Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Stability of GARCH models for prediction the exchange rate based on machine learning with time-varying

To cite this article: Tahir R Dikheel and sura H Sami 2021 *J. Phys.: Conf. Ser.* **1897** 012013

Stability of GARCH models for prediction the exchange rate based on machine learning with time-varying

Tahir R Dikheel[1] and sura H Sami [2]

[1] University of Al- Qadisiyah, College of Management and Economics, AL-Qadisiyah, Iraq.

[2] University of Al- Qadisiyah, College of Management and Economics, AL-Qadisiyah, Iraq.

Email: Tahir.dikheel@qu.edu.iq and surahassan154@gmail.com

Abstract. Recently, machine learning methods such as regression trees, Support Vector Machine and lasso have been widely used in many scientific fields where the variance of the random errors is homogenous or constant. This issue is disagreement with the data that is based on the time, such as financial data which is analysis by using time series. Unfortunately, the financial time series is instable as a result of fluctuations that happen during different time periods between relative calm and high turbulence. A great effort was paid in the literature to tackle this problem by combining machine learning with GARCH model. In this paper, we suggest estimating the variance of varying-time and parameters of machine learning methods to predict the exchange rate of the Iraqi dinar (IQD) against (USD) with using an iterative procedure. Real data and simulation study are curried out to know the performance of our proposed procedure with others. The result shows that our proposed procedure outperforms than others.

1. Introduction

Time series is a one of the important statistical tools that is widely used to analysis financial data, such as detection of random stock movement in the financial markets and predicting the exchange rates of foreign currencies against local currency. Unfortunately, the random walk of exchange rates results in the heteroscedasticity of variance and time series instability as a result of fluctuations that happen during different time periods between relative calm and high turbulence. Engle (1982) proposed Autoregressive Conditional Homoscedasticity ($ARCH$) models to estimate the fluctuations that have been considered for the description of the characteristics of financial markets where can be defined as follows [5]:

$$r_t = \mu + e_t \quad (1)$$

$$e_t = \sigma_t^{1/2} \varepsilon_t \quad (2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 \quad (3)$$

where the r_t is the return series, $\alpha_0, \alpha_1, \alpha_2, \ldots\ldots, \alpha_i$ the model parameters and $\alpha \geq 0$, $e_t \sim iid\ N(0,1)$ and σ_t^2 the volatility (Conditional variance is a linear function of the preceding squares of variance and observations).

Unconditional variance is known for e_t as follows.

$$V(e_t) = \frac{\alpha_0}{1-\sum_{i=1}^{p}\alpha_i} > 0 \qquad (4)$$

where $= 1$, the model is of the first order $ARCH(1)$ that can be written the equation as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-i}^2 \qquad (5)$$

Bollerslev (1986) noted that increasing the value of p produce negative α which is have to be positive as that mentioned above. To overcome this problem, Bollerslev (1986) suggested Generalized Autoregressive Conditional Homoscedasticity $(GARCH)$ by adding another parameter to the Eq (3) as follows [3] [9]:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i e_{t-i}^2 + \sum_j^q \beta_j\, \sigma_{t-j}^2 \qquad (6)$$

and then unconditional variance in Eq (4) can be written as follows:

$$V(e_t) = \frac{\alpha_0}{1-(\sum_{i=1}^{p}\alpha_i + \sum_j^q \beta_j)} > 0 \qquad (7)$$

where $\beta_1, \beta_2, \dots \beta_j$ the new parameter and $\beta_j \geq 0$,p and q is the order of the $GARCH$ model .

2. Estimation methods

2.1 Regression Tree (RT)

The regression tree is a supervised learning algorithm which is used in the machine learning to solve the regression and classification problems as decision tree. Decision tree is one of the methods of predictive modeling in statistics which is used to support the decision through a graphical representation that resembles the tree, as it consists of internal nodes that represent a test all independent variables and terminal nodes are used for making the final decision or prediction. Linear regression analysis is the most important statistical tool for describing the linear relationship between the dependent variable and one or more independent variables. Linear regression analysis is concerned with the study and analysis of the effect of several quantitative independent variables on a quantitative dependent variable, where the regression model is used as a means of predicting future values. When the data includes explanatory variables that interact in complicated, non-linear fashion the regression model has poor predictive performance. Therefore, we need an alternative approach to nonlinear regression to deal with nonlinearity and the interactions in the data, it is regression tree approach. A Regression tree is a binary tree where each node in a tree has two nodes and its benefits give a visual representation to facilitate the user's decision-making in the node that was used to predict [2].

The goal of the decision tree algorithm is to obtain the best explanatory variables through splitting, where the root node splits into sub-nodes, then these sub-nodes splits into sub-branches provided that you pass the test on the basis of which it moves to the right or left sub-branch and then reach to the terminal nodes (leaves), where these splits is used to reduce the division. The leaf node is making the prediction as follows [9]:

$$m_q = \frac{1}{n_q} \sum_{i \in q} y_i \qquad (8)$$

Where $n_q$is the number of the observation in terminal nodes. y_i is the response variables and build a function at the end of the regression tree, it must have a constant m_q as following:

$$g(x) = \sum_{i=1}^{a} m_q \, I\left(x \in R_q\right) \qquad (9)$$

2.2 Support Vector Machine (SVM)

SVM have been suggested by Vapnik et al. (1995) to identify patterns. It is one of the most popular the machine learning algorithms that used for the purpose to solve regression and classification problems. In this paper, it is used the time series forecasting by data is mapped to a higher-dimensional space. It is a powerful tool for separating data in many scientific disciplines. In regression, it is an extension of the SVM for classification proposed by (Boser et al1992) and its aim is to reduce an upper bound on the expected risk rather than reduce error on training data [2]. Moreover, it avoids the over fit. The main idea of the SVM algorithm is the mapping a hypothetical space in a high-dimensional space for linearly separable data, based on the closest points of the support vectors and thus build the hyperplane [6]. The optimal hyperplane is chosen from a group of hyperplanes that increase the marginal of the hyper plane, which marginal is the distance between the nearest points of the support vectors and the hyperplanes [2]. For data that cannot be separated linearly, the Kernel function is used. It uses the Kernel function to implicit mapping of input data into a high dimensional feature space. There are several types of kernel functions such as linear, nonlinear, polynomial, Gaussian kernel, Radial basis function (RBF), sigmoid etc., the best way to choose a kernel function is through trial and error. In regression analysis, SVM uses the $\varepsilon - insensitive$ loss function, i.e.[2]

$$\|y - f(x)\| = \max\{0, \|y - f(x)\| - \varepsilon\} \qquad (10)$$

2.3 The Least Absolute Shrinkage and Selection Operator ($LASSO$)

consider usual the linear regression model:

$$y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k + \varepsilon \qquad (11)$$

Where y is the dependent variable, $X_1, \dots, X_k$ are the explanatory variables, $\beta_0, \beta_1, \dots, \beta_k$are the regression coefficient, ε is the random error. The ordinary least squared (OLS) is a linear function used to estimate unknown parameters in a linear regression model. OLS is used to minimize the residuals squared error. Therefore, it is the best unbiased estimation method. Unfortunately, in high-dimensional data, the performance of the OLS method is poor, due to the high variance that affects prediction accuracy and may be difficult to interpret (Brown,1993) [2]. Sometimes the accuracy of prediction can be improved by shrinkage at 0 some parameters. Moreover, the penalization techniques proposed techniques to improve the OLS estimator as ridge regression and LASSO etc. In this paper, we focus on using LASSO. Tibshirani (1996) suggested LASSO method to estimate the regression coefficients depend on l_1-norm penalized least squares criterion [12]. The LASSO algorithm is shrinkage coefficient and variable selection simultaneously, which it minimizes the mean squared error MSE.LASSO performs shrinkage some the coefficients and forces others to be zero, which it provides the interpretable results. LASSO can be written as [10]:

$$\widehat{\beta_{lasso}} = \min \frac{1}{n}(y - X\beta)^T(y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (12)$$

Subject to

$$\sum_{j=1}^{p} |\beta_j| \leq t \qquad (13)$$

where t is the constant value that called tuning parameter, which it controls a mount shrinkage to be applied to estimate the regression coefficient. We use the k-folds to select the best value of tuning parameter.

3. Research and Method

The algorithm of our paper for predicting of returns can be written in the following steps,

Let

$$\widehat{y_t} = f(\odot, x_t) \qquad (14)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 \ (15)$$

Where $e_t = y_t - \widehat{y_t}$

1. Split he data set into two subsets which are denoted as training and test sets
2. Compute the residuals of machine learning model using training set.
3. Find the residuals of machine learning model (step 2) with test set
4. Combining the residuals in the 2 and 3 step in one vector.
5. Estimate the time varying variance of GARCH model by using the residuals vector in step 4, and then the standard error of time varying variance $\sqrt{\sigma_t^2}$ is estimated.
6. Reduce the volatility clustering effect on y_t,the transformation of y_t is carried out by using $\sqrt{\sigma_t^2}$ as follows

$$y_t^* = y_t / \sqrt{\sigma_t^2}$$

7. The transformed y_t^* is used with machine learning model again, and then $RMSE$ is calculated.

4. Results

4.1 Simulation Study

The simulation study has been carried out (1000) iteration to know the performance of competitions methods. For this purpose, (100) *and* (200)observations are generated from three GARCH (1, 0), GARCH (1, 1) and GARCH (1, 2) models with different parameters. The error term e_{t-1}^2 is generated randomly from standard normal distribution,

Model_1 $\sigma_t^2 = 0.5 + (0.4)\varepsilon_{t-1}^2$

Model_2 $\sigma_t^2 = 0.5 + (0.3)\varepsilon_{t-1}^2 + (0.5)\sigma_{t-1}^2$

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

The previous models are used to generate three Autoregressive Distributed Lag (ADL) models with fixed order $(5,3,3)$ [11]

$$y_t = \omega + \sum_{k=0}^{5} \beta_{(0,k)} y_{(t-k)} + \sum_{k=0}^{3} \beta_{(1,k)}\, x_{(1,t-k)} + \sum_{k=0}^{3} \beta_{(2,k)}\, x_{(2,t-k)} + e_t \quad (14)$$

where ω is the intercept , $\beta_{0,k}, \beta_{1,k}, \ldots \beta_{p,k}$ is the regression coefficient , k is the lag operator and independent variables (x_1, x_2) are generated from normal distribution with zero mean and 25 variance. The random errors e_t of the ADL model are generated from normal distribution with zero mean and 9 variance

The simulated dataset is divided into (30%) training and (70%) test sets and the performance predictive of a three machine learning methods, LASSO, SVM and RT is compared. The Root of Mean Squared Errors $(RMSE)$ over (1000) datasets is computed for each method. The best method is the one that possess the smallest$RMSE$.

In case: sparse case.

Table 1. Default values of parameters

y_{t-1}					$x_{1,t-1}$				$x_{2,t-1}$			
$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{0,3}$	$\beta_{0,4}$	$\beta_{0,5}$	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$	$\beta_{2,3}$
0.3	-0.2	0.1	0	0	0.9	0.7	0.5	0	1	-0.7	0.5	0

Table 2. The $RMSE$ of three methods over all 1000 simulation data sets

Methods used	RMSE					
	Model_1		Model_2		Model_3	
	n=100	n=200	n=100	n=200	n=100	n=200
LASSO	1.0215	1.0125	1.0308	1.0123	1.0092	1.0110
SVM	0.5723	0.9691	0.5857	0.9978	0.5748	1.0255
RT	0.8768	0.6381	0.8405	0.6280	0.8729	0.6322

Table 2 is shown that SVM perform comparatively better than LASSO and RT with n=100 and different models as Model_1, Model_2 and Model_3. Then the forecasting performances of RT is better than SVM and LASSO when n=200 with models. Unfortunality, as the sample size increases, we notice a higher $RMSE$ which means that the SVM is very poorly with large sample sizes, whilst RT performance is poorly with small sample sizes.

In case: very sparse case

Table 3. Default values of parameters

y_{t-1}					$x_{1,t-1}$				$x_{2,t-1}$			
$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{0,3}$	$\beta_{0,4}$	$\beta_{0,5}$	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$	$\beta_{2,3}$
0.2	0	0	0	0	0.9	0	0	0	1	0	0	0

Table 4. The $RMSE$ of three methods over all 1000 simulation data sets

Methods used	RMSE					
	Model_1		Model_2		Model_3	
	n=100	n=200	n=100	n=200	n=100	n=200
LASSO	1.0199	1.0135	1.0225	1.0133	1.0281	1.0153
SVM	0.5717	0.9652	0.5889	0.9567	0.5767	1.0363
RT	0.8569	0.6327	0.8569	0.6301	0.8674	0.6350

Table 4 is shown that SVM perform comparatively better than LASSO and RT with n=100 and the models as Model_1, Model_2 and Model_3, whilst RT is the best when n=200 with the models as Model_1, Model_2 and Model_3. Unfortunality, as the sample size increases, we notice a higher $RMSE$ which means that the SVM is very poorly with large sample sizes. Whilst RT performance is poorly with small sample sizes.

4.2 Real Data

The exchange rate (IQD/USD) dataset has been collected from Central Bank of Iraq for the period $(Jan, 2005)$ to$(Dec, 2018)$. It is well known that Iraq is instable country, therefore, many events have been happened during this period. These events result in volatility of monthly returns series as follows,

$$y_t = \ln(p_t) - \ln(p_{t-1}) \quad (15)$$

where y_t is monthly returns at time t, p_t is the exchange rate at time t.

We consider the first 117 observation as training set and the remaining observations as test set. The series of monthly returns of (IQD/USD) is depicted in Figure 1. It is clear Figure 1 shows the series of returns is stationary in mean. But it displays the typical volatility clustering phenomenon with periods of relative tranquility followed by period's large and small volatility.

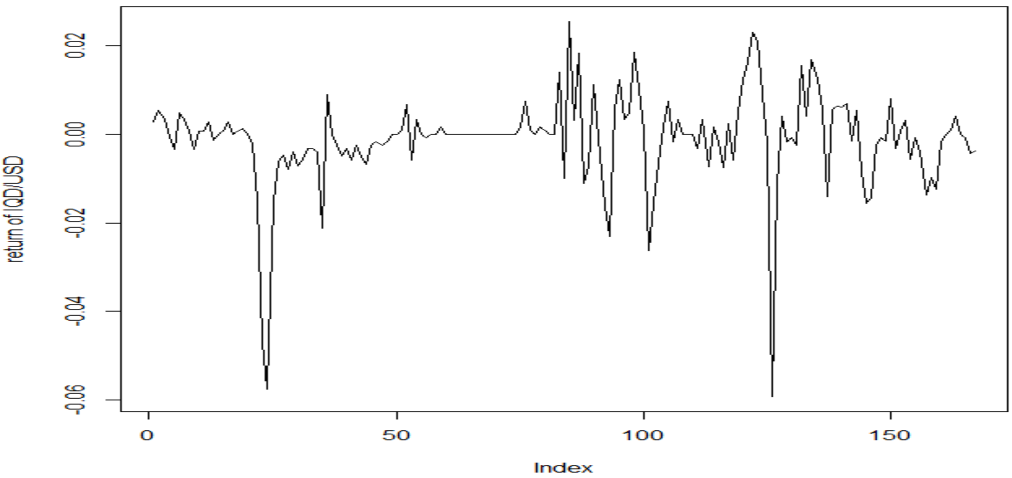

Figure 1. IQD/USD index: Jan, 2005- Dec, 2018

The autocorrelation function (ACF) of the returns and squared returns series of (IQD/USD) are depicted in Figures 2 .In Figure 2 (a), we note that the autocorrelation of the return series is present only in lag 1, 2 and lag 6 .In the anther word, it indicate that the returns series is more affected significant by the volatility clustering than the squared returns series. Whereas in **(b)** most of all the spikes are inside the boundary, it means there is little or no correlation.

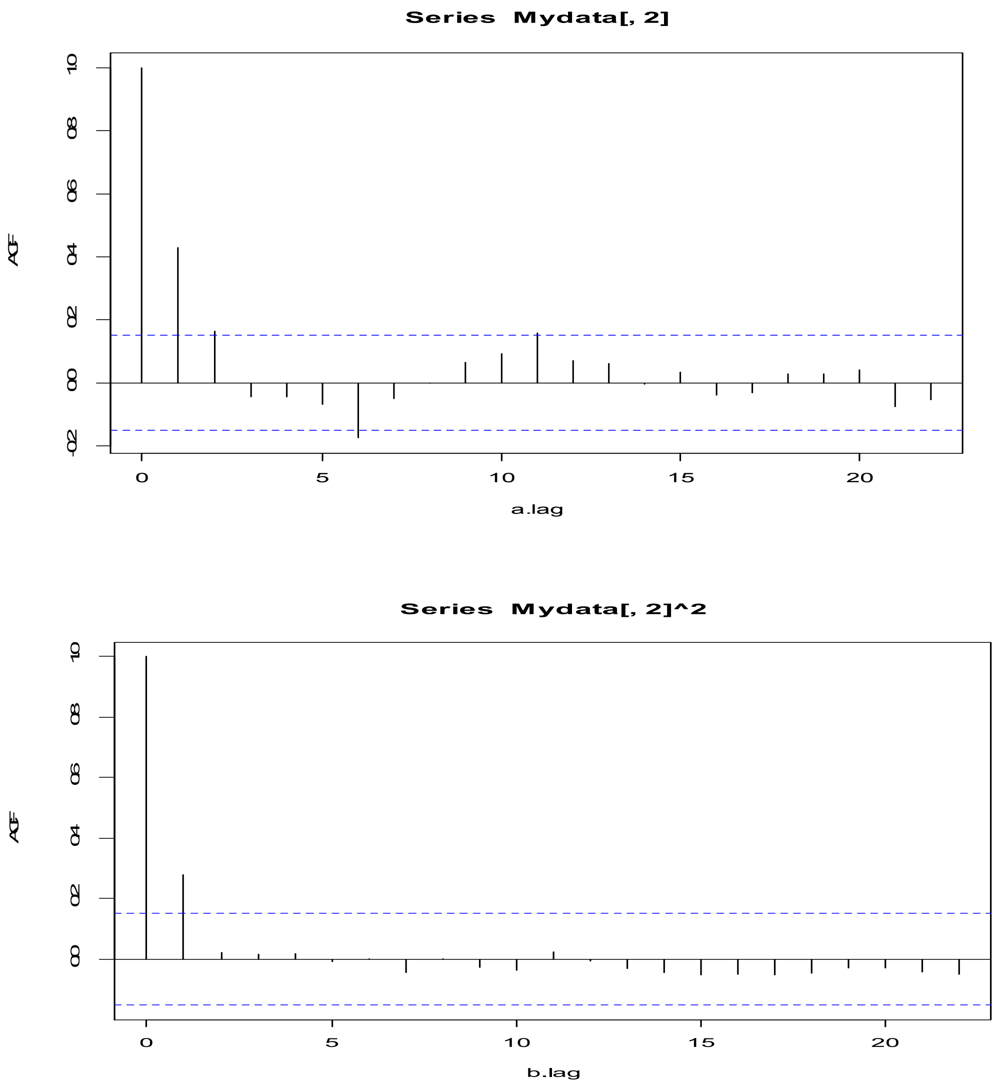

Figure 2. a. ACF for the return series and b. ACF for the squared returns.

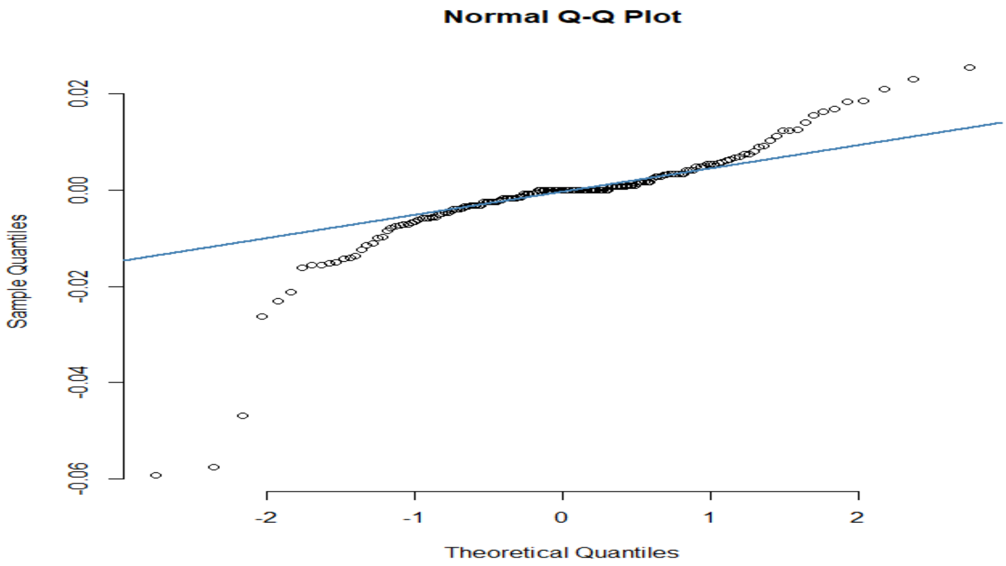

Figure 3. Q-Q plot for return of IQD/USD.

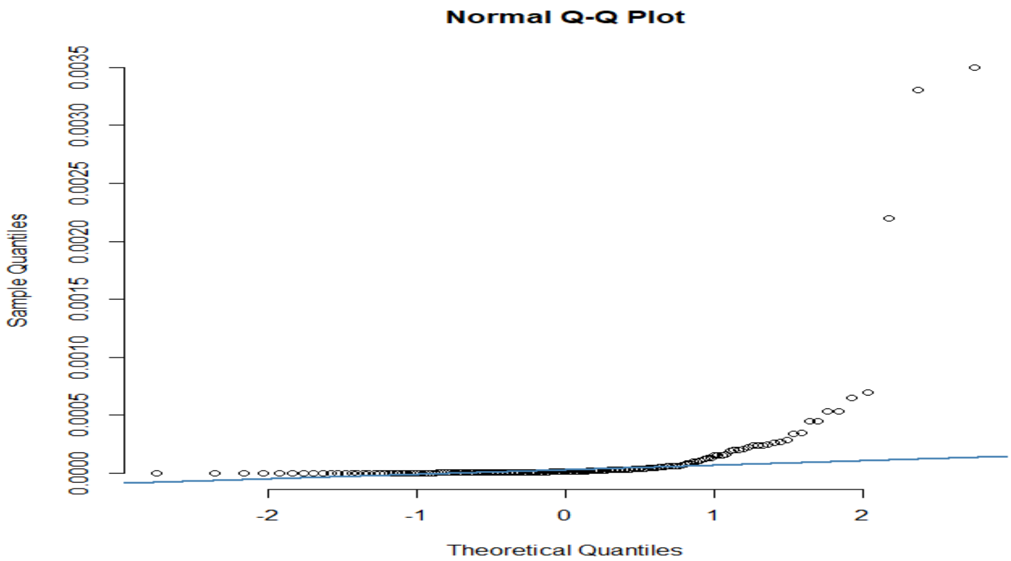

Figure 4. Q-Q plot for the squared return of IQD/USD.

Figure 3 shown Q-Q plot for return series of IQD/USD which we noted that some observations is not assuming normality, perhaps they are outliers. Figure 4 appears the assumption of normality has been violated seeming of three outliers or unusual observations for IQD/USD returns.

Table 5. Descriptive statistic for the returns of IQD/USD.

Return	Mean	Stander Division	Median	Maximum	Minimum	Skewness	Kurtosis
IQD/USD	-0.0011	0.0106	0	0.0254	-0.0591	-2.1835	13.7359

Table 6. Tests of GARCH models for the returns of IQD/USD.

Return	Jargue-Bera test	Ljung-box test Q (10)	Ljung-box test Q (15)	Ljung-box test Q (20)	Ljung-box test Q (10) *	Ljung-box test Q (15) *	Ljung-box test Q (20) *	LM Arch Test
IQD/USD	237.9615 [0.0000]	25.6767 [0.0042]	38.677 [0.0007]	41.6976 [0.0030]	2.8029 [0.9856]	6.7362 [0.9645]	8.6744 [0.9863]	3.0389 [0.9952]

Note: p-values are in brackets.

Table 5 shows the descriptive statistics for the monthly returns and essential tests of the diagnostics stage of GARCH. It is evident from Table 5; the mean of the returns is different from its median, so returns distribution is asymmetric. It is notable that the skewness coefficient is negative which refers to the error distribution with long tail on left side of the shape. It is clear, the statistical indicators in the table 5 lead us to conclude the random distribution of errors is non-normal. The confirmation of that conclusion is evident in the value of kurtosis test which exceeds the threshold (3) and in table 6 the Jarque-Bera statistics rejects the normality hypothesis of returns at a significant level 5%. Moreover, we note that the p-values of $Q\ (10), Q\ (15)$ and $Q\ (20)$ are less than 0.05, so the null hypothesis of no autocorrelation is rejected. In another word, the autocorrelation of return is significant and variance of errors is not constant. Whereas, the p-values of $Q\ (10)\ *, Q\ (15)\ *$ and $Q\ (20)$ * of the squared returns are greater than 0.05, so, it refers to the autocorrelation is no significant at the squared returns. Whilst, $p-values$ of LM Arch test is greater than 0.05 that means not significant for GARCH effect on return series. Despite the previous evidence Figure 1 shows clear fluctuations over time, in addition to statistics (Ljung-Box statistic, Jarque-Bera statistics) and Kurtosis that prove the presence of the GARCH effect. This problem is produced due to the outlier detected by the graphical test (Figure 3: Q-Q).However, we can conclude the return series of IQD/USD offer clustering volatility and leptokurtosis. Then, we use the Maximum Likelihood Method to estimate the coefficient of GARCH model and then we select the best model for the returns. The model which parameters have significant effect is the best model. The GARCH $(1, 0)$ is very appropriate to forecast for the return series because of his parameter significant.

Table 7. Show the $RMSE$ values for monthly returns of IQD/USD

Methods	$RMSE$
LASSO	1.1579
SVM	0.8888
RT	0.8990

Table 7 presents the predictive performance for monthly return is obtained by $RMSE$ that is shown the SVM outperform better than LASSO and TR.

5. Conclusion

In this paper, we measured the performance of the machine learning models used with the GARCH model to predict monthly exchange rate returns, which the models are evaluated using $RMSE$ criterion .simulation study shows that a preference of SVM model with n=100. In the anther word, SVM model is more stable than comparison of the other models in sparse and very sparse cases with small sample size, whilst RT is the better than the SVM and LASSO model with large sample sizes in sparse and very sparse cases. As for the real data shows GARCH (1,0) is very appropriate to forecast for the return series. SVM gave the best result comparison of the other models to improve performance predictive for monthly returns of IQD/USD.

Reference

[1] Altaf H. and Mohammed N. (2011). Recurrent Support and Relevance Vector Machines Based Model with Application to Forecasting Volatility of Financial Returns. *Journal of Intelligent Learning Systems and Applications,* **3**, 230-241.

[2] Ankita G. (2012). Forecasting exchange rates using machine learning models with time-varying volatility. University Linköping, division of Statistics, *Department of Computer and Information Science.*

[3] 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**.

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

[5] Bollerslev,T.,R.,Y.,Chou and K.F. Kroner (1992),ARCH Modeling in Finance :A Selective Review of the Theory and Empirical Evidence *.Journal of Econometrics* **,52**,5-59.

[6] Engle, R., F.,(2001).The Use of ARCH/GARCH Models in Applied**,** *Journal of Econometrics Perspectives* –Volume 15, Number **4**, (157-168).

[7] Fernando Perez-Cruz, Julio A Afonso-Rodr´ıguezan and Javier G.(2003). Estimating GARCH models using support vector machines. *Quantitative Finance Volume 3 1–10.*

[8] Francq, C., and Zakoian J.M., (2010) **.**GARCH Models *.John Wiley &Sons.*

[9] Heston, S.L., Nandi, S., (2000).A Closed-Form GARCH Option Valuation Model Review of Financial, *pp.585-625.*

[10] Onur E. (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

[11] Pascalis K. Matthew and Abubakar Yahaya(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**) 93-99.

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

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