Forecasting the Pharmaceutical Stock Prices in China in the Context of the Coronavirus Crisis Based on ARIMA-GARCH Model

Lingxian Zhu(✉)

Department of Mathematical Science, University of Durham, Durham, UK
Zelaine1998@gmail.com

Abstract. This paper researches the future trend of the Chinese pharmaceutical company's stock prices in the Covid-19 pandemic. Through the optimized ARIMA-GARCH model, this paper investigates the China National Medicines Corporation stock representing the Chinese pharmaceutical stocks. The data analysis provides reliable support for future investment strategies in Chinese pharmaceutical stocks for investors.

Keywords: COVID-19 · forecasting · stock price · time-series · ARIMA-GARCH · China

1 Introduction

Covid-19, a series of acute atypical respiratory diseases, occurred in Wuhan, China, in January 2020 and quickly spread to other areas. It was declared as "Public Health Emergency of International Concern" and alarmed worldwide. Vaccination is always a standard method to confront the virus and control the spread of the epidemic. Four companies have produced most of the coronavirus vaccines approved by governments in China which are the China National Medicines Corporation, Sinovac Biotech, CanSino Biologics, and Zhifei Biological Products. Apart from Sinovac Biotech, the remaining three pharmaceutical companies have all gone public. At the early stage of the Covid-19 pandemic spreading (i.e., from 2020 to June 2021), the Chinese government and citizens emphasized coronavirus prevention and invested a large amount of capital in the pharmaceutical industry (especially in vaccine production and promotion), so stock prices of these three companies fluctuated and increased significantly. However, the three stocks have shown downward trends in varying degrees from July 2021, perhaps because of the high vaccination rate and the effective control of the coronavirus diffusion. Therefore, will these pharmaceutical stocks no longer have investment value shortly? While the spread of coronavirus has been effectively controlled, small-scale infections of viruses like delta variant still occur as the virus continues to mutate. Under this situation, this paper investigates what investment attitude investors should hold toward pharmaceutical stocks in China?

The time series analysis is the well-known statistical method used for the stock prices analysis. In terms of popular time series models, the Autoregressive Moving Average

Z. Zeng et al. (Eds.): ECIT 2022, AHE 11, pp. 786–795, 2023.
https://doi.org/10.2991/978-94-6463-005-3_79

Model (ARMA) usually emphasizes the stationarity of time series data. In contrast, the Autoregressive Integrated Moving Average Model (ARIMA), which extends the ARMA model, and the Autoregressive Conditional Heteroskedasticity Model (ARCH), which includes the Generalized Autoregressive Conditional Heteroskedasticity Model (GARCH), are mainly for forecasting the non-stationary time series data [12]. Because of the constant conditional variance of the ARMA model, it cannot capture the process with volatility, a time-varying conditional variance [10]. The ARIMA model is a method that delivers better prediction results for long-term financial forecasting [1]. Under normal distribution assumption, Frimpong and Oteng-Abayie (2006) showed that the GARCH (1, 1) model performs better in predicting [4]. However, though the GARCH model can capture characteristics of financial time series and is widely used for stock volatility analysis, its hypothesis ignores the symbol of new information [15]. Furthermore, the high-dimensional volatility modeling problem cannot be overlooked since the interconnected stock markets are more than three for volatility analysis. But many researchers showed the GARCH model is limited from solving high-dimensional (over three-dimensional) problems [6, 8, 14, 16, 19]. Mixed models such as the ARMA-GARCH/ARIMA-GARCH models have been investigated [10, 15]. In sum, the optimal forecasting model is various for different focuses of the research object, such as long-term or short-term, different sizes of the companies, and different types of securities [17].

This research uses time series analysis to analyze and forecast the future trend of China's pharmaceutical stock prices. Since stock price data are usually considered as the non-stationary time series data and this article hopes to predict the long-term tendency of China's pharmaceutical stocks, the main model adopted will be the ARIMA model. It is necessary to do the residual analysis of the fitted model. If the residuals present a skewness normal distribution instead of a standard normal distribution, the ARCH-type models need to be considered to optimize the volatility of the fitting model. Since GARCH (1, 1) usually presents an ideal fitting and prediction effect [4], it is common to take the GARCH model to fit the residual model. Combining the main data model and the residual fitting model is the optimized ARIMA-GARCH model. Carry out the residual test on the optimized ARIMA-GARCH model. If the residuals conform to the standard normal distribution and the significance is within the set range, it means that the model fits well and can be used for prediction. Otherwise, the model needs to continue to be adjusted. In general, the optimized ARIMA-GARCH model combines the advantages of the ARIMA model and the GARCH model and has a better performance for the long-term financial market forecasts overall.

As the Sinovac Biotech and CanSino Biologics went public in December 2020 and August 2020, there is no such data available for analysis during the early stage of the coronavirus. Market participants have become accustomed to the coronavirus, so despite the existence of the coronavirus's macro impact, investors always depend on the long-term performance of stocks. The China National Medicines Corporation, a Fortune 500 company, came on the market in 2002 and became a leading pharmaceutical company with outstanding long-term market performance in China. While the Sinovac Biotech and CanSino Biologics will benefit investors, this research focuses on analyzing and forecasting the future trend of the China National Medicines Corporation, which is more representative of China's pharmaceutical industry.

The dataset used for this research is downloaded from the trading data on the Shanghai Stock Exchange. The close data of all trading days since the listing date are used to determine the best-fitting model. It is unrealistic to forecast the daily stock price data for the long-term (3–5 years) due to the large calculation. However, predicting a small amount of daily stock price data cannot serve as a reasonable investment recommendation. In order to simplify the calculation and at the same time enable the forecast results to have a longer-term investment strategy, this article calculates the annual average closing price as the new time series data and predicts the 3–5 subsequent time-series data to determine the general trend of the future stocks.

The forecasting results can provide the data support of investment strategies in pharmaceutical stocks for investors. The rest of the paper is organized as follows. Section 2 introduces the model and estimation process of the time series analysis and the relevant productions. The investment strategy and forecasting result discussion are given in Sect. 3. Section 4 concludes the paper.

2 Methodology

Suppose the stock prices of the China National Medicines Corporation are a non-stationary time series from Fig. 1. Let the ARIMA model be the basic analysis model to achieve optimal forecasting for a long-term period. The GARCH model may be needed to consider depending on the subsequent test results. The significance level used in this research is $\alpha = 0.05$.

Consider the ARIMA model

$$\begin{aligned} y_t &= ARIMA(p, d, q), \\ \Delta^d y_t &= x_t = \sum\nolimits_{i=1}^{p} \varphi_i x_{t-i} + u_t + \sum\nolimits_{i=1}^{q} \theta_i u_{t-i}. \end{aligned}$$

The Augmented Dickey-Fuller (ADF) test confirmed that the data form a non-stationary time series since the p-value $= 0.3551 > \alpha = 0.05$. The first-order difference of our data gives the p-value $< 0.01 < \alpha = 0.05$, providing a stationary time series and leading to the parameter d $= 1$. The remaining parameters p and q can be estimated through the Akaike Information Criterion (AIC) with $AIC = -2\log(\Lambda) + 2k$ and estimated coefficients can be obtained. The best-fitted model is, therefore, the ARIMA (1, 1, 2) model with the expression as

$$\begin{aligned} ARIMA(1, 1, 2) &= \hat{y}_t - \hat{y}_{t-1} \\ &= 0.6739\big(\hat{y}_{t-1} - \hat{y}_{t-2}\big) + \hat{\epsilon}_t \\ &\quad - 0.6108\hat{\epsilon}_{t-1} - 0.1166\hat{\epsilon}_{t-2}. \end{aligned}$$

The test on the residuals of the fitted ARIMA (1, 1, 2) model is necessary since it can provide sufficient evidence on whether the ARIMA (1, 1, 2) model needs to be optimized via the GARCH model.

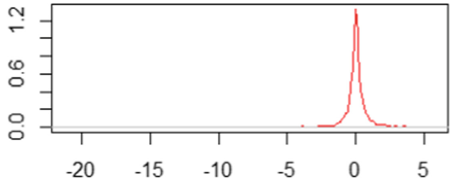

Fig. 1. The historical stock prices of the China National Medicines Corporation based on the close prices

Except for some extreme outliers on the left, the residual density distribution of the fitted ARIMA (1, 1, 2) model seems to follow the standard normal distribution by Fig. 2. However, it is distributed under a non-Gaussian case according to the graphs of the Autocorrelation Function (ACF) shown in Fig. 3 and Partial Autocorrelation Function (PACF) given as Fig. 4, and the Ljung-Box (LB) test with $p = 0.0008665 < \alpha$, when LAG = 22. From the Chi-squared test, we got p-value $< 2.2e{-}16 < \alpha$, so utilizing the GARCH model will be necessary for optimization.

Consider the GARCH model as

$$\begin{aligned} GARCH(m, s) &= \sigma_t^2 \\ &= w + \sum\nolimits_{i=1}^{m} \alpha_i \epsilon_{t-i}^2 + \sum\nolimits_{j=1}^{s} \beta_j \sigma_{t-j}^2. \end{aligned}$$

The GARCH (1, 1) model fits the residual sequence giving results as

$$\begin{array}{lllll} \multicolumn{5}{l}{Coefficient(s):} \\ & Estimate & Std.Error & t\ value & \Pr(>|t|) \\ a0 & 1.999e-03 & 7.061e-05 & 28.31 & < 2e-16\ *** \\ a1 & 8.205e-02 & 3.535e-03 & 23.21 & < 2e-16\ *** \\ b1 & 9.297e-01 & 2.175e-03 & 427.35 & < 2e-16\ *** \end{array}$$

$$\begin{array}{l} \qquad\qquad Jarque\ Bera\ Test \\ data: Residuals \\ X - squared = 1841770, df = 2, p - value < 2.2e - 16 \\ \qquad\qquad Box - Ljung\ test \\ data: Squared.Residuals \\ X - squared = 0.060849, df = 1, p - value = 0.8052. \end{array}$$

All the p-values of the estimated coefficients are less than α, showing statistical significance. Furthermore, the p-value of the LB test is greater than α, giving the GARCH (1, 1) model is sufficient to present the residuals of the ARIMA (1, 1, 2) model. Taking the estimated coefficient, the GARCH (1, 1) model is

$$\begin{aligned} GARCH(1, 1) &= \hat{\sigma}_t^2 \\ &= 1.999e - 03 + 8.205e - 02\ \hat{\epsilon}_{t-1}^2 \\ &\quad + 9.297e - 01\ \hat{\sigma}_{t-1}^2. \end{aligned}$$

Fig. 2. The residual density distribution of the fitted ARIMA (1, 1, 2) model

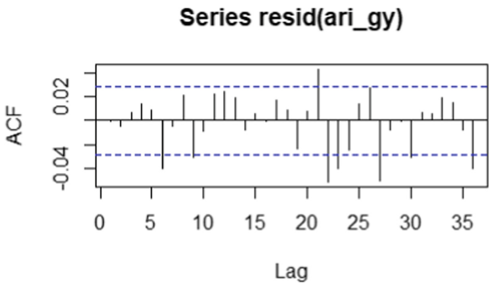

Fig. 3. The Autocorrelation Function of the residuals of the fitted ARIMA (1, 1, 2) model

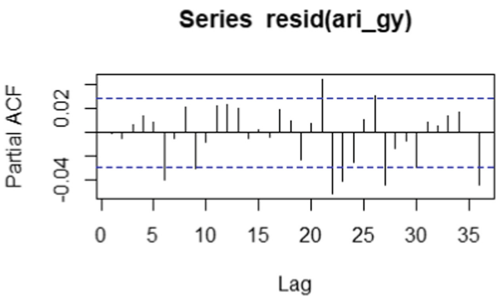

Fig. 4. The Partial Autocorrelation Function of the residuals of the fitted ARIMA (1, 1, 2) model

The optimized ARIMA (1, 1, 2) - GARCH (1, 1) model can be determined by combining the ARIMA (1, 1, 2) model and the GARCH (1, 1) model with the formula of

$$\begin{aligned} & ARIMA(1, 1, 2) - GARCH(1, 1) \\ &= \hat{y}_t - \hat{y}_{t-1} \\ &= 0.6739\left(\hat{y}_{t-1} - \hat{y}_{t-2}\right) - 0.6108\hat{\epsilon}_{t-1} - 0.1166\hat{\epsilon}_{t-2} \\ &\quad + 1.999e - 03 + 8.205e - 02\ \hat{\epsilon}_{t-1}^2 + 9.297e - 01\hat{\sigma}_{t-1}^2. \end{aligned}$$

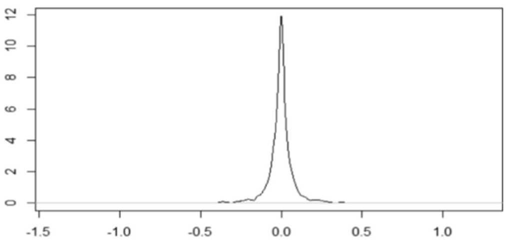

Fig. 5. The residual density distribution of the ARIMA (1, 1, 2) - GARCH (1, 1) model

Fig. 6. The real stock price (blue) of the China National Medicines Corporation compared with the estimated price (red) based on the ARIMA (1, 1, 2) - GARCH (1, 1) model

Figure 5 shows the residual density distribution of the fitted ARIMA (1, 1, 2) - GARCH (1, 1) model. The estimations for stock prices of our fitted ARIMA (1, 1, 2) - GARCH (1, 1) model are shown in Fig. 6, where the blue line is the actual stock prices, and the red line is the estimated stock prices.

The residual density distribution of the ARIMA (1, 1, 2) - GARCH (1, 1) model follows a well Gaussian distribution, and Fig. 6 shows it has been well-fitted as well. Therefore, the optimized ARIMA-GARCH model can be utilized to predict the stock price trend of the China National Medicines Corporation.

3 Results and Discussion

3.1 Forecasting Results

Under the ARIMA-GARCH model, the future forecasting of stock prices for the China National Medicines Corporation is given in Fig. 7.

3.2 Investment Strategy

Figure 7 shows that the price curve rises slightly in the following year but decreases significantly afterward. Until about the latter 2024, a turning point shows a remarkable

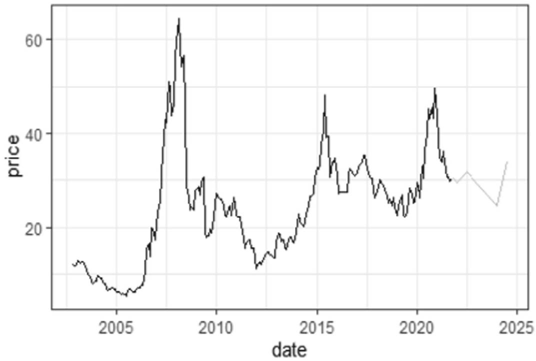

Fig. 7. The forecasting results of the China National Medicines Corporation stock price

upward trend. However, whether stock prices will continue to rise substantially is still doubtful.

From the perspective of investment income, investors are not eager to buy shares in the China National Medicines Corporation stock shortly. Experienced short-term investment operations may still bring benefits but are not outstanding. Buying stocks in later 2024 may obtain excellent investment returns because of the turning point.

3.3 Discussion

The optimized ARIMA (1, 1, 2) - GARCH (1, 1) model provides a well-fitted prediction for stock prices of the China National Medicines Corporation. However, the accuracy and adequate forecasting time may be limited by applying the basic prediction algorithm. Therefore, the application of the model will be significantly restricted due to the lack of efficient high-precision forecasting algorithm optimization.

In recent years, besides the statistical models, a significant contribution has been made in developing and applying machine learning techniques to predict stock prices. A wide range of machine learning methods can be adopted for stock price forecasting: regression analysis (logistic, lasso, ridge, etc.), cluster analysis (K-means, hierarchical clustering, etc.), support vector machines (SVM), decision trees, random forests, extreme gradient boosting (XGBoost), and neural networks [1, 2]. Compared with traditional Machine Learning models, Deep Learning, also called the Artificial Neural Network, performs better in stock prices analysis and prediction. Sezer et al. (2020) summarized the contribution of the recent popular Deep Learning models such as Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), and Long-Short Term Memory (LSTM) in financial time series forecasting problems from 2005 to 2019 [11]. Reinforcement Learning is a relatively new approach in Machine Learning, and Dang (2019) believed that Reinforcement Learning has a lot of potentials in decision-making for stock trading [13]. The LSTMs models, which use special gates to control information flow, are currently the type of neural network that gives the preferred performance in financial forecasting [18].

Using Machine Learning to forecast stock price volatility highly depends on the scale and pre-processing techniques of the dataset. Since the coronavirus is an unprecedented collective crisis with insufficient data for model training and sentiment analysis, the overfitting problems, lack of historical stock data and unavailability of information such as government policies and market sentiment will lead to biased results [2]. However, the machine learning techniques have a high potential prediction accuracy and numerous researchers have adopted a hybrid model combining statistics and machine learning techniques [21] for forecasting algorithm optimization. The multi-layer LSTM optimization algorithm for time series forecasting is widely used to improve prediction accuracy [7].

Furthermore, considering the volatility of the financial market, an analysis of the overall performance of stocks and market participants' sentiments should be included. And it is necessary to adopt technical analysis methods flexibly rather than obstinately apply forecasting algorithms.

From the perspective of stock price trends, the China National Medicines Corporation pattern presents a cyclical triangle. Pattern Recognition, a popular technical method, can aid forecasting and algorithm optimization [9]. Meanwhile, pharmaceutical stocks are unlikely to fall sharply considering the impact of major epidemic diseases. In contrast, the performance of pharmaceutical companies in epidemic diseases will affect the sentiment of market participants and thus stock prices. During the SARS-CoV outbreak (2002–2003) and the Influenza A virus subtype H1N1 episode (2009–2010), the China National Medicines Corporation did not significantly contribute to the research and development for related vaccinations. Due to the market impact, its stock price declined, but not dramatically during the epidemic disease period, considering people tend to rely on pharmaceutical companies psychologically in the disease [22]. With the remarkable achievements in researching and developing the vaccination for the Covid-19, the stock price of the China National Medicines Corporation has shown a significant rise. Market participants' mental and emotional factors can drive market changes in the short term and even cause unexpected price fluctuations in the financial market. Therefore, sentiment analysis, a relatively novel approach, is unique in financial market forecasting by analyzing and measuring data sources' potential emotions and text corpus [5] implemented by machine learning technique [23].

Moreover, despite our normal-distributed residuals, the normal distribution is ideal, and most time-series data tend to show skewness. For non-Gaussian considerations, using various copulas with the flexible specification of the marginal distributions is an extension that performs remarkably well in dynamic volatility forecasting with copula-based multivariate models [3, 20].

Furthermore, with a series of severe social and environmental issues such as global warming, health awareness growth, and the Chinese government's strong support for the pharmaceutical industry, pharmaceutical stocks still have long-term solid investment significance.

4 Conclusion

The optimized ARIMA (1, 1, 2) - GARCH (1, 1) model shows a well-performance in predicting stock prices for the China National Medicines Corporation and provides investors with data-based investment strategies.

Future research can focus on the optimization of the forecasting algorithms. Combining the machine learning technique such as the preferred multi-layer LSTM models, Pattern Recognition and Sentiment Analysis may provide more accurate long-term prediction results and bring more significant investment returns.

Acknowledgment. The author would like to acknowledge the anonymous referee for his careful reading of the paper and constructive suggestions.

References

1. Shah D, Isah H, Zulkernine F (2019) Stock market analysis: a review and taxonomy of prediction techniques. Int J Financ Stud 7(2):26
2. Gandhmal DP, Kumar K (2019) Systematic analysis and review of stock market prediction techniques. Comput Sci Rev 34:100190
3. Liu HH, Wang TK, Li W (2019) Dynamical volatility and correlation among us stock and treasury bond cash and futures markets in presence of financial crisis: a copula approach. Res Int Bus Financ 48:381–396
4. Frimpong JM, Oteng-Abayie EF (2006) Modelling and forecasting volatility of returns on the Ghana stock exchange using GARCH models
5. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8
6. Bauwens L, Laurent S, Rombouts JVK (2006) Multivariate GARCH models: a survey. J Appl Econ 21(1):79–109
7. Peng L, Zhu Q, Lv S-X, Wang L (2020) Effective long short-term memory with fruit fly optimization algorithm for time series forecasting. Soft Comput 24(19):15059–15079. https://doi.org/10.1007/s00500-020-04855-2
8. Caporin M, McAleer M (2012) Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. J Econ Surv 26(4):736–751
9. Velay M, Daniel F (2018) Stock chart pattern recognition with deep learning
10. Emenogu NG, Adenomon MO, Nwaze NO (2019) Modeling and forecasting daily stock returns of Guaranty Trust Bank Nigeria Plc using ARMA-GARCH models, persistence, half-life volatility and backtesting. Sci World J 14(3):1–22
11. Sezer OB, Gudelek MU, Ozbayoglu AM (2020) Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. Appl Soft Comput 90:106181
12. Brockwell PJ, Davis RA (2009) Time series: theory and methods. Springer
13. Dang QV (2019) Reinforcement learning in stock trading. In: International conference on computer science, applied mathematics and applications, December 2019. Springer, Cham, pp 311–322
14. Engle R (2002) Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J Bus Econ Stat 20(3):339–350
15. Chai S, Zhang Z, Du M, Jiang L (2020) Volatility similarity and spillover effects in G20 stock market comovements: an ICA-based ARMA-APARCH-M approach. Complexity

16. Basher SA, Sadorsky P (2016) Hedging emerging market stock prices with oil, gold, VIX, and bonds: a comparison between DCC, ADCC and GO-GARCH. Energy Econ 54:235–247
17. Poon SH, Granger CWJ (2003) Forecasting volatility in financial markets: a review. J Econ Lit 41(2):478–539
18. Kamolov S, Iskhakov D, Ziyaev B (2021) Machine learning methods in time series forecasting: a review. Ann Math Comput Sci 2:10–14
19. Ahmad W, Sadorsky P, Sharma A (2018) Optimal hedge ratios for clean energy equities. Econ Model 72:278–295
20. Chen X, Huang Z, Yi Y (2021) Efficient estimation of multivariate semi-nonparametric GARCH filtered copula models. J Econ 222(1):484–501
21. Wang Y, Guo Y (2020) Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. China Commun 17(3):205–221
22. Wang Y, Wang J, Wang X (2020) COVID-19, supply chain disruption and China's hog market: a dynamic analysis. China Agric Econ Rev
23. Xu Y, Cohen SB (2018) Stock movement prediction from tweets and historical prices. In: Proceedings of the 56th annual meeting of the association for computational linguistics (Volume 1: Long Papers), July 2018