# 1 Introduction

In recent years, as economies have rapidly developed, the interconnections among various financial assets have become increasingly close, with their structures growing more complex and diverse. The frequent occurrence of financial crises has drawn heightened attention from scholars both domestically and internationally to the relationships among financial assets. Traditionally, researchers studying the distribution of financial time series returns have assumed these to follow a normal distribution. However, a plethora of empirical evidence has demonstrated that the assumption of normality does not hold, as financial time series often exhibit skewness, kurtosis, and heavy tails. Furthermore, the relationships among different financial assets are not linear, hence the complex interdependencies between two financial assets cannot be adequately measured using simple correlation coefficients. Additionally, the relationships among financial assets evolve over time.

The onset of the Russia-Ukraine conflict in February 2022 marked a significant geopolitical event with widespread implications, not least on global financial markets, including those of Hong Kong. This conflict provides a pertinent backdrop for examining changes in the financial sector due to its potential to disrupt economic stability and influence market behaviors drastically. Hong Kong's financial markets, known for their dynamic nature and integral role in the global financial landscape, present a unique case study for observing the effects of such geopolitical tensions on sector-specific indices like the Hang Seng Composite Index's Energy and Financial sectors.

The use of Copula functions addresses the limitations observed in traditional models by allowing the study of dependencies between financial assets without the need to assume specific types of distribution functions for asset returns. This overcomes the traditional model's shortfall, which requires solving the joint distribution function's density function of two sequences. Copulas can measure the non-linear relationships between financial time series and are capable of adapting to changes over time with time-varying Copula models.

This report investigates the impact of the Russia-Ukraine conflict on the Hang Seng Composite Index sectors, specifically focusing on the Energy and Financial sectors, using a combination of time-series models and Copula functions. By doing so, it aims to reveal not only the changes in market behavior pre- and post-conflict but also the evolving interdependencies between these crucial sectors, providing insights into the broader implications of geopolitical conflicts on financial markets.

# 2 Literature Review

The concept of Copula functions was first introduced by Sklar in 1959. He discovered that any joint distribution function could be decomposed into its marginal distributions and a Copula function, which serves as a connection function to describe the correlations between different variables [1]. Initially, due to various constraints, Copula functions did not receive significant attention until nearly four decades later.

In 1998, the significance of Copulas was brought into the spotlight by Nelsen in his seminal book, "An Introduction to Copula", which elaborated on the properties and mathematical foundations of Copula functions [2]. The following year, Embrechts pioneered the application of Copula models in finance, using them to measure correlations between different financial return series [3]. This marked a pivotal moment in the integration of Copula functions into financial analysis. In 2001, Patton explored the application of Copula functions in time-series modeling, suggesting the integration of GARCH models with Copula models to form Copula-GARCH models for studying financial time series [4]. This was further advanced by Jondeau and Rockinger in 2006, who demonstrated that t-distribution Copula functions could better capture the correlation between financial indices [5]. The same period saw the application of ARMA-GARCH models combined with Copula functions by Roch and Alcgre to study the dependencies in the Spanish stock market [6]. The field saw significant advancements in 2017 with Mokni and Mansouri, who constructed GARCH-t-Copula models for estimating market risks based on relationships between major international stock markets [7]. Patton's work on dynamic Copula models particularly highlighted the changing correlations between the yen and the dollar, and the pound and the dollar around the introduction of the euro system [8-12]. The necessity of incorporating time-varying Copula functions to account for temporal changes in conditional correlations was advocated by Andersen and Bauwens [13]. This was supported by empirical studies such as those by Garcia and Tsafack, who investigated the interrelationship between bond and stock markets of two countries using Copula functions [14], and Taylor and Bartram, who applied time-varying Copula models to European stock market data [15].

Compared to international research, China's engagement with Copula functions started later but has gained momentum rapidly in recent years. Zhang Yaoting first used Copula functions in 2002 while exploring correlation metrics, and further elaborated on the definition and properties of Copula functions [16]. We Yanyan and Zhang Shiying in 2004, and subsequently Liu Xibo and others, demonstrated the effectiveness of t-Copula in fitting the daily return series of Shanghai and Shenzhen stock markets [19]. More recently, Fu Qiang and Li Zhe used a time-varying SJC-Copula function to study the tail dependencies in these markets, finding an increasing trend in their interdependence and significant lower tail correlations [21]. Song Xue-lian[22] uses a Copula-GARCH model to demonstrate the impact of crude oil price fluctuations on the stock market, important for risk management and economic planning. Cong Yingnan[23] identifies significant interdependencies among China's financial assets using the Copula Approach. Wang Yiming[24] shows long-term positive correlations between Chinese and Western stock markets, aiding in investment strategies with a Wavelet Analysis and GARCH-Copula Method. Zhao Haitao[25] discusses the biased and asymmetric correlation between China's gold and silver futures markets using a GJR-MRS-SJC-Copula Model, providing insights for investors and policymakers.

This literature review establishes a robust foundation for understanding how Copula functions have evolved and been applied across different facets of financial analysis, highlighting their importance in capturing dynamic and complex dependencies in financial markets.

# 3 Questions

## 5.1 General Analysis

### 5.1.1 Financial Market

In this section, the weekly closing prices of Hang Seng Composite Index-Financial Industry from February 1, 2020 to February 29, 2024 are selected as the original data, a total of 213 sets of data. After the data conversion of closing prices with , from February 1, 2020 to February 29, 2024 is selected as the training data of seasonal time series analysis, and from March 1, 2024 to April 30, 2024 is selected as the test data of seasonal time series analysis, with 212 sets of training data and 9 sets of test data, where represents the closing price.

The following figure reflects the trend of the closing price and the fluctuation of the return rate of the index since 2020. It can be obviously observed that the closing price has a sharp decline in 2020, 2021 and the middle of 2022. The logarithmic return rate shows an obvious seasonal effect and a certain synchronizing in general. We will first choose to decompose the seasonal effect of this data in 5.1.2.

图表, 折线图

描述已自动生成

折线图

描述已自动生成

Next, we make a statistical analysis of the log return rate. The following figure is the frequency histogram of the log return rate. It can be seen that the log return rate has a high peak, and is biased and fat-tailed. The skewness is -0.06924217<0, the kurtosis is 3.066223>3, showing left-skewed and spiky properties.

图表, 直方图

描述已自动生成

#### Seasonality Checking

After the seasonal effect decomposition of log returns, we observe that this time series has a clear seasonal effect with a period of 52.

文本

中度可信度描述已自动生成

#### Data Distribution Test

First, we test the distribution of the log rate of return to determine the distribution of the data, so as to pave the way for the fitting of the ARMA model.

|  |  |  |
| --- | --- | --- |
|  | KS test | AD test |
| norm | 0.6064 | 2.83e-06 |
| t | 0.6377(df=65.137, m= -0.001080151  , s= 0.02740782) | 0.7061(df=65.137) |

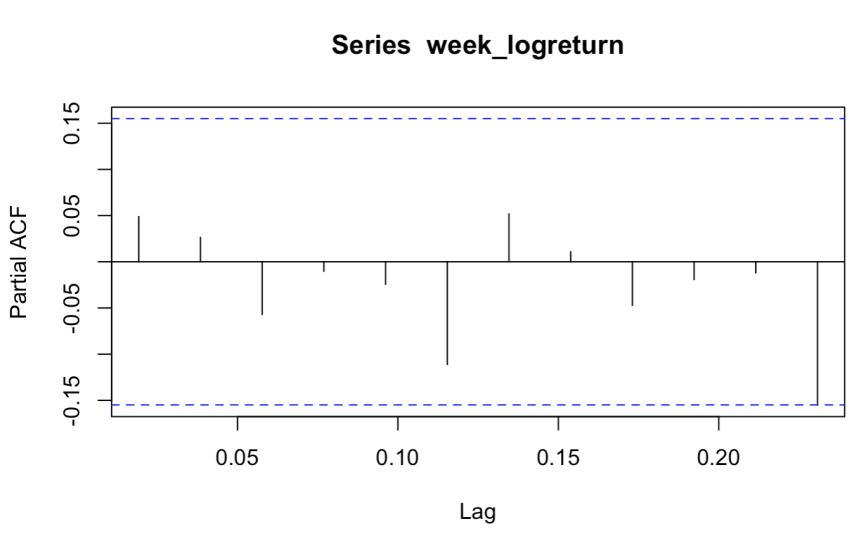
Therefore we consider the log return series to follow a T-distribution.

#### Stationary Test

After the 52-order difference of the data, we use the adfTest function in the fUnitRoots package to test the stationarity of the data.

图表

描述已自动生成



#### Model Selection

We tried many seasonal ARIMA models, and finally got five different models after modification, and tested the residuals of the models. In the case of passing the residual test, we compared their AIC. According to the principle of minimum AIC value, The final selected model is ARMA(0,0)\*seasonal(0,0,1) without mean.

|  |  |  |
| --- | --- | --- |
| Model | AIC | p |
| ARMA(0,0)\*seasonal(0,0,1) without mean | -622.71 | 0.9712547 |
| ARMA(0,0)\*seasonal(1,0,0) without mean | -621.01 | 0.9321794 |
| ARMA(1,1)\*seasonal(0,0,1)  without mean | -620.08 | 0.9616206 |
| ARMA(3,3)\*seasonal(1,0,0)  without mean and ma2=0 | -614.41 | 0.8789814 |

After that, we conduct ARCH test on the data, p-value=0.7394, and there is no heteroscedasticity, so this section does not fit the GARCH model on the data.

#### Model prediction

Based on the above analysis, we selected the model with the best effect, namely ARMA(0,0)\*seasonal(0,0,1) without mean, and used this model to forecast the next nine periods and compare it with the real value. The following table shows the predicted value and the real value, and at the same time, we visualized the table:

|  |  |  |
| --- | --- | --- |
| Date | Estimator | True value |
| 2024/3/3 | 0.005472210 | -0.006347363 |
| 2024/3/10 | 0.004010567 | -0.001459779 |
| 2024/3/17 | 0.003177568 | -0.003362752 |
| 2024/3/24 | -0.002383072 | -0.006858488 |
| 2024/3/31 | 0.002263810 | 0.00594277 |
| 2024/4/7 | -0.018490888 | -0.004698289 |
| 2024/4/14 | -0.009492635 | -0.002842948 |
| 2024/4/21 | -0.006904419 | 0.027988764 |
| 2024/4/28 | 0.024714593 | 0.011280416 |

图表, 折线图

描述已自动生成

As can be seen from the above table and the figure above, our prediction is very close to the real value, and the prediction interval completely covers the real value, which proves the practicality of the model.

### 5.1.2 Energy Market

In this section, the weekly closing prices of Hang Seng Composite Index-Energy Industry from February 1, 2020 to February 29, 2024 are selected as the original data, a total of 213 sets of data. After the data conversion of closing prices with , Rt from February 1, 2020 to February 29, 2024 is selected as the training data of seasonal time series analysis, and R\_t from March 1, 2024 to April 30, 2024 is selected as the test data of seasonal time series analysis, with 212 sets of training data and 9 sets of test data, where Pt represents the closing price.

The following figure reflects the trend of the closing price and the fluctuation of the return rate of the index since 2020. It can be obviously observed that the closing price has kept fluctuating and rising since 2021. The logarithmic return rate shows an obvious seasonal effect and a certain synchronizing in general. We will first choose to decompose the seasonal effect of this data.

图形用户界面, 图表, 散点图

描述已自动生成

Energy Fig 1

` Next we make a statistical analysis of the log return rate. The following figure is the frequency histogram of the log return rate. It can be seen that the log return rate has a high peak, and is biased and fat-tailed. The skewness is -0.4179<0, the kurtosis is 4.7626>3, showing left-skewed and spiky properties.

图表, 直方图

描述已自动生成

Energy Fig 2

#### Seasonality Checking

After the seasonal effect decomposition of log returns, we observe that this time series has a clear seasonal effect with a period of 52.

文本, 信件

描述已自动生成

Energy Fig 3

#### Data distribution test

First, we test the distribution of the log rate of return to determine the distribution of the data, so as to pave the way for the fitting of the ARMA model.

|  |  |  |
| --- | --- | --- |
|  | KS test | AD test |
| norm | 0.6518 | 0.263 |
| t | 0.6377(df=7.884, m= -0.0022  , s= 0.0338) | 0.9943(df=7.884) |

Therefore we consider the log return series to follow a T-distribution more likely than norm distribution.

#### Stationary Test

After the 52-order difference of the data, we use the adfTest function in the fUnitRoots package to test the stationarity of the data. The p value is 0.01 which means that the series id stationary.

图表, 直方图

描述已自动生成

Energy Fig 4

#### Model Selection

We tried many seasonal ARIMA models, and finally got four different models after modification, and tested the residuals of the models. In the case of passing the residual test, we compared their AIC. According to the principle of minimum AIC value, The final selected model is ARMA(0,0)\*seasonal(0,0,1) without mean.

|  |  |  |
| --- | --- | --- |
| Model | AIC | p |
| ARMA(0,0)\*seasonal(1,0,0) without mean | -523.5 | 0.8124 |
| ARMA(0,0)\*seasonal(0,0,1) without mean | -517.59 | 0.4835 |
| ARMA(1,1)\*seasonal(0,0,1)  without mean | -518.77 | 0.6733 |
| ARMA(1,1)\*seasonal(1,0,0)  without mean | -524.58 | 0.9220 |

After that, we conduct ARCH test on the data, p-value=0.8013, and there is no heteroscedasticity, so this section does not fit the GARCH model on the data.

#### Model prediction

Based on the above analysis, we selected the model with the best effect, namely ARMA(1, 1)\*seasonal(1, 0, 0) without mean, and used this model to forecast the next nine periods and compare it with the real value. The following table shows the predicted value and the real value, and at the same time, we visualized the table:

|  |  |  |
| --- | --- | --- |
| Date | Estimator | True value |
| 2024/3/3 | -0.04122104 | 0.05262229 |
| 2024/3/10 | -0.01572714 | -0.01193646 |
| 2024/3/17 | 0.03652257 | -0.00274599 |
| 2024/3/24 | 0.01666539 | 0.01656787 |
| 2024/3/31 | 0.00384030 | 0.03889836 |
| 2024/4/7 | -0.01249377 | 0.01776898 |
| 2024/4/14 | -0.03547989 | 0.00430402 |
| 2024/4/21 | -0.01263808 | 0.01165081 |
| 2024/4/28 | 0.00155090 | -0.00421109 |

图表, 折线图

描述已自动生成

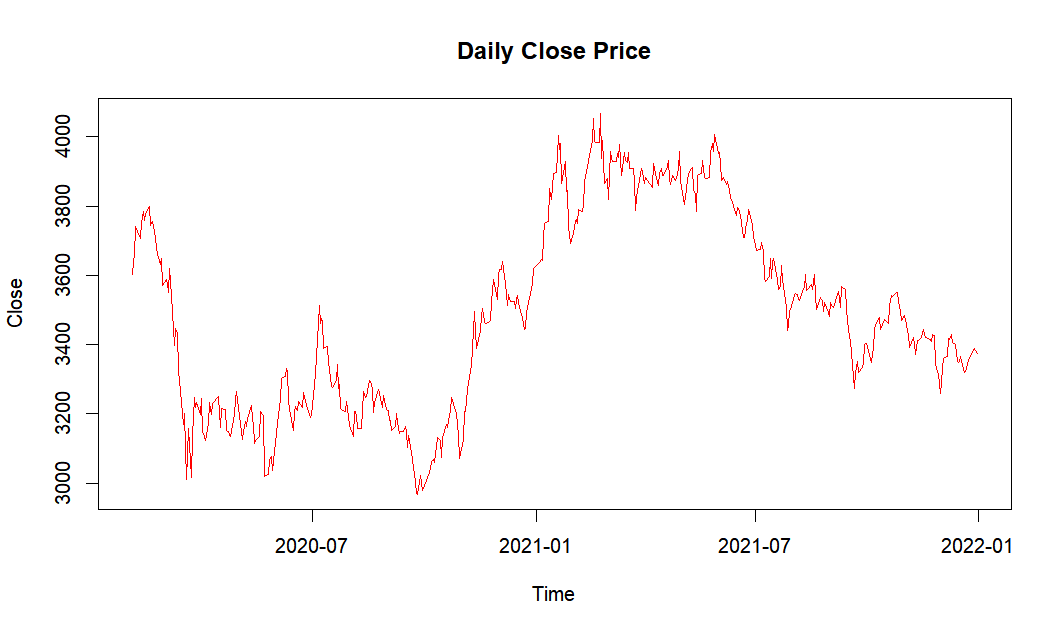
As can be seen from the above table and the figure above, our prediction is very close to the real value, and the prediction interval completely covers the real value, which proves the practicality of the model.

## 5.2 Model Before War

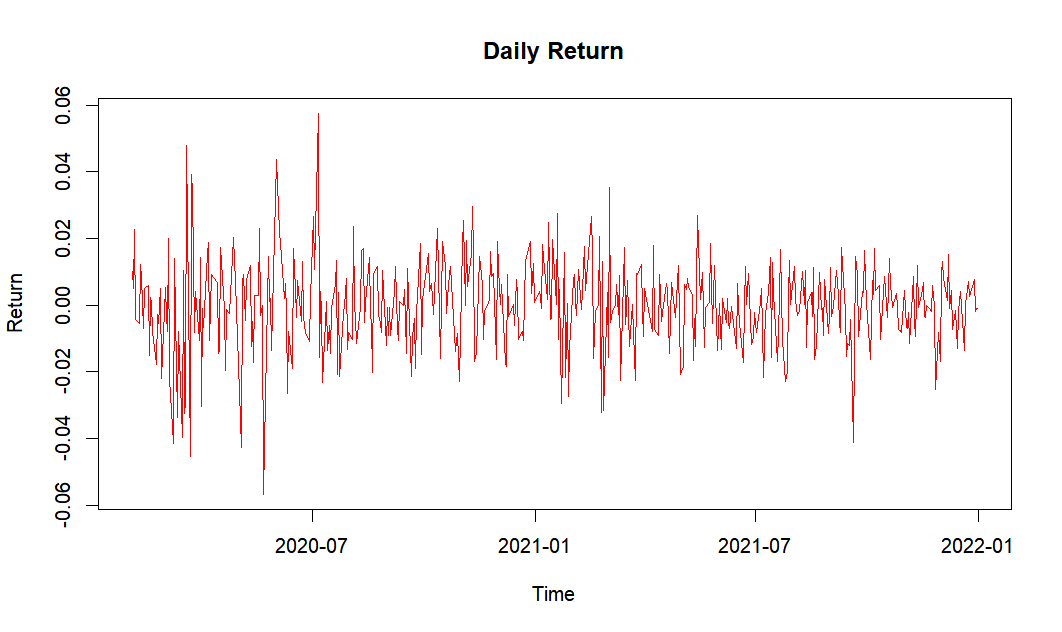
### 5.2.1 Financial Market

#### Model 1: ARMA-GARCH Model

Considering the effect of such a significant event and the market normal dynamic, the financial index price has this performance shown at the figure.



For analyzing the log return part, it has the following performance.



According to this figure, we could find that the log-return part seems like stationary one and mean-reverting, which is fluctuate around the 0 approximately.

##### Previous Assessment:

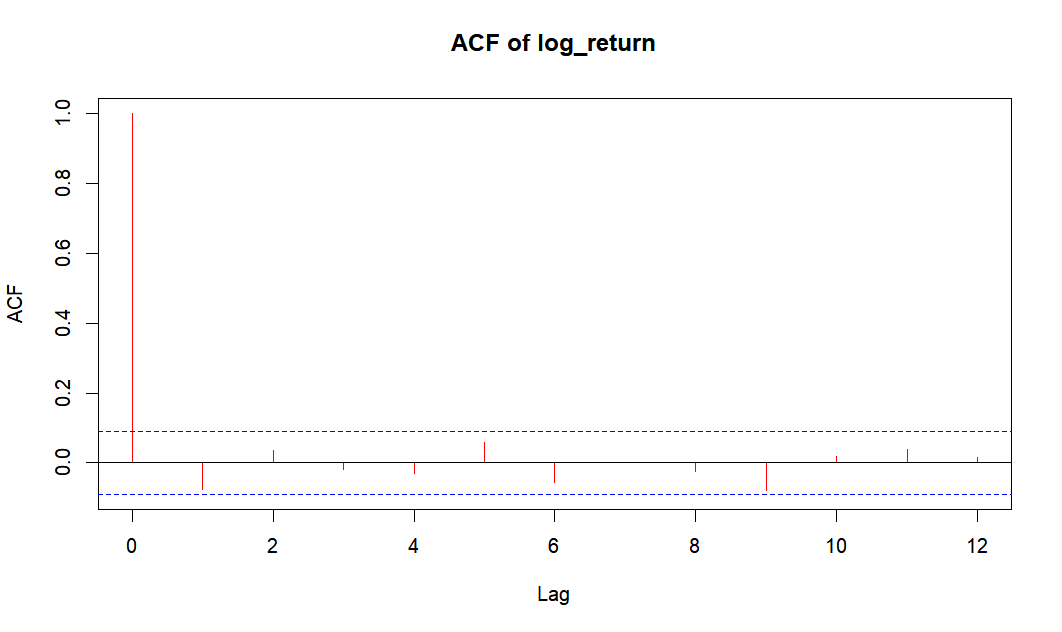
The testing result is shown at the following table.

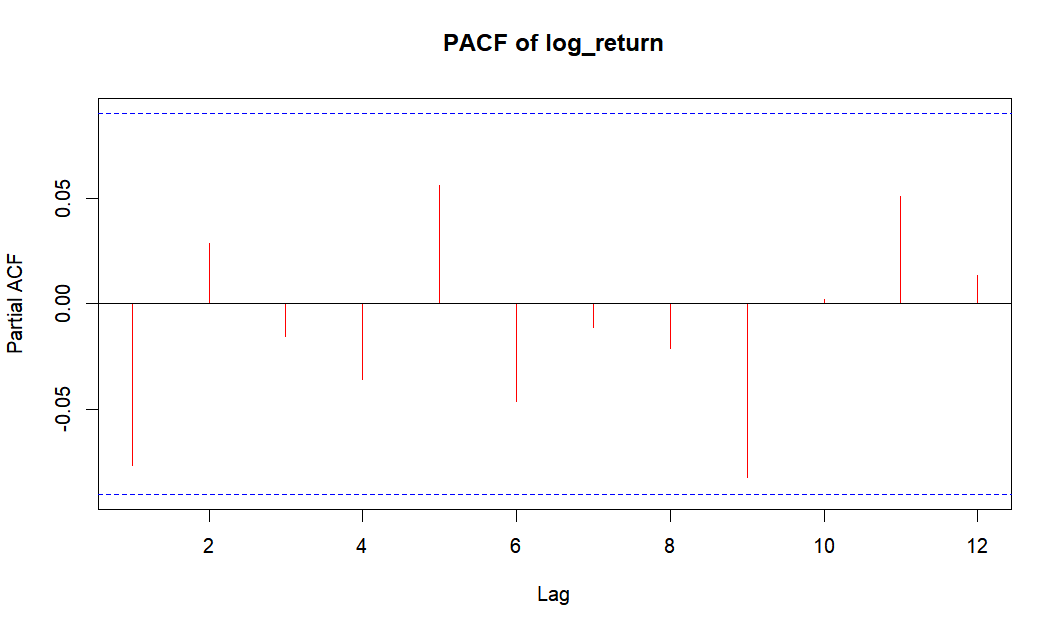
|  |  |  |  |
| --- | --- | --- | --- |
|  | Seasonality | Seasonality | ARCH Effect |
| Statistics | - | -6.4172 | 84.255 |
| p-value | - | 2.164e-08 | 6.327e-13 |

1. Checking Seasonality  
   Because it is daily trading information, we do not need to consider the seasonal analysis.
2. Checking Stationary  
   By applying ADF Test, we could find, for 12 lags, the p-value is 2.164e-08 (<0.05), thus, we could reject H0. It is a good signal that there is no unit root, indicating the movement is stationary right now.
3. Test ARCH Effect  
   Applying Box-Ljung Test on the centralized log return, because of p-value is 6.327e-13 (<0.05), we could reject H0, indicating the heteroscedastic variance.

##### Order Defining Part:

1. ARMA Order Define  
   Applying a Box test for a simple testing, we could find that, because the statistics considering 12 lags is 0.4765(>0.05), we could guess, there is no serial correlations in the log return.  
   In the formal testing, we apply ACF, PACF for specifying an accurate guess.





Thus, we could get the following result.

* 1. For ACF, excepting the lag 0, which is rationally one, the rest values are at the confidence interval based on zero-value assumption. We assume there is no MA term;
  2. For PACF, still flowing the same logistics, we assume there is no AR term;

Considering the above phenomenon, we could guess it follows ARMA(0,0), but To err on the side of caution, we tried ARMA(0,0), ARMA(0,1), ARMA(1,0), ARMA(1,1):

|  |  |  |  |
| --- | --- | --- | --- |
|  | P-Value | AIC | BIC |
| ARMA(0,0) | 0.4765 | -2727.529553 | -2719.215595 |
| ARMA(0,1) | 0.7458 | -2728.159369 | -2715.688432 |
| ARMA(1,0) | 0.757 | -2728.317011 | -2715.846074 |
| ARMA(1,1) | 0.7708 | -2726.663090 | -2710.035174 |

We still choose ARMA(0,0)

1. GARCH Order Define  
   In the text book, 3.5.1 (P116) mentions that "Specify the order of GARCH model is not easy. Only lower order GARCH Models are used in most applications, says, GARCH(1,1), GARCH (2,1), and GARCH(1,2) Models." Thus, in summary, we will test GARCH, ARCH, iGARCH, tGARCH, GARCH-M, and so on with low order case.
2. Constant Part Checking  
   Here we try to apply hypothesis testing, which is defined the null hypothesis is that the mean is zero. The statistics is like , where is the sample mean of the log return and s is the sample volatility of the log return. And for the significant level fixed at 5%, we could get the following result.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Statistics | Upper bound | Lower Bound |
| Value | -0.2233512 | 1.965013 | -1.965013 |

Thus, the statistics is lying at the range, we could not reject H0, which means we could assume, in the most cases, the intercept part should be fixed at the zero. However, that does not means, in all cases, we could make sure it is zero constantly. Please be careful about this point.

1. Residual Distribution Checking  
   For GARCH part, for improving the accuracy, we should check the distribution of residuals by KS test and AD test. Because in reality, log return usually follows normal distribution or t-distribution, hence, we get the results like this.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Normal Distribution | | T-distribution | |
| Method | KS test | AD test | KS test | AD test |
| Value | 0.1306 | 3.805e-05 | 0.9323 | 0.9621 |

Because we pass all t-distribution testing, we use t-distribution in the following GARCH process.

##### Model Construction:

Combine those ARMA and GARCH together, and testing their results, we could get the following results for those successfully passing the tests and being composed by significant lag-components.

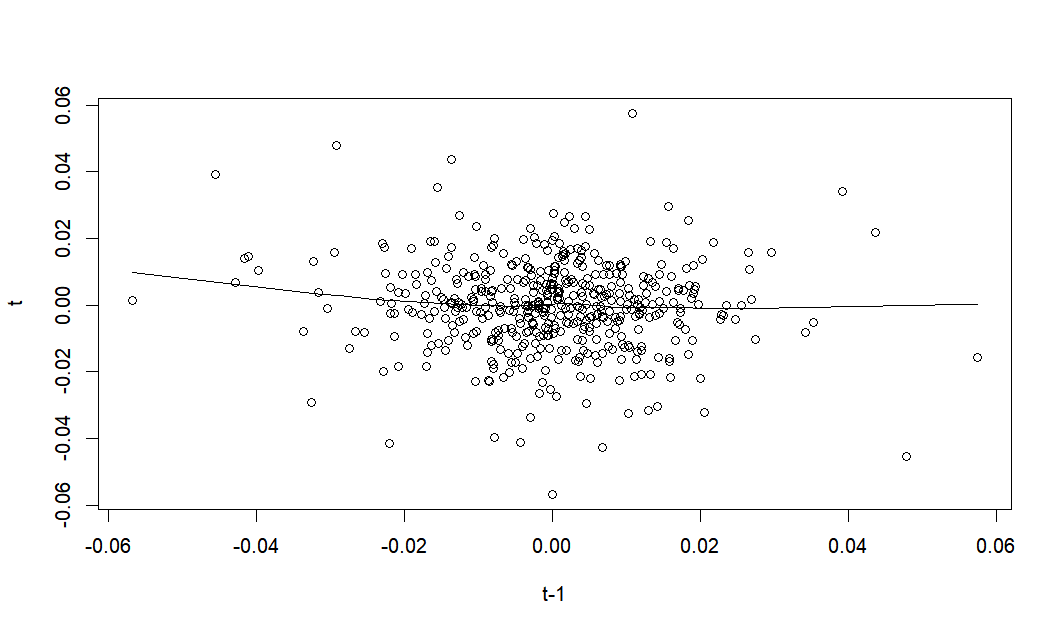
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | White Noise Test | ARCH Test | AIC | BIC |
| ARMA(0,0)-GARCH(1,1) no mean and omega | 0.8027 | 0.6642 | -5.886253 | -5.859831 |
| ARMA(0,0)-eGARCH(1,1) no alpha1 | 0.8203 | 0.3313 | -5.88976 | -5.854531 |
| ARMA(0,0)-tGARCH(1,1) no gamma1 | 0.8392 | 0.6227 | -5.895494 | -5.860265 |

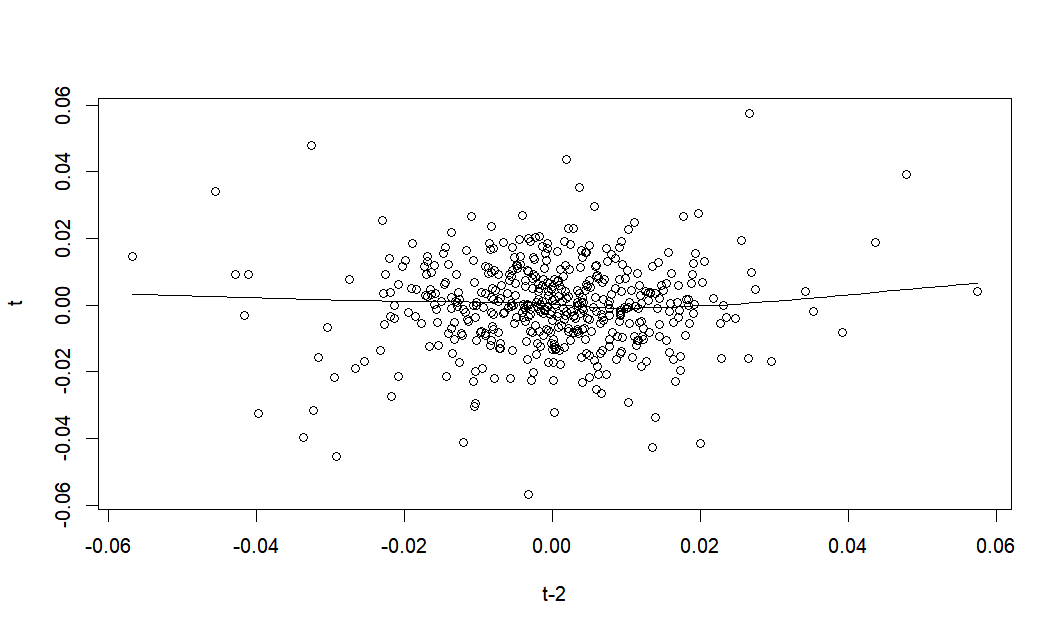
Because all pass the tests with 5% significant level, we conclude that ARMA(0,0)-tGARCH(1,1) without gamma1 is the best one, which is:  
where .

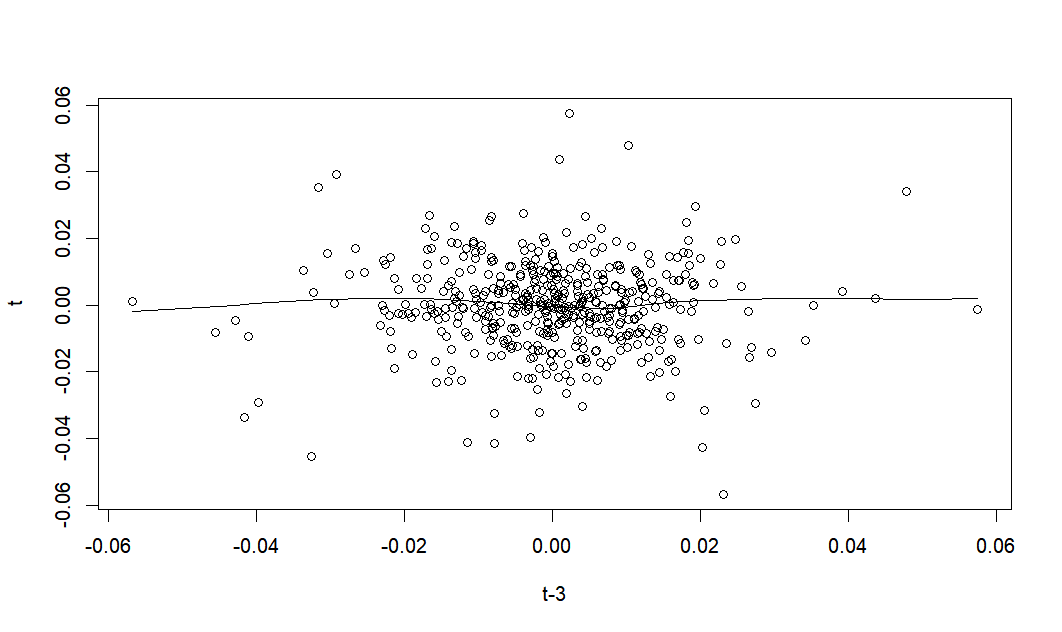
#### Model 2: Threshold AR Model

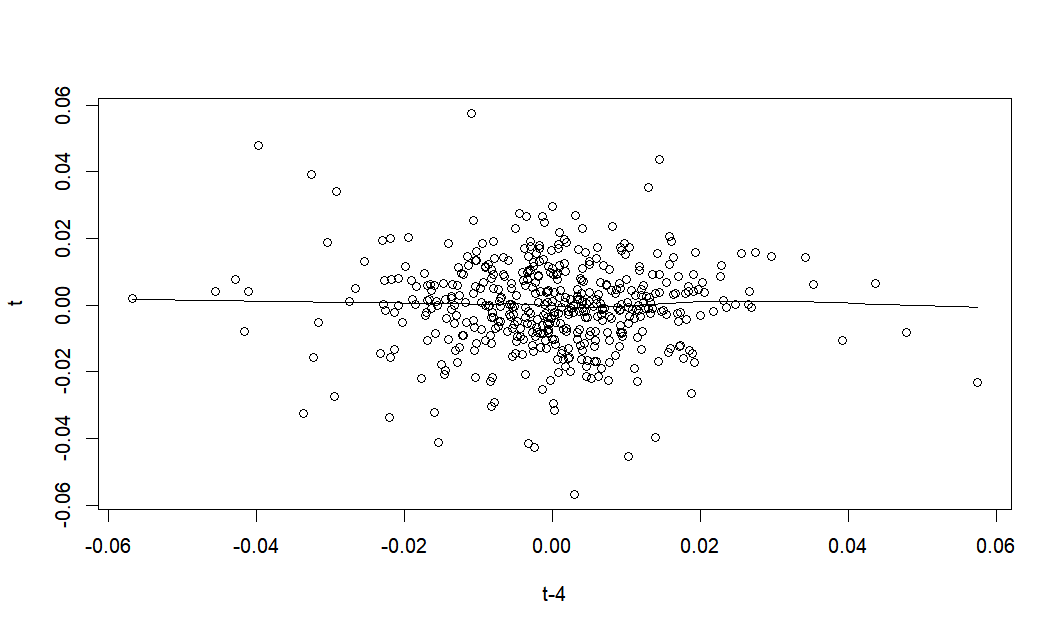
For better capturing the stock movement, we consider that whether there are different power influencing the stock movement. Thus, we want to apply Threshold AR model on it.

##### Rough Testing Method

1. Scatter-line Plot  
   At the beginning, we do the scatter-line plot and get the following result.  
   

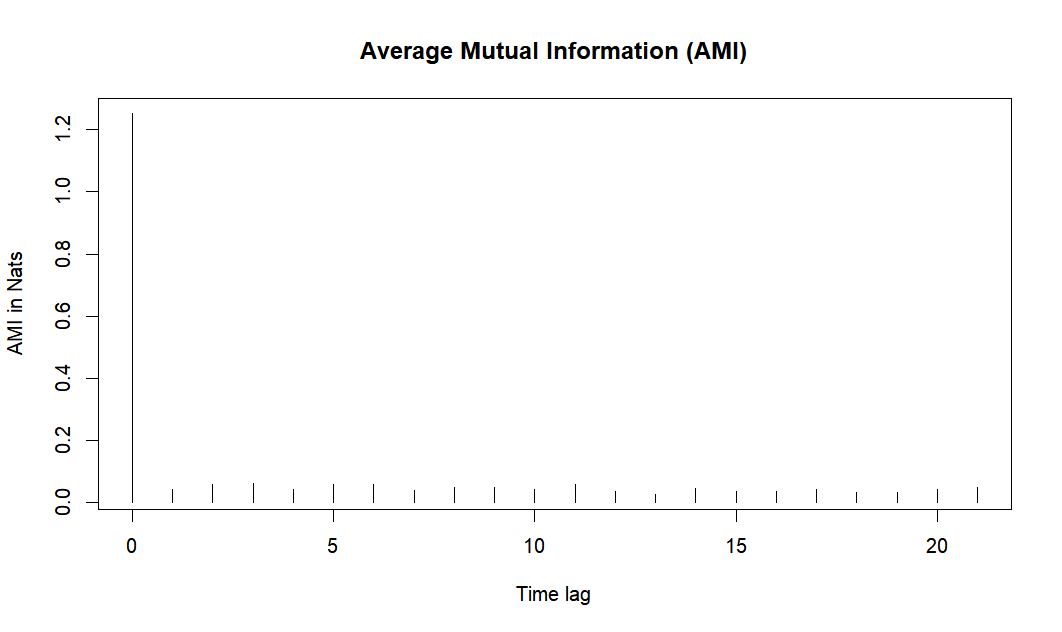






On the one hand, it seems like a good evidence of white noise. But because it seems like there is no non-linear relationship between lags.

1. Mutual Information Testing  
   For Average Mutual Information, we could get the following result.



It also shows that there is no cross or shared information between lags.

Sometimes however it happens so, that it’s not that simple to decide whether this type of nonlinearity is present. In this case, we’d have to run a statistical test — this approach is the most recommended by both Hansen’s and Tsays procedures.

##### Accurate Testing Method

Here we apply the threshold testing by using grid search method, which searches the lag of the threshold (d) from 1 to 3 and orders of models (p) form 1 to 3, and we get the following F-statistics and p-value.

|  |  |  |  |
| --- | --- | --- | --- |
| d  p | 1 | 2 | 3 |
| 1 | 2.019124  (0.1340315) | 2.47345  (0.06112494) | 1.902326  (0.109106) |
| 2 | 1.172376  (0.3106228) | 0.8728394  (0.4551147) | 0.8072895  (0.5209817) |
| 3 | 1.510815  (0.2219075) | 0.9510905  (0.4158064) | 1.580358  (0.1785389) |

According to the above result, there are no possible models.

#### Model 3: Stochastic Volatility Model

If we apply the Stochastic Volatility Model to describe the log return, we could get anormal distribution and the following result.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.00032 | 0.01007 | -0.0171 | 0.0004 | 0.016 | 10000 |
| phi | 0.99947 | 0.00044 | 0.9986 | 0.9996 | 1.000 | 289 |
| sigma | 0.16859 | 0.04743 | 0.0976 | 0.1636 | 0.252 | 75 |
| Exp(mu/2) | 0.99985 | 0.00503 | 0.9915 | 0.9998 | 1.008 | 10000 |
| Sigma^2 | 0.03067 | 0.01735 | 0.0095 | 0.0268 | 0.063 | 75 |

We could find that:

* The estimated mean (mu) is essentially zero, closely aligning with its prior distribution. This suggests minimal overall movement in the central tendency of log returns, indicating a stable average level over time without significant trends.
* The phi parameter is very close to 1, indicative of exceptionally high persistence in the volatility process. This high value suggests that once the market experiences a change in volatility, it tends to persist in this new state for a prolonged period, reflecting strong autocorrelation in the volatility data.
* The parameter sigma, which represents the volatility of the volatility, shows significant variability as evidenced by its relatively wide credible interval and standard deviation. This indicates that the model captures a substantial degree of uncertainty and fluctuation in market volatility.
* The exponential transformation of half the mean, exp(mu/2), is very close to 1, reinforcing the stability of mu and suggesting no significant exponential growth or decay in the volatility scale as a function of the mean level.
* The squared volatility sigma^2 provides further insight into the extent of variability in the volatility process. The estimates highlight a noticeable range of potential volatility states, crucial for understanding and managing financial risk.

The stochastic volatility model elucidates several critical aspects of the log returns' behavior:

Stability in Mean: The near-zero mean and its stable estimate suggest little long-term drift, important for models assuming mean-reversion or similar characteristics in financial series.

High Persistence: The near-unity phi underscores a market characteristic where volatility states are highly persistent, affecting strategies around volatility clustering.

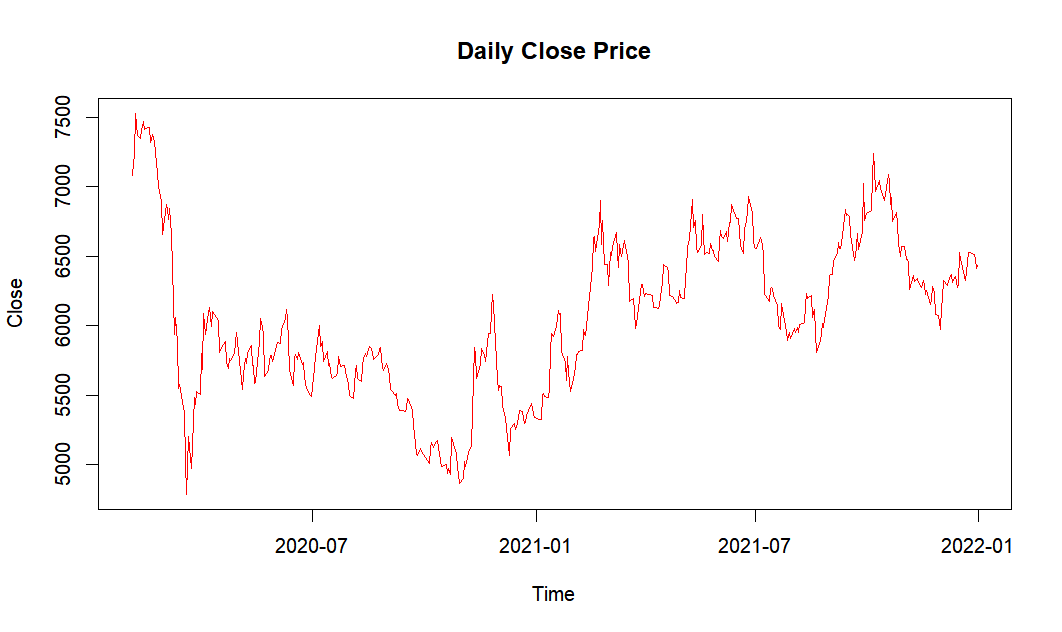
Variable Volatility of Volatility: The significant variability in sigma indicates that the volatility itself is highly unstable, which is vital for pricing derivatives and managing portfolio risk.

These results are invaluable for financial data analysis, offering deep insights into the underlying dynamics of market movements and assisting in the effective prediction and management of future market volatility.

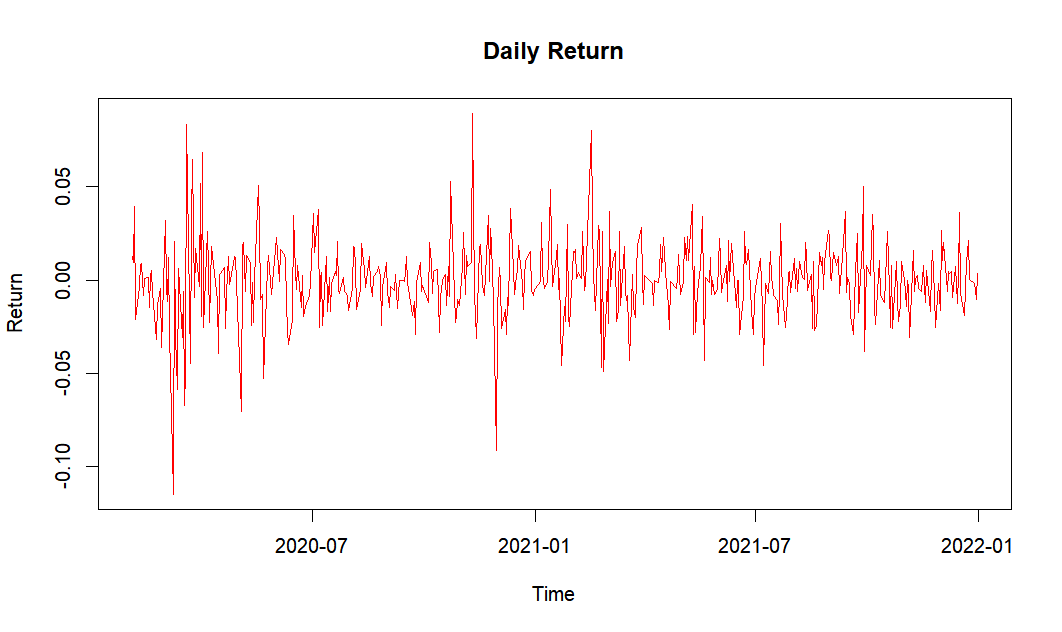
### 5.2.2 Energy Market

#### Model 1: ARMA-GARCH Model

Considering the effect of such a significant event and the market normal dynamic, the energy index price has this performance shown at the figure.



For analyzing the log return part, it has the following performance.



According to this figure, we could find that the log-return part seems like stationary one and mean-reverting, which is fluctuate around the 0 approximately.

**Previous Assessment:**

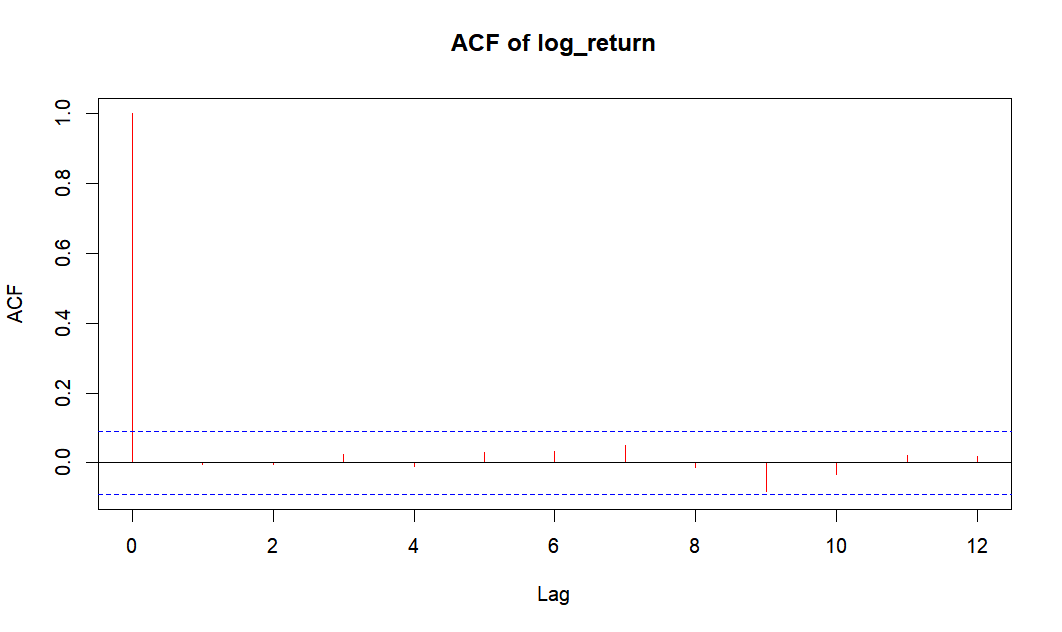
The testing result is shown at the following table.

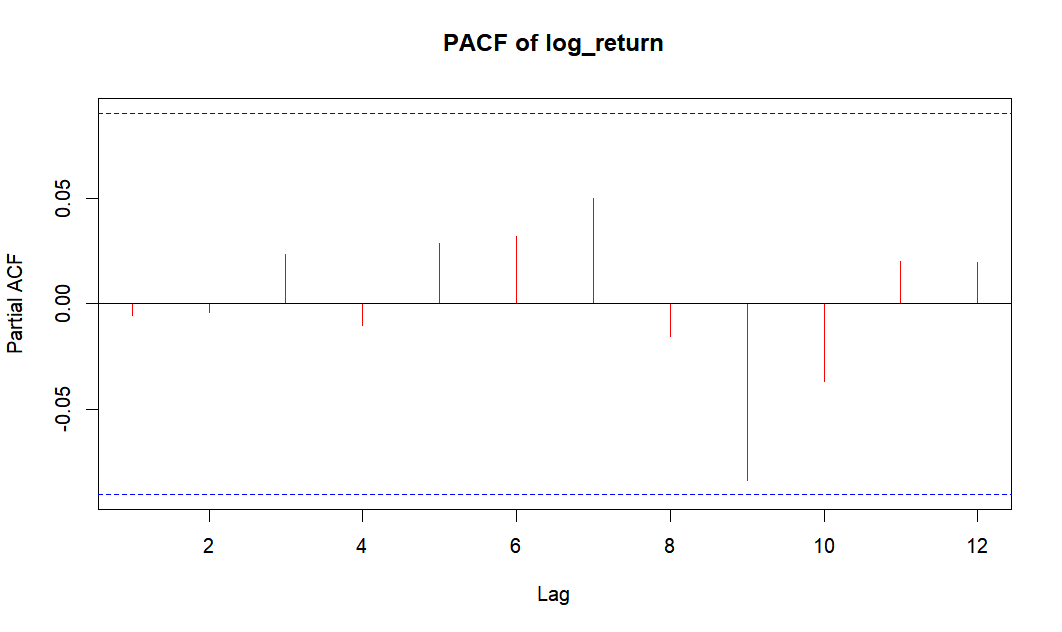
|  |  |  |  |
| --- | --- | --- | --- |
|  | Seasonality | Seasonality | ARCH Effect |
| Statistics | - | -6.5271 | 76.745 |
| p-value | - | 1.186e-08 | 1.717e-11 |

1. Checking Seasonality  
   Because it is daily trading information, we do not need to consider the seasonal analysis.
2. Checking Stationary  
   By applying ADF Test, we could find, for 12 lags, the p-value is 1.186e-08 (<0.05), thus, we could reject H0. It is a good signal that there is no unit root, indicating the movement is stationary right now.
3. Test ARCH Effect  
   Applying Box-Ljung Test on the centralized log return, because of p-value is 1.717e-11(<0.05), we could reject H0, indicating the heteroscedastic variance.

**Order Defining Part:**

1. ARMA Order Define  
   Applying a Box test for a simple testing, we could find that, because the statistics considering 12 lags is 0.8783(>0.05), we could guess, there is no serial correlations in the log return.  
   In the formal testing, we apply ACF, PACF for specifying an accurate guess.





Thus, we could get the following result.

* 1. For ACF, excepting the lag 0, which is rationally one, the rest values are at the confidence interval based on zero-value assumption. We assume there is no MA term;
  2. For PACF, still flowing the same logistics, we assume there is no AR term;

Considering the above phenomenon, we could guess it follows ARMA(0,0), but To err on the side of caution, we tried ARMA(0,0), ARMA(0,1), ARMA(1,0), ARMA(1,1):

|  |  |  |  |
| --- | --- | --- | --- |
|  | P-Value | AIC | BIC |
| ARMA(0,0) | 0.8783 | -2306.761790 | -2298.447832 |
| ARMA(0,1) | 0.8767 | -2304.775645 | -2292.304708 |
| ARMA(1,0) | 0.8767 | -2304.775537 | -2292.304600 |
| ARMA(1,1) | 0.8788 | -2302.782678 | -2286.154762 |

We still choose ARMA(0,0)

1. GARCH Order Define  
   In the text book, 3.5.1 (P116) mentions that "Specify the order of GARCH model is not easy. Only lower order GARCH Models are used in most applications, says, GARCH(1,1), GARCH (2,1), and GARCH(1,2) Models." Thus, in summary, we will test GARCH, ARCH, iGARCH, tGARCH, GARCH-M, and so on with low order case.
2. Constant Part Checking  
   Here we try to apply hypothesis testing, which is defined the null hypothesis is that the mean is zero. The statistics is like , where is the sample mean of the log return and s is the sample volatility of the log return. And for the significant level fixed at 5%, we could get the following result.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Statistics | Upper bound | Lower Bound |
| Value | -0.2111928 | 1.965013 | -1.965013 |

Thus, the statistics is lying at the range, we could not reject H0, which means we could assume, in the most cases, the intercept part should be fixed at the zero. However, that does not means, in all cases, we could make sure it is zero constantly. Please be careful about this point.

1. Residual Distribution Checking  
   For GARCH part, for improving the accuracy, we should check the distribution of residuals by KS test and AD test. Because in reality, log return usually follows normal distribution or t-distribution, hence, we get the results like this.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Normal Distribution | | T-distribution | |
| Method | KS test | AD test | KS test | AD test |
| Value | 0.007663 | 1.271e-06 | 0.7919 | 0.9426 |

Because we pass all t-distribution testing, we use t-distribution in the following GARCH process.

**Model Construction:**

Combine those ARMA and GARCH together, and testing their results, we could get the following results for those successfully passing the tests and being composed by significant lag-components.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | White Noise Test | ARCH Test | AIC | BIC |
| ARMA(0,0)-sGARCH(1,1) no mean and omega | 0.9109 | 0.3742 | -5.091906 | -5.065485 |
| ARMA(0,0)-iGARCH(1,1) no mean and omega | 0.9113 | 0.381 | -5.096835 | -5.079221 |
| ARMA(0,0)-eGARCH(1,1) no mean and alpha1 | 0.9394 | 0.8689 | -5.103539 | -5.068311 |
| ARMA(0,0)-tGARCH(1,1) no omega and gamma1 | 0.9109 | 0.3742 | -5.091906 | -5.065485 |

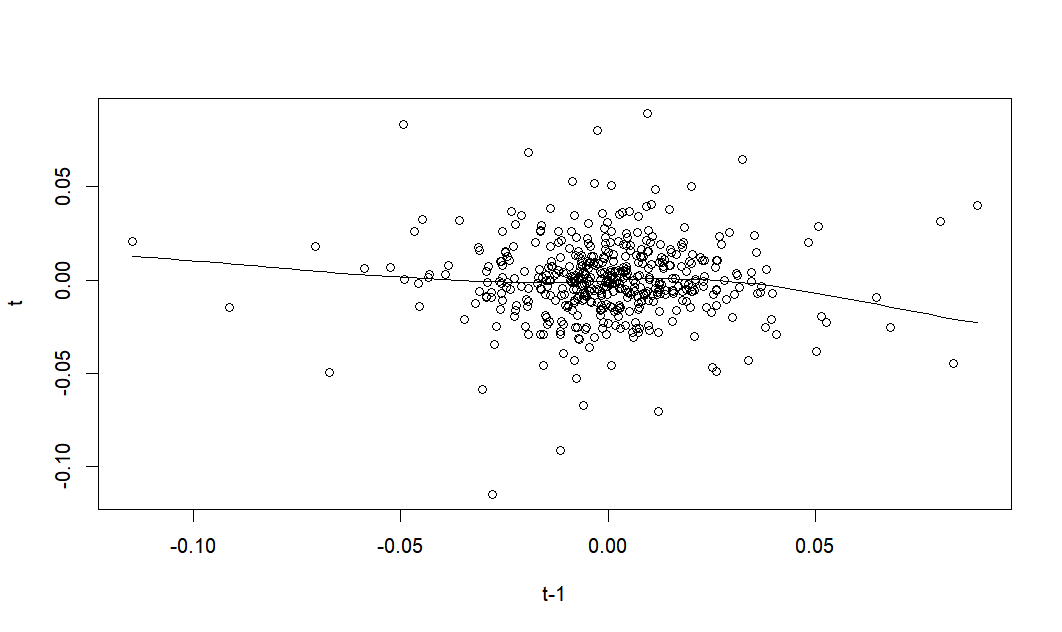
Because all pass the tests with 5% significant level, we conclude that ARMA(0,0)-eGARCH(1,1) without gamma1 is the best one, which is:

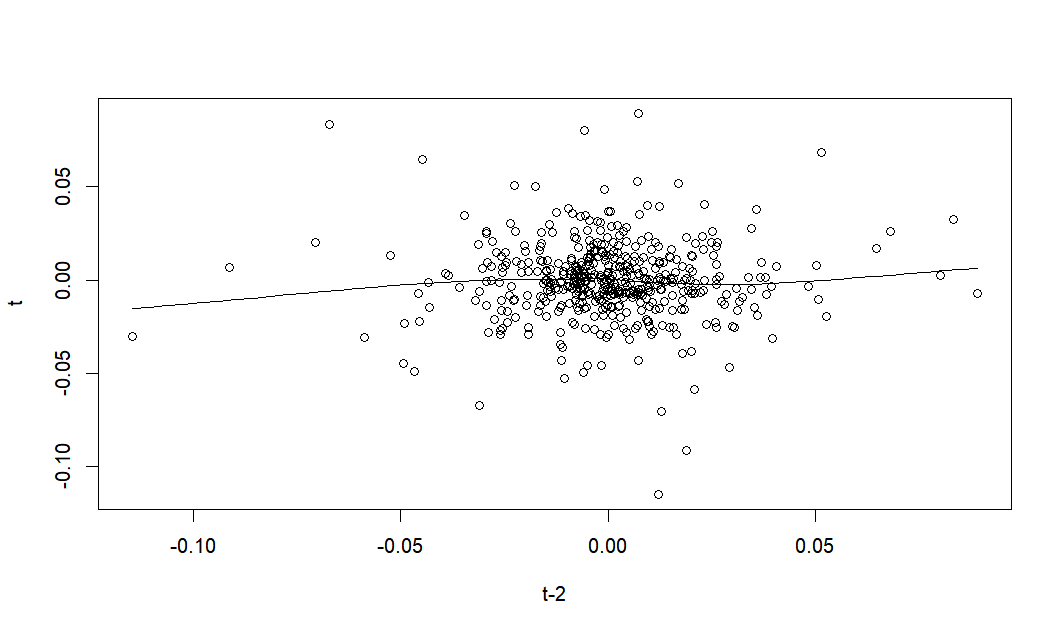
Where ,

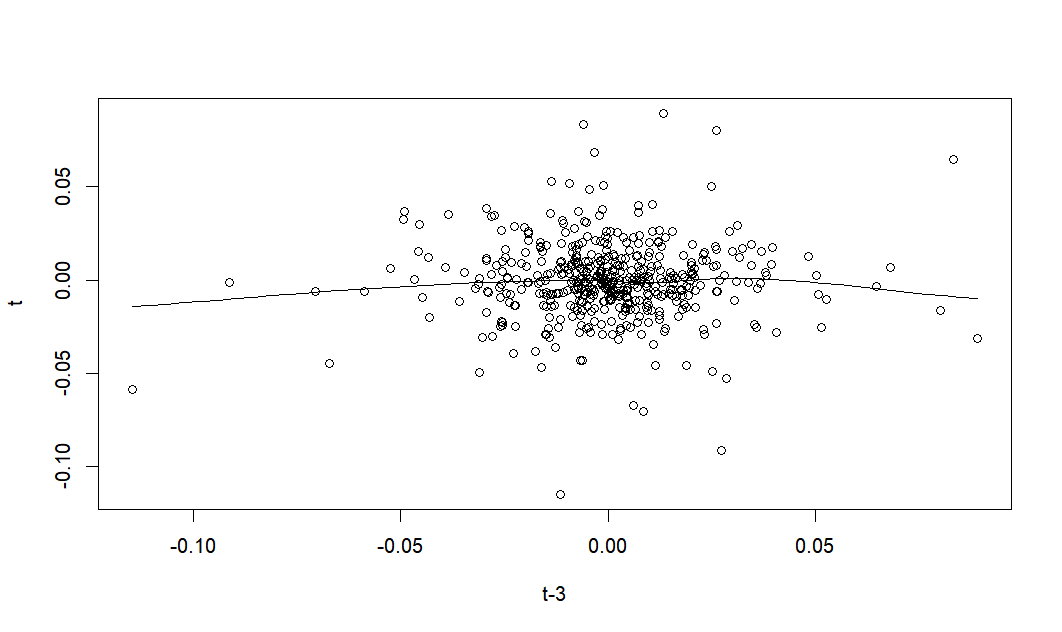
#### Model 2: Threshold AR Model

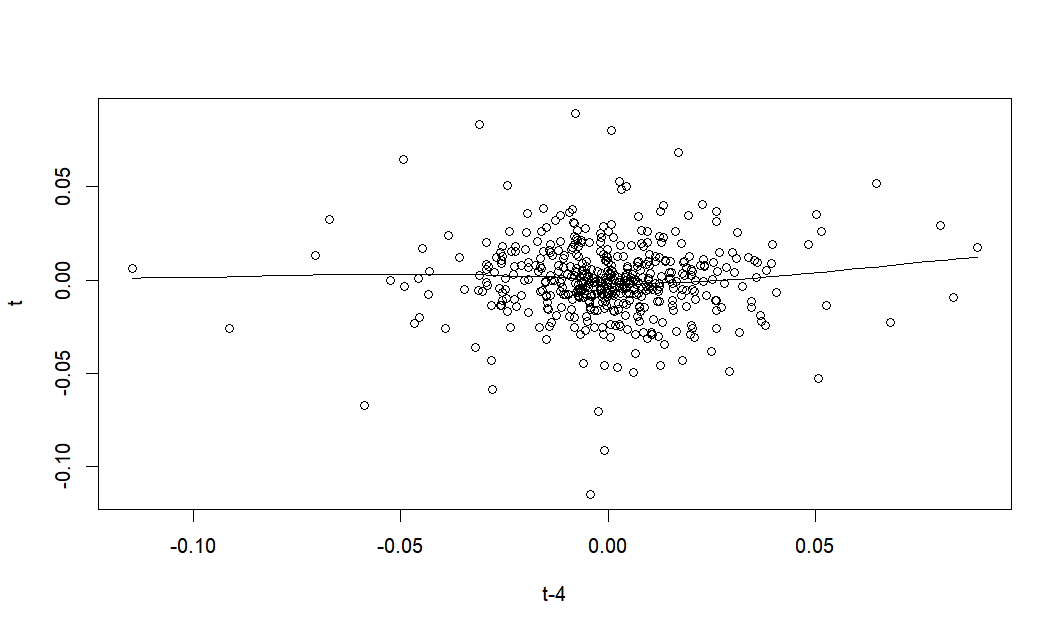
For better capturing the stock movement, we consider that whether there are different power influencing the stock movement. Thus, we want to apply Threshold AR model on it.

**Rough Testing Method**

1. Scatter-line Plot  
   At the beginning, we do the scatter-line plot and get the following result.  
   

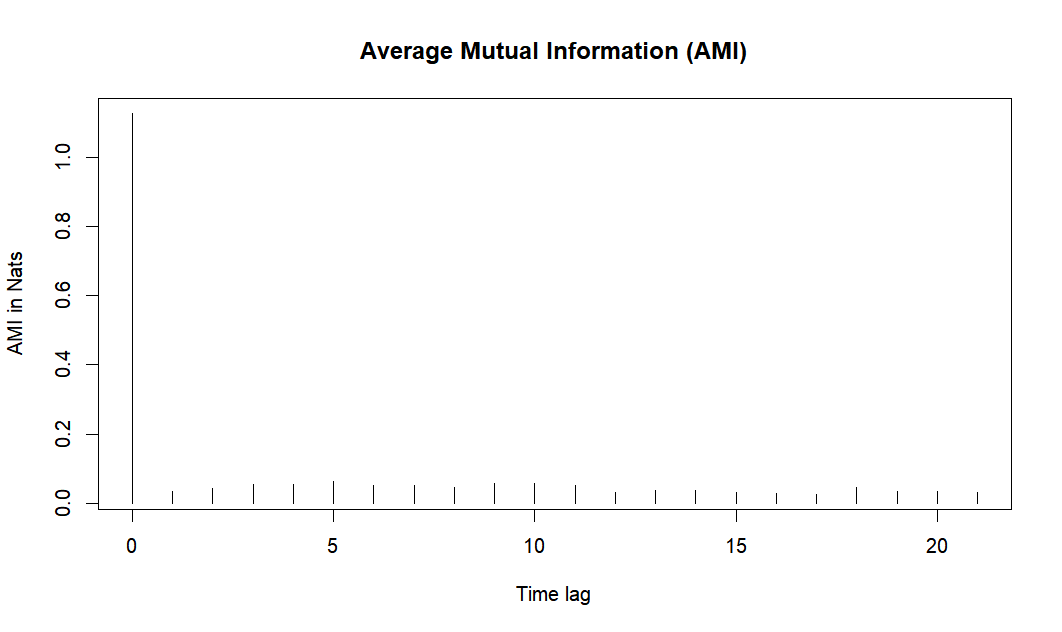






On the one hand, it seems like a good evidence of white noise. But because it seems like there is no non-linear relationship between lags.

1. Mutual Information Testing  
   For Average Mutual Information, we could get the following result.



It also shows that there is no cross or shared information between lags.

Sometimes however it happens so, that it’s not that simple to decide whether this type of nonlinearity is present. In this case, we’d have to run a statistical test — this approach is the most recommended by both Hansen’s and Tsays procedures.

**Accurate Testing Method**

Here we apply the threshold testing by using grid search method, which searches the lag of the threshold (d) from 1 to 3 and orders of models (p) form 1 to 3, and we get the following F-statistics and p-value.

|  |  |  |  |
| --- | --- | --- | --- |
| d  p | 1 | 2 | 3 |
| 1 | 2.434543  (0.08884773) | 1.822489  (0.1423263) | 1.27214  (0.2801686) |
| 2 | 2.798873  (0.06199333) | 1.823731  (0.1421006) | 1.620607  (0.1680586) |
| 3 | 0.04453816  (0.9564435) | 0.6421966  (0.588224) | 1.507162  (0.1991229) |

According to the above result, there are no possible models.

#### Model 3: Stochastic Volatility Model

If we apply the Stochastic Volatility Model to describe the log return, we could get anormal distribution and the following result.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.00024 | 0.01003 | 0.017 | -0.00017 | 0.016 | 10707 |
| phi | 0.99877 | 0.00084 | 0.997 | 0.99895 | 1.000 | 646 |
| sigma | 0.25504 | 0.05144 | 0.178 | 0.25200 | 0.347 | 128 |
| Exp(mu/2) | 0.99989 | 0.00501 | 0.992 | 0.99992 | 1.008 | 10707 |
| Sigma^2 | 0.06769 | 0.02780 | 0.032 | 0.06350 | 0.120 | 128 |

We could find that:

* The estimated mean (mu) is close to zero, suggesting minimal overall movement in the central tendency of log returns. This indicates a stable average level over time without significant trends, aligning with the often stationary nature of financial return series.
* The phi parameter is very close to 1, indicative of exceptionally high persistence in the volatility process. This suggests that once the market experiences a change in volatility, it maintains this new state for an extended period, reflecting strong autocorrelation in the volatility data.
* The parameter sigma, which represents the volatility of volatility, shows significant variability, as evidenced by its relatively wide credible interval and standard deviation. This indicates that the model captures a substantial degree of uncertainty and fluctuation in market volatility.
* The exponential transformation of half the mean, exp(mu/2), is very close to 1. This reinforces the stability of mu and suggests no considerable exponential growth or decay in the volatility scale as a function of the mean level.

The squared volatility sigma^2 offers further insight into the extent of variability in the volatility process. The estimates highlight a noticeable range of potential volatility states, crucial for understanding and managing financial risk.

The stochastic volatility model elucidates several critical aspects of the log returns' behavior:

Stability in Mean: The near-zero mean and its stable estimate suggest little long-term drift, which is important for models assuming mean-reversion or similar characteristics in financial series.

High Persistence: The near-unity phi underscores a market characteristic where volatility states are highly persistent, affecting strategies around volatility clustering.

Variable Volatility of Volatility: The significant variability in sigma indicates that the volatility itself is highly unstable, which is vital for pricing derivatives and managing portfolio risk.

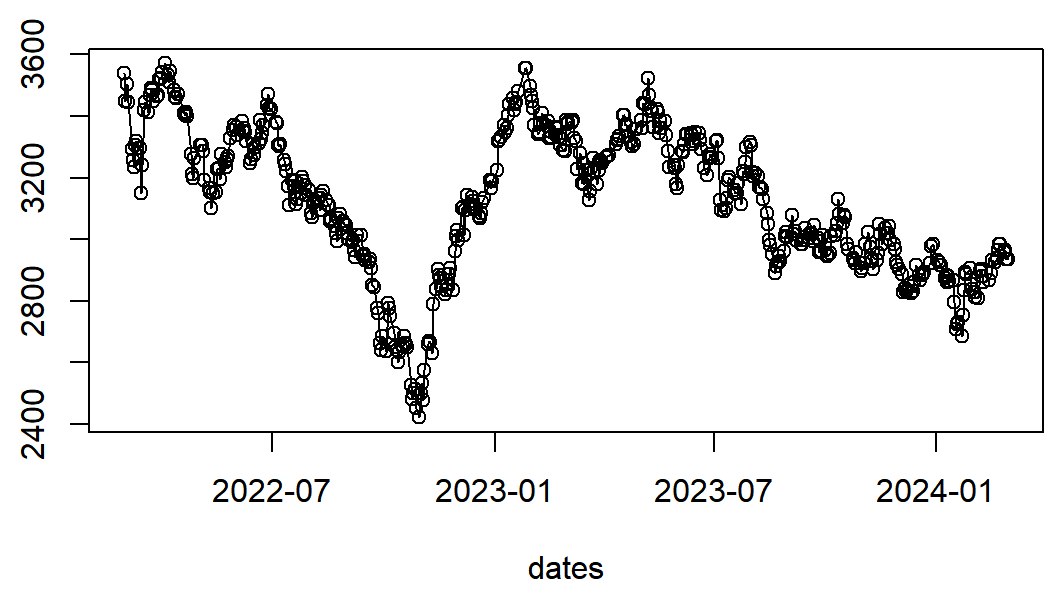
These results are invaluable for financial data analysis, offering deep insights into the underlying dynamics of market movements and assisting in the effective prediction and management of future market volatility.

## 5.3 Model After War

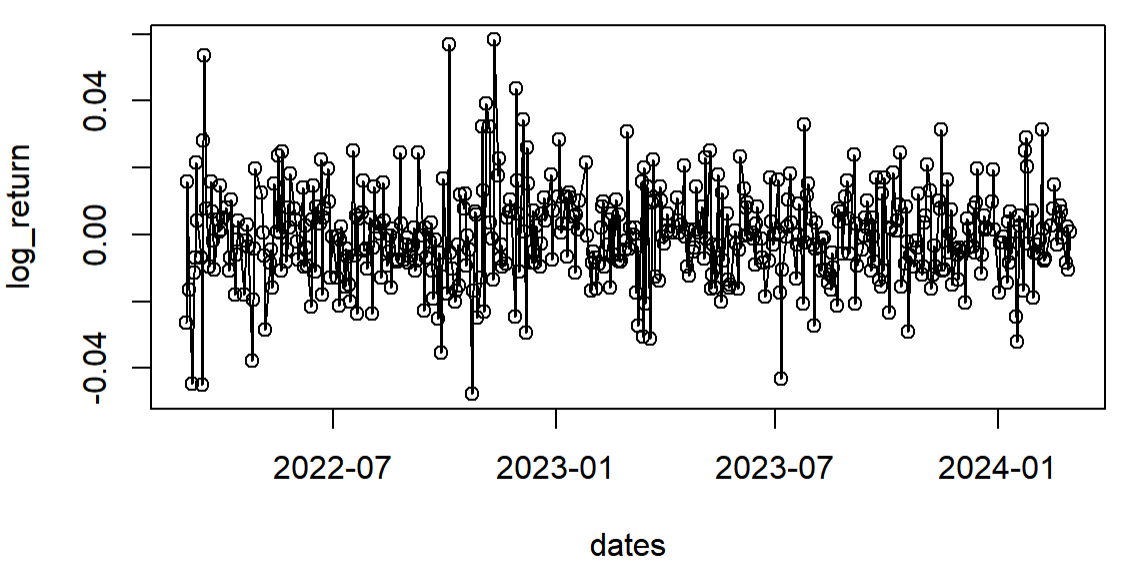
### 5.3.1 Financial Index

#### Model 1: ARMA-GARCH Model

Considering the effect of such a significant event and the market normal dynamic, the financial index price has this performance shown at the figure.



For analyzing the log return part, it has the following performance.



According to this figure, we could find that the log-return part seems like stationary one and mean-reverting, which is fluctuate around the 0 approximately.

##### Previous Assessment:

The testing result is shown at the following table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Seasonality | Seasonality | ARCH Effect |
| Statistics | - | -3.384 | 50.616 |
| p-value | - | 0.01275 | 0.01069 |

1. Checking Seasonality

Because it is daily trading information, we do not need to consider the seasonal analysis.

1. Checking Stationary

By applying ADF Test, we could find, for 30 lags, the p-value is 0.01275 (<0.05), thus, we could reject H0. It is a good signal that there is no unit root, indicating the movement is stationary right now.

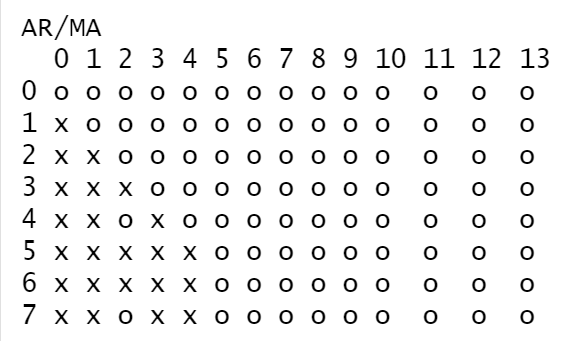
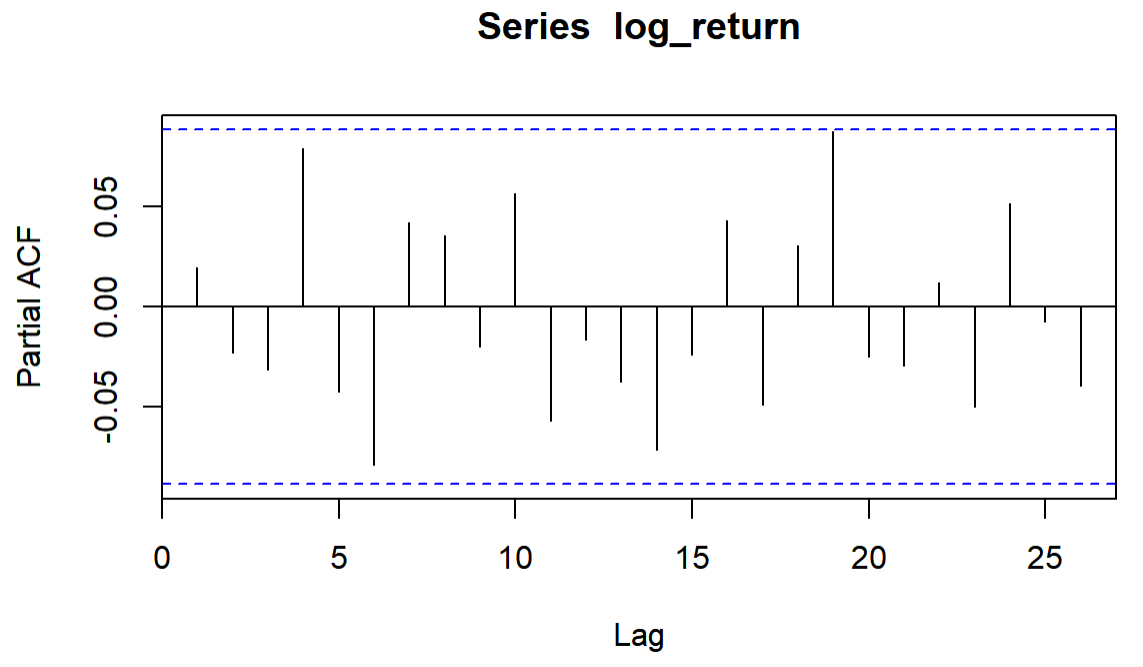
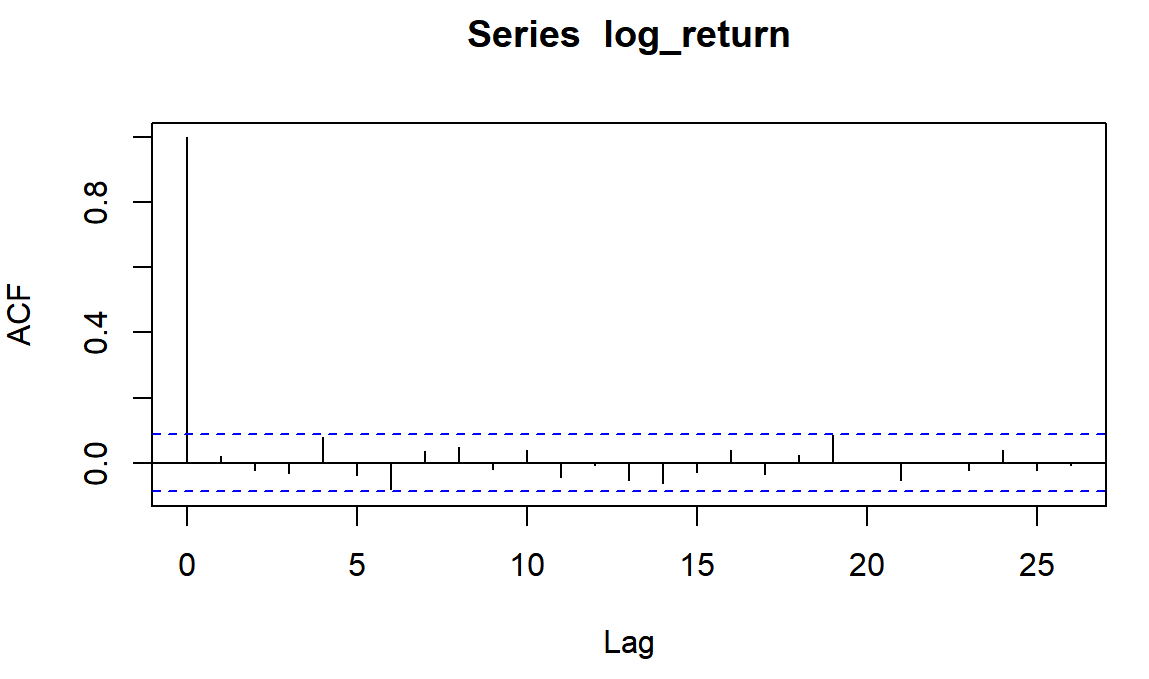
1. Test ARCH Effect

Applying Box-Ljung Test on the centralized log return, because of p-value is 0.01069 (<0.05), we could reject H0, indicating the heteroscedastic variance.

##### Order Defining Part:

1. ARMA Order Define

Applying a Box test for a simple testing, we could find that, because the statistics considering 12 lags is 0.4658 (>0.05), we could guess, there is no serial correlations in the log return.  
In the formal testing, we apply ACF, PACF and EACF for specifying an accurate guess.



Thus, we could get the following result.

* For ACF, excepting the lag 0, which is rationally one, the rest values are at the confidence interval based on zero-value assumption. We assume there is no MA term;
* For PACF, still flowing the same logistics, we assume there is no AR term, even though the lag 19 is very close to the boundary;
* For EACF, following its identifying method, which could form a perfect upper-o triangle starting its vertex at (0,0).

Considering the above phenomenon, we could guess it follows ARMA(0,0).

1. GARCH Order Define

In the text book, 3.5.1 (P116) mentions that "Specify the order of GARCH model is not easy. Only lower order GARCH Models are used in most applications, says, GARCH(1,1), GARCH (2,1), and GARCH(1,2) Models." Thus, in summary, we will test GARCH, ARCH, iGARCH, tGARCH, GARCH-M, and so on with low order case.

1. Constant Part Checking

Here we try to apply hypothesis testing, which is defined the null hypothesis is that the mean is zero. The statistics is like , where is the sample mean of the log return and s is the sample volatility of the log return. And for the significant level fixed at 5%, we could get the following result.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Statistics | Upper bound | Lower Bound |
| Value | -0.5951015 | 1.964827 | -1.964827 |

Thus, the statistics is lying at the range, we could not reject H0, which means we could assume, in the most cases, the intercept part should be fixed at the zero. However, that does not means, in all cases, we could make sure it is zero constantly. Please be careful about this point.

1. Residual Distribution Checking

For GARCH part, for improving the accuracy, we should check the distribution of residuals by KS test and AD test. Because in reality, log return usually follows normal distribution or t-distribution, hence, we get the results like this.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Normal Distribution | | T-distribution | |
| Method | KS test | AD test | KS test | AD test |
| Value | 0.1396 | 6.369e-05 | 0.9442 | 0.9402 |

Because we pass all t-distribution testing, we use t-distribution in the following GARCH process.

##### Model Construction:

1. Model 1: ARMA-GARCH Model

Combine those ARMA and GARCH together, and testing their results, we could get the following results for those successfully passing the tests and being composed by significant lag-components.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | White Noise Test | ARCH Test | AIC | BIC |
| ARMA(0,0)-GARCH(1,1) no mean and omega | 0.3492 | 0.08452 | -5.7192 | -5.6936 |
| ARMA(0,0)-iGARCH(1,1) no mean and omega | 0.3718 | 0.5921 | -5.6862 | -5.6777 |
| ARMA(0,0)-eGARCH(1,1) no mean | 0.488 | 0.7609 | -5.7404 | -5.6976 |
| ARMA(0,0)-GJRGARCH(1,1) no mean and alpha1 | 0.5069055 | 0.6141958 | -5.7499 | -5.7156 |
| ARMA(0,0)-CGARCH(0,1) no mean | 0.3770878 | 0.5344655 | -5.7115 | -5.6772 |

Because all pass the tests with 5% significant level, we conclude that ARMA(0,0)-GJRGARCH(1,1) without mean and alpha1 is the best one, which is:  
where .

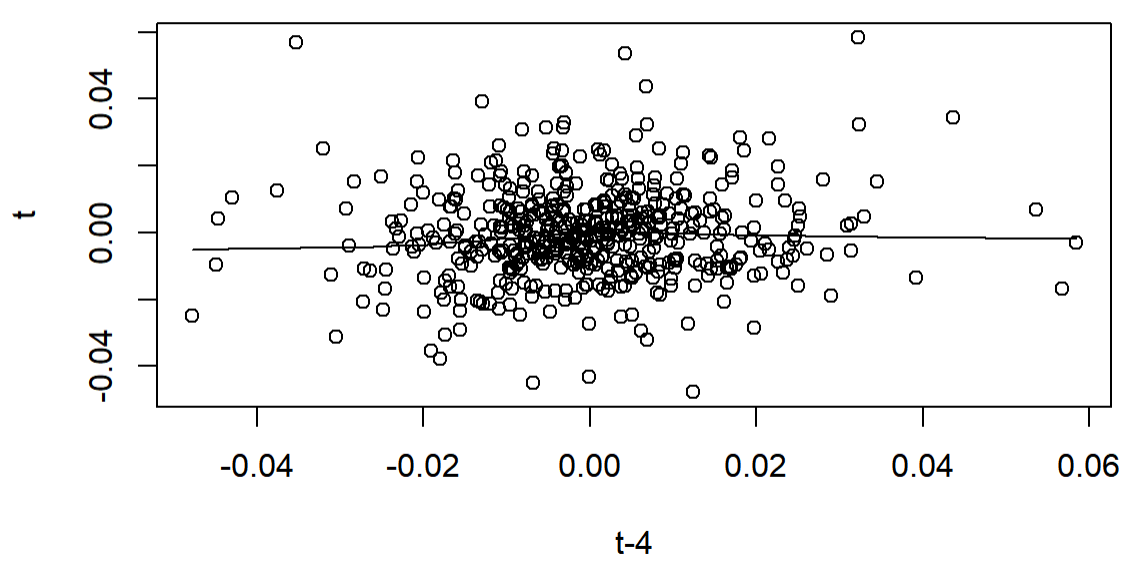
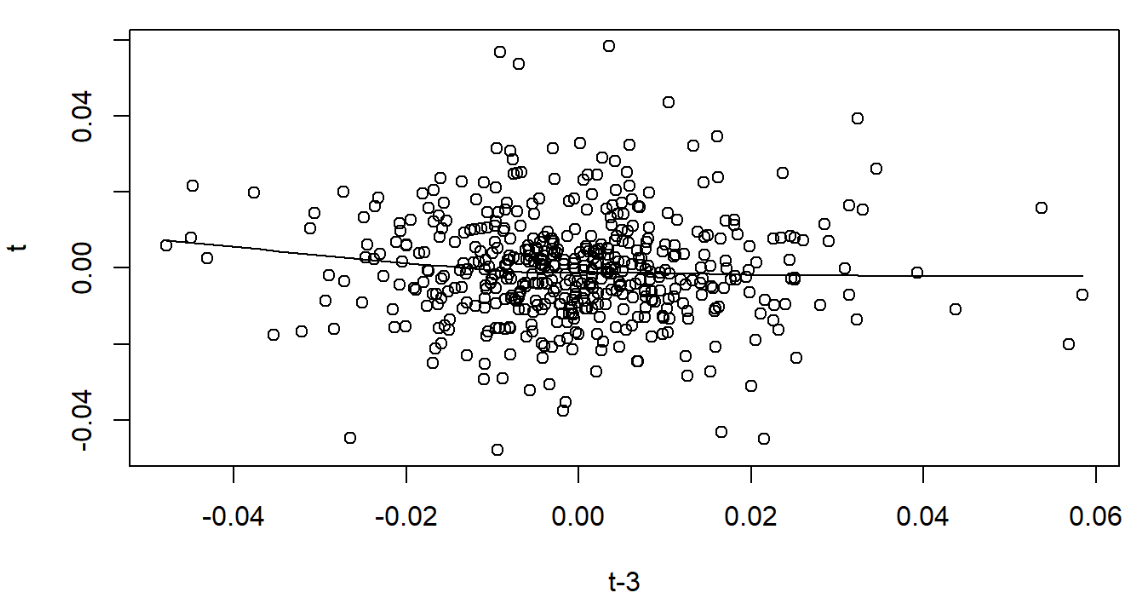
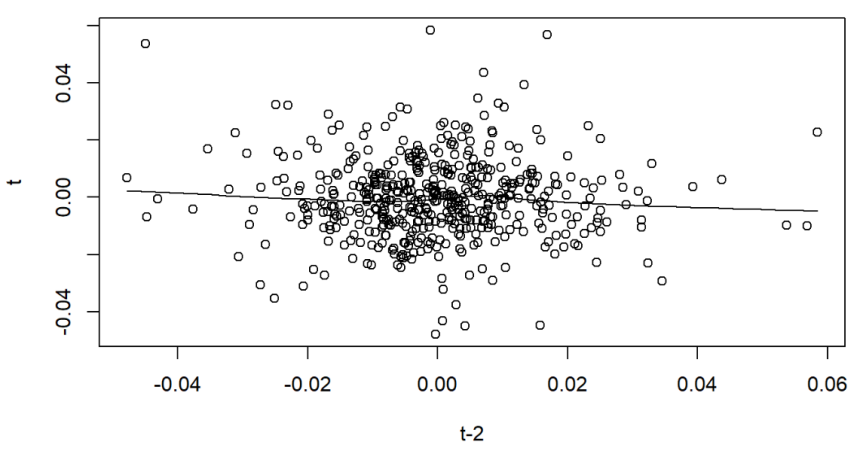
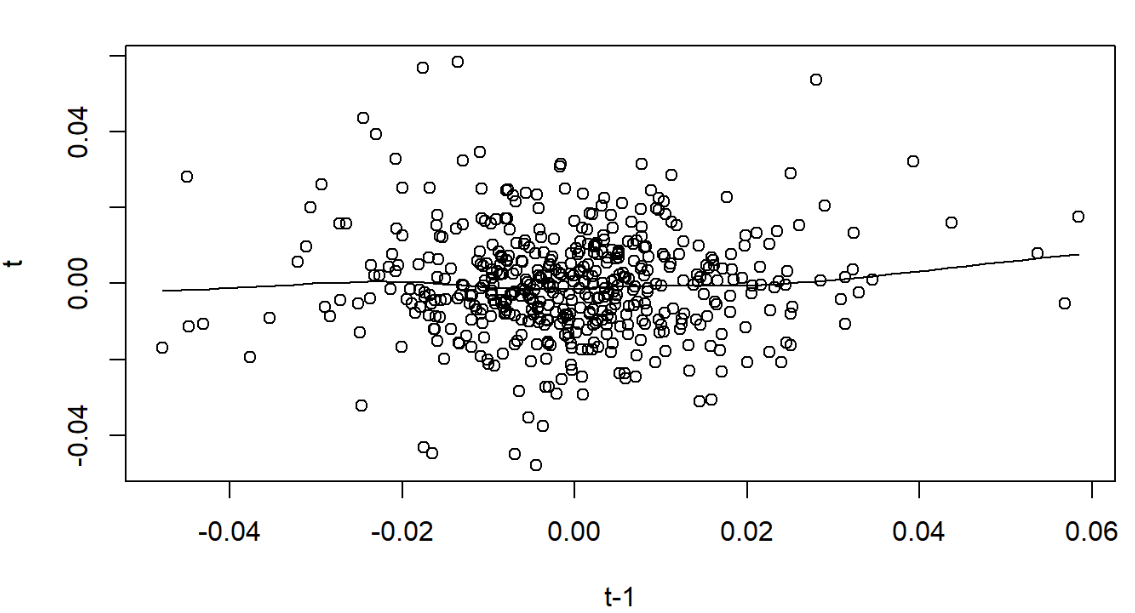
#### Model 2: Threshold AR Model

For better capturing the stock movement, we consider that whether there are different power influencing the stock movement. Thus, we want to apply Threshold AR model on it.

1. Rough Testing Method

* Scatter-line Plot

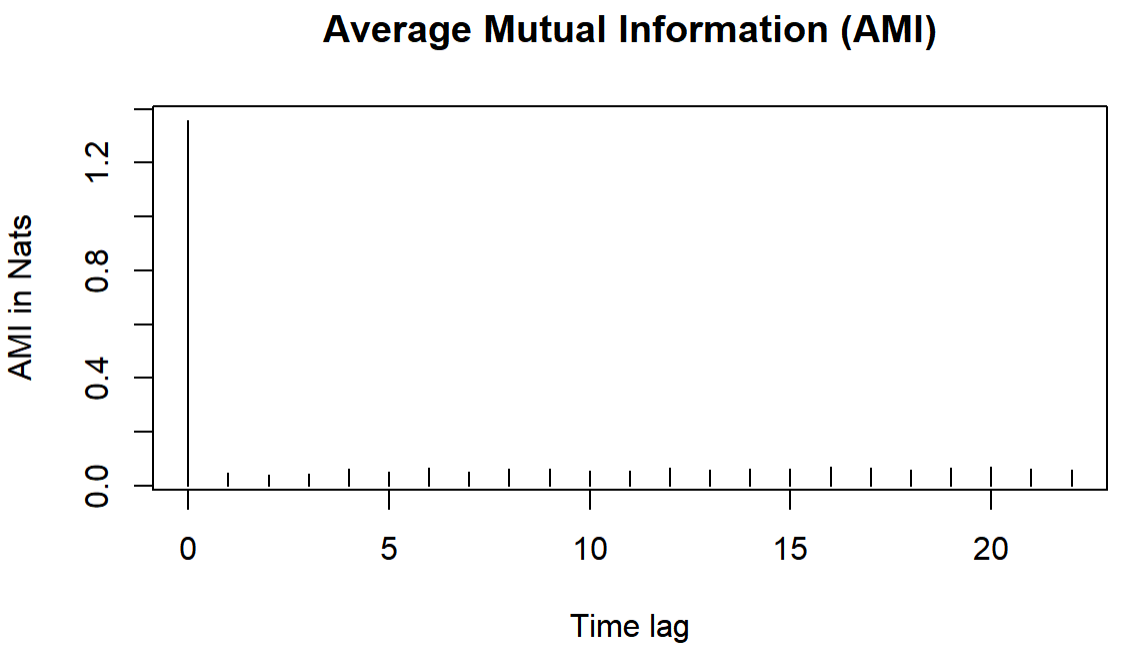
At the beginning, we do the scatter-line plot and get the following result.



On the one hand, it seems like a good evidence of white noise. But because it seems like there is no non-linear relationship between lags.

* Mutual Information Testing

For Average Mutual Information, we could get the following result.



It also shows that there is no cross or shared information between lags.

Sometimes however it happens so, that it’s not that simple to decide whether this type of nonlinearity is present. In this case, we’d have to run a statistical test — this approach is the most recommended by both Hansen’s and Tsays procedures.

1. Accurate Testing Method

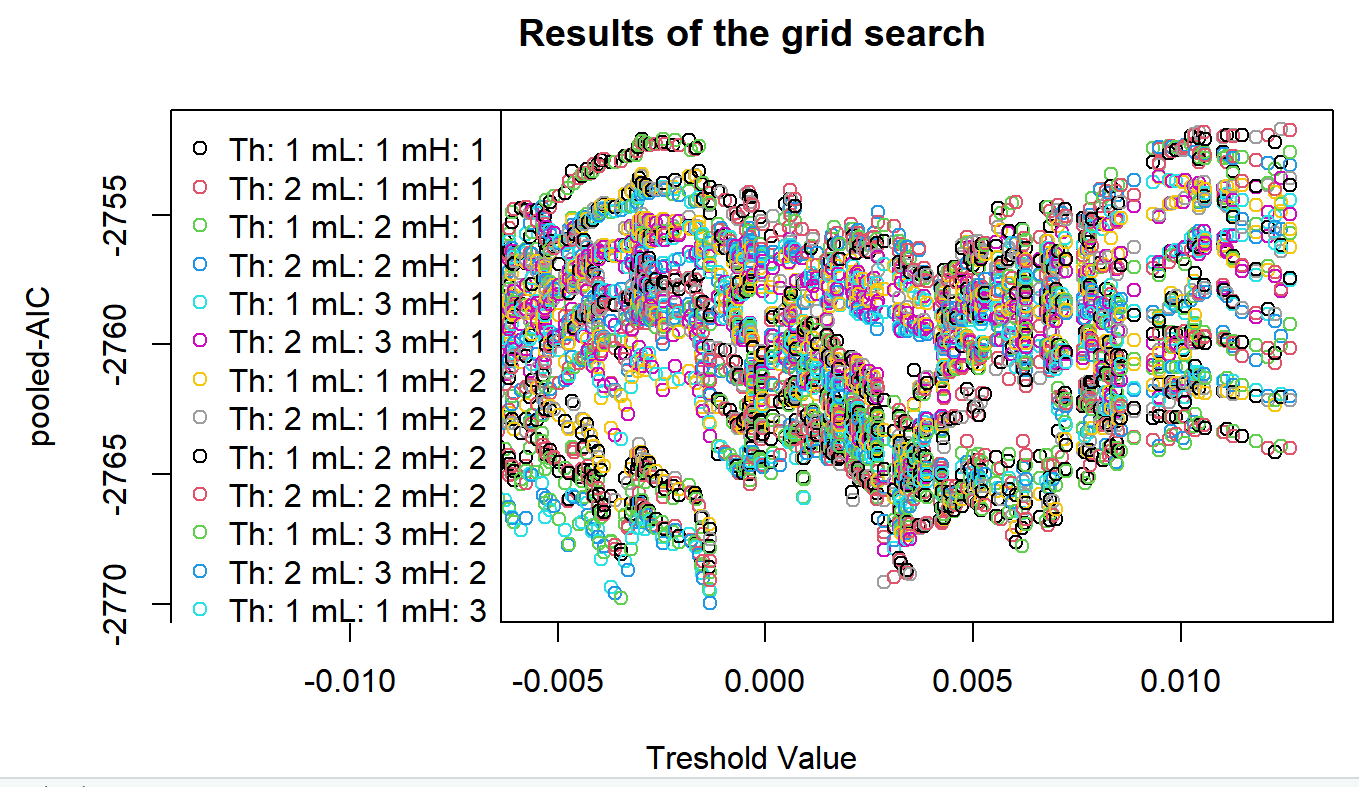
Here we apply the threshold testing by using grid search method, which searches the lag of the threshold (d) from 1 to 3 and orders of models (p) form 1 to 3, and we get the following F-statistics and p-value.

|  |  |  |  |
| --- | --- | --- | --- |
| d  p | 1 | 2 | 3 |
| 1 | 4.997893  (0.007134374) | 3.438502  (0.01686917) | 4.135606  (0.002668473) |
| 2 | 1.254824  (0.2861308) | 2.347373  (0.07206336) | 1.94638  (0.1017711) |
| 3 | 1.569419  (0.209316) | 1.401027  (0.2418645) | 1.003366  (0.4054115) |

According to the above result, TAR(1,1), TAR(1,2) and TAR(1,3) are possible models.

For much more accurate model specification, which means different order for different scenarios, we get the following result.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold Lag Order | Order 1 | Order 2 | Threshold Value | AIC |
| 1 | 2 | 1 | -0.001324173 | -2769.956 |
| 1 | 2 | 1 | -0.003473832 | -2769.775 |
| 1 | 2 | 1 | -0.003607720 | -2769.600 |
| 1 | 2 | 1 | -0.001325956 | -2769.498 |
| 1 | 2 | 1 | -0.001348086 | -2769.411 |
| 1 | 2 | 1 | -0.003718638 | -2769.349 |
| 2 | 1 | 2 | 0.002842765 | -2769.140 |
| 1 | 1 | 1 | -0.001324173 | -2769.092 |
| 1 | 2 | 1 | -0.001511602 | -2769.004 |
| 2 | 1 | 2 | 0.003096540 | -2768.983 |



Considering low AIC, TAR(2,1,2) with threshold 0.003096540 is the best one.

1. Model Fitting and Checking

After fitting TAR(2,1,1), we could get the model like this.

For Model Checking part, we could get p-value for white noise checking is 0.4471 and that for ARCH test is 0.03417. Thus, we still need GARCH model to specify the volatility part.

#### Model 3: Stochastic Volatility Model

If we apply the Stochastic Volatility Model to describe the log return, we could get anormal distribution and the following result.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.0003 | 0.0100 | -0.017 | -0.00017 | 0.016 | 10000 |
| phi | 0.9995 | 0.0004 | 0.999 | 0.99955 | 1.000 | 587 |
| sigma | 0.1578 | 0.0349 | 0.105 | 0.15565 | 0.218 | 109 |
| Exp(mu/2) | 0.9999 | 0.0050 | 0.992 | 0.99992 | 1.008 | 10000 |
| Sigma^2 | 0.0261 | 0.0117 | 0.011 | 0.02423 | 0.048 | 109 |

We could find that:

* Mean is equal to -0.0003 that is approximate to that part in the prior distribution,indicating there is no huge movement of the central;
* The standard deviation of mean is 0.0100, which shows the unconfidence level, in this case, it’s very small that shows great stable of our result;
* The mean of phi is close to 1 greatly, which indicates high self correlation; and the standard deviation is also very small, indicating high-level accuracy score;
* Sigma is averagely stayed at 0.1578 and exists 0.0349 fluctuation. And the confidence interval for 95% is [0.105, 0.218];
* Because exp(mu/2) is close to 1, it indicates there is no significant change at mu;

The model demonstrates that the estimates for the parameter `mu` and `exp(mu/2)` are very stable, while the value of `phi` being very close to 1 indicates a strong auto-correlation in the time series. The volatility in `sigma` and `sigma^2` is relatively high, indicating that the estimation of volatility in the model is not as stable as other parameters. These insights are particularly important for financial data analysis as they help in analyzing and predicting future market volatility.

### 5.3.2 Energy Market

#### Model 1: ARIMA-GARCH model

##### Previous Assessment:

The adfTest function in the fUnitRoots package is used to test the stationarity of the data, and the p value is 0.01 < 0.05, so we reject H\_0. It is a good signal that there is no unit root, indicating the movement is stationary right now.

Next, we check the ARCH effect.

Firstly, ARCH test was conducted to determine whether the time series data of this study had heteroscedasticity characteristics, and the calculated p value was 1.885\*10^{-6}<0.05, Therefore, we need to modify the ARMA model with GARCH model.

Model Construction Part

1. Residual Distribution Checking

We test the distribution of the log rate of return to determine the distribution of the data, to pave the way for the fitting of the ARMA model.

|  |  |  |
| --- | --- | --- |
|  | KS test | AD test |
| norm | 0.169 | 8.931e-06 |
| t | 0.9992(df=4.929, m= 0.0006316485  , s= 0.01347647) | 0.9967(df=4.929) |

Therefore we consider the log return series to follow a T-distribution.

1. ARMA Order Define

First, we draw the acf and pacf of the dataset.

图表

描述已自动生成

图表, 箱线图

描述已自动生成

We tried many ARIMA models, and finally got five different models after modification, and tested the residuals of the models. In the case of passing the residual test, their AIC is compared, and according to the principle of minimum AIC value, the model ARMA(2,2) without mean is finally selected.

|  |  |  |
| --- | --- | --- |
| Model | AIC | p |
| ARMA(1,2) without mean | -2595.6 | 0.9999135 |
| ARMA(2,1) without mean | -2594.39 | 0.288745 |
| ARMA(2,2) without mean | -2597.66 | 0.9208 |
| ARMA(3,3) without mean | -2597.26 | 0.7579754 |

1. GARCH Order Define

In this section, we try almost all GARCH(1,1) models, conduct residual test on the model, and select the model with the smallest AIC value as in 5.3.2.2.3. Here we choose the sgarch(1,1) model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | AIC | BIC | White Noise Test | ARCH Test |
| Sgarch(1,1)without mean | -5.3735 | -5.3050 | 0.1368443 | 0.2323223 |
| egarch(1,1)without omega, mean, ar2,ma2 | -5.3616 | -5.2845 | 0.5356058 | 0.3054561 |
| Gjrgarch(1,1)without omega, omega,gamma1, beta1 | -2.85138 | -2.7999 | 0.570898 | 0.9391288 |
| apgarch(1,1)without omega and gamma1 | -5.3713 | -5.3028 | 0.223174 | 0.1098353 |

Finally, for this dataset, we construct a ARIMA(2,0,2)-SGARCH(1,1) model:

Prediction

We choose from March 1, 2024 to April 30, 2024 as the test data of seasonal time series analysis, and this study calculates the mean square error loss of the test data to judge the superiority and inferiority of the model, the formula is as follows。

|  |  |  |  |
| --- | --- | --- | --- |
| date | True value | Estimate | MSE |
| 2024/3/3 | -0.006347363 | -0.006178 | 0.001663068 |
| 2024/3/10 | -0.001459779 | -0.006178 |
| 2024/3/17 | -0.003362752 | -0.006178 |
| 2024/3/24 | -0.006858488 | -0.006178 |
| 2024/3/31 | 0.00594277 | -0.006178 |
| 2024/4/7 | -0.004698289 | -0.006178 |
| 2024/4/14 | -0.002842948 | -0.006178 |
| 2024/4/21 | 0.027988764 | -0.006178 |
| 2024/4/28 | 0.011280416 | -0.006178 |

#### Model 2: Threshold AR Model

For better capturing the stock movement, we consider that whether there are different power influencing the stock movement. Thus, we want to apply Threshold AR model on it.

Rough Testing Method

1. Scatter-line Plot  
   At the beginning, we do the scatter-line plot and get the following result.

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

These plots indicate that the data is a white noise.

Mutual Information Testing

For Average Mutual Information, we could get the following result.

图表

描述已自动生成

It also shows that there is no shared information between lags.

However, to prove the linearity exists, we’d have to run a statistical test — this approach is the most recommended by both Hansen’s and Tsays procedures.

Accurate Testing Method

Here we apply the threshold testing by using grid search method, which searches the lag of the threshold (d) from 1 to 3 and orders of models (p) form 1 to 3, and we get the following F-statistics and p-value.

|  |  |  |  |
| --- | --- | --- | --- |
| d  p | 1 | 2 | 3 |
| 1 | 0.6732497  (0.5105648) | 0.2932374 (0.7459887) | 3.372043 (0.03519869) |
| 2 | 0.4397831 (0.7246558) | 2.352712 (0.07156076) | 2.79907  (0.03969529) |
| 3 | 0.315296 (0.8677502) | 1.978843 (0.09672164) | 0.7432428 (0.5629059) |

According to the above result, TAR(3,1), TAR(2,2) are possible models.

For much more accurate model specification, which means different order for different scenarios, we get the following result.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold Lag Order | Order 1 | Order 2 | Threshold Value | AIC |
| 2 | 1 | 1 | -0.007462866 | -2600.413 |
| 2 | 1 | 1 | -0.007436058 | -2599.924 |
| 2 | 1 | 1 | -0.007436353 | -2599.916 |
| 2 | 1 | 2 | -0.007462866 | -2599.350 |
| 2 | 3 | 1 | -0.007462866 | -2599.304 |
| 2 | 1 | 1 | -0.007294616 | -2599.278 |
| 2 | 1 | 2 | -0.007436058 | -2598.962 |
| 2 | 3 | 1 | -0.007436058 | -2598.908 |
| 2 | 1 | 2 | -0.007436353 | -2598.898 |
| 2 | 3 | 1 | -0.007436353 | -2598.787 |

图表, 散点图

描述已自动生成

Considering low AIC, TAR(2,1,1) with threshold 0.003096540 is the best one.

Model Fitting and Checking

Through TAR(2,1,1), we could get the model like this.

For Model Checking part, we could get p-value for white noise checking is 0.7367 and that for ARCH test is 0.001376. The results shows that the model isn’t perfectly accurate. In the future time, we should try to combine TAR model with GARCH model.

#### Model 3: Stochastic Volatility Model

We apply the Stochastic Volatility Model and got the following result.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.00091 | 0.01001 | -0.0174 | -0.00075 | 0.015 | 9642 |
| phi | 0.99954 | 0.00041 | 0.9987 | 0.99966 | 1.000 | 152 |
| sigma | 0.13148 | 0.04332 | 0.0661 | 0.12725 | 0.214 | 48 |
| Exp(mu/2) | 0.99956 | 0.00500 | 0.9914 | 0.99963 | 1.008 | 9642 |
| Sigma^2 | 0.01916 | 0.01291 | 0.0044 | 0.01619 | 0.046 | 48 |

We conclude:

* Mean is equal to -0.00091 that is approximate to that part in the prior distribution, indicating there is no huge movement of the central;
* The standard deviation of mean is 0.01001. It’s very small that shows great stable of our result;
* The mean of phi is close to 1 greatly, which indicates high self-correlation; and the standard deviation is also very small, indicating a high accuracy;
* Sigma is averagely stayed at 0.13148 and its standard deviation is 0.04332. At the same time, the confidence interval for 95% is [0.0661, 0.214];
* Because exp(mu/2) is close to 1, it indicates there is no significant change at mu;

The model demonstrates that the estimates for the parameter `mu` and `exp(mu/2)` are very stable, while the value of `phi` being very close to 1 indicates a strong auto-correlation in the time series. The volatility in `sigma` and `sigma^2` is relatively high. This implies that the estimation of volatility in the model is not as stable as other parameters.

# 6 Analysis

## 6.1 Event Effect Analysis

### 6.1.1 Financial Market

#### Model Analysis

According to the previous analysis, we get following models.

* + - 1. Short-term Effect Analysis: ARMA-GARCH Model

We can find that,

* For the ARMA part, it still seems like white noise feature.
* For the GARCH part, it changes greatly. Originally, it follows GARCH, which gives same power of lag1 element that controlled by ; After the war, it changes to GJRGARCH one, which provide different power effect of lag1 element. Considering the magnitude, negative lag1 effect changes from to ; and positive one changes from 0.092988 to 0. Thus, we could find that this war let the market becomes more and more negatively fluctuated and unstable compared with the previous one, especially for the downward fluctuation. It is rational.
* For intercept at GARCH part, even though there is a sightly increasing after this event, this is may because of huge fluctuation and adjustment effect.
  + - 1. Short-term Movement Analysis: Threshold AR Model

After fitting TAR(2,1,1), we could get the model like this.

We can find that,

* Before the event, there is no TAR model that is significant. It means there is no non-linearity before event.
* After the event, TAR(2,1,1) could fit the phenomenon in a perfect way. It shows that the threshold is approximately 0.13%. If the log return is lower than 0.13%, the effect of lag1 is positive, which push the log return gradually climb close to the threshold; if log return is higher than 0.31%, the effect of lag1 is negative, which gives a negative effect on log return, which push it decrease and back to the threshold. That means, after the war, there exist the central reverting property, which is also rational especially combined with the high fluctuation.
  + - 1. Long-term Effect Analysis: Stochastic Volatility Model

Before the war,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.00032 | 0.01007 | -0.0171 | -0.0004 | 0.016 | 10000 |
| phi | 0.99947 | 0.00044 | 0.9986 | 0.9996 | 1.000 | 289 |
| sigma | 0.16859 | 0.04743 | 0.0976 | 0.1636 | 0.252 | 75 |
| Exp(mu/2) | 0.99985 | 0.00503 | 0.9915 | 0.9998 | 1.008 | 10000 |
| Sigma^2 | 0.03067 | 0.01735 | 0.0095 | 0.0268 | 0.063 | 75 |

After the war,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.0003 | 0.0100 | -0.017 | -0.00017 | 0.016 | 10000 |
| phi | 0.9995 | 0.0004 | 0.999 | 0.99955 | 1.000 | 587 |
| sigma | 0.1578 | 0.0349 | 0.105 | 0.15565 | 0.218 | 109 |
| Exp(mu/2) | 0.9999 | 0.0050 | 0.992 | 0.99992 | 1.008 | 10000 |
| Sigma^2 | 0.0261 | 0.0117 | 0.011 | 0.02423 | 0.048 | 109 |

In comparing the pre-war and post-war estimates from the Stochastic Volatility Model, we can derive the following conclusions:

* The mean parameter remains very similar before and after the war, indicating that the long-term mean level of the process has not changed significantly due to the war; And for exp(mu/2), this indicates very minimal changes in the price level, which are nearly negligible.
* The phi value remains near 1 both before and after the war, suggesting a very strong persistence in the time series. The increase in ESS post-war indicates more stability and reliability in the estimation of this parameter;
* There is a slight decrease in the average level of volatility and its standard deviation post-war, which is from 0.16859 to 0.1578, suggesting a reduction in market volatility and more stable estimates; Besides, the mean and standard deviation of volatility squared decrease post-war, indicating a lower level of market volatility and more stable estimates.

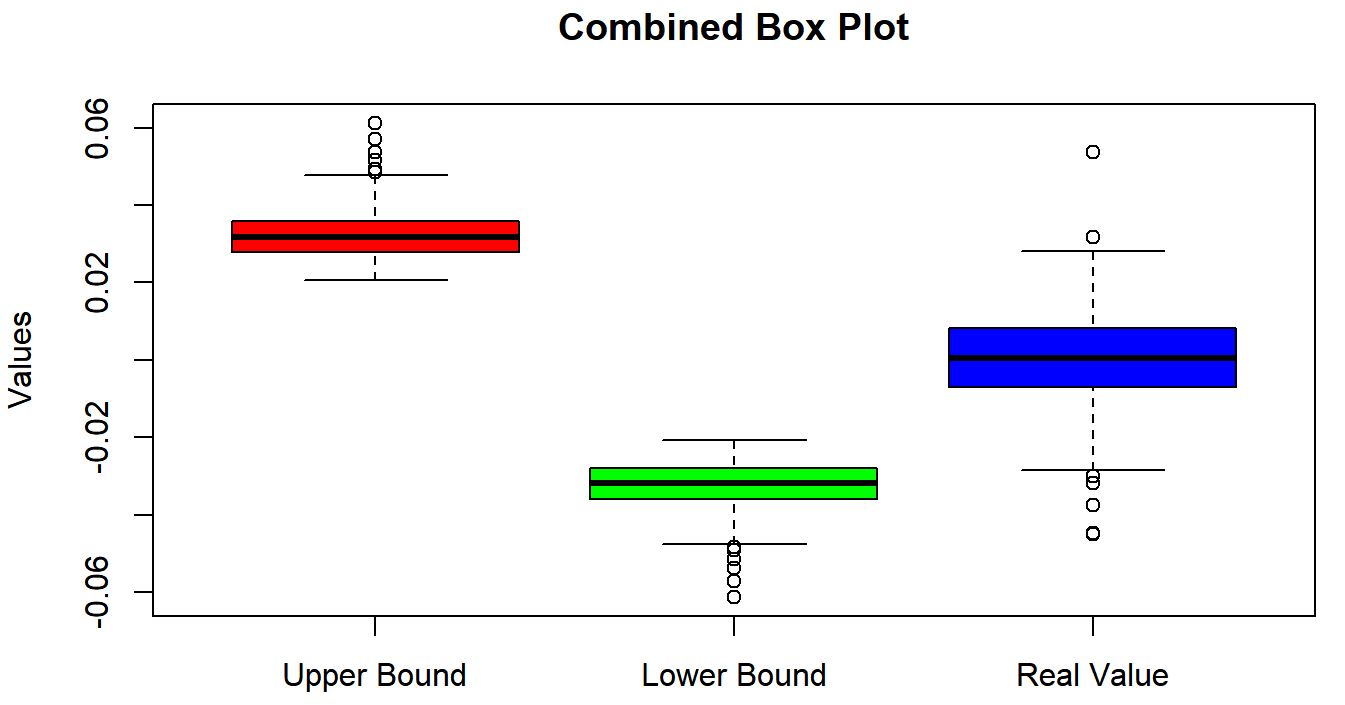
Overall, the post-war model exhibits lower and more stable market volatility compared to pre-war. The autocorrelation coefficient (phi) near 1 remains unchanged, indicating high persistence in the time series data. The ESS for parameters phi and sigma improved post-war, suggesting enhanced stability and reliability of the model's estimates.

#### Quantitative Analysis

For quantifying the influence of the war, we try to use the optimal model before the event, and then predict the return after the war, which results show the corresponding returns if the war has not happened. Here we select the period from January 2022 to May 2022, and the following is the result.

If confidence level is 5%, we get the following result, where predict result is based on t-distribution with df=7.996378.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Predict Mean | Predict Sigma | Upper Bound | Lower Bound |
| 2022-01-03 | 0 | 0.009161331 | 0.02112773 | -0.02112773 |
| 2022-01-04 | 0 | 0.008965546 | 0.02067622 | -0.02067622 |
| 2022-01-05 | 0 | 0.009205598 | 0.02122982 | -0.02122982 |
| 2022-01-06 | 0 | 0.009304129 | 0.02145705 | -0.02145705 |
| 2022-01-07 | 0 | 0.009024303 | 0.02081172 | -0.02081172 |

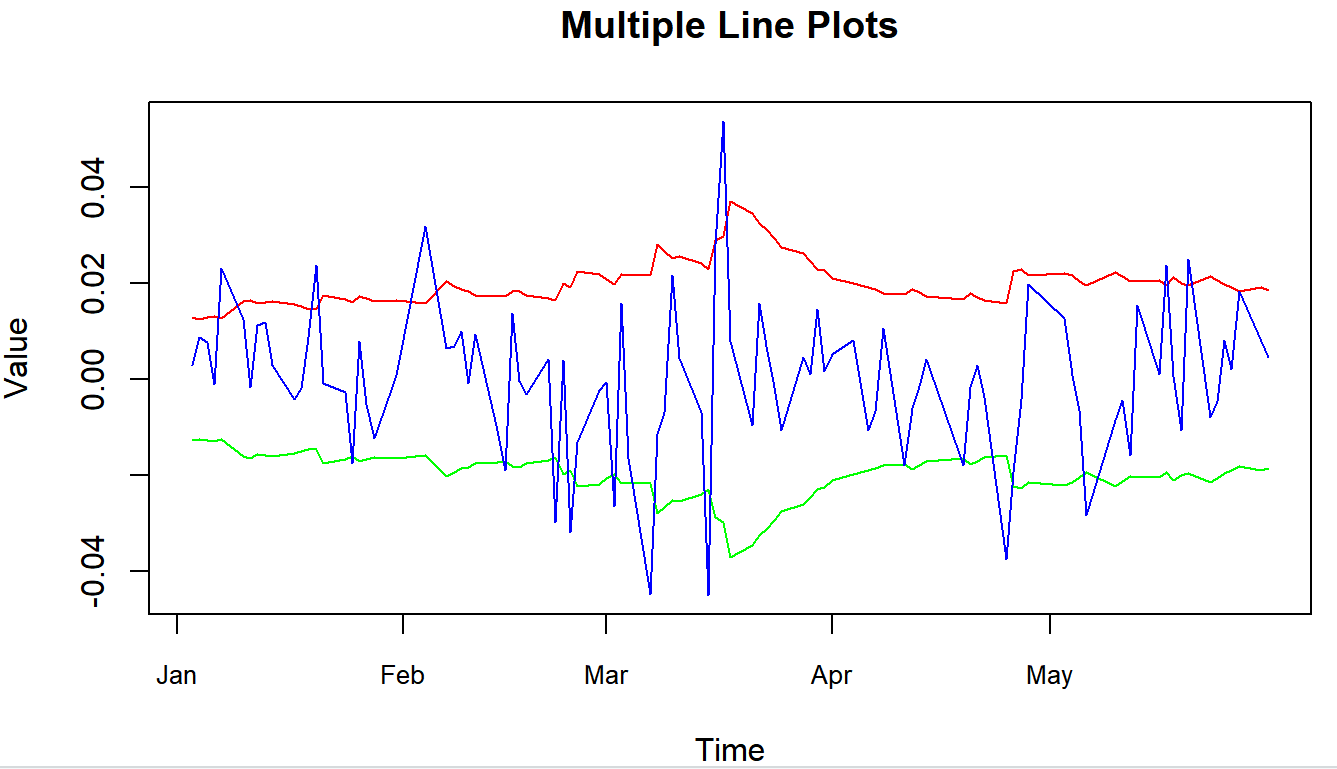
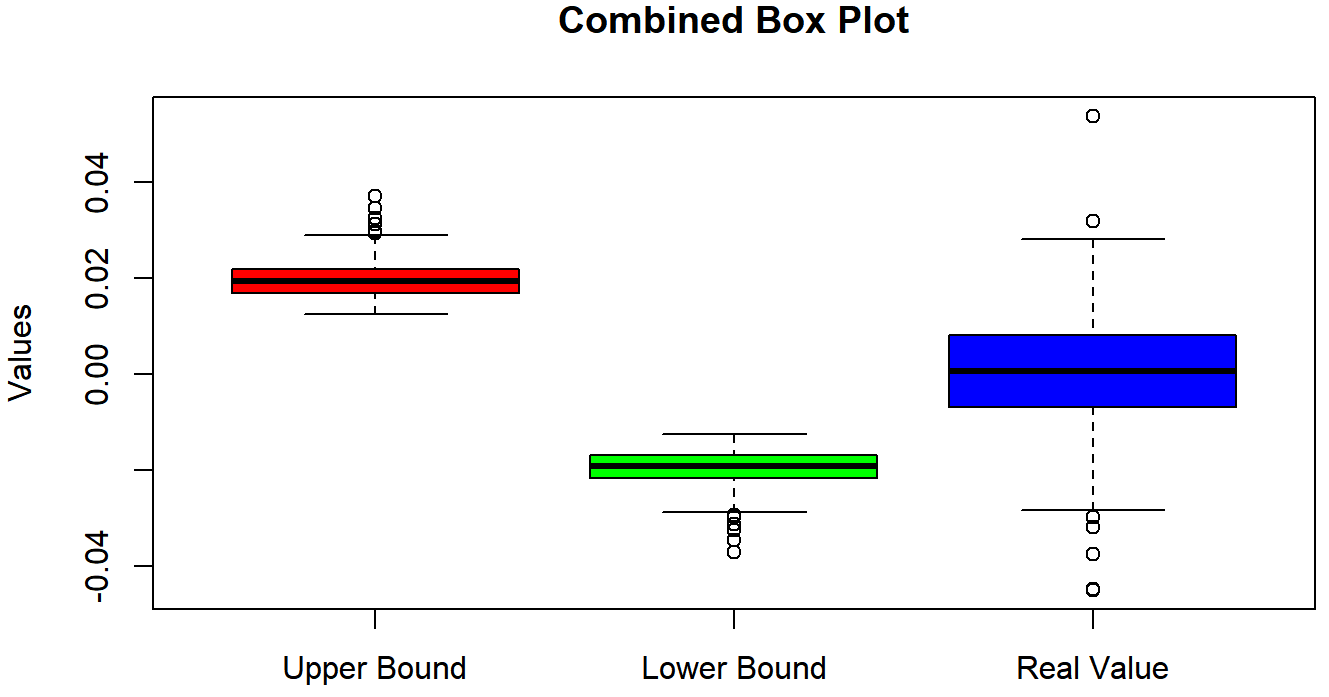


According to the above results, we could know that,

* There is 5 time points where are out of the confidence interval;
* Usually, in 5% confidence level, most of these real value is located at the confidence interval, which means, in this level, the effect of the war is predictable;

If confidence level is 20%, we get the following result, where predict result is based on t-distribution with df=7.996378.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Predict Mean | Predict Sigma | Upper Bound | Lower Bound |
| 2022-01-03 | 0 | 0.009161331 | 0.01279721 | -0.01279721 |
| 2022-01-04 | 0 | 0.008965546 | 0.01252372 | -0.01252372 |
| 2022-01-05 | 0 | 0.009205598 | 0.01285904 | -0.01285904 |
| 2022-01-06 | 0 | 0.009304129 | 0.01299668 | -0.01299668 |
| 2022-01-07 | 0 | 0.009024303 | 0.01260580 | -0.01260580 |



According to the above results, we could find that,

* There is 11 time points where are out of the confidence interval;
* From Box plot, we could find that, there is no such relative large fluctuation for boundaries; but there are the big fluctuation and more extreme cases for real returns;
* According to the line change, we could find that there are some time points where the fluctuation is so large that the previous model cannot capture it, and its corresponding prediction has slow adjustment, which means it tries to get a larger interval when we meet such cases, even though it did not capture previously.

The analysis of the impact of war on financial markets, utilizing ARMA-GARCH, Threshold AR, and Stochastic Volatility models, provides insights into both short-term and long-term market behaviors. In the immediate aftermath of the war, the switch from GARCH to GJRGARCH models indicates increased negative fluctuation and market instability, particularly with more significant downward movements. This reflects typical market responses to geopolitical uncertainty, characterized by heightened volatility. The Threshold AR model suggests that post-war market dynamics adopted a central reverting behavior, where returns oscillate around a newly established threshold, indicative of the market's attempt to stabilize amidst the turmoil.

In terms of long-term effects, the Stochastic Volatility Model reveals that while the overall mean level of the market remains stable, indicating resilience in the fundamental market structure, there is a noticeable reduction in volatility post-war. This reduced volatility points towards a gradual stabilization of the market over time. The persistence of high autocorrelation (phi near 1) both before and after the war further underscores this resilience, suggesting that the core characteristics of the market are largely unaffected in the long run.

Quantitative predictions using pre-war models to forecast post-war returns indicated that actual market returns frequently deviated from these predictions, especially at tighter confidence levels. This mismatch highlights the challenges predictive models face in accounting for the extreme market behavior typically triggered by unexpected geopolitical events like wars.

Overall, the market's response to the war—initial increased volatility followed by a stabilization phase—is a rational reaction to the heightened uncertainty and risk, demonstrating the market's inherent capacity to adapt and regain equilibrium. This analysis not only illustrates the immediate shocks and subsequent adjustments within financial markets following geopolitical disturbances but also highlights the limitations of predictive models in times of such extreme uncertainty.

### 6.1.2 Energy Market

#### Model Analysis

According to the previous analysis, we get following models.

* + - 1. Short-term Effect Analysis: ARMA-GARCH Model

We can find that,

* For the ARMA part, before the event, the log return of the closing price of the energy index is white noise and does not depend on the previous data. This simplified model may be due to the fact that the market was relatively stable during this period with no apparent trend. However, after the event, the log return depends not only on its own lagged values and but also on the white noise and its lagged values and . This model is more complex than the previous ones. One of conjectures of this study as to the possible reasons for the changes occurring is **the geopolitical risk**. The log yield of the previous order is positive, and people are more inclined to believe that the yield will not continue to rise, which is most likely because the Russia-Ukraine war may lead to intensified geopolitical risk, trigger international concerns about the region, and people are not optimistic about the Hong Kong stock market, thus affecting market sentiment and leading to a downward trend in stock prices. Besides, **volatility in energy markets** is also important. Russia is one of the largest exporters of natural gas in the world, and Ukraine is one of the transit routes for natural gas, so the war may have an impact on the gas market and the energy market, which can lead to lower stock prices. What's more, the **economic sanctions** is also important. The war may lead to the implementation of economic sanctions against Russia by the international community. As the second largest exporter of crude oil and the first largest exporter of natural gas, Russia's energy export will not be smooth, which will indirectly affect the energy market of Hong Kong.
* For GARCH part, Change of model: Prior to the conflict, the EGARCH(1,1) model was employed to characterize volatility's heteroscedasticity, excluding the alpha1 term. This omission suggests a reduced volatility symmetry, where market participants perceive and react differently to price increases versus decreases. This modeling choice likely stemmed from insights into prior data volatility patterns and market environment cognizance. Post-Russia-Ukraine War (SGARCH(1,1)-STD): Following the war's outbreak, market volatility may have shifted, necessitating model adjustments to better capture volatility dynamics. Thus, the optimal model transitioned to SGARCH(1,1), reflecting the belief that market participants are more sensitive to absolute price changes rather than price fluctuations. This shift indicates a move towards more symmetric volatility.
* Change of intercept. Before the war, the intercept pf the model is 0.6269114, while it changed to after the war. Change of effect of lag1 element. The impact of lag1 volatility is 2.560161 before the event, while the impact after the event has occurred is 0.902453, which means that the volatility is relatively more independent and less affected by the volatility of lag1. ~~In summary, changes in the ARIMA and GARCH models likely aimed to enhance adaptation to the post-Russia-Ukraine war market environment, facilitating the capture of dynamic data characteristics and volatility fluctuations.~~

1. Short-term Movement Analysis: Threshold AR model

After fitting TAR(2,1,1), we could get the model like this:

We can find that,

* Before the event, there is no TAR model that is significant. It means there is no non-linearity before event.
* After the event, TAR(2,1,1) could fit the phenomenon to some extent. It shows that the threshold is approximately -0.75%. If the log return is lower than -0.75%, the effect of lag1 is positive, which push the log return gradually climb close to the threshold; if log return is higher than -0.75%, the effect of lag1 is negative and more smaller, which gives a negative effect on log return, which push it decrease and back to the threshold. That means, after the war, there exist the central reverting property, which is also rational especially combined with the high fluctuation.
* But we use the white noise test and ARCH test to check the TAR model, finding that it passed the white noise test but failed the ARCH test. There may exist some Heteroscedasticity phenomenon. This illustrates the need to find models that combine GARCH and TAR models.

1. Long-term Effect Analysis: Stochastic Volatility Model

Before the war,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.00024 | 0.01003 | -0.017 | -0.00017 | 0.016 | 10707 |
| phi | 0.99877 | 0.00084 | 0.997 | 0.99895 | 1.000 | 646 |
| sigma | 0.25504 | 0.05144 | 0.178 | 0.25200 | 0.347 | 128 |
| Exp(mu/2) | 0.99989 | 0.00501 | 0.992 | 0.99992 | 1.008 | 10707 |
| Sigma^2 | 0.06769 | 0.02780 | 0.032 | 0.06350 | 0.120 | 128 |

After the war,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | 5% | 50% | 95% | ESS |
| mu | -0.00091 | 0.01001 | -0.0174 | -0.00075 | 0.015 | 9642 |
| phi | 0.99954 | 0.00041 | 0.9987 | 0.99966 | 1.000 | 152 |
| sigma | 0.13148 | 0.04332 | 0.0661 | 0.12725 | 0.214 | 48 |
| Exp(mu/2) | 0.99956 | 0.00500 | 0.9914 | 0.99963 | 1.008 | 9642 |
| Sigma^2 | 0.01916 | 0.01291 | 0.0044 | 0.01619 | 0.046 | 48 |

By comparing the pre-war and post-war estimates of stochastic volatility models, we can draw the following conclusions:

* The mean parameters remained very similar before and after the war, suggesting that the long-run mean level of the process did not undergo significant change due to the war. Specifically, for exp(mu/2), this implies that the change in the price level is minuscule and almost negligible.
* The parameter phi remains around 1 both before and after the war, indicating strong persistence in the time series.
* Both the mean level and standard deviation of volatility experienced a slight decrease after the war, moving from 0.25504 to 0.13148. This suggests lower market volatility and more stable estimates post-war. Furthermore, the mean and standard deviation of squared volatility also decreased after the war, indicating a lower level of market volatility and more stable estimates during this period.

In summary, the post-war model demonstrates lower and more stable market volatility compared to the pre-war period. The consistent autocorrelation coefficient (phi) around 1 indicates enduring persistence in the time series data. However, it's noteworthy that the effective sample size (ESS) of each parameter has significantly declined post-war, suggesting that the current model's stability is diminished compared to its pre-war counterpart. This could possibly be attributed to the substantial impact on the post-war energy market, subdued investor sentiment, and heightened sensitivity to new information.

#### Quantitative Analysis

If confidence level is 5%, we get the following result, where predict result is based on t-distribution with df= 3.393.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Predict Mean | Predict Sigma | Upper Bound | Lower Bound |
| 2022-03-01 | 0 | 0.01942009 | 0.03806338 | -0.0380634 |
| 2022-03-02 | 0 | 0.01867660 | 0.03660614 | -0.0366061 |
| 2022-01-03 | 0 | 0.01950266 | 0.03822521 | -0.0382252 |
| 2022-01-04 | 0 | 0.02103786 | 0.04123421 | -0.0412342 |
| 2022-01-07 | 0 | 0.02018524 | 0.03956307 | -0.0395631 |

图表, 箱线图

描述已自动生成

The chart above shows that the true value deviated from the forecast range by only three points, indicating that log returns remained relatively predictable even in the event of a war.

For quantifying the influence of the war, we try to use the optimal model before the event, and then predict the return after the war, which results show the corresponding returns if the war has not happened. Here we select the period from January 2022 to May 2022, and the following is the result.

图表, 折线图, 直方图

描述已自动生成

• From Box plot, we could find that, there is no such relative large fluctuation for boundaries, but there are big volatility for real returns;

• According to the line change, we could find that there are some time points where the fluctuation is so large that the previous model cannot capture it, and its corresponding prediction has slow adjustment, which means it tries to get a larger interval when we meet such cases, even though it did not capture previously.

The analysis of the impact of war on energy markets, utilizing ARMA-GARCH, Threshold AR, and Stochastic Volatility models, provides insights into both short-term and long-term market behaviors.

In the immediate aftermath of the war, the switch from EGARCH to SGARCH models indicates the vanishing asymmetry.

The Threshold AR model suggests that post-war market dynamics adopted a central reverting behavior, where returns oscillate around a newly established threshold, indicative of the market's attempt to stabilize amidst the turmoil.

In terms of long-term effects, the Stochastic Volatility Model reveals stable mean returns and lower volatility, suggesting that the core characteristics of the market are largely unaffected in the long run.

Quantitative predictions using pre-war models to forecast post-war returns indicated that actual market returns frequently deviated from these predictions, especially at tighter confidence levels. This mismatch illustrates the importance of adjusting the model in a timely manner in the face of unexpected conditions, which may cause large deviations from the forecast. Several factors may account for this outcome in light of market changes:

1. Market Sentiment: If market sentiment experiences a shift, investors may adjust their portfolios, leading to changes in asset price volatility. With heightened risk aversion, investors may lean towards holding safe-haven assets, affecting the returns of risk assets and rendering the model ineffective.

2. Investor Behavior: Market changes may alter investor trading behaviors. For instance, amidst market turbulence, investors may prefer rapid trading or adopt a more conservative investment strategy. Regardless of market direction, investors may perceive neither an optimistic nor pessimistic signal in the long term, favoring swift trades to lock in gains or minimize losses, thereby accentuating symmetry.

3. Political Events: Geopolitical tensions such as the Russia-Ukraine war can elevate market sentiment volatility, particularly concerning Russia's significance as an energy powerhouse. Investors weigh heavily on adverse factors, leading to market turmoil and increased asset price volatility, deviating from the model's predictions.

Overall, the relatively accurate mean forecast shows that the market has the ability to adapt and restore equilibrium, but the frequent occurrence of deviation values shows that the model with fixed coefficients has certain limitations after encountering unexpected events.

## 6.2 Correlation Analysis

How can we estimate the probability and risk measures of extreme black swan events, such as Extreme Value at Risk (EVaR) or tail risk, to enhance portfolio efficiency and robustness? Copula can be used to estimate extreme risk measures such as extreme value at risk or tail risk. Copula helps understand the dependency relationships between different assets, particularly in terms of tail risk. By modeling the dependency structure between upper and lower tails, a better understanding of the correlation between assets is achieved, facilitating risk diversification. This is crucial for constructing portfolios with lower tail risk or managing financial risks effectively.

Copulas are the part of a multivariate distribution function that fully captures the cross sectional dependence between the variables of interest and they have become a very popular tool to model dependencies different from the linear correlation of elliptical distributions. Copulas are popular in high-dimensional statistical applications as they allow one to easily model and estimate the distribution of random vectors by estimating marginals and copula separately.

Before copula, we should introduce Sklar’s theorem. Sklar's theorem, named after Abe Sklar, provides the theoretical foundation for the application of copulas. Sklar's theorem states that every multivariate cumulative distribution function of a random vector can be expressed in terms of its marginals , we have

And copula is unique if the marginals are continuous.

We can give the definition of copula in probabilistic terms, is a d-dimensional copula if C is a joint cumulative distribution function of a d-dimensional random vector on the unit cube with uniform marginals.

Next, we give some classical notation and properties of bivariate Elliptical and Archimedean Copula families included in CDVine.

图形用户界面, 文本, 表格

描述已自动生成

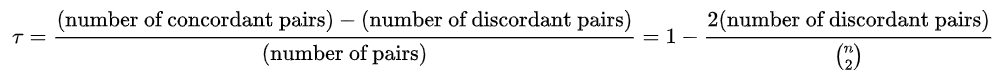
Kendall rank correlation coefficient is a statistic used to measure the ordinal association between two measured quantities. It is a measure of rank correlation: the similarity of the orderings of the data when ranked by each of the quantities. It is named after Maurice Kendall, who developed it in 1938, though Gustav Fechner had proposed a similar measure in the context of time series in 1897.

Now let’s introduce the definition of Kendall rank correlation. Let be a set of observations on the joint random variables X and Y, such that all the values are unique.

Then let’s introduce concordant and discordant. Any pair of observations and , it is concordant if the sort order of and agrees.

图片包含 图示

描述已自动生成



Intuitively, the Kendall correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) rank (i.e. relative position label of the observations within the variable: 1st, 2nd, 3rd, etc.) between the two variables, and low when observations have a dissimilar (or fully different for a correlation of −1) rank between the two variables.

### 6.2.1 Whole Point

For the residuals obtained from the weekly closing prices of Hang Seng Composite Index-Energy Industry and finance Industry from February 1, 2020 to February 29, 2024. The histogram figures are as follows:

图表, 散点图

描述已自动生成

图表, 直方图

描述已自动生成

Cor Fig 6

Then we plot the residuals. The lower off diagonal draws scatter plots, the diagonal histograms, the upper off diagonal reports the Pearson correlation (with pairwise deletion).

To do distribution test for both of them. For energy residuals, we use MLE of fitting the student t distribution and gained the maximum loglikelihood results with degree of freedom equal to 29.75. For finance residuals, we set different degrees of freedom and the best parameters rise as the max degree of freedom rises. The infinite degree of freedom means that the distribution of residuals may be norm distributed. The adtest result p value of the finance residuals is 0.9566 which means we can not reject the hypothesis that the data follows the norm distribution.

|  |  |  |
| --- | --- | --- |
| distributions | parameters | adtest p-value |
| Energy-student t dist | df=29.75 | 0.971 |
| Finance-student t dist | df=100 | 3.243e-05 |
| Finance-student t dist | df=500 | 2.107e-05 |
| Finance-norm dist | - | 0.7998 |

Then we try to fit the desired distribution to the corresponding data and calculate the CDF of the data. For the CDF values, we plot. The spearman correlation is 0.49.

图表, 直方图

描述已自动生成

图表, 散点图

描述已自动生成

Cor Fig 7

We can see that the corresponding CDF values are uniformly lying between [0, 1]. The density plot is as follows:

图表

描述已自动生成

Cor Fig 8

We try different types of copula functions to fit the CDF value list.

|  |  |  |
| --- | --- | --- |
| Copula types | Parameters | AIC |
| Gaussian copula | Upper TD=0; Lower TD=0 | -44.61 |
| Student t copula (t-copula) | Upper TD=0; Lower TD=0 | -42.28 |
| Clayton copula | Upper TD=0; Lower TD=0.38 | -37.51 |
| Gumbel copula | Upper TD=0.27; Lower TD=0 | -34.71 |
| Frank copula | Upper TD=0; Lower TD=0 | -40.43 |
| Joe copula | Upper TD=0.4; Lower TD=0 | -21.77 |
| Clayton-Gumbel copula | Upper TD=0.21; Lower TD=0.26 | -41.01 |
| Joe-Clayton copula | Upper TD=0.24; Lower TD=0.32 | -40.27 |

The best model is Gaussian copula with par equal to 0.5. It decide the shape of the copula. Kendall's tau, a measure of correlation, has a value of 0.34. It indicates the degree of correlation between variables, with values close to 1 representing strong positive correlation, close to -1 representing strong negative correlation, and close to 0 indicating no correlation. Both Upper TD and Lower TD have values of 0, indicating very weak dependence in the extreme tails, and the variables are nearly independent.

### 6.2.2 Before Event Point

For the residuals obtained from the daily log return residuals from the finance and energy dataset ranging from 2020-2-3 to 2021-12-31.The histogram figures are as follows:

图表, 直方图

描述已自动生成

图表, 直方图

描述已自动生成

Cor Fig 6

We do not need to build MLE models for them because we have already assigned the distribution of residuals in the model we build in ARMA-GARCH. They are both student t distributions.

Then we try to fit the desired distribution to the corresponding data and calculate the CDF of the data.

图表, 散点图

描述已自动生成

图表, 折线图

描述已自动生成

Cor Fig 7

We can see that the corresponding CDF values are uniformly lying between [0, 1]. The density plot is as follows:

图表, 散点图

描述已自动生成

Cor Fig 8

We try different types of copula functions to fit the CDF value list.

|  |  |  |
| --- | --- | --- |
| Copula types | Parameters | AIC |
| Gaussian copula | Upper TD=0; Lower TD=0 | -256.68 |
| Student t copula (t-copula) | Upper TD=0.25; Lower TD=0.25 | -267.05 |
| Clayton copula | Upper TD=0; Lower TD=0.59 | -255.53 |
| Gumbel copula | Upper TD=0.5; Lower TD=0 | -214.73 |
| Frank copula | Upper TD=0; Lower TD=0 | -234.03 |
| Joe copula | Upper TD=0.52; Lower TD=0 | -135.24 |
| Clayton-Gumbel copula | Upper TD=0.27; Lower TD=0.52 | -279.63 |
| Joe-Clayton copula | Upper TD=0.31; Lower TD=0.56 | -276.66 |

From the table above we can see that the best copula is Clayton-Gumbel copula with par1 and par2 equal to 0.83 and 1.27. Kendall's tau, a measure of correlation, has a value of 0.44. It indicates the degree of correlation between variables. The Upper TD (upper tail dependence) has a value of 0.27, while the Lower TD (lower tail dependence) has a value of 0.52. These values suggest that there is moderate dependence in the extreme tails of the variables. And the lower TD is larger than the upper TD which means that This indicates that when one variable has extreme low values, the other variable is more likely to have extreme low values than that of extreme high values.

### 6.2.3 After Event Point

For the residuals obtained from the daily log return residuals from the finance and energy dataset ranging from 2021-3-1 to 2024-2-29.The histogram figures are as follows:

图表

描述已自动生成

图表, 直方图

描述已自动生成

Cor Fig 6

We do not need to build MLE models for them because we have already assigned the distribution of residuals in the model we build in ARMA-GARCH. They are both student t distributions.

Then we try to fit the desired distribution to the corresponding data and calculate the CDF of the data.

图表, 散点图

描述已自动生成

图表, 折线图

描述已自动生成

Cor Fig 7

We can see that the corresponding CDF values are uniformly lying between [0, 1]. The density plot is as follows:

图表, 散点图

描述已自动生成

Cor Fig 8

We try different types of copula functions to fit the CDF value list.

|  |  |  |
| --- | --- | --- |
| Copula types | Parameters | AIC |
| Gaussian copula | Upper TD=0; Lower TD=0 | -237.49 |
| Student t copula (t-copula) | Upper TD=0.04; Lower TD=0.04 | -236.99 |
| Clayton copula | Upper TD=0; Lower TD=0.5 | -188.17 |
| Gumbel copula | Upper TD=0.47; Lower TD=0 | -208.71 |
| Frank copula | Upper TD=0; Lower TD=0 | -226.77 |
| Joe copula | Upper TD=0.52; Lower TD=0 | -151.52 |
| Clayton-Gumbel copula | Upper TD=0.35; Lower TD=0.32 | -230.86 |
| Joe-Clayton copula | Upper TD=0.39; Lower TD=0.41 | -222.67 |

The best model is Gaussian copula with upper and lower td equal to 0. In the Gaussian Copula model, we separate the modeling of marginal distributions from the modeling of their correlations. By doing so, we have the flexibility to choose different distributions for the marginals, rather than being restricted to normal distributions. The Kendall’ tau is 0.42. A TD of 0 for both the upper and lower tails indicates that the dependence between variables is extremely weak in the extreme tail regions. It suggests that the variables are nearly independent, especially in extreme events.

### 6.2.4 Conclusion and Reasoning

From the copula results for the whole point, before war and after war. We can see that taking a longer-term perspective, the correlation between the upper tail and lower tail is not significant. Before the war, there was a certain degree of correlation between the upper and lower tails, with a higher probability in the lower tail. However, after the war, the correlation between the upper and lower tails is not significant.

Before the Russia-Ukraine war, they exhibited certain upper and lower tail effects. While the overall situation may not be apparent, there is reason to believe that after the impact of the Russia-Ukraine war subsides, we should continue to monitor their mutual relationship. Employing various financial instruments and derivatives such as options, futures, and derivatives products can help hedge against risks arising from extreme events. For example, purchasing protective option contracts can provide downside protection during market declines, reducing portfolio losses. Employing risk models and conducting stress tests allows for risk assessment of the portfolio and simulation of the impact under different extreme scenarios. By identifying potential risk factors and extreme conditions, appropriate risk management strategies can be implemented.

To dig into the reasons why tail properties are changed. Firstly, we check the distribution of energy and finance in whole, before and after. We can see that the distributions of different stages have a lot uncommon. QQ plots indicate that only the distribution before the war exhibits a strong upper and lower tail, while it is less pronounced during other time periods. The overall data for the weekly observations is not very clear, which may be due to the smoothing effect of the weekly aggregation, obscuring the extreme values.

图表

描述已自动生成

图形用户界面, 图表, 折线图

描述已自动生成

# Appendix

2.1 Financial Time Series Marginal Distribution Models

Financial time series analysis involves understanding the properties of historical financial data to forecast future values and make informed financial decisions. One of the primary aspects of this analysis is modeling the marginal distribution of time series data, which refers to understanding and modeling the distribution of values at any point in time, independent of other time points.

2.1.1 Autoregressive Moving Average Model (ARMA)

The Autoregressive Moving Average (ARMA) model is a cornerstone of financial time series analysis, combining aspects of autoregressive (AR) models and moving average (MA) models. This model is particularly useful for modeling time series data showing autocorrelations, where future values are influenced by past values and random disturbances. The general form of an ARMA model, denoted as ARMA(p, q), is given by :

where is the time series value at time , are the parameters of the AR part, are the parameters of the MA part, is the backward shift operator, , is the white noise error term at time .

2.1.2 Conditional Variance Models

2.1.2.1 ARCH (Autoregressive Conditional Heteroskedasticity)

The ARCH model, introduced by Engle in 1982, captures time-varying volatility by allowing the variance of the current error term to be a function of the actual sizes of previous time periods' errors. The model assumes that the error terms are heteroskedastic, meaning their variance changes over time. The ARCH model can be mathematically represented as:

where is the conditional variance of the error term at time , are the residuals at time , are parameters to be estimated, is the lag order of the ARCH model.

2.1.2.2 GARCH (Generalized Autoregressive Conditional Heteroskedasticity)

GARCH models, extending ARCH, include both past squared residuals and past variances to model current variance. This extension provides a more comprehensive framework for volatility modeling.

Basic GARCH: This model is the foundational GARCH model, typically denoted as GARCH(p, q), where 'p' and 'q' are the orders of the autoregressive and moving average parts, respectively.

where is the number of lagged squared residuals, is the number of lagged conditional variances, are coefficients for squared residuals, are coefficients for lagged conditional variances.

SGARCH (Stable GARCH): Assumes a stable process in the long-term variance, Same as Basic GARCH with stability constraints on parameters.

EGARCH (Exponential GARCH): Modifies the GARCH model to account for asymmetric effects of shocks of different signs.

Where ,

TGARCH (Threshold GARCH):

Allows for different responses based on whether the shock is positive or negative, useful for modeling leverage effects.

Where is an indicator variable such that, . One expects to be positive so that prior negative returns have higher impact on the volatility.

IGARCH (Integrated GARCH):

Implies a unit root in the GARCH process, suggesting a persistent, time-dependent volatility.

where is the forecast origin. The effect of on future is persistent, and the volatility forecasts form a straight line with slope .

2.2 Definition of Copula Functions

A copula is a function from the unit square [0,1]×[0,1] to the interval [0,1], and it is designed to capture the dependence structure between random variables. Its definition conforms to the following properties:

1. For any the volume of the copula over the rectangle defined by these points must be non-negative: .

2.3 Common Bivariate Copula Functions

1）Bivariate Normal Copula: Assumes a normal distribution of marginals and is symmetrical.

where is the linear correlation coefficient, and is the inverse function of the standard normal distribution function .

1. Bivariate t-Copula: Similar to the normal Copula but with heavier tails, providing a better model for extreme co-movements.

where is the linear correlation coefficient, and is the inverse function of the univariate t-distribution function with degrees of freedom.

1. Bivariate Gumbel Copula: Focuses on modeling the upper tail dependencies.

where [1, ) is the correlation parameter, and .

1. Bivariate Clayton Copula: Well-suited for modeling lower tail dependencies.

where is the correlation parameter.

5）Bivariate Frank Copula: Used for modeling dependencies without a focus on tail behavior.

where is the correlation parameter, .

2.4 Definition of Time-varying Copula Functions

Time-varying copulas are extensions of the classical copula functions that allow for dynamic changes in the dependency structure between random variables over time. These models are particularly useful in financial applications where correlations between assets can vary significantly during different market conditions.

2.5 Common Bivariate Time-varying Copula Functions

Time-varying Copula functions possess the following properties:

1. For any , we have

.

is the conditional joint distribution of the variables X, given by .

1) Bivariate Time-varying Normal Copula:

The time-varying normal copula modifies the correlation parameter to be a function of time, , often modeled using a dynamic process such as a GARCH model:

The functional expression is:

The parameter evolution equation is:

where the adjustment function

2）Bivariate Time-varying SJC Copula:

The Symmetric Joe-Clayton (SJC) copula introduces time-varying tail dependencies. The parameters controlling the upper and lower tail dependence and can be modeled dynamically:

The functional expression is:

The parameter evolution equation is:

where,

ith and where and represent the upper and lower tail correlation coefficients, respectively, and the adjustment function

2.3 ARMA-GARCH-t-Copula model

For the return series , where and , we establish an ARMA model, model its conditional heteroskedasticity using a GARCH-t model, then select an appropriate Copula function, and ultimately formulate the ARMA-GARCH-t-Copula model:

2.4 ARMA-GARCH-SJC-Copula model

For the return series , where and , we establish an ARMA model, model its conditional heteroskedasticity using a GARCH model, then select the Symmetrized Joe-Clayton Copula function, and ultimately formulate the ARMA-GARCH-SJC-Copula model:

where and are the marginal transformations (typically Student's t-distributions in the GARCH-t framework), and represents the Symmetrized Joe-Clayton Copula characterized by its parameters (upper tail dependence) and (lower tail dependence). These parameters can be modeled dynamically or kept static depending on the complexity and needs of the analysis.

The Symmetrized Joe-Clayton Copula is particularly useful for modeling asymmetric dependencies between the upper and lower tails of the distributions of and . This is crucial in financial applications where extreme co-movements (both positive and negative) can have significant implications.

The parameters for the SJC Copula, and can be modeled as functions of past data to incorporate dynamics in tail dependence, similar to how GARCH parameters are updated:

where and are possibly transformed or standardized residuals, and is a transformation function, often logistic, ensuring the parameters remain within appropriate bounds.

2.5 Evaluation and Measures of Correlation in Copula Models

2.5.1 Evaluation of Copula Models

The evaluation of Copula models focuses on determining the appropriateness of the model in describing the interdependencies among variables. To assess the fitting quality, several indicators are employed: the maximum likelihood value, Akaike Information Criterion (AIC), and Euclidean distance. The process begins by calculating the maximum likelihood function value and the AIC, as depicted in Equation (10). Here, LLF symbolizes the log-likelihood function and \(k\) denotes the number of parameters within the Copula model. Following this, the Euclidean distance between the empirical Copula function and the fitted Copula function is computed to measure the closeness of the fit.

The selection of the optimal Copula model is based on a comparative analysis of these indicators. A lower AIC value indicates a better fit, a smaller Euclidean distance suggests a higher degree of fit, and a lower value of the maximum likelihood function signifies that the model is more suitable. The formula for AIC is given by:

This methodology aids in selecting the most effective Copula model to represent the data accurately, ensuring robust statistical analysis.

2.5.2 Measures of Correlation in Copula Models

In analyzing the correlation within Copula models, two main categories are identified: rank-based correlation measures and tail-dependence measures.

Rank-based Correlation Measures: These measures include Kendall's tau and Spearman's rho. They are designed to evaluate the consistency of trends between two time series. These coefficients are crucial in understanding the monotonic relationships between the variables, which do not rely on the linear attributes of the data.

Tail-dependence Measures: These measures are particularly focused on the extremities of the distributions. They include the upper tail and lower tail dependence coefficients, which quantify the probability that extreme values in one variable coincide with extreme values in another. The expressions for these coefficients are detailed in Table 1. The function , used in these expressions, is a logistic transformation defined as:

This transformation ensures that the tail dependence coefficients remain within the bounded range of (0, 1), making the measures both stable and interpretable.

These correlation measures provide a comprehensive toolset for understanding both the general and extreme dependencies among variables in multivariate data sets, making them indispensable in the application of Copula models to real-world data analysis.

Table 1: Expressions for Upper and Lower Tail Dependence Coefficients in Regular Copula Models

| Copula Model | Upper Tail Dependence Coefficient | Lower Tail Dependence Coefficient |
| --- | --- | --- |
| Normal- Copula | 0 | 0 |
| t - Copula |  |  |
| Bivariate Time-varying SJC Copula: |  |  |