in the financial returns series. Both the results of the simulation and the real data show that ALASSO-SVR and EN-SVR methods are more able to improve the forecast of monthly exchange rate returns in comparison with other methods. Particularly, the elastic net –SVR method gives the best results in comparison with the adaptive LASSO–SVR methods.

5.2 Recommendations

1. Extend to other methods such as group lasso and fused lasso to combine them with SVR model.

2. Use another GARCH models such as EGARCH, IGARCH و GARCH-M.

3. Apply these methods ALASSO-SVR and EN-SVR to other parameter values of GARCH models

4. Studying very large or small sample sizes and comparing the results with those used by the researcher in this thesis

5. More attention to test the suitability of the data for the GARCH models, because most of the phenomenon suffers from the problem of homoscedasticity

6. Use another data fields such as environmental or medical or other natural phenomena.

References

Abu-Ahmahd, R., S. (2002). Banks Management, Amman: *Dar Alfikr for publishing.*

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. vol. 19, no. 6, pp. 716-723.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*. 66, 237-242.

Alamili, M. (2011). Exchange Rate Prediction using Support Vector Machines a comparison with Artificial Neural Networks. Master thesis of Science in Management of Technology, Deleft Uuniversity of Technology, Department of Section of Information and Communication Technology

Alfons, A., Croux, C. & Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *Ann. Appl. Stat*. 7 (1) 226 – 248.

Ashfaq A. K, Sultan A. and Syed N. Hussain(2013). Comparative Prediction Performance with Support Vector Machine and Random Forest Classification Techniques. *International Journal of Computer Applications* (0975 – 8887) Volume 69– No.11.

Atto, S. (2002). The development of the financial sector and the instruments of monetary and banking policy to enable the Iraqi economy. Third Scientific Conference of the Department of Economic Studies entitled: Iraqi economy, Arab and international environment variables.

Awad, M. & Khanna, R. (2015). *Efficient learning machines*: Theories, Concepts and Applications for Engineering and System Designer. Apress.

Bollerslev,T., (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics***.** Vo1.31, pp.307-327.

Box G. E. & Jenkins G. M. (1976). *Time series analysis: forecasting and control, revised ed.* Holden –Day.

Box G. E., Jenkins G. M. & Reinsel, G. C. (2008). Forecasting. *Time series analysis*, 137-191.

Box G. E., Jenkins G. M., Reinsel, G. C. and Ljung, M. G. (2016). *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Breiman, L. (2001). *Random Forests*. Machine Learning. Vol 45, 5-32.

Breiman, L., Friedman, J. H. , Olshen, R. A. & Stone C. J. (1984). Classification and Regression Trees. *Wadsworth*.

Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). Introduction to Time Series and Forecasting.Volum 2. *New York: Springer*.

Carlos M. Jarque & Anil K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters.* Volume 6, Issue 3, Pages 255-259.

Chai, T. & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geosci. Model Develoment. 7, 1247–1250.

Chatfield, D. C. (1995). *The analysis of time series: an introduction*. CRC press.

Chen, TH., Chen, MY. & Du, GT. (2020). The determinant of Bitcoin's price: Utilization of GARCH and machine learning approach. *Comput Econ*.

Cont R., (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*. pp. 1–14.

Cordenm, M. W., (1994). Exchange rate policy in developing countries. *IMF*. pp 1-45.

Ekici, O. (2014). Applicability and Parameterization of Machine Learning Approaches for Time Series Modeling .University Dresden , Dep. of Computer Science Institute for System Architecture, Database Technology Group.

Engle, R., F., (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *The Econometric Society*. Vol. 50, No. 4 (Jul., 1982), pp. 987-1007.

Evgeniou, T. & Pontil, M. (2014). *Workshop on support vector machines: Theory and Applications*. Center for Biological and Computational Learning, Artificial Intelligence Laboratory, MIT, E25-201, Cambridge, MA 02139, USA.

Fan, J., & Li, R. (2001). Variable selection via non concave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360.

Fernando Pérez-cruz., Julio A. Afonso-Rodriguez and Rubio, G. J. (2003). Estimating GARCH models using support vector machines. *Quantitative Finance*. Vol. 3, issue 3, 163-172.

Francq, C., & Zakoian J. M. (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. Bernouli. 10: (605-637).

Francq, C., & Zakoian J. M. (2010). GARCH models. *John Wiley & Sons*.

G. M. ljung & G. E. P. box (1978). On a measure of lack of fit in time series models. *Biometrika*. Volume 65, Issue 2, Pages 297–303.

Garg A. (2012) . Forecasting exchange rates using machine learning models with time-varying volatility. Master Thesis of division of Statistics, *Department of Computer and Information Science* of Statistics University Linköping.

Hang, Chang-Ha & Shin, Sa-lm (2010). Estimating GARCH models using kernel machine learning. *Journal of the Korean Data and Information Science Society*. Vol. 21, issue 3, 419-425.

Hannan, E. J. and Quinn, B. G. (1979). The Determination of the Order of an Auto regression. *ROYAL STATISTICAL SOCITY*. Volume 41, Issue 2, Pages 190-195.

Hassan M., (2016). Estimating and Analyzing the Impact of Some Economic Variables on the Fluctuations of Iraqi Dinar Exchange Rate during the Period (1995-2015). *International Journal Humanities and Cultural Studies* ISSN 2356-5926.

Hastie T, Tibshirani R & Friedman J (2001). *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*. Springer.

Haydee L. Roman, (2008). Forecasting the Volatility of Philippine Inflation using GARCH Models. Center for monetary and financial policy, working paper series #1, BANGKO SENTRAL NG PILIPINAS.

Hesterberg, T., Choi, N. H., Meier, L., & Fraley, C. (2008). Least angle and l_1 penalized regression: A review. *Statistics Surveys*, 2, 61-93.

Hoerl A. E. & Kennard R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technimetrics*. 12:1, 55-67.

Ince, H. & Rafalis, TH. B. (2006). A hybrid model for exchange rate prediction. *Decision support system.* Vol. 42, issue 2, 1054-1062.

Kibria, B. M. G. & Banik, S. (2016). Some Ridge Regression Estimators and Their Performances. Journal of Modern Applied Statistical Methods. Vol. 15, No. 1, 206-238.

Kirchgässner, G. & Wolters, J. (2007). Introduction to modern time series analysis. *Springer*.

Lawrence R. Glosten, Ravi Jagannathan and Davids E. Runkle (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess

Return on Stocks. *Journal of the American finance association*. Volume 48, Issue 5**,** Pages 1779-1801.

Leng, C., Lin, Y. & Wahba, G. (2006). A note on the lasso and related procedures in model selection. *Statistica Sinica* 16, 1273-1284.

Maïnassara, Y. B. & Kokonendji, C. C. (2016). Modified Schwarz and Hannan–Quinn information criteria for weak VARMA models. *Stat Inference Stoch Process* 19:199–217.

Matthew, P. K. & Yahaya, A. (2015). Performance analysis on least absolute shrinkage selection operator, elastic net and correlation adjusted elastic net regression methods. International Journal of Advanced Statistics and Probability. 3 (1) (2015) 93-99.

Mohsin, M. (2018). Modeling Exchange Rate Volatility Using GARCH Models: Empirical Evidence From Pakistan. *European Journal of Research*. № 1 (2).

Montgomery, D. C., Jennings, C. L. & Kulahci, M. (2007). *Introduction to Time Series Analysis and Forecasting*. Wiley Series in Probability and Statistics.

Nason, G. P. (2006). Stationary and Non-stationary Time Series.

Nelson, D. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica, 59*(2), 347-370. doi:10.2307/2938260.

Okasha, M. K., (2014). Using Support Vector Machines in Financial Time Series Forecasting. *International Journal of Statistics and Applications*. 4(1): 28-39.

Park, H., Sakaori, F. (2013). Lag weight lasso for time series model. *Comp Stat*. 28, 493-504.

Pesaran, M.H., &Shin, Y. (1998). An Autoregressive Distributed-Lag Modelling approach to cointegration analysis .*Econometric Society Monographs.* 31,371-413.

Rajaratnam, B., Roberts, S., Sparks, D. & Yu, H. (2019). Influence Diagnostics for High-Dimensional Lasso Regression. *Journal of Computational and Graphical Statistics*, ISSN: 1061-8600 (Print) 1537-2715.

Rofael D. and Hosn R. (2015), Modeling Exchange Rate Dynamics in Egypt: Observed and Unobserved Volatility, *Modern Economy*, 2015, 6, 65-80.

Sami, Kh. (2007). The international economy. *Cairo: Dar Arabian Al-Nahtha for Publishing.*

Sampson, M. (2001). *Time series analysis*. Loglinear publishing.

Sartori, S. (2011). *Penalized regression*: Bootstrap confidence intervals and variable selection for high- dimensional data sets.

Shumway, R. H. & stoffer, D. S. (2011). *Time series analysis and its application*. Springer, New York Dordrecht Heidelberg London.

Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*. 105–146.

Takaishi, T. (2018). GARCH Parameter Estimation by Machine Learning. *International Journal of Engineering and Applied Sciences* (IJEAS), ISSN: 2394-3661, Volume-5, Issue-8.

Thadewald, T. & Büning, H. (2004). *Jarque-Bera Test and its Competitors for Testing Normality: A Power Comparison, Diskussionsbeiträge*, No. 2004/9, Freie Universität Berlin, Fachbereich Wirtschaftswissenschaft, Berlin.

Tibshirani, R .J. & Tibshirani, R. (2009). A bias correlation for the minimum error rate in cross validation. *The Annals of Applied Statistics*, Vol. 3, No. 2, 822-829.

Tibshirani, R.(1996) .Regression Shrinkage and Selection via LASSO**.** *Journal of the Royal Statistical Society.*

Tizro, A. T., Ghashghaie, M., Georgiou, P. & Voudouris, K. (2014). Time series analysis of water quality parameters. *Journal of Applied Research in Water and Wastewater* 1(2014) 40-50.

Tsay, R. (2002). *Analysis of financial time series*. John Wiley & Sons, Canada.

Vapnik , V. Boser, B.E., Guyon, I.M. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*. Pages 144–152.

Walker, G. (1931). On periodicity in series of related terms. *Proceedings of the Royal Society*. vol. A 13, pp. 518.

Wold, H. (1938). A study in the Analysis of Stationary Time Series. Almgrist and Wiksell, Stockhom.

Yaseen, A. Q. (2019). *Feature selection in ARMA models with application*. Master thesis of statistics, statistics department of Al-Qadisiyah University.

Yule, GU. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society 89*, pp. 1–64.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of Royal Statistical Society*, B67, pp: 301 – 320.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 1418-1429.

- المستخلص

يتعتبر نموذج متجه الانحدار الساند SVR جزء مهم من تقنيات تعليم الالة التي طورت لحل مشاكل الانحدار. يستخدم نموذج SVR في مساحات عاليه الابعاد لتحديد المستوى الفائق الأمثل لنمذجة الانماط غير الخطي في الانحدار. تم تطوير نموذج SVR في هذه الأطروحة من خلال دمجه مع الطرق المنتظمة مثل لاسو التكيفي (ALASSO) والشبكة المرنة (EN) لتقليل أبعاد النموذج وتحسين أدائه حيث نشير إلى الأساليب المقترحة بواسطة ALASSO- SVR و EN - SVR. ومع ذلك ، تتم مقارنة الطرق المنتظمة مع نماذج التعلم الآلي (الغابات العشوائية وأشجار الانحدار) لمعرفة الأداء التنبئي الأفضل لهذه النماذج. تتميز السلاسل الزمنية المالية بوجود تقلبات تحدث بشكل عشوائي خلال فترات زمنية مختلفة ، وهذا لا يتوافق مع الأساليب المنظمة ونماذج التعلم الآلي التي تفترض ثبات التباين. علاوة على ذلك ، نستخدم نماذج الانحدار الشرطي العام (GARCH) مع هذه النماذج لتقدير التباين الشرطي ومعلمات طرق ALASSO-SVR و EN-SVR ونماذج التعلم الآلي. ثم يتم تنفيذ الأداء التنبئي لهذه النماذج من خلال إجراء تكراري حيث ان المعلمات المقدرة بواسطة نموذج GARCH لجعل التنبؤ بخطوة واحدة للأمام عن طريق التقدير العودي. وماعدا ذلك ، يتم تحديث هذه المعلمات بواسطة المعلومات الجديدة. نستخدم المتغيرات الاقتصادية والمتغيرات المتأخرة للتنبؤ بعوائد سعر الصرف الشهرية للدينار العراقي / الدولار الأمريكي. تظهر نتائج المحاكاة والبيانات الحقيقية أن الطرق المقترحة أفضل من نماذج التعلم الآلي ، ولا سيما طريقة EN-SVR قادرة على التنبؤ بشكل أفضل بعائدات سعر الصرف الشهرية وكذلك تحسين الأداء التنبئي.

جامعة القادسية

كلية الادارة والاقتصاد

قسم الاحصاء

استخدام نماذج التعليم الالة متغيرة الزمن في تقدير نماذج GARCH

رسالة ماجستير مقدمة الى مجلس كلية الاداره والاقتصاد/جامعة القادسية وهي جزء من متطلبات نيل درجة الماجستير في قسم الاحصاء

من طلبة الماجستير

سرى حسن سامي عريبي

اشراف

أ.د طاهر ريسان دخيل

1442هـ

2021