

Analysis and Predication of Stock Price of China Eastern Airlines Based on ARIMA-GARCH Model

Yang Yang^{1,*}, and Huashan Tan²

¹School of Economy and Management, Chongqing Normal University, Chongqing, China

²School of Computer and Information Science, Chongqing Normal University, Chongqing, China

Keywords: Time series, ARMA-GARCH model, EViews software

Abstract: This paper takes the daily data of the historical closing price of China Eastern Airlines (600115) as the sample series, and uses some models in time series to analyze and forecast its price. First, the ARMA model is established by analyzing the sample series. Secondly, the obvious conditional heteroscedasticity in the residual error of the model is found after the ARCH testing, and a more reasonable ARIMA-GARCH model which can eliminate the conditional heteroscedasticity is set up. Finally, this paper forecast the closing price in the next three days. The result of empirical analysis shows that the model is effective and accurate. This has a certain guiding significance for the decision-making of investors.

1. Introduction

It is difficult to predict the stock price accurately because of many factors such as market supply and demand. At present, there are more qualitative researches than quantitative ones on the prediction of stock price. Time series analysis is a statistical method which uses the past historical data to predict the future development trend according to the continuous law of the development of objective things. Because the price of stock is sensitive to time, it is a feasible method to establish time series model to predict stock price [1].

There are two hypotheses to establish the time series model to predict the stock price. One is that the stock price will not change suddenly, but will change in a relatively small range. The other is that the past and current prices may predict the trend of the price in future. Only time factor is considered here, and other factors affecting stock price fluctuation are ignored. This determines that in general, time series analysis is significant for short-term prediction and has great limitations for long-term prediction.

2. ARIMA-GARCH Model

2.1. ARIMA (p, d, q) Model

ARIMA (p, d, q) model is suitable for non-stationary time series. In theory, any non-stationary time series can be fitted with ARMA model if it is stationary after difference. ARMA model, an autoregressive moving average model, is a kind of time series model. It was proposed by box and Jenkins in 1970s. The model is expressed as follows:

$$X_t = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (1)$$

In this equation, $\{\varepsilon_t\}$ is white noise sequence and $\phi_p \neq 0, \theta_q \neq 0$. ARMA model includes both autoregressive AR term and moving average MA term. In the selection of the model, the characteristics of ACF (Autocorrelation coefficient function) and PACF (Partial autocorrelation coefficient function) of stationary series samples should be examined [2].

2.2. ARCH effect and GARCH model

After eliminating the influence of the deterministic non-stationary factors, the volatility of the residual series is stable in most periods, but it will continue to be large in some periods and small in some periods, showing the cluster effect. That is to say, the sequence of residual is error uncorrelated, but its square or absolute value sequence is related. In 1982, Engle proposed an important arch effect test method: Lagrange multiplier test. Arch (q), an autoregressive model, was used to fit the square sequence of residuals. The model is as follows:

$$\varepsilon_t^2 = a_0 + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 + e_t \quad (2)$$

In this equation, $\{\varepsilon_t\}$ is white noise sequence and $a_0 > 0, \lambda_i \geq 0$. If the equation holds obviously, the square sequence of residual error has autocorrelation, otherwise, the assumption of homogeneity of variance cannot be rejected.

The assumptions of Lagrange multiplier test are as follows:

$$H_0: \lambda_1 = \lambda_2 = \dots = \lambda_q = 0 \leftrightarrow H_1: \lambda_1, \lambda_2, \dots, \lambda_q \text{ are not all } 0 \quad (3)$$

$$SST = \sum_{t=q+1}^T \varepsilon_t^2$$

Note that the sum of the squares of the total errors is $SST = \sum_{t=q+1}^T \varepsilon_t^2$, and the Sum of squared

residuals is $SSE = \sum_{t=q+1}^T e_t^2$, and the sum of regression squares is $SSR = SST - SSE$. Then we can

construct test statistics as $LM(q) = \frac{SSR/q}{SSE/(T-2q-1)}$ [3-5].

The LM statistic approximately obeys the chi square distribution with degree of freedom $q-1$. When the p value of the LM statistic is less than the significance level α , the original hypothesis is rejected and the variance of the sequence is considered to be non-homogeneous.

The distribution of the a-statistic obeys the chi square of the degree of freedom B approximately. In practice, it is found that the heteroscedasticity function of the residual sequence sometimes has long-term autocorrelation. If we continue to use arch model to fit the heteroscedasticity function, it will produce a very high order of moving average. The difficulty of parameter estimation will increase because of the high order of moving average, and the fitting effect of arch model will become worse[6-8].

Then, in order to correct this problem, T. Bollerslev proposed the GARCH (generalized autoregressive conditional heteroscedasticity) model in 1986, which is abbreviated as GARCH (P, q). The GARCH model is formed by adding the p-order autoregression of heteroscedasticity function on the basis of arch model, which is more obvious in the fitting of heteroscedasticity function with long-term autocorrelation. The advantage of GARCH model is also reflected in the study of volatility [9].

The specific structure of GARCH model is as follows:

$$a_t = \sigma_t \varepsilon_t, \sigma_t^2 = a_0 + \sum_{j=1}^q \lambda_j a_{t-j}^2 + \sum_{i=1}^p \eta_i \sigma_{t-i}^2 \quad (4)$$

The model has two constraints: $a_0 > 0, \eta_i \geq 0, \lambda_j \geq 0$ and $\sum_{i=1}^p \eta_i + \sum_{j=1}^q \lambda_j < 1$.

3. The stock price model and forecast of China Eastern Airlines

3.1. Preliminary analysis and pre-processing of data

This paper selects 431 days closing price of China Eastern Airlines (600115) from November 3, 2017 to August 8, 2019 as sample data for analysis and prediction, and the closing price of the last three days is test data. First, we look at the image distribution of the data, as shown in Figure 1.

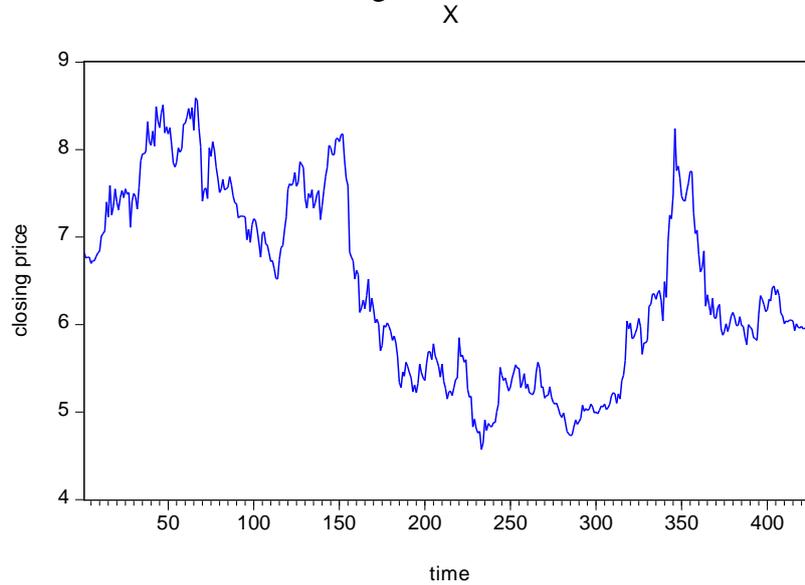


Figure 1 (a) Closing price sequence chart of China Eastern Airlines

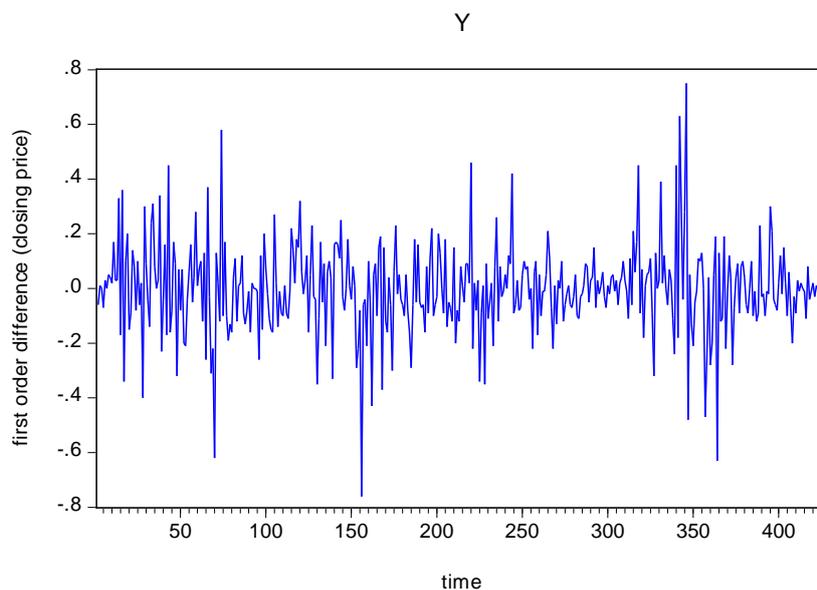


Figure 1 (b) First order difference timing chart of closing price of China Eastern Airlines

It can be seen from the trend chart that the original sequence and the first-order difference sequence may be stationary, so the ADF unit root test method. is used to test if they are stationary.

3.2. Data stationary test

The unit root test is carried out for the original series and the first-order difference series of closing price of China Eastern Airlines to test their stationarity. There are three types of unit root tests in ADF test: the p -order autoregressive process with no constant and no trend; the p -order autoregressive process with constant and no trend; and the p -order autoregressive process with both constant and linear trend. The main test indexes are shown in Table 1 and Table 2.

It can be seen from Table 1 that the P value is far greater than the significance level α

($\alpha = 0.01, 0.05, 0.1$). The original hypothesis cannot be rejected, so the closing price time series is not stationary. It can be seen from Table 2 that the P value is less than significance level α . The original hypothesis can be rejected, so the closing price time series is stationary. We can use the ARIMA model to fit the closing price series.

Table 1 ADF unit root test of the original series

| ADF unit root test | None | Intercept | Trend and intercept |
|--------------------|-----------|-----------|---------------------|
| T Statistic | -0.669241 | -1.467688 | -2.047736 |
| P value | 0.4269 | 0.5492 | 0.5729 |

Table 2 ADF unit root test of the first-order difference series

| ADF unit root test | None | Intercept | Trend and intercept |
|--------------------|-----------|-----------|---------------------|
| T Statistic | -22.15083 | -22.13431 | -22.11502 |
| P value | 0.0000 | 0.0000 | 0.0000 |

3.3. Model recognition and modelling

We get the autocorrelation coefficient graph (ACF) and partial autocorrelation coefficient graph (PACF) after the first order difference series of closing price, as shown in Figure 2.

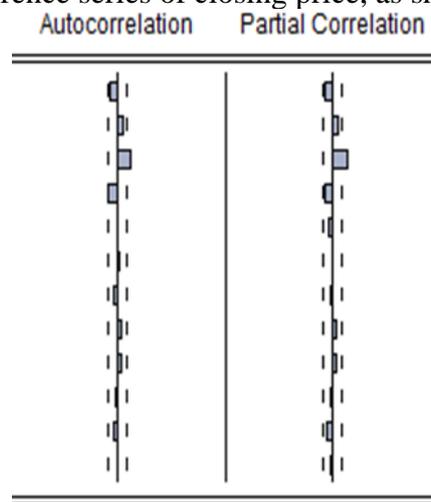


Figure 2 Autocorrelation coefficient and partial autocorrelation coefficient

It can be seen from Figure 2 that the ACF and PACF of the difference series have certain tailing, and the autocorrelation coefficient and partial autocorrelation coefficient are significant when they are in the third order, so we try to use AR (3) and MA (3) models to fit the series.

It is found that the coefficients of AR (1), AR (2), MA (1) and MA (2) are not significant, so the AR (3) and MA (3) models with missing coefficients are adopted. Since the AIC values of the two models are not significantly different, and there is no significant difference in model order determination and selection, the AR (3) model with missing coefficients is considered to fit the difference sequence. Assuming the closing price series is X_t , the fitting model is as follows:

$$Y_t = 0.132254 Y_{t-3} + \varepsilon_t \quad (5)$$

In the equation, $Y_t = X_t - X_{t-1}$. The AR (3) model with missing coefficient is shown in Figure 3.

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| AR(3) | 0.132254 | 0.048421 | 2.731355 | 0.0066 |

Figure 3 The AR (3) model with missing coefficient

After getting the mean equation, we need to check whether the residual sequence $\{\varepsilon_t\}$ is a white

noise sequence. If the residual sequence is not a white noise sequence, the useful information of the sequence has not been fully extracted by the model, and the validity of the model is not strong. The Q statistic test of box Ljung is conducted for the residual sequence to check whether it is irrelevant. The test results are shown in Table 3.

Table 3 White noise test of residual sequence of AR (3) model

| Lag period | Q statistic | P value |
|---------------|-------------|---------|
| 6 lag phases | 6.0650 | 0.300 |
| 12 lag phases | 8.6572 | 0.654 |

Because the p value is far greater than the significance level α ($\alpha = 0.01, 0.05, 0.1$), the original hypothesis of pure random can not be rejected, and the residual sequence is considered as white noise data without autocorrelation.

Further, we draw the diagram of the residual sequence and the residual square sequence, as shown in Figure 4.

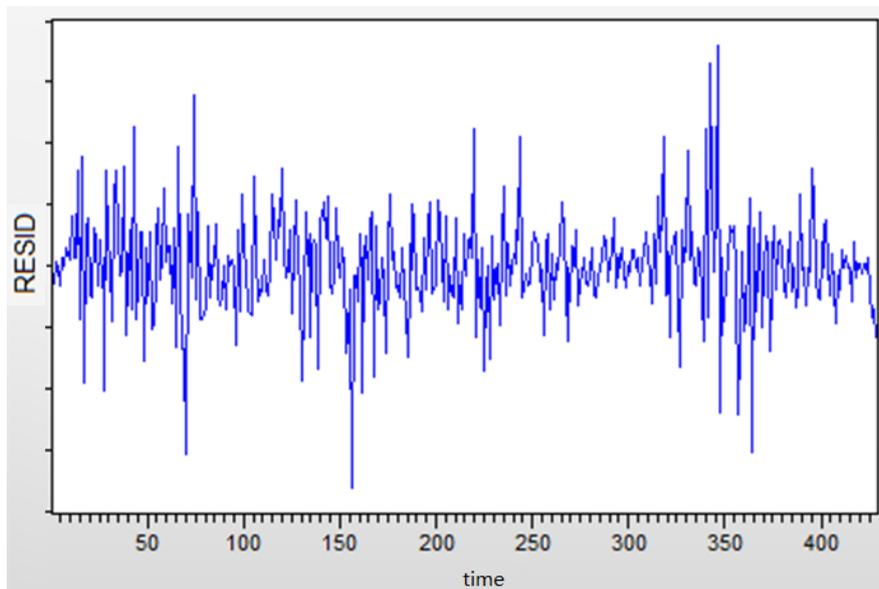


Figure 4 (a) Diagram of the residual sequence

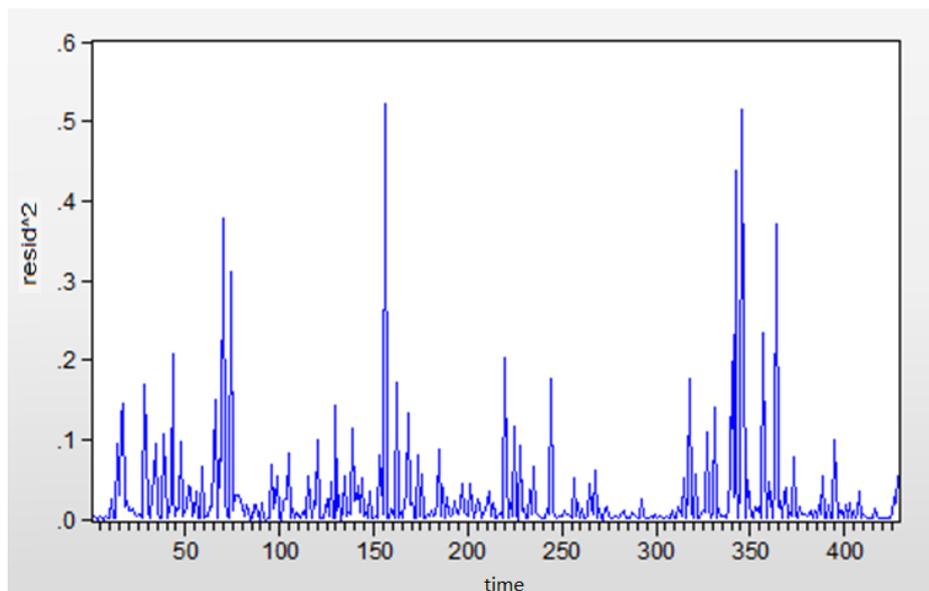


Figure 4 (b) Diagram of the residual square sequence

From Figure 4, it can be seen that the residual sequence has obvious volatility clustering effect, that is to say, the volatility in some periods is continuously large, while the volatility in some periods is continuously small. so it is considered that there may be conditional heteroscedasticity on

the residual sequence change.

3.4. The arch effect test

The arch effect is tested by Lagrange multiplier, and the test results are shown in Figure 5.

| F-statistic | 3.913193 | Prob. F(1,421) | 0.0486 |
|---------------|----------|---------------------|--------|
| Obs*R-squared | 3.895574 | Prob. Chi-Square(1) | 0.0484 |

Figure 5 Arch effect test of the residual sequence

The p value of LM statistic is less than the significance level $\alpha = 0.05$, so the variance of the sequence is not homogeneous, and there is conditional heteroscedasticity on the residual sequence change.

3.5. Establishment of comprehensive model

Because of the conditional heteroscedasticity of the model, GARCH model is considered to fit the residual sequence. The model AR (3) with missing coefficients and GARCH (1, 1) are used for joint estimation. The model is shown in Figure 6.

| Variable | Coefficient | Std. Error | z-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|-----------|
| AR(3) | 0.090498 | 0.051051 | 1.772711 | 0.0763 |
| Variance Equation | | | | |
| C | 0.001357 | 0.000474 | 2.864720 | 0.0042 |
| RESID(-1)^2 | 0.124009 | 0.025932 | 4.782142 | 0.0000 |
| GARCH(-1) | 0.836758 | 0.032292 | 25.91221 | 0.0000 |
| R-squared | 0.015193 | Mean dependent var | | -0.003396 |
| Adjusted R-squared | 0.015193 | S.D. dependent var | | 0.166472 |
| S.E. of regression | 0.165202 | Akaike info criterion | | -0.843286 |
| Sum squared resid | 11.54441 | Schwarz criterion | | -0.805081 |
| Log likelihood | 182.7767 | Hannan-Quinn criter. | | -0.828192 |
| Durbin-Watson stat | 2.139322 | | | |

Figure 6 Comprehensive model

We get the comprehensive model as follows:

$$\text{Mean value equation: } Y_t = 0.132254 Y_{t-3} + a_t \quad (6)$$

$$\text{GARCH (1, 1) model: } a_t = \sigma_t \varepsilon_t, \sigma_t^2 = 0.01593 + 0.124009 a_{t-1}^2 + 0.836758 \sigma_{t-1}^2 \quad (7)$$

When the significance level is taken as 0.1, the t-test of model coefficients is passed, and then the white noise test of residual sequence of joint model is considered. Here, the q-statistic test of box Ljung is still used to test whether it is irrelevant. The test results are shown in Table 4.

Table 4 White noise test of residual sequence of comprehensive model

| Lag period | Q statistic | P value |
|---------------|-------------|---------|
| 6 lag phases | 4.7655 | 0.445 |
| 12 lag phases | 7.2721 | 0.777 |

Because the p value is far greater than the significance level α ($\alpha = 0.01, 0.05, 0.1$), the residual sequence is white noise data, and the validity of the joint model can be tested.

Next, judge whether there is arch effect in the comprehensive model. Here, Lagrange multiplier test is still used. The test results are shown in Figure 7.

Heteroskedasticity Test: ARCH

| | | | |
|---------------|----------|---------------------|--------|
| F-statistic | 0.221363 | Prob. F(1,421) | 0.6382 |
| Obs*R-squared | 0.222298 | Prob. Chi-Square(1) | 0.6373 |

Figure 7 Arch effect test of comprehensive model

The p value of LM statistic is far greater than the significance level, which indicates that there is no Heteroscedasticity in the joint model, and the adequacy of the joint model can be tested.

3.6. Forecasting of the model

Next, use the joint model to make short-term forecast on the closing price of August 6, 7 and 8, 2019, and compare it with the actual data. The results are shown in Table 5.

Table 5 The actual value and forecast value of the closing price

| date | actual value | forecast value | absolute value of error |
|----------------|--------------|----------------|-------------------------|
| August 6, 2019 | 5.23 | 5.314615 | 0.084615 |
| August 7, 2019 | 5.12 | 5.217330 | 0.09733 |
| August 8, 2019 | 5.14 | 5.097375 | 0.042625 |

4. Conclusion

This paper uses the modelling idea of time series analysis to analyze the closing price series of China Eastern Airlines through EViews 8.0. In the process of modelling using ARIMA model, it is found that some items in AR model have no obvious autocorrelation, so the AR model with missing coefficient is used. In the mean equation modelling process, we found that the residual sequence has ARCH effect, and finally fitted the ARIMA-GARCH joint model. The joint model has a good effect on the short-term prediction of stock closing price. The prediction results of the model have certain guiding significance for investors to make full use of market information to make investment decisions.

References

- [1] Deng Jun, Yang Xuan, Wang Jin, et al. An Empirical Study on the prediction of stock price using ARMA model. *Business Herald*, 2010 (6): 266-268.
- [2] Chen Yiheng, Huang Changquan. *Time series and financial data analysis*. Beijing: China Statistics Press, 2004.
- [3] Wu Yuxia, Wen Xin. Short term stock price prediction based on ARIMA model. *statistics and decision*, 2016 (23): 82-86.
- [4] Wang Yan. *Applied time series analysis (Fourth Edition)*. Beijing: China Renmin University Press, 2015.
- [5] Zhang Xiaodong. *Fundamentals of econometrics*. Tianjin: Nankai University Press, 2001.
- [6] Wu Chan, Zhao Yu, Cai Jiannan, Li Wen. Comparison of RMB exchange rate prediction based on ARMA model and GARCH model. *Economic perspective*, 2013 (16): 243-245.
- [7] He Shuyuan. *Applied time series analysis*. Peking University Press, 2009.
- [8] Yang Qi, Cao Xianbing. Stock price analysis and prediction based on arma-garch model. *Practice and understanding of mathematics*, 2016 (3): 80-86.
- [9] Song Bo, Chen Wanyi. Prediction of RMB exchange rate trend based on HP Filtering and arma-garch model. *Practice and understanding of mathematics*, 2017 (1): 70-77.