Utilizing Time Series Models to Capture and Forecast Different Markets' Volatility

Yufei Jing/yj2640

Zi Fang/zf2258

Zikun Dong/zd2268

Xuehan Yu/xy2478

Statistical Methods in Finance (GR5261/GU4261)

Professor Zhiliang Ying

19 April 2021

1 Introduction

1.1 Topic of Study

Since the very end of 2019, people have experienced unprecedented changes posted by coronavirus pandemic and other social events, such as the presidential election and BLM movements. The pandemic slowdown has deeply affected companies and jobs. According to the World Bank COVID-19 Business Pulse Surveys, conducted between May and August 2020, companies were forced to reduce hours and wages and, to a lesser extent, fire workers. With less money, people will be forced to make a trade-off that could harm health and learning for generations. Those changes have not only influenced our daily living routines, but also altered the way people look at the investment world and the function of businesses, which turned out to cause a huge amount of irregular or abnormal volatilities in the financial markets. This report will investigate the adaptability of various time series models, such as ARMA and GARCH models in adjusting for and predicting sudden changes made in different markets' fluctuation that was caused by various social issues. Specifically, this report would also analyze various market behaviors existing in different sectors and how those differences could be identified with differently parametrized Time Series Models.

1.2 Data Selection & Preprocessing

The data used in our study is extracted from Yahoo Finance, which is one of the largest online providers for financial news, data and commentary including stock quotes, press releases, financial reports. To capture the fluctuation in returns within different sectors, historical prices of certain Select Sector SPDR Funds are chosen as our research target, namely, XLV (Health Care Select Sector SPDR Fund), XLE (Energy Select Sector SPDR Fund), XLF (Financial Select Sector SPDR Fund), XLK (Technology Select Sector SPDR Fund), XLI (Industrial Select Sector Fund). In the data preprocessing steps, 10-year daily closing prices, ranging from April 2011 to April 2021 of the selected funds mentioned above, is used to calculate net and log returns, and are assessed on their practicability. With these, we would like to see how industry performed and will perform in a short future in terms of volatility.

2 Main Objectives

This study would seek for advanced Time Series models which can capture unexpected and irregular volatilities existing in different financial markets and show investors and stakeholders the proper Time Series models to use when evaluating volatility of risks within diverse sectors. To achieve this, this report contains our detailed steps in data selection, fitting proper ARIMA and GARCH models according to data patterns, validating fitted models by conducting tests such as weighted Ljung-Box Tests and weighted ARCH LM tests, and checking coverage rates by conducting Interval testing and VaR testing. A model forecast is also performed. Lastly, we will present a comparison between different funds' performance based on our models.

3 Analytical Process

3.1 Model selection

Before doing further calculations or fitting models, we need to choose a set of proper returns. By plotting net return and log return for each sector (Appendix Figure-6.2.1), we can easily find that they have similar distribution patterns, which should make sense since log returns are approximately equal to net returns when net returns are small. They have the same sign as well, and the magnitude of log returns is larger than net returns if both are positive. We choose to use log returns because they have well-defined domains while net returns are lower bounded by -1, which may cause some trouble. Also, log returns rather than net returns are normally used in practice.

In building an ARIMA model, the first step is to plot and explore the time series data to see if there are some specific features, such as trend, seasonal component, and changing variability. If trends exist, we can either do differencing or classical decomposition to remove them. Which method would work better depends on the data, and unit root has important implications for modeling. Checking whether the data is stationary or not is essential as well. The stationary process usually represents weak stationarity with unchanged first and second order moment properties over time. We use the Augmented Dickey-Fuller (ADF) test, which has the same testing procedure as the Dickey-Fuller test, but is applied to the model:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p-1} + \varepsilon_t$$

where is a constant, is the coefficient on a time trend, and p is the lag order of the auto regressive process. We apply ADF tests on the mean-corrected data for each sector to test the null hypothesis that a unit root is present. The alternative hypothesis is that the given time series is stationary. By performing ADF tests on all the five sectors, we get p-values of less than 5%, meaning we have enough evidence to reject the null hypothesis that there is a unit root. The test results can be found in the Appendix Table-6.1.1. Therefore, our time series mean-corrected data on XLI, XLV, XLK, XLF, XLE are stationary.

An Auto-Regressive Integrated Moving Average (ARIMA) model is characterized by three key components: The number of auto regressive terms (p), the number of differencing required to get stationary time series (d), and the number of the moving average terms (q). Due to the stationarity, we determine that d should be equal to zero, then we only need to focus on p and q. By looking at the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots (Appendix Figure-6.2.2), we can find some possible q values from the ACF plot and possible p values from PACF plots by counting lags out of the 95% confidence interval. We also use the AR function in R to determine the optimal p-value if treating the data as an auto-regressive model. After that, we can apply auto.arima function on mean-corrected data with d equals to zero, max.q equals to greatest possible q value that we find previously plus two, max.p equals to the greatest possible p value plus two, and IC equals to Akaike Information Criterion (AIC). When deciding max p/q values, we add another two orders to provide some flexibility. Given that d equals zero, the auto.arima function will search for all the possible combinations of p and q values and return the optimal one with the lowest AIC value. By performing the same procedure, we can get the optimal ARIMA models for all five sectors, and the optimal models can be found in Appendix Table-6.1.1.

After getting the optimal ARIMA models' parameters, we fit the model and use residuals from the optimal models to check whether the models are appropriate or not, which can be separated into two parts: test for randomness and test for normality. In the randomness part, we first plot the residuals, ACF and PACF plots (Appendix Figure-6.2.3), and they all look great. We perform the Box-Pierce test and Box-Ljung test for quantitatively testing randomness. In the test for normality part, we first plot the Q-Q plot (Appendix Figure-6.2.4) and find that the residuals have a significant probability of being not normally distributed. Then we use the Shapiro-Wilk normality test and Jarque Bera test for quantitatively testing the normality assumption. We use two different tests for each part to doubly verify the results by seeing whether they are consistent with each other or not. The corresponding p-values for each test can be found in the Appendix Table-6.1.1. When the p-value is less than or equal to 0.05, we have enough evidence to reject the null hypothesis (randomness/normality). If the p-value is greater than 0.05, we do not have enough evidence to reject the null hypothesis. Therefore, randomness assumptions can be satisfied while the normality assumptions are failed.

Now that we determined the ARIMA parameters used in the mean model for each sector fund, we should move on to the error distribution and the variance model. Specifically, for distribution model, we first assume normality for error term distributions and fit the GARCH model with the returns; then, as we choose to check the normality from the visualization, we plot the histogram for standardized residual values, which show that in each case errors turn out to be (skewed) t distribution, as they show fat tails and some of the distributions also show skewness. We then decide to change the original "norm" in the distribution model and set "sstd"/ "std" instead.

For variance model selection, we introduce information criteria (AIC and BIC). In general, information criteria balance the likelihood that the model fits the data and the number of parameters or the degree of model complexity:

$$IC = 2k - Pln(L)$$

with the number of parameters k, the penalty factor P and log-likelihood L. Based on the rule of thumb, we prefer the model with lower information criteria for each fund and determine the model this way. Based on the above, we choose the best GARCH specifications that fit the data.

3.2 Model Validation

3.2.1 Hypothesis tests for autocorrelation and ARCH effect

To test the autocorrelations and the ARCH effects, we used weighted Ljung-Box Tests for standard residuals and weighted ARCH LM tests for squared residuals, respectively.

The Box-Ljung test is a diagnostic tool used to test the lack of fit of a time series model. The test is applied to the residuals of a time series after fitting an ARMA(p,q) model to the data. Given a time series Y of length n, the test statistic is defined as:

$$Q = n(n + 2) \sum_{k=1}^{m} \frac{\hat{r}_k^2}{n-k},$$

where $\hat{r}_k^2$ is the estimated autocorrelation of the series at lag k, and m is the number of lags being tested. With a chosen significant level α, the Box-Ljung test rejects the null hypothesis and indicate that the model has significant lack of fit if $Q > \chi^2_{1-\alpha,h}$, where $h = m - p - q$ denotes degrees of freedom in chi-square distribution and suggests that the autocorrelations are very small.

However, an uncorrelated time series can still be serially dependent due to a dynamic conditional variance process and is said to exhibit conditional heteroscedasticity, which is also known as ARCH effect. Therefore, we also introduced the ARCH LM test, a Lagrange multiplier test to assess the significance of ARCH effects. Define the residual series $e_t = y_t - \widehat{\mu}_t$ and regression in squared residual $e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 +...+ \alpha_m e_{t-m}^2 + \mu_t$. The null hypothesis states no autocorrelation in the squared residuals, which is H_0: $\alpha_0 = \alpha_1 =... = \alpha_m = 0$ while the alternative hypothesis states that at least two squared residual terms are having autocorrelation.

3.2.2 Confidence Interval Coverage Rate

To evaluate the ultimate performance of selected models, we followed back testing techniques and constructed 95% confidence intervals to calculate the coverage rates of the models. We decided to use the first 8 years data (2011-2018) as training set and the last 2 years data as test set because more irregular market behaviors are contained in the latter years' range compared to former years' range. We used the parameter-settings discussed in the model selection section to estimate the corresponding GARCH models on a rolling window by refitting the model after every 5 days. The 'ugarchroll()' function in 'rugarch' package is used during the back testing process. In the following step, we used estimated means and standard errors of daily returns to fit confidence intervals beginning from the middle of 2018.

3.2.3 VaR Coverage

Also, we backtest our models from the VaR coverage perspective. For each series, we calculate the estimated VaR for the “in-sample” predicted data (rolling estimation is used) by using the chosen GARCH specifications, and then calculating the frequency of exceedances when the actual returns are lower than the predicted GARCH VaR. If the model is valid, then the VaR coverage should be close to the alpha set beforehand (in this project, the alpha is 5%).

3.3 Out-of-sample forecasting

Now that we define the proper GARCH model for each fund, we then start out-of-sample forecasting. Our plan is to forecast the volatility for the next 13 days based on the previous data value, namely the one-ahead forecast. ‘ugarchforecast()’ method in rugarch package helps us with it. To be more specific, we set the “n.roll” parameter (i.e. the number of times to roll the forecast) is set to be 13, as well as the “n.ahead”. It gives us both the future series and their volatility measured by sigma. Figure-6.1.1 shows the visualizations of the forecasting results. Based on our forecast, we also do the 5% VaR calculations.

4 Results

4.1 Model Selection

Figure-6.2.5 shows the (skewed) t distribution, as there is fat tails and skewness. The results are shown in Table-6.1.2 in Appendix. This table also shows the information criteria results. An interesting phenomenon during this process is the results show that all our selected returns fit the GJR GARCH instead of standard GARCH, and an economically significant reason for this fact is that financial returns usually show leverage-effect. This means news impact is asymmetric to conditional variance: bad news brings larger effects, which means the predicted variance will be higher after a negative surprise for financial returns. Leverage effect is also shown in Figure-6.1.2 in Appendix. When the prediction error is negative, which means there is bad news, the predicted variance will be bigger than when the error is positive. To be more specific, the GJR GARCH (1,1) model is:

$$R_t = \sigma_t \epsilon_t, \epsilon_t \sim N(0, 1)$$

$$\sigma_t^2 = \omega + (a + \gamma I_{t-1}) R_{t-1}^2 + \beta \sigma_{t-1}^2$$

with the log returns R_t, the dummy variable I_{t-1} is 0 if there is good news ($R_{t-1} > 0$) and 1 otherwise. Mathematically, for financial returns, bad news has a higher coefficient ($\alpha + \gamma$) multiplying the squared returns, and γ before the dummy variable I_{t-1} is statistically significant for most of the time.

4.2 Model Validation

For the log return data, there is no significant statistical evidence showing signs of autocorrelation and ARCH effect in our models, indicating no obvious lack of fit. Detailed test results are included in the Appendix 6.1 Table-6.1.3.

For CI coverage, all coverage ratios are around 92% - 94%, which are close to the 95% threshold. The plot of the CI coverage intervals is shown in Table-6.2.6 in Appendix 6.2. It is shown that during the initial months starting from the middle of 2018, since the volatility level in the market is relatively low, the prediction interval has a narrower band width. Since the beginning of 2020, with an increasing amount of unpredicted volatility, the prediction band width increases significantly to account for the changes, and gradually settles and shrinks in size after late 2020. The performance of our models could potentially be improved if we set the refit window to a shorter length. Currently we refit the model every 5 days for easier converges of the models.

For VaR coverage, we found that all the coverage ratio is close to 5%, which is the significance level we choose, as shown in Table-6.1.3 in Appendix 6.1, and Figure-6.2.7 in 6.2 shows the VaR coverage for XLV data. As a result, the VaR coverage indicates the models are valid.

4.3 Forecasting Future data

As Figure-6.1.1 in Appendix shows, for XLV, XLI, and XLF data, we expect an upward trend in their volatility in the next 13 days, while a reverse trend for XLK and XLE. On the other hand, comparing the absolute value of the sigma, for the next 13 days, XLE, XLK and XLF are suffering bigger volatility. This means energy, technology and financial industries are going to be more volatile than healthcare, industrial and financial sectors in a short future. Although we forecast for a short period, the results do resonate with some facts in history and may inspire the future investment decision-making process. In 2020, there were two big events that hit the U.S. stock markets: COVID-19 pandemic and the presidential election. For each event we analyze the volatility during a specific period (3/2/2020-5/29/2020 for the pandemic and 10/30/2020-11/6/2020 for the election), the data shows that in general XLE, XLK, and XLF returns perform more unstable than their counterparts, as shown in Table-6.1.5 in Appendix 6.1. We conclude that energy, technology, and financial sectors stocks tend to underperform when the stock markets are facing risks. To forecast the volatility for a longer period, we should introduce the mean-reversion characteristic for variance. For GARCH (1,1):

$$R_t = \mu + e_t, e_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2$$

the long-run variance equals $\frac{\omega}{1-\alpha-\beta}$ when $\alpha + \beta < 1$(i.e., the GARCH (1,1) is covariance stationary). Since we are using ARMA-GARCH for the sector funds returns, we do expect different mean-reverting patterns for their volatility. Although not knowing the exact formula to calculate the long-run stable volatility, using the sample volatility from the ten-year historical data is a reasonable alternative. For example, from the observations of XLV for the past ten years, we set its reasonable long-run variance to be 0.011%, so the sigma (sample standard deviation) is 1.043%. In our out-of-sample forecast, for the next 13 days, the XLV volatility is below this long-run level, so we expect the future returns to be more volatile than the short-term prediction does.

5 Conclusion

Volatility is an essential topic in investment and future decision-making. Knowing how different sectors react when exposed to risks can help take investment opportunities. In this project, we select proper ARMA-GARCH models for the SPDR funds in Healthcare, Technology, Financial, Energy, and Industrial sectors. With these sector funds, we can see the picture of different sector stocks performance. The model selection part involves determining ARIMA orders for each fund, defining distribution for standard residuals and variance models. For the model validation part, our models pass hypothesis testing for autocorrelation and ARCH effect, and we prove their validness by checking the CI coverage and VaR coverage with rolling estimations using in-sample data. The forecast shows Energy, Technology and Financial are more volatile when exposed to events (pandemic and presidential election), and we use sample standard deviation to approximate the long-run stable volatility to predict the future movement of sector funds sigma. The future study will focus on finding a better mean-reversion formula for ARMA-GARCH volatility and using machine learning methods (e.g., LSTM) to predict future data and compare it with our results in this report.

6 Appendix

6.1 Figures and Tables in Results Section

Funds Sector	Augmented Dickey-Fuller Test (p-value)	Best Model	Test for Randomness (p-value)		Test for Normality (p-value)	
			Box-Pierce Test	Box-Ljung Test	Shapiro-Wilk Normality Test	Jarque Bera Test
XLV	0.01	ARIMA(6,0,5)	0.9208	0.9183	< 2.2e-16	< 2.2e-16
XLI	0.01	ARIMA(4,0,5)	0.3448	0.3371	< 2.2e-16	< 2.2e-16
XLK	0.01	ARIMA(0,0,4)	0.7203	0.714	< 2.2e-16	< 2.2e-16
XLF	0.01	ARIMA(3,0,5)	0.8593	0.855	< 2.2e-16	< 2.2e-16
XLE	0.01	ARIMA(6,0,6)	0.994	0.9938	< 2.2e-16	< 2.2e-16

Table-6.1.1 Optimal ARIMA Model parameters with test statistics by sectors

Funds Sector	Distribution Model	Variance Model	AIC-sGARCH	AIC-gjrGARCH	BIC-sGARCH	BIC-gjrGARCH
XLV	t distribution	gjr-GARCH	-6.635318	-6.649919	-6.598228	-6.610511
XLI	t distribution	gjr-GARCH	-6.382521	-6.404804	-6.350068	-6.370033
XLK	skewd t distribution	gjr-GARCH	-6.37118	-6.388925	-6.347999	-6.363425
XLF	skewd t distribution	gjr-GARCH	-6.161593	-6.174483	-6.129139	-6.139711
XLE	t distribution	gjr-GARCH	-5.756375	-5.773557	-5.716967	-5.731831

Table-6.1.2 Selected Distribution/Variance model parameters and IC values by sectors

	XLV	XLI	XLK	XLF	XLE
CI Coverage Rate	93.25%	93.10%	93.70%	92.95%	92.80%
VaR Coverage	6.17%	6.50%	6.54%	5.70%	7.12%

Table-6.1.3 Coverage Rates by sectors

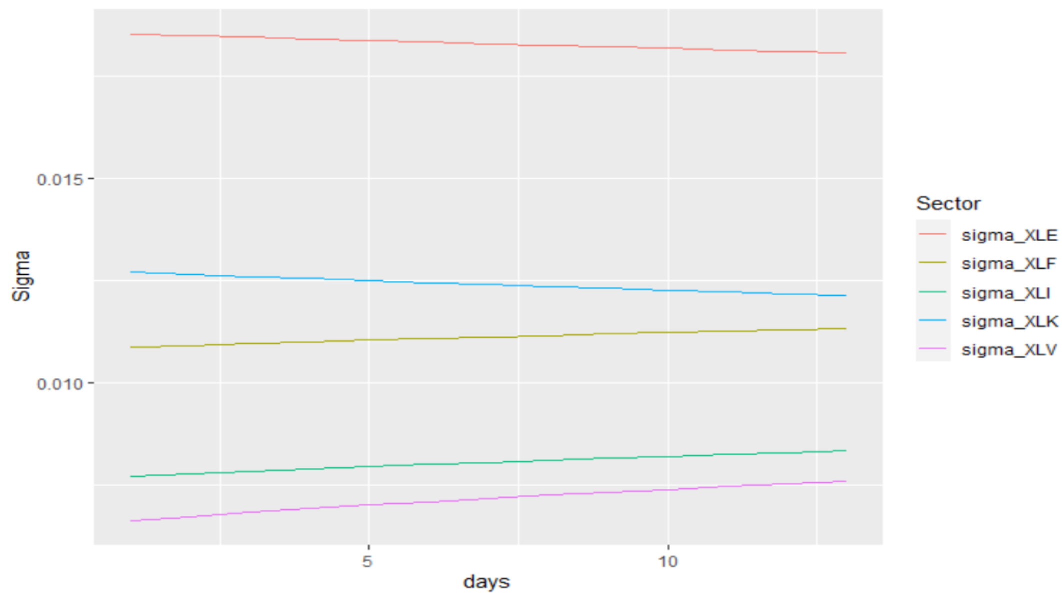


Figure-6.1.1 forecasted sigma by sector for the next 13 days

Funds Sector	Sector	Pandemic	Election
XLV	Healthcare	0.0258175	0.0194584
XLI	Industrial	0.03459038	0.01208321
XLK	Technology	0.03491854	0.02126747
XLF	Financial	0.04175718	0.01520514
XLE	Energy	0.05446975	0.02719546

Table-6.1.5 Funds standard deviations for different time period

Note: 3/2/2020-5/29/2020 for the Pandemic and 10/30/2020-11/6/2020 for the Election

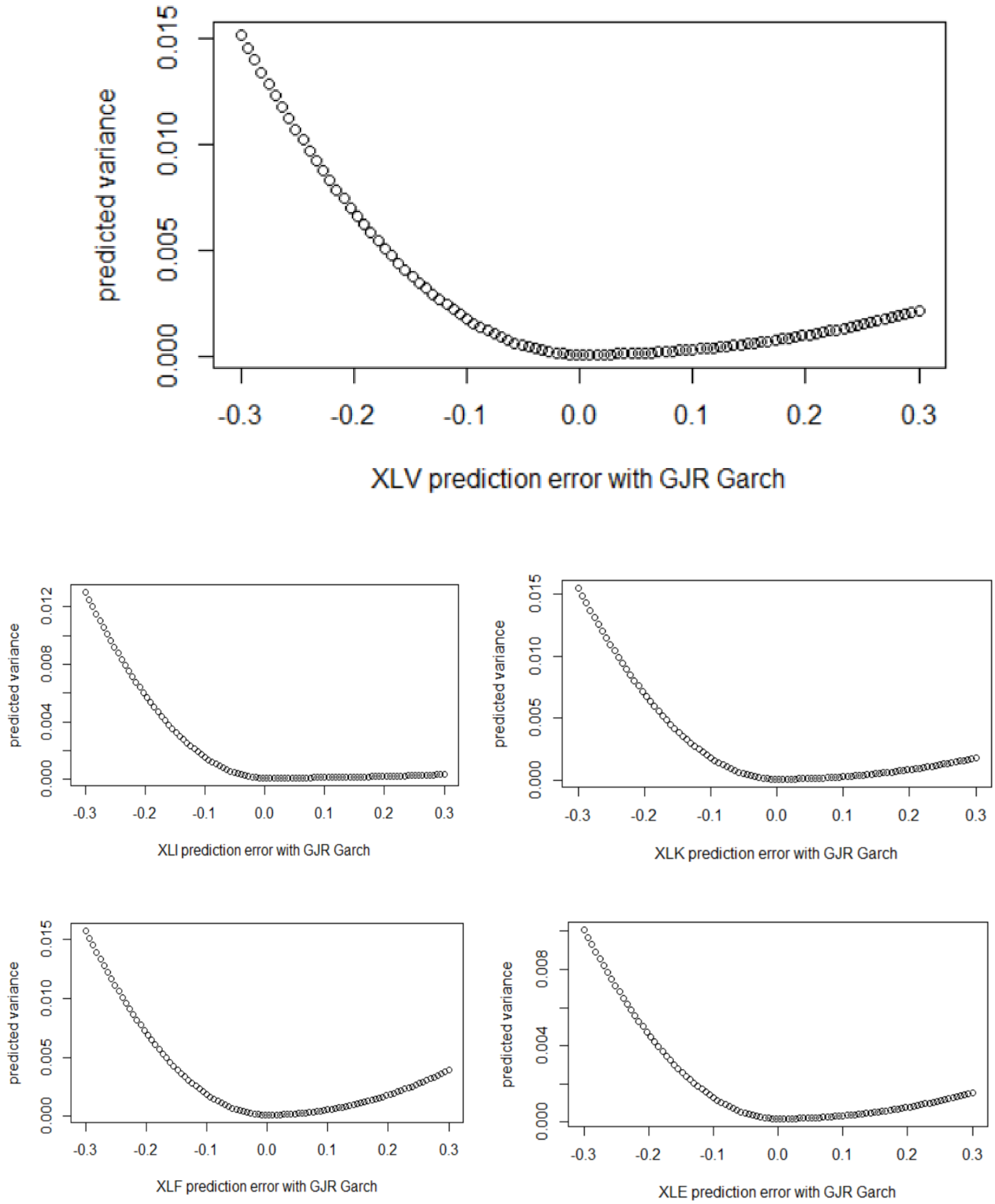


Figure-6.1.2 Visualization of GJR-GARCH Leverage Effect

6.2 Additional Figures Used in the Project

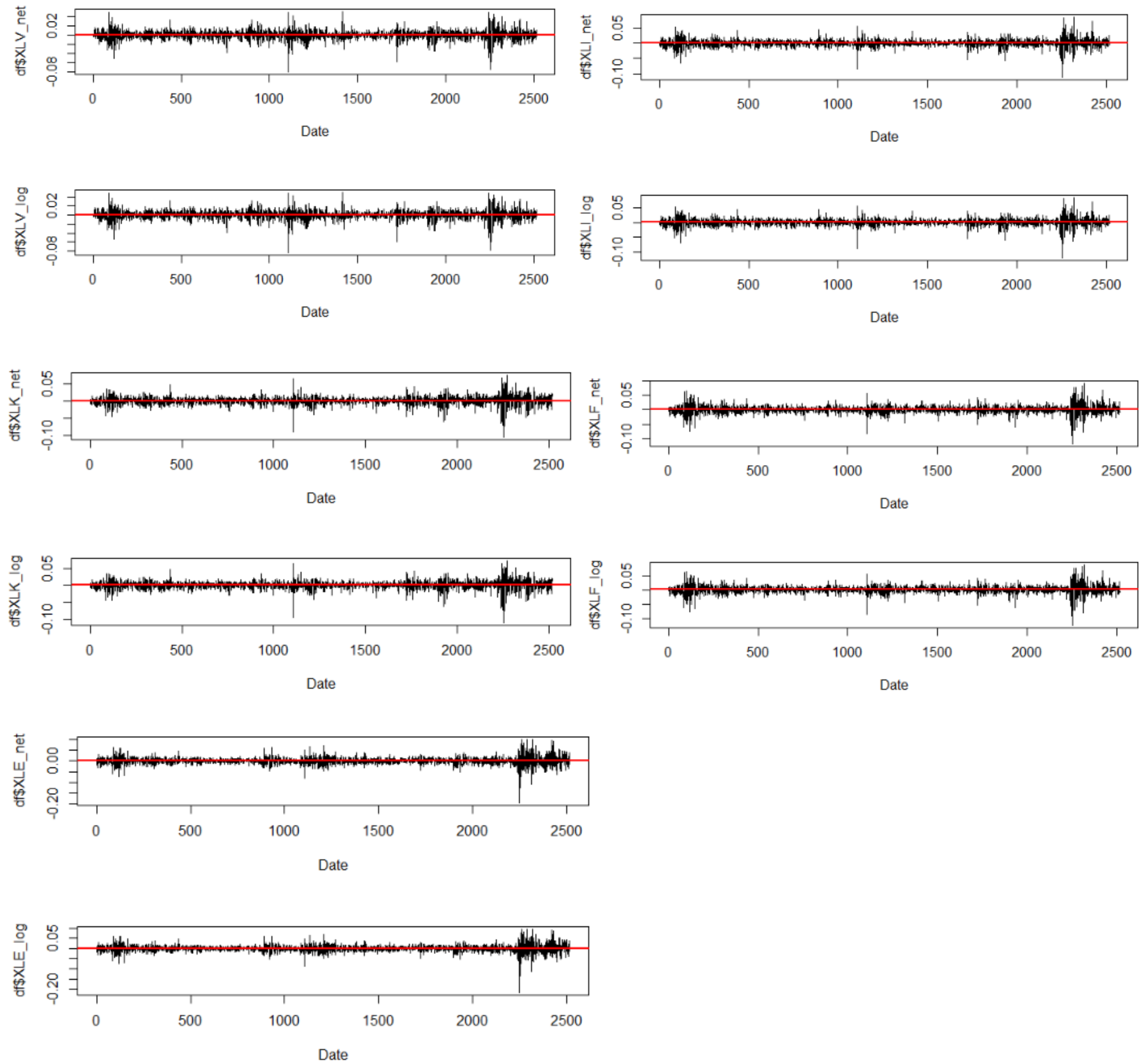


Figure -6.2.1 Net Return vs Log Return

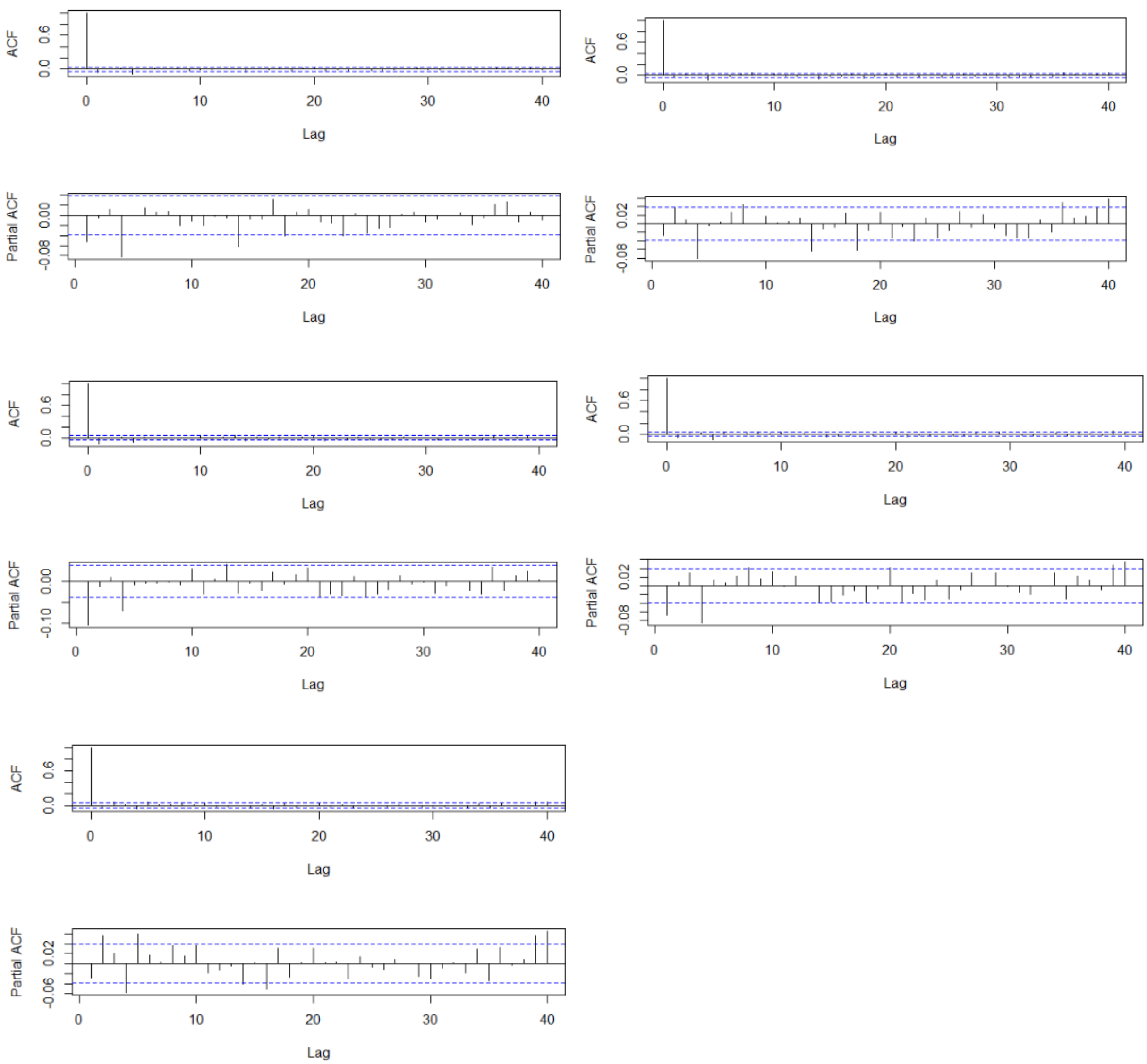
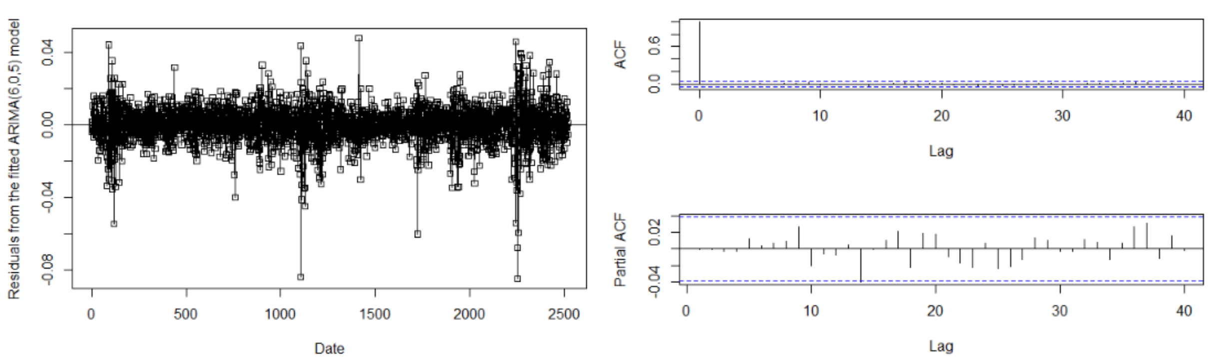


Figure-6.2.2 ACF & PACF



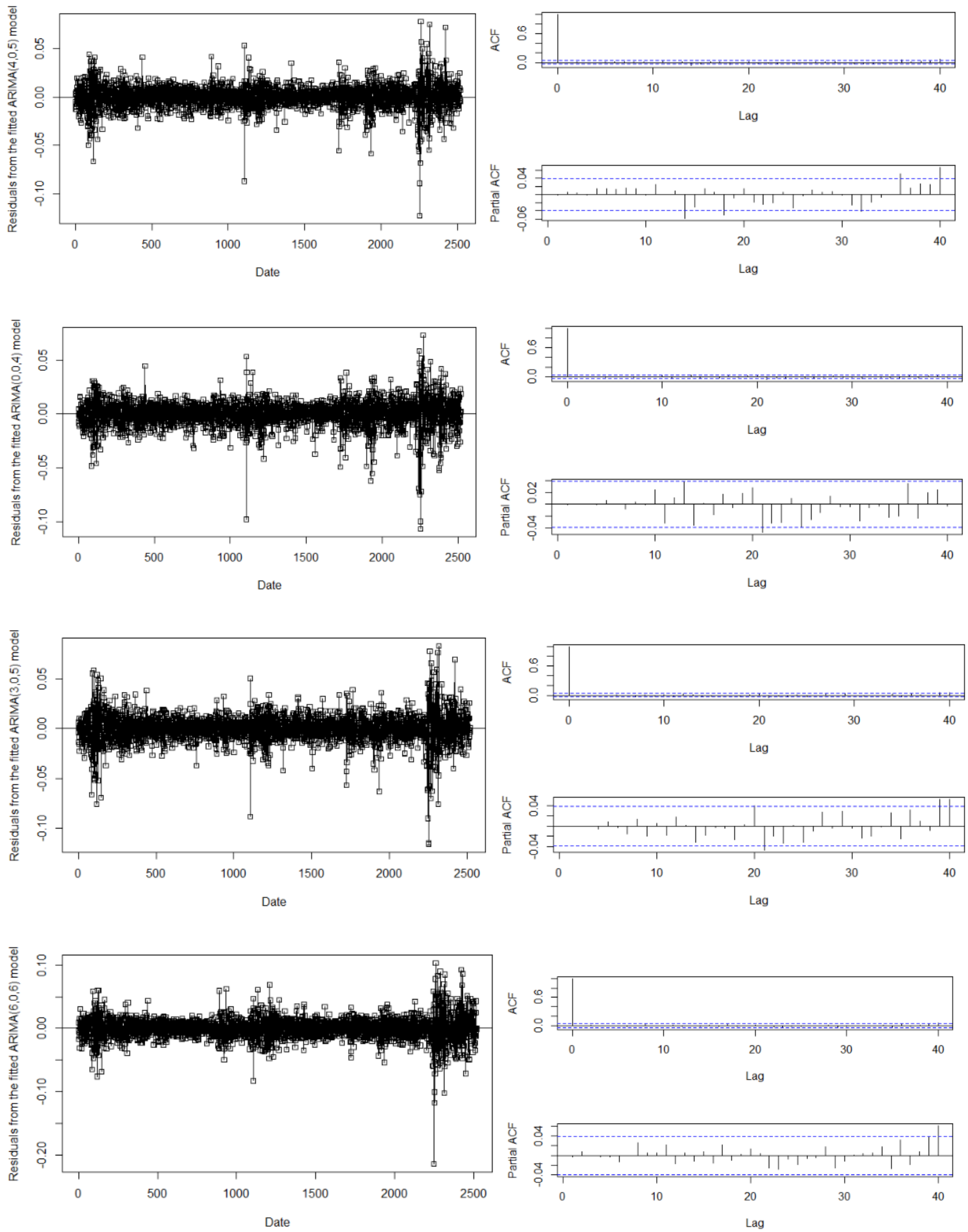


Figure-6.2.3 Residuals; Residuals' ACF & PACF

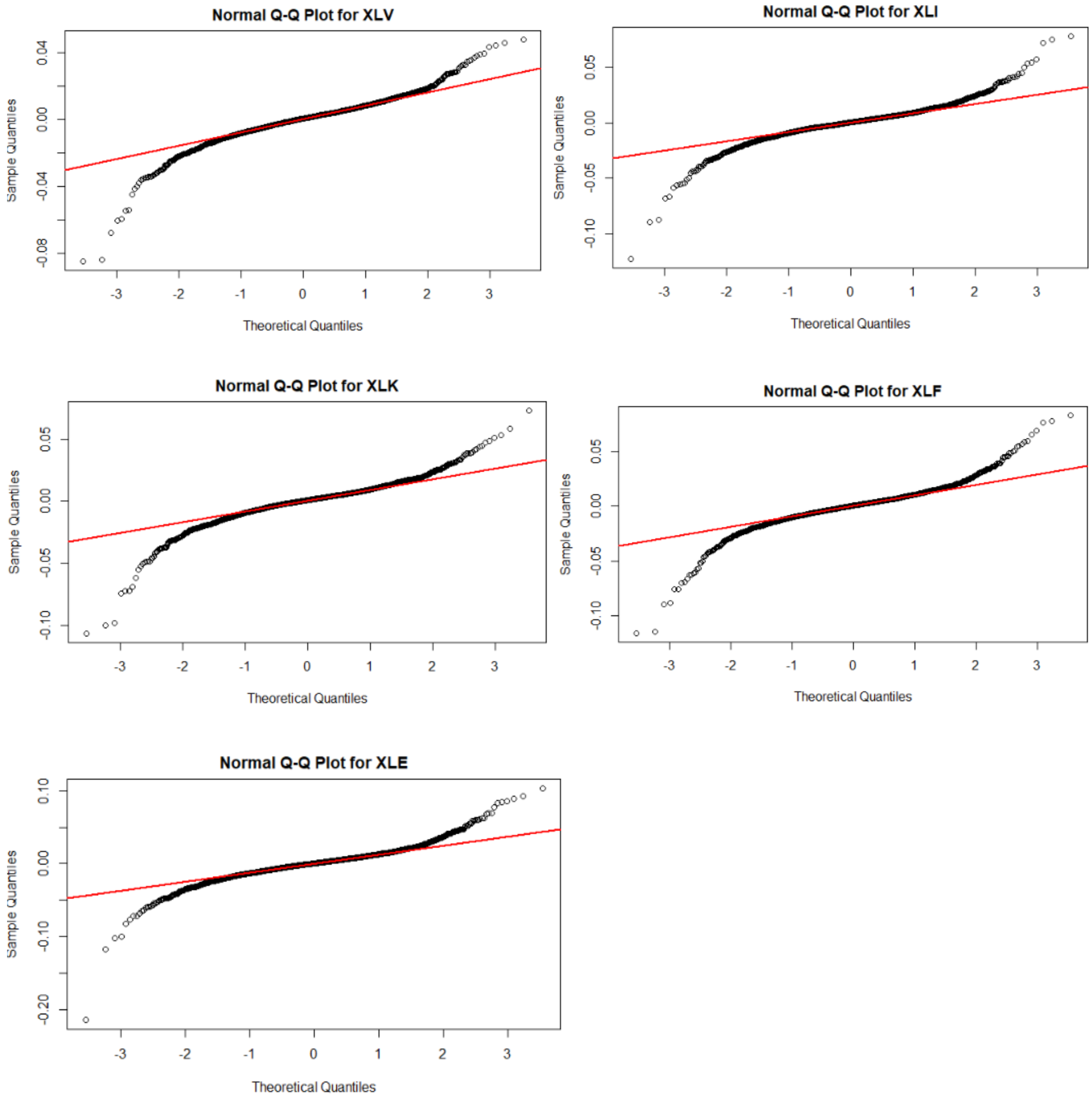


Figure-6.2.4 Normal Q-Q Plots

Note: For Figure-6.2.1 to Figure-6.2.4, the corresponding sectors from left to right are XLV, XLI, XLK, XLF, XLE.

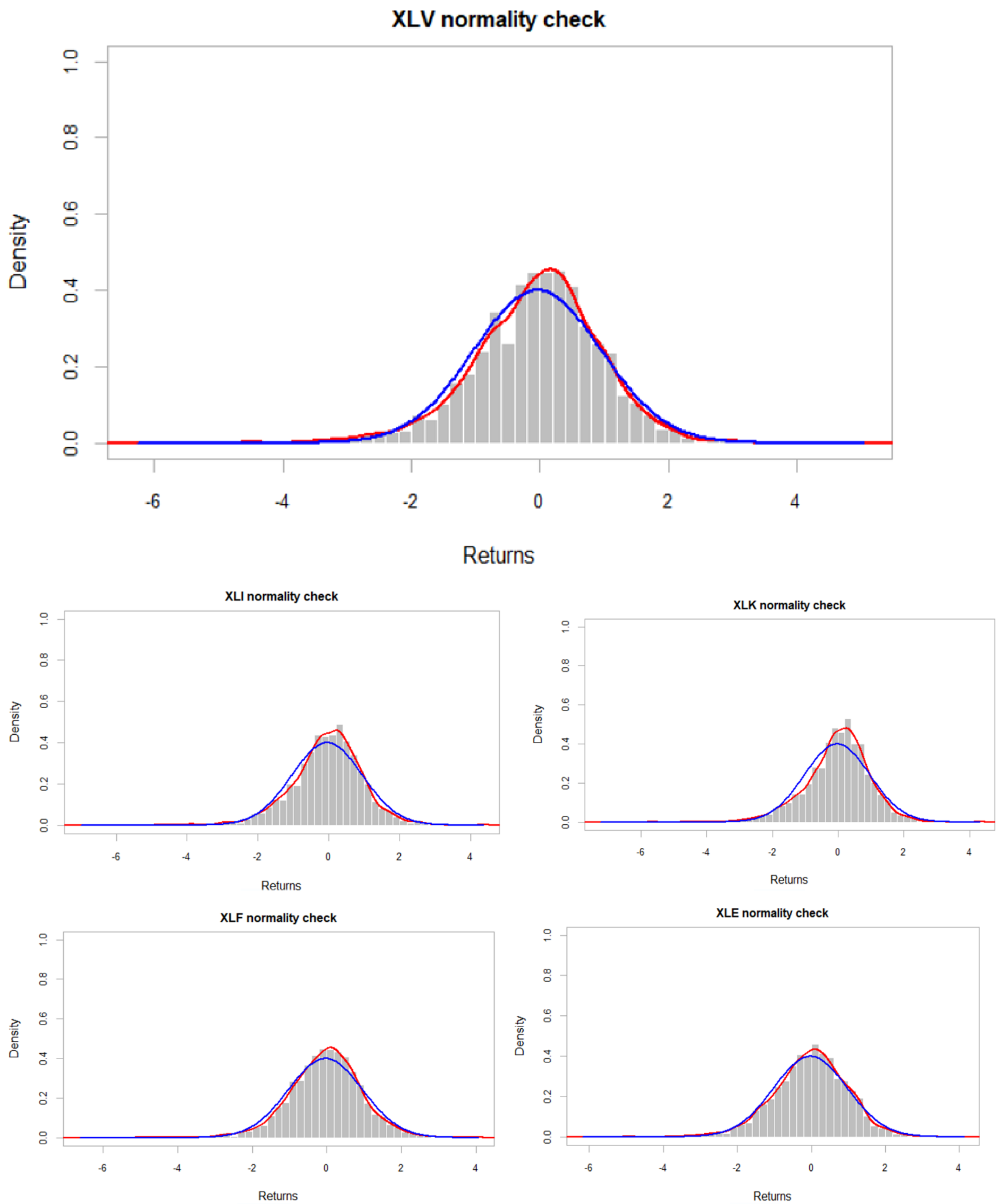


Figure-6.2.5 Histogram for standardized residual values comparing to standard normal density

Note: blue line describes normal distribution, red line represents the specific fund

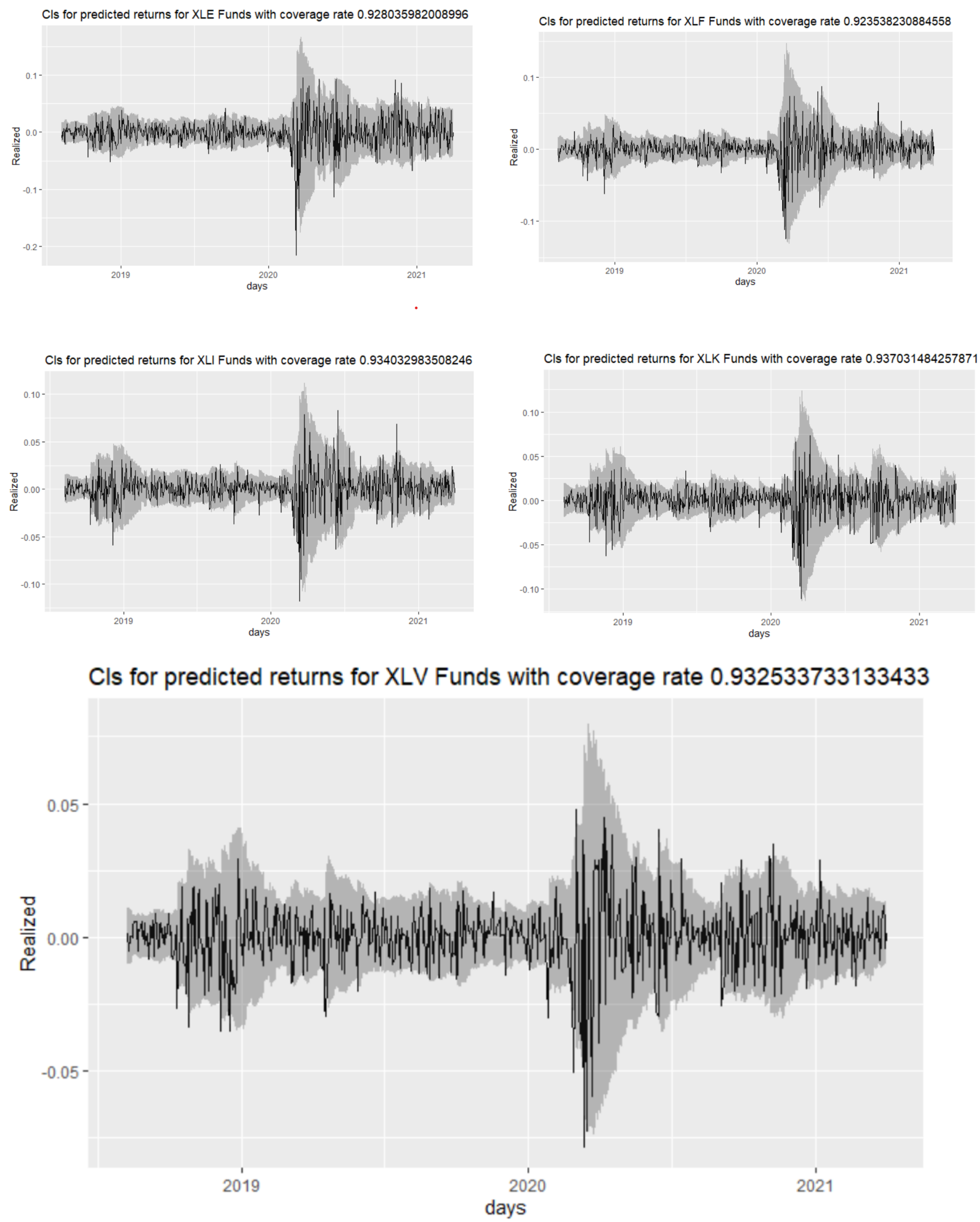


Figure-6.2.6: 95% Confidence Interval Coverage Plots

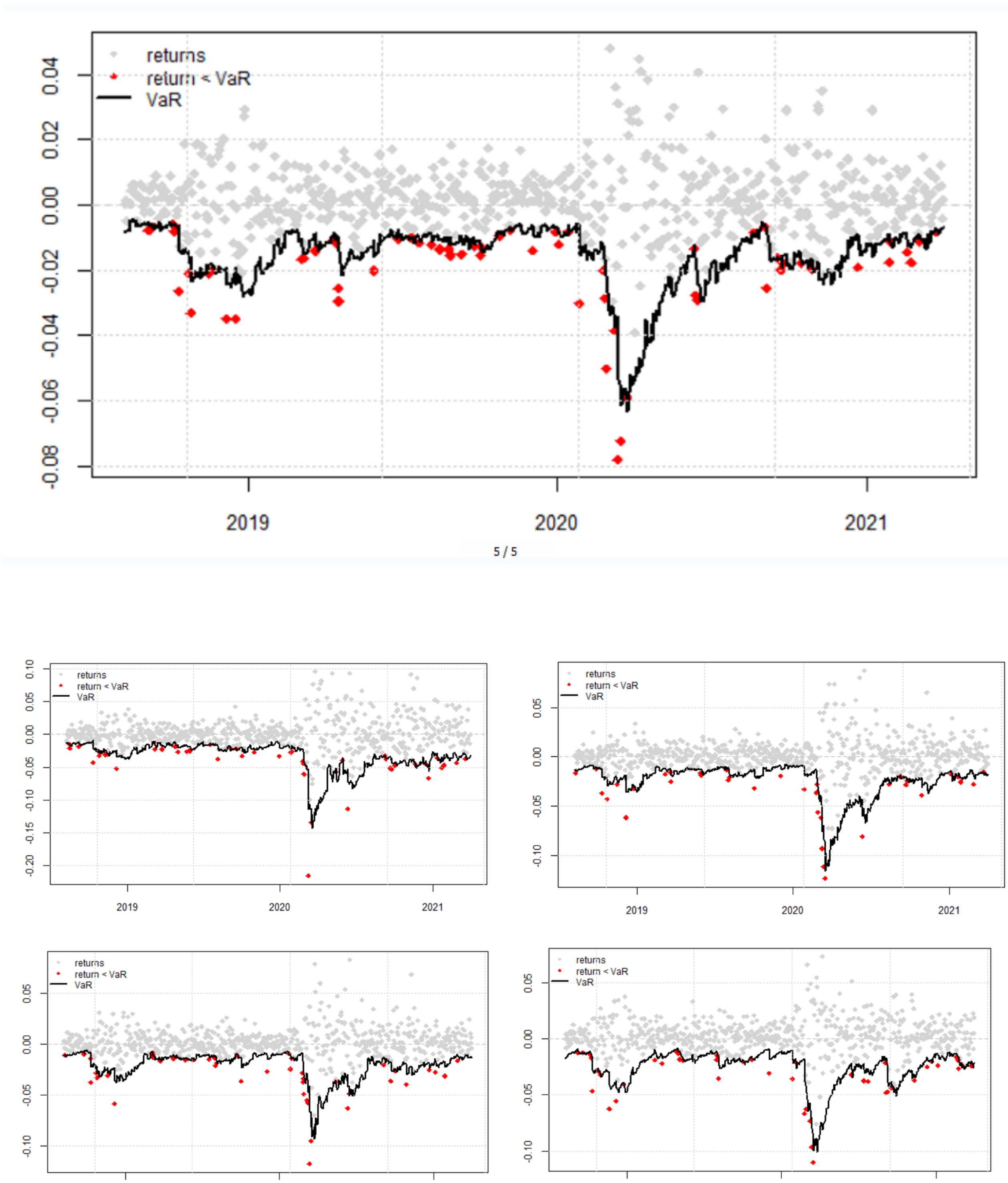


Figure-6.2.7 Visualization of the VaR Coverage (from top to bottom, from left to right in order of XLV, XLE, XLF, XLI, XLK)

6.3 Computer Codes

Please check the coding in our zip file.

6.4 References

1. D B Nugroho (2019). *Empirical Performance of GARCH, GARCH-M, GJR-GARCH and log-GARCH Models for Returns Volatility.* J. Phys.: Conf. Ser. 1307 012003
2. I M Md Ghani and H A Rahim (2019). *Modeling and Forecasting of Volatility using ARMA-GARCH: Case Study on Malasia Natural Rubber Price.* IOP Conf. Ser.: Mater. Sci. Eng. 548 012023
3. Marius Hofert (2021) *Fitting and Predicting VaR based on an ARMA-GARCH Process*. Retrieved from https://cran.r-project.org/web/packages/qrmtools/vignettes/ARMA_GARCH_VaR.html
4. Dueker, M. (1997). *Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility*. Journal of Business & Economic Statistics, 15(1), 26-34. doi:10.2307/1392070
5. Rizwan Raheem Ahmed, Jolita Vveinhardt, Dalia Streimikiene & Zahid Ali Channar (2018). *Mean reversion in international markets: evidence from G.A.R.C.H. and half-life volatility models*. Economic Research-Ekonomska Istraživanja, 31:1, 1198-1217, doi: 10.1080/1331677X.2018.1456358