# Time Series Analysis of Regional GDP and Stock Market Volatility Using ARIMA and ARMA-GARCH Models

Mingrui Ma *

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

* Corresponding Author Email: mmrmingruima@163.com

**Abstract.** This study applies ARIMA and ARMA-GARCH models to analyze quarterly GDP data (2007-2019) for Beijing, Tianjin, and Hebei, and daily Shanghai Composite Index prices (2002-2021). The ARIMA(1,0,0)(0,1,0) model demonstrates predictive accuracy for regional GDP (MAE = 8.5%), while the ARMA(0,0,3)-GARCH(1,1) model effectively captures the "peak and fat tail" characteristics of stock returns. Findings provide actionable insights for economic policymakers and financial risk managers, highlighting the models' effectiveness in addressing non-stationarity and volatility clustering in economic and financial time series.

**Keywords:** ARIMA, GARCH, GDP forecasting, stock volatility, time series analysis.

## 1. Introduction

Financial time series prediction is an important method for studying the fluctuations in financial markets. By using historical data generated in financial markets to establish predictive models, it aims to uncover the inherent fluctuation patterns and predict future trends.[1]

The research objects of financial time series prediction are diverse, and the research methods have undergone extensive evolution, from statistical models to intelligent models, and from single models to combined models. Statistical models, represented by the Autoregressive Integrated Moving Average (ARIMA) model and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, have been widely applied. [2] The model descriptions for the volatility of financial assets mainly fall into the following three categories: The descriptive method of historical volatility models (HV for short). The construction of such volatility models is based on historical return data, and the time scale of this historical return data is generally long, usually daily, weekly, or even monthly. Representative models among historical volatility models include the Autoregressive Conditional Heteroskedasticity (ARCH) model, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, and Stochastic Volatility (SV) model. [ 3~5] For example, comparing the accuracy of improved GARCH family models in stock market forecasting, among other aspects [6].

This study presents two complementary analyses of Chinese economic time series: (1) regional GDP forecasting for the Beijing-Tianjin-Hebei area using ARIMA modeling, and (2) volatility analysis of the Shanghai Composite Index through ARMA-GARCH framework. The research addresses critical gaps in understanding both macroeconomic trends and financial market behavior in China's developing economy.

## 2. Case Study 1: Regional GDP Forecasting with ARIMA

### 2.1. ARIMA Model Specification

#### 2.1.1. ARMA (p, q) model

ARMA (p, q) model is a hybrid form of autoregressive model (AR) and moving average model (MA), so it is also called autoregressive moving average hybrid model. The equation is as follows:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (1)$$

Where c is the constant, $\phi_1, \phi_2, \cdots, \phi_p$ are the coefficient of the autoregressive model AR, p is the order of AR, $\theta_1, \theta_2, \cdots, \theta_q$ are the coefficient of the moving average model MA, q is the order of MA, $e_t$ is the white noise sequence with mean 0 and variance $\sigma^2$.

### 2.1.2. ARIMA (p, d, q) model

ARMA model can only be used in stationary time series, but for non-stationary time series, ARMA model is no longer applicable, so a new model, namely ARIMA model, needs to be introduced. ARIMA model mainly solves the problem of non-stationary time series.

$\{Y_t\}$Arima model is an unstable time series, after the D-difference operation, the series gradually tends to be stable, so it is called the autoregressive moving average summation hybrid model. $\{Y_t\}$If the differencing sequence satisfies the ARMA (p, q) model, it is called the ARIMA (p, d, q) process, and the equation of the model is as follows:

$$W_t = \nabla^d Y_t \{Y_t\} \tag{2}$$

Let, Wt =Yt-Yt-1:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{3}$$

To denote by the sequence symbol Yt:

$$\begin{aligned} Y_t = \quad & (1+\phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + (\phi_3 - \phi_2)Y_{t-3} + \cdots \\ & +(\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \end{aligned} \tag{4}$$

Call it the difference equation form of the model.

The ARMA model assumes that the variance is constant. However, if the heteroscedasticity test of the residual series after the optimal ARMA model is fitted shows that the series has heteroscedasticity, the further GARCH model is widely used in the analysis of heteroscedasticity. ARIMA (p, d, q) -GARCH (u, v) model is defined as the ARIMA (P, D, Q) -GARCH (U, V) model

$$\begin{gathered} \nabla^d X_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i (\nabla^d X_{t-i}) + u_t + \sum_{j=1}^{q} \theta_i u_{t-j} \\ \nabla X_t = X_t - X_{t-1} \\ \{ \quad u_t = \sigma_t \varepsilon_t \\ \varepsilon_t \sim N(0,1) \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{\mu} \alpha_i u_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2 \\ \sum_{i=1}^{\mu} \alpha_i + \sum_{j=1}^{v} \beta_j < 1 \end{gathered} \tag{5}$$

Where, $\mu$ and v are orders of GARCH model and ARCH model, respectively; $\mu\varepsilon_t$Is the error term; ${\sigma_t}^2$ is the volatility of the time series; $\alpha_i$ and $\beta_j$ are the parameters to be determined in the nonnegative constraint, respectively. The BIC criterion is used to select the order. The optimal ARIMA model based on BIC criterion was determined as ARIMA (p, d, q). The residual autocorrelation test was performed, and the residual sequence diagram was observed to test whether the residual of ARIMA model was white noise. If there is significant autocorrelation in the residual sequence. The LM test method was selected to test the ARCH effect of the error series. The null hypothesis of the LM test is that the autocorrelation coefficient is 0, so there is no ARCH effect. The p-value is the probability of accepting the null hypothesis.

## 2.2. Data Preparation

Database: quarterly data by province; Indicators: regional GDP _ cumulative value (100 million yuan); Time: 2007-2019. The quarterly data from 2018 to 2019 are revised according to the data of the Fourth National Economic Census. Data source: National Bureau of Statistics

Test whether each value is missing, TRUE is missing; To determine whether the sample is complete, TRUE is complete. All values have been verified to be non-missing and complete

### 2.3. Statistical characteristics of the data

Stick to the data set and calculate the mean and variance of Beijing, Tianjin and Hebei respectively are 12098.48, 7512.19 and 15598.23. The results of mean and variance were the highest in Hebei and the lowest in Tianjin.

Then, the sticky data set was disentangled.

Then draw a boxplot. The boxplot depicts that the middle thick line of the data is the average, the upper and lower borders are the upper and lower quartiles, and the top and bottom are the maximum and minimum values; There are two outliers for Beijing.

Since boxplots cannot show the data distribution, violin plots are drawn. It is a combination of boxplot and kernel density plot. The figure shows that Beijing presents a thin pear shape, Tianjin is shorter, and Hebei is more uniform.

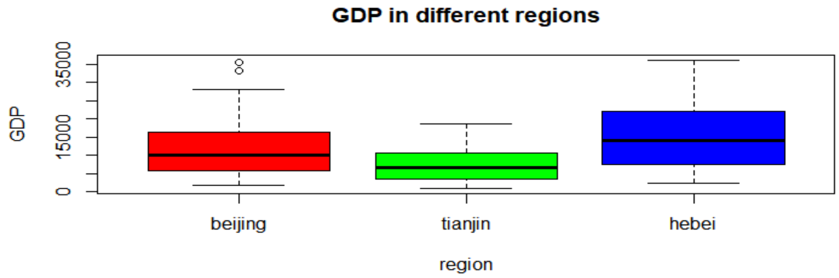


**Fig 1.** Boxplot of GDP.

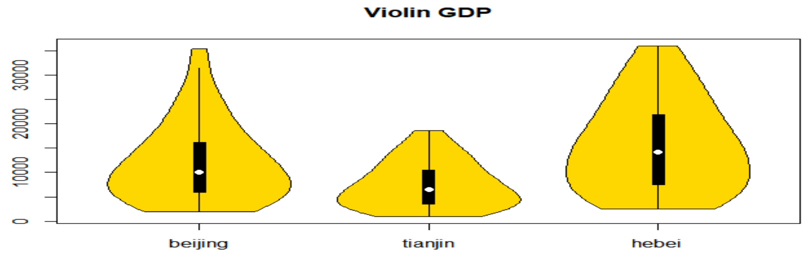


**Fig 2.** Violin Plot of GDP.

### 2.4. Visualization of data

Scatter plot matrix: the main diagonal is the kernel density curve of Beijing, Tianjin and Hebei, and the rest contains linear and smooth fitting curves.

Draw a heat map: smaller values in blue and larger values in purple. The values are all from small to large, but Tianjin has a decreasing trend near 2019.

Draw the Facebook graph: from face 1 to 52, all the numbers are getting bigger, indicating that the GDP has basically changed from small to large

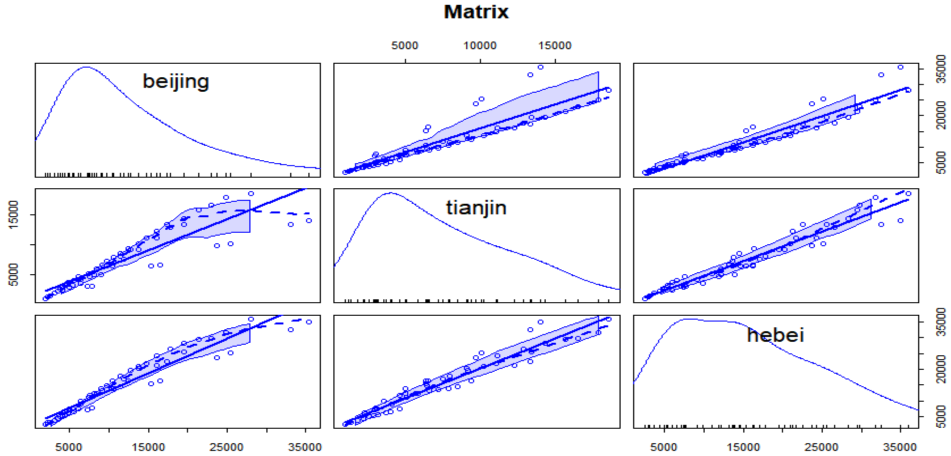


**Fig 3.** Scatter Plot Matrix.

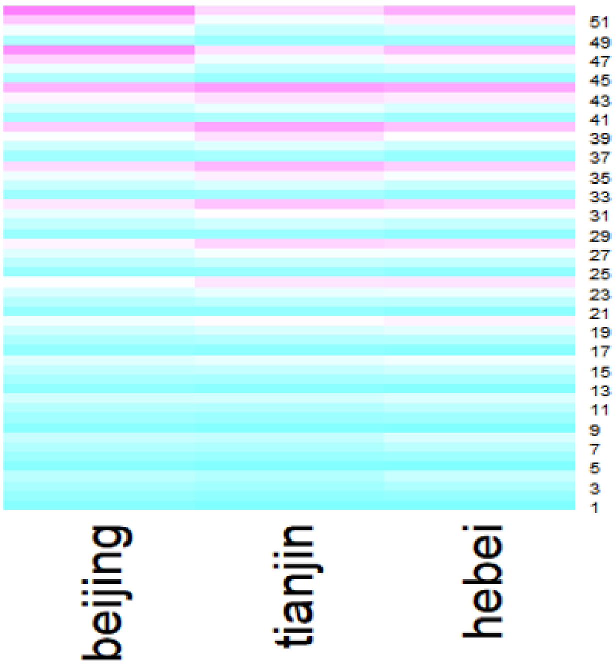


**Fig 4.** Heat map.

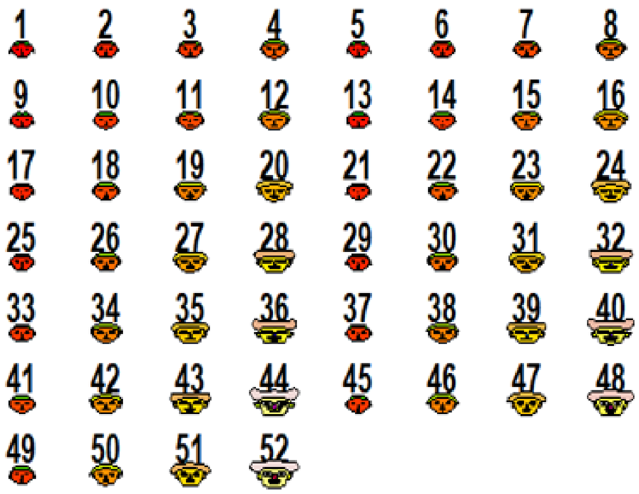

**Fig 5.** Facebook graph.

### 2.5. Data were selected and transformed into time series

Data were selected and transformed into time series. Firstly, the outliers in Beijing are removed, and the "Hebei" region is selected to transform the data into time series data.

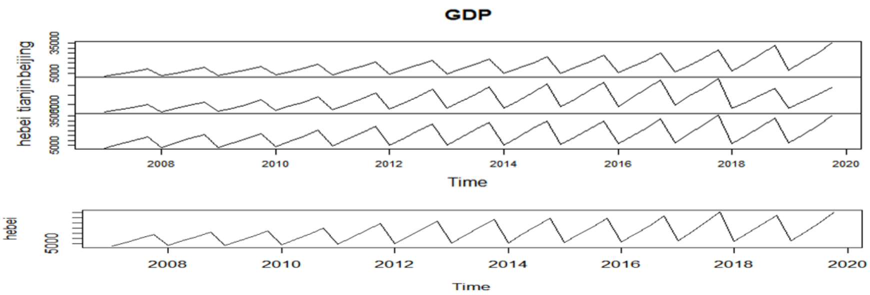


**Fig 6.** Time series.

Smoothed and predicted with ARIMA. Calculate the minimum and maximum values are 2450.1 and 35964.0.

Simple moving average for smoothing. 3, 7 and 12 items of simple moving average were applied to the original sequence graph. The larger k was, the smoother the graph was.

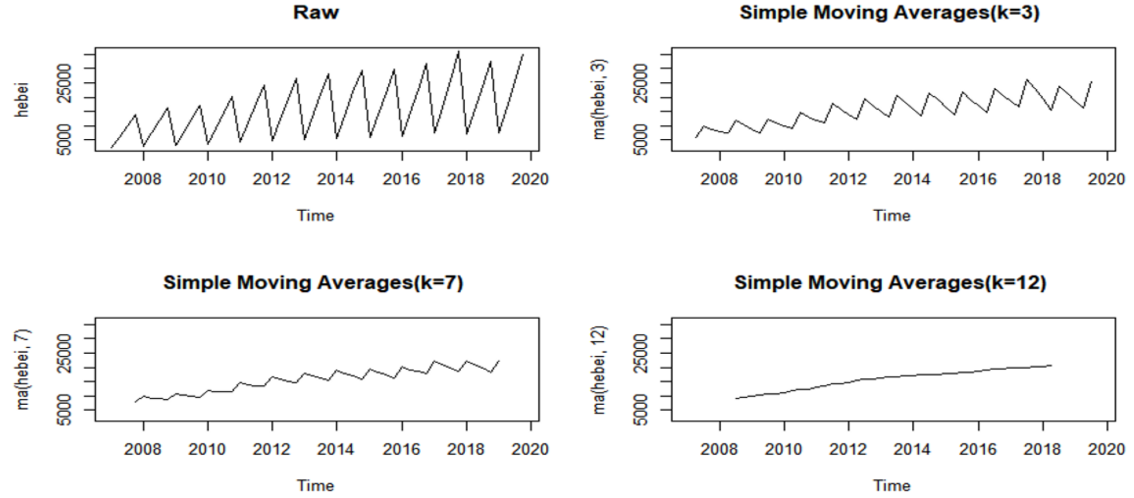


**Fig 7.** Simple moving average for smoothing.

## 2.6. Forecast with an ARIMA model

Test for stationarity. Observation of the timing diagram shows that it is not stationary, and automatic difference is performed. The time series requires first-order differencing to achieve stationarity.

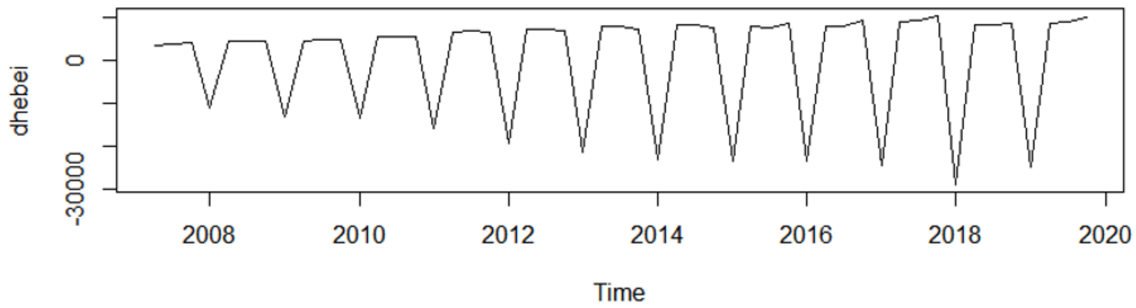


**Fig 8.** First-order differencing to achieve stationarity.

ADF test: $p < 0.05$, the null hypothesis is rejected and considered to be stationary.

```
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0  -9.66  0.0100
[2,]   1  -9.20  0.0100
[3,]   2 -63.63  0.0100
[4,]   3  -2.46  0.0166

Type 2: with drift no trend
     lag    ADF p.value
[1,]   0  -9.59  0.0100
[2,]   1  -9.15  0.0100
[3,]   2 -82.04  0.0100
[4,]   3  -3.43  0.0165

Type 3: with drift and trend
     lag    ADF p.value
[1,]   0  -9.49  0.0100
[2,]   1  -9.05  0.0100
[3,]   2 -83.56  0.0100
[4,]   3  -3.54  0.0468
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

**Fig 9.** ADF test.

## 2.7. Model order fitting and model diagnosis

Method 1: Draw the auto correlogram and partial auto correlogram: the auto correlogram is tailed, and the partial auto correlogram is censored at order 3 or 4. Select (3,1,0) or (4,1,0).

**Table 1.** Model diagnosis.

| ACF | PACD | Choice model |
|---|---|---|
| Trailing tail | P-order censoring | ARMA (p, 0) |
| q order censoring | Tailing | ARMA (0, q) |
| Trailing tail | Tailing | ARMA (p, q) |

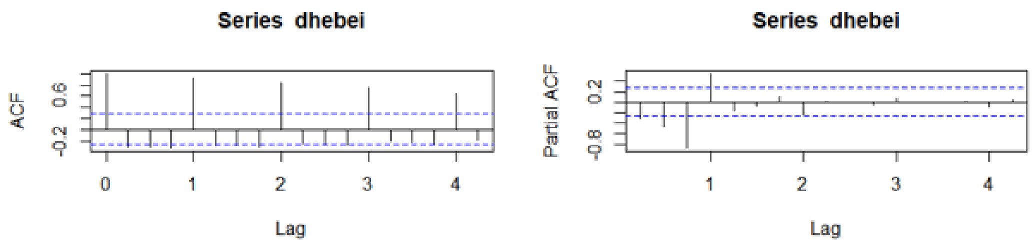


**Fig 10.** Autocorrelogram and partial autocorrelogram.

Method 2: Automatically identify the form of: SARIMA (p, d, q) (P, D, Q) S, where: p, d, q are the parameters in the non-seasonal ARIMA model as described above; P is the order of seasonal autoregression; D is the number of seasonal differences; Q is the order of the seasonal moving average; S is the time span over which the seasonal pattern repeats.

The coefficient value divided by the standard deviation is greater than 1.96 and is significant.

**Table 2.** Automatically identify.

| Series: hebei |
|---|
| ARIMA (1,0)(0,1,0)[4] with drift |

**Table 3.** Coefficients.

| Coefficients | | |
|---|---|---|
| | ar1 | drift |
| | 0.4902 | 274.1864 |
| s.e. | 0.1250 | 81.8492 |
| sigma^2 = 1450313 | log likelihood = -407.72 | |
| AIC = 821.44 | AICc = 821.99 | BIC = 827.05 |

MASE (0.61) is relatively the most important index. When the mean absolute error (MAE) is used as the relative index of model prediction accuracy, if MASE>1, it indicates that the out-of-sample prediction is worse than the naive prediction based on the sample itself. Root means square error (RMSE); MAE (mean absolute error); MPE mean relative error. If the MAPE is 8.503, the average deviation of the predicted result from the real result is 8.503 %.

**Table 4.** Model Forecasting Performance Metrics.

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | 7.504064 | 1132.682 | 851.0219 | -3.403766 | 8.503687 | 0.6167809 | 0.07691341 |

QQ plot for model test: because the residual is assumed to be in line with normal distribution, QQ plot is basically consistent.

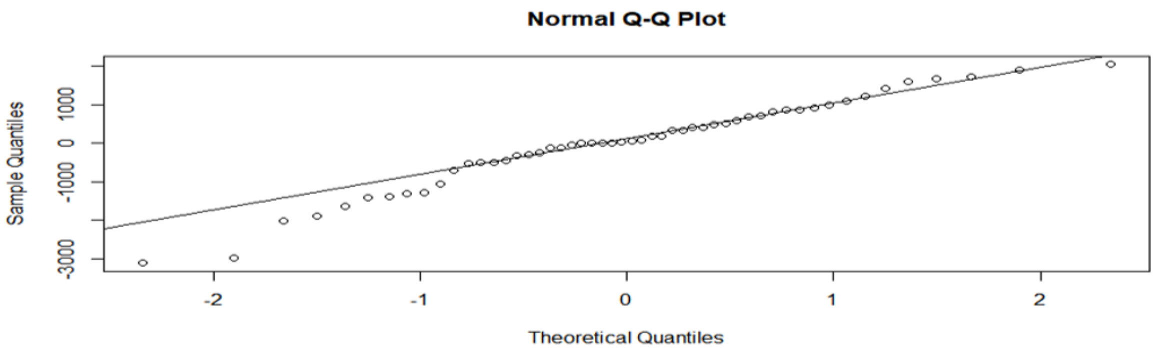


**Fig 11.** QQ plot for model test.

According to Box-Ljung test, if the p-value>0.05, the null hypothesis H0 is accepted, the result is not significant and the sequence is not correlated, and the sequence is considered as white noise sequence.

### 2.8. Conclusions

All the residuals in the figure are basically within the critical value of significance, which is like white noise. The Ljung-Box test also shows that there is no autocorrelation among the residuals.

**Table 5.** Ljung-Box test.

| | | |
|---|---|---|
| Ljung-Box test | | |
| data: Residuals from ARIMA (1,0,0) (0,1,0) [4] | | |
| Q* = 8.1901 | df = 7 | p-value = 0.3161 |
| Model df: 1. | Total lags used: 8 | |

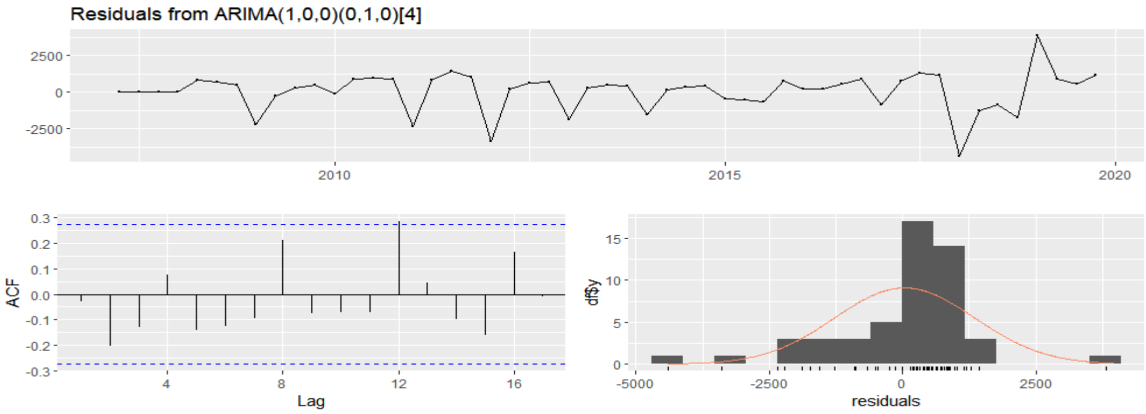


**Fig 12.** Residuals from ARIMA.

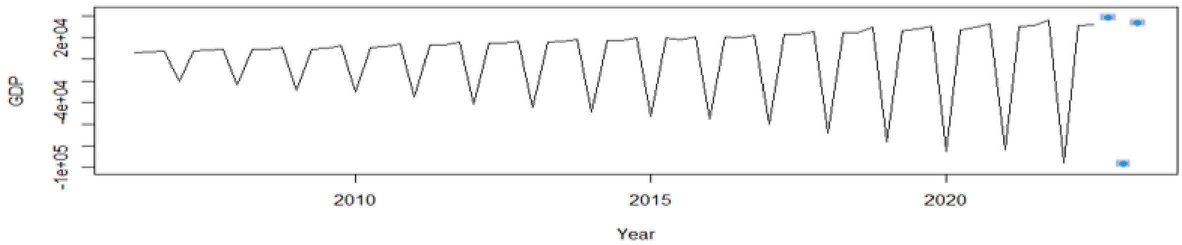


**Fig 13.** Forecast of GDP.

## 3. Case Study 2: Stock Market Volatility with ARMA-GARCH

### 3.1. Sample selection and basic statistical analysis

### 3.1.1. Basic statistical analysis

Determine the file position as the correct position, then import the closing price data of SSE index, observe the sequence trend. Log the closing price of SSE index.

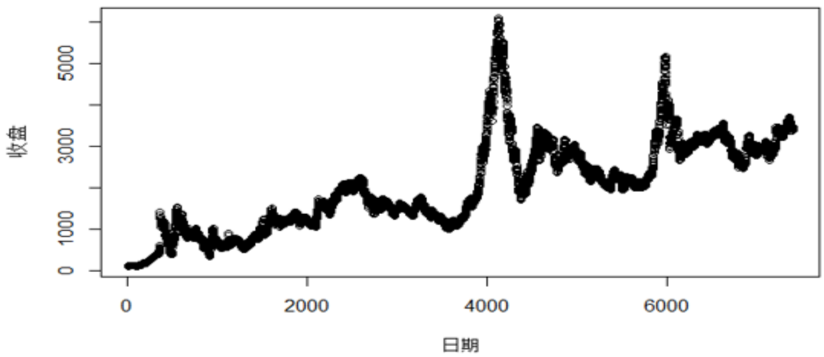


**Fig 14.** Observe the sequence trends.

### 3.1.2. Preprocessing of the data

After time processing, change to import stockdata. Observe the time series chart, load the time series and forecast package, assign the time according to the day/month/year, assign the closing price to the stock price, and become a time series.

## 3.2. ARMA model

### 3.2.1. Select local data because the amount of data is beyond the system's capacity (5000)

Due to the capacity limitation of data, the stocks from February 1, 2002 to January 22, 2021 are selected, and the abscissa is "time" and the ordinate is "index". The title is "Shanghai Composite Index" time series chart of Shanghai Composite Index.


**Fig 15.** Time series chart of Shanghai Composite Index.

### 3.2.2. Original series unit root test

ADF test, also called unit root test. It is used to test whether there is a unit root in the series. If there is a unit root, it is a non-stationary time series.

p=0.3628: P value greater than 0.05, the null hypothesis is accepted, indicating that the original series is non-stationary.

**Table 6.** Augmented Dickey-Fuller Test.

| Augmented Dickey-Fuller Test |
|---|
| data: selectdata |
| Dickey-Fuller = -2.5089, Lag order = 16, p-value = 0.3628 |
| alternative hypothesis: stationary |

### 3.2.3. Log the closing price index and multiply by 100

Log the closing price index and multiply by 100. Many time series data need to be logarithmically processed. The main purpose of this is to remove the exponential trend and turn the multiplicative relationship into an additive one. In general, prices (such as stock prices) are lognormal because they are greater than zero. Since real data is often heteroskedastic in nature, legitimization eliminates heteroscedasticity.

First, the selected stock price data is log-processed, and then the graph is drawn.

Then, the first order difference of the logarithm of the Shanghai Composite index is multiplied by 100 to convert the data. It is named "index(rlogdiff)". coredata () gets the matrix part of the object. Index () gets the original vector of the index.

### 3.2.4. Statistical features

Draw the distribution chart and QQ chart of Shanghai Composite index in one row and three columns.

The first picture: The role of par (oma) is to adjust the distance of the outer frame line from the drawing boundary. The function of hist is to draw a histogram. probability is the logical value and the function of freq parameter is the opposite. TRUE stands for frequency, FALSE stands for frequency; col is the fill color; xlab is the label for the X-axis, ylab is the label for the Y-axis; xlim and ylim are used to set the range of the graph's X-axis to the Y-axis. The function of density is to find the data density; lwd specifies the line width, the default value is lwd=1; rug indicates the frequency with which an element appears on the axis. Once it appears, there will be a vertical bar. The denser the rug, the higher the density.

The second is a Q-Q chart:

Quantile-Quantile Plots. It is a probability plot method that compares two probability distributions by comparing their quantiles. By mapping the values of the same quantile points of two probability distributions to the x and y axes, if the two distributions are relatively similar, then the points on the graph roughly fall on the y=x line; If the two distributions are linearly related, then the points are roughly distributed on a straight line, but not necessarily on y=x.

The distribution of log returns has obvious characteristics of negative skewness, fat tail and sharp peak, and the Q-Q plot shows that it does not follow the normal distribution.

The third chart:

Abline adds vertical lines, lty=2 for solid lines. If the expectation is 0, the variance is fixed. This means that the image basically fluctuates around y = 0, with no increasing or decreasing trend. But Figure 3 is not, indicating that the variance is not homogeneous.

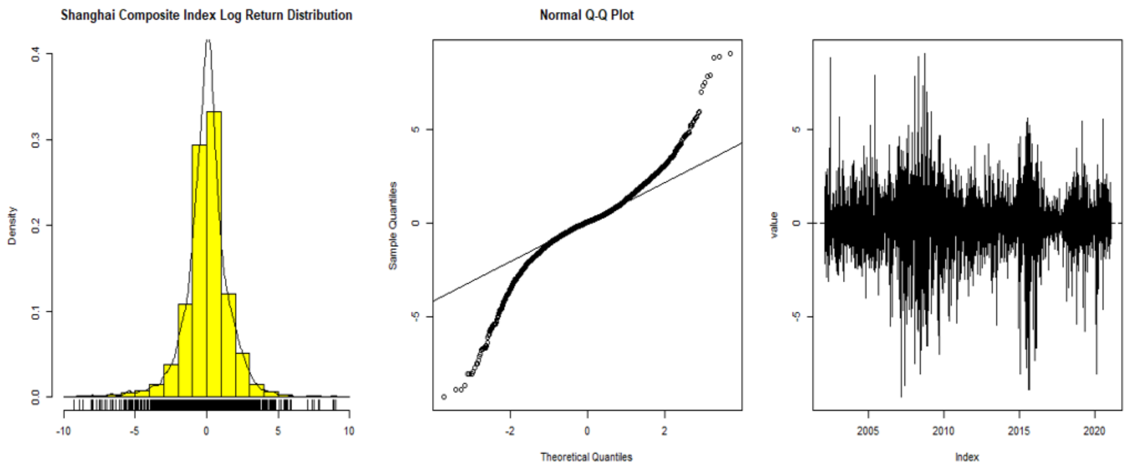


**Fig 16.** Distribution chart and Quantile-Quantile Plots.

Test for non-normal distribution: the larger the value, the more it is not normal distribution; The smaller the p-value, the less normal the distribution

Shapiro-Wilk test: It is used to test whether the data are normally distributed. It is the residual of the regression curve. Results: W = 0.93393, p = $2.2*10^{-16}$< 0.05, which rejected the null hypothesis, indicating that the data did not conform to the normal distribution.

**Table 7.** Shapiro-Wilk normality test.

| Shapiro-Wilk normality test |
| --- |
| data: rlogdiffdata<br>W = 0.93393, p-value < 2.2e-16 |

The ADF test is performed on the differenced data: p is small and the null hypothesis is rejected.

**Table 8.** Augmented Dickey-Fuller Test.

| Augmented Dickey-Fuller Test |
| --- |
| data: rlogdiffdata<br>Dickey-Fuller = -15.189, Lag order = 16, p-value = 0.01<br>alternative hypothesis: stationary |

### 3.2.5. ARIMA model fit

ARIMA model was used for fitting, and LM test was performed on the residuals.
Based on AIC minimum criterion, the best fitting model is automatically found:

**Table 9.** Model fit ARIMA (0,0,3) with zero mean

| ma1 | ma2 | ma3 |
|---|---|---|
| -0.0516 | -0.0151 | 0.0878 |

Based on the minimum BIC criterion, the best fitting model was automatically found, and the results were consistent.

### 3.2.6. ARCH effect test

Test the ARCH effect of residuals: all are below the dotted line, indicating that the test has not been passed, indicating that the ARCH effect of residuals is significant:

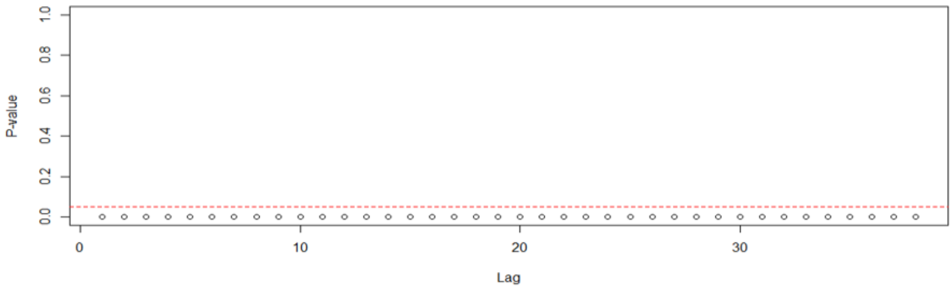


**Fig 17.** ARCH effect test.

The LM test showed that the variance of the series was significantly non-homogenous, and the squared residuals were significantly correlated, so the ARCH models of order 1 to order 5 were significantly valid.

## 3.3. Arima-garch model

To fit the eGARCH model:

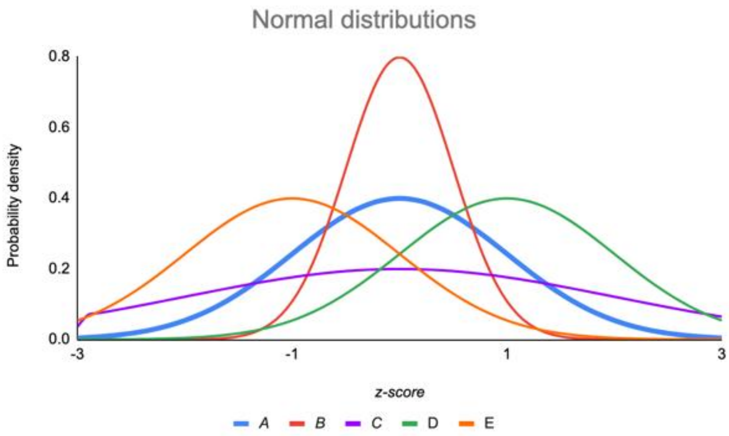


**Fig 18.** Arima-garch model

Normality test of residuals: T distribution

**Table 10.** Normality test of residuals: T distribution.

| Curve of curve | Position or shape (relative to standard normal distribution) |
|---|---|
| A (M= 0, SD= 1) | Standard normal distribution |
| B (M= 0, SD= 0.5) | Squeeze because SD< 1 |
| C (M= 0, SD= 2) | Stretch because SD> 1 |
| D (M= 1, SD= 1) | Move right because M > 0 |
| E (M=-1, SD= 1) | Shift left because M< 0 |

Shapiro. test: A larger value means it is not normal, and a smaller P means it is not normal.

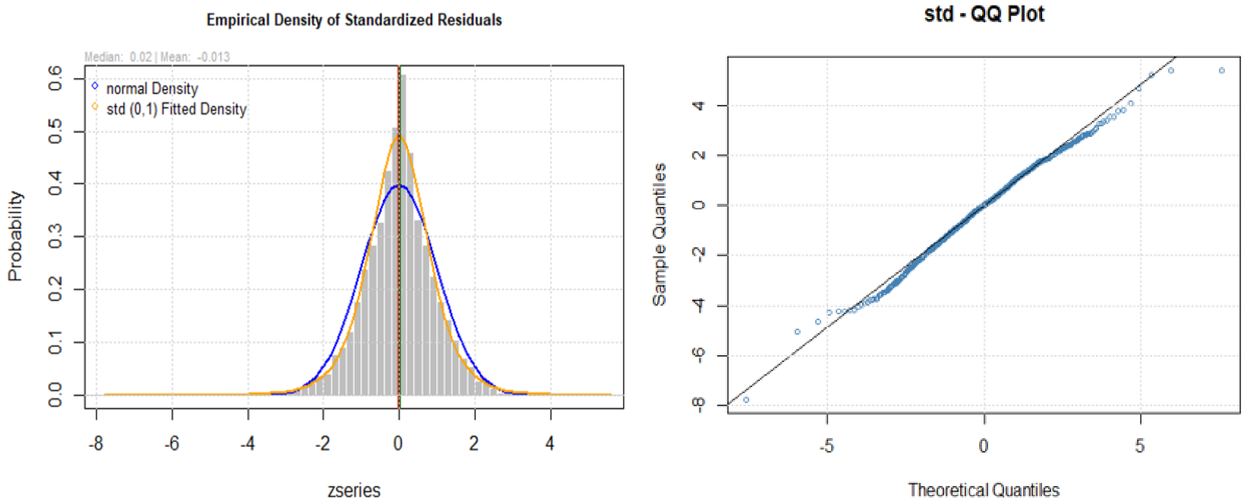


**Fig 19.** Normality test of residuals: T distribution.

Correlation test of residuals:

The P value was observed to show whether the coefficient was significant. And the residual effect was observed.

Shapiro. test: A p value greater than 0.05 was defined as conforming to normal distribution.

Hypothesis: subjects with a certain sample size n always conform to normal distribution. The samples with a sample size of n were arranged in order of size, and then the value of the statistic W was calculated according to the formula. When the value was closer to 1 and the significance level was greater than 0.05, the null hypothesis could not be rejected.

**Table 11.** Shapiro-Wilk normality test.

| Shapiro-Wilk normality test |
|---|
| data: coredata (residuals(myfit)) |
| W = 0.93393, p-value < 2.2e-16 |

**Table 12.** Adjusted Pearson Goodness-of-Fit Test

| group | statistic | p-value(g-1) |
|---|---|---|
| 20 | 60.30 | 3.474e-06 |
| 30 | 67.59 | 6.429e-05 |
| 40 | 80.90 | 9.326e-05 |
| 50 | 89.16 | 3.980e-04 |

Asymmetric model ARIMA-EGARCH model, which is matched with the log return of Shanghai Composite Index, shows that the Shanghai Composite index shows the characteristics of sharp peak, fat tail and asymmetry from February 1, 2002 to January 1, 2021.22.

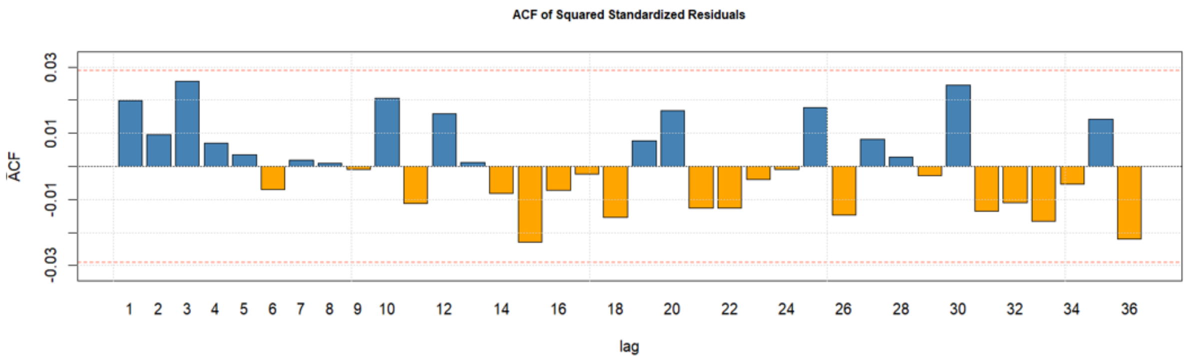


**Fig 20.** ARIMA-EGARCH model.

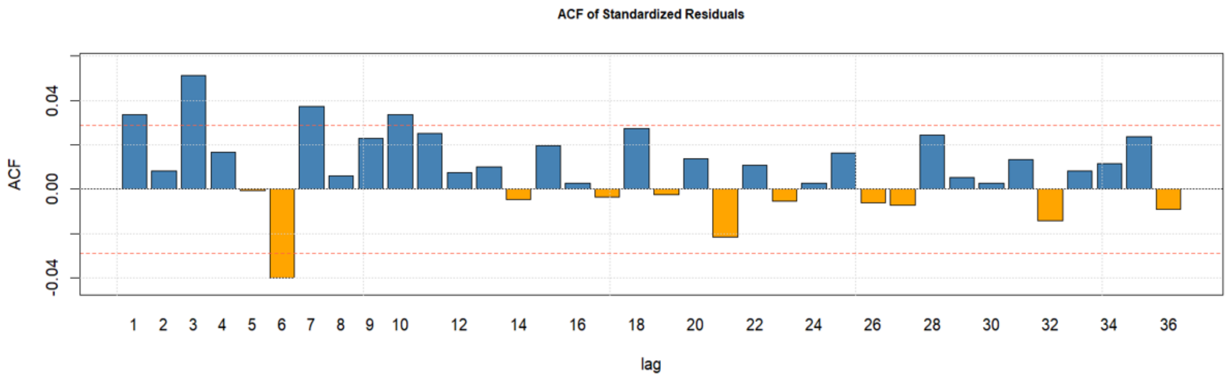


**Fig 21.** ARIMA-EGARCH model.

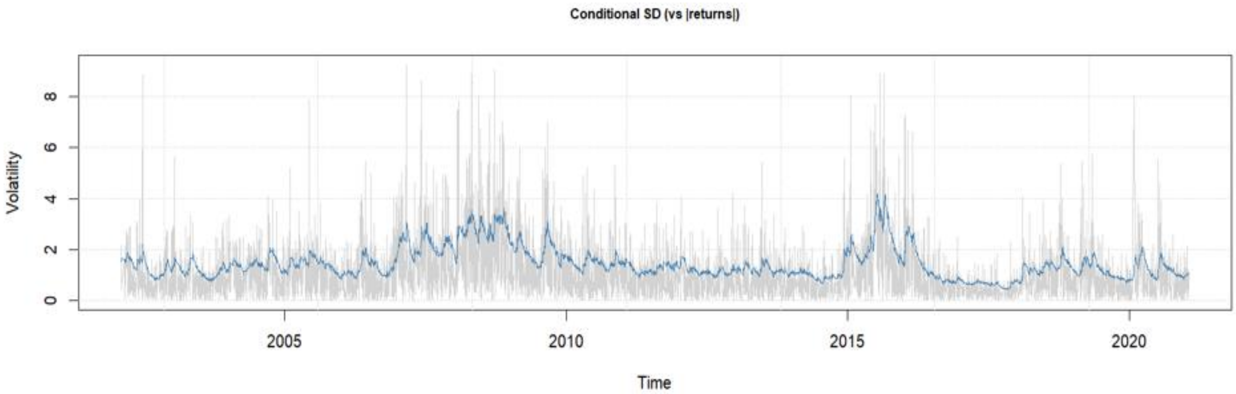


**Fig 22.** ARIMA-EGARCH model.

## References

[1] He Y, Li J M, Ruan S M, et al. A Hybrid Model for Financial Time Series Forecasting-Integration of EWT, ARIMA With the Improved ABC Optimized ELM [JJ. IEEE Access,2020, (8).

[2] Cibicong, Zhang Pinyi. Financial Time Series Prediction Based on ARIMA-LSTM Model [J]. Statistics and Decision, 2022, 38(11): 145-149. DOI: 10.13546/j.cnki.tjyjc.2022.11.029.

[3] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. Econometrica, 50(3), 987-1007.

[4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(5), 307-327.

[5] Taylor, S. J. (1986). Modeling financial time series. Chichester: John Wiley and Sons, pp. 35-40.

[6] Kim H Y, Won C H. Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM With Multiple GARCH-type Models[J]. Expert Systems with Applications,2018, (103).