Proceedings of the 7th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/37/20231874

A Comparative Study on Stock Price Forecasting after the COVID-19 Pandemic

Li Sun[1,a,*]

[1]*Northwestern Polytechnical University, Shaanxi Province, China, 710129*
a. sl17366014359@163.com
**corresponding author*

Abstract: Nowadays, stock price forecasting is still a difficult problem plaguing people, due to the uncertainty of the market. We contrast the effectiveness of three models in this research.: ARIMA, SVM, and BPNN model. To get the outperformed model in predicting after the COVID-19 pandemic, the data set is the stock price of Citic Securities (600030) on the Shanghai Stock Exchange from 2022.4.1 to 2022.12.30. Experiment results show that all three models predict outcomes with very little error from the actual values. And BPNN model has the best accuracy because of its great ability to portray the non-linear characteristic of our data.

Keywords: stock price prediction, ARIMA model, support vector machine, BP neural network

1. Introduction

Driven by the wave of modern finance, more and more individuals enter the stock market and make investments in an effort to reap lucrative returns, considerably enhancing the market's success. Additionally, behind this investment strategy, more and more investors are gradually realizing the value of stock market forecasting. Many researchers have created and improved prediction models in an effort to lower investment risk.

However, there is no such model that always performs best. Therefore, for different data sets, we need to assess the superiority of the predictions made by various models in terms of accuracy. In this essay, we'll make a share price prediction of the Shanghai Stock Exchange during and after the epidemic, adopting ARIMA, SVM, and BPNN models. The experimental results of this paper are based on the stock price of Citic Securities coded 600030 from 2022.4.1 to 2022.12.30.

2. Literature review

Early time series prediction models were mainly single-series linear models such as ARIMA. This model fits each sequence separately [1-4]. On the basis of ARIMA, the optimization of introducing nonlinearity and external features is proposed. Engle's initial proposal for heteroskedasticity scenarios was the auto-regressive conditional heteroskedasticity model. [5]. The fundamental presumption of the ARCH model is that the variance represents a normal analysis of the quantity changing over time and that zero is the mean value of all noise at once. A linear combination of the squares of the previous values in the finite term sequence yields a variance that varies with time. Since the GARCH model

DOI: 10.54254/2754-1169/37/20231874

may more accurately capture the long-term memory characteristics of actual data, Bollerslev (1986) expanded ARCH.[6-8].

Artificial neural networks (ANN), a soft computing technique, are also applied to stock prediction problems. Numerous studies have demonstrated that ANN can more accurately model heteroscedasticity, or data with erratic variations and excessive volatility, due to its ability to uncover latent links in the data. [9-11]. This is helpful in predicting stock prices when the data is very volatile. Moreover, time-delay neural networks (TDNN), back propagation neural networks (BPNN), probabilistic neural networks (PNN), and recurrent neural networks (RNN) are proposed by researchers in the follow-up studies on ANN [12, 13].

A generalized linear classifier called the Support Vector Machine (SVM) uses supervised learning to categorize data. [14-16]. The hyperplane that has the learning sample's biggest margin of error is the decision boundary. SVM was gradually theorized and incorporated into statistical learning theory as a result of the theoretical investigation into the pattern recognition maximum margin decision boundary, the development of slack-based programming technique for solving problems, and the suggestion of the Vapnik-Chervonenkis dimension (VC dimension). [17, 18].

3. Methodology

3.1. AR (Auto Regression) Model

The variable's history temporal data is used by the autoregressive model to forecast itself and to define the connection between the present value and the past value. The generalized P-order autoregressive model AR is shown as follows: $P_t = a1Pt{-}1 + a2Pt{-}2 + \cdots + apPt{-}p + ut$

When the random disturbance term is white noise ($u_t = \sigma_t$), the process is referred to as a pure AR (p) process and is represented by : $P_t = a_1 P_{t-1} + a_2 P_{t-2} + \cdots + a_p P_{t-p} + \sigma_t$

The autoregressive model requires establishing an order p, which calls for the utilization of past values across a number of periods to forecast the present value.

3.2. MA (Moving Average) Model

In the AR model, if u_t is not a white noise, it's usually thought of as a q-order moving average. And the equation is denoted as: $u_t = \sigma_t + b_1 \sigma_{t-1} + \cdots + b_q \sigma_{t-q}$
with σ_t representing the white noise sequence.

3.3. ARMA Model

A generic autoregressive moving average model (ARMA) (p, q) is created by combining the two coefficients of AR (p) and MA (q): $P_t = a_1 P_{t-1} + a_2 P_{t-2} + \cdots + a_p P_{t-p} + \sigma_t + b_1 \sigma_{t-1} + \cdots + b_q \sigma_{t-q}$

The degree of correlation between the current value of the sequence and its previous value is expressed by the full autocorrelation function, or ACF. Time series might include trend, seasonality, periodicity, and residual components. When searching for correlations, the ACF takes into account all of these factors.

PACF stands for partial autocorrelation function. Instead of finding a connection between the two, like in ACF, it finds a correlation between the following lag value and the residual. In contrast to intuition, PACF accounts for the impacts of other shorter lag terms by describing simply the direct link between the observed value y_t and its lag term y_{t-k}.

DOI: 10.54254/2754-1169/37/20231874

3.4. SVM (Support Vector Machine)

Models for binary classification include support vector machines (SVM). In contrast to the perceptron, its core model is a linear classifier constructed on the feature space with the largest margin. SVM is essentially a nonlinear classifier due to additional kernel methods. Margin maximization, which may be considered to be a convex quadratic programming issue. Convex quadratic programming may be solved using the learning process of the SVM.

3.5. BPNN (Back Propagation Neural Networks)

Using error back propagation neural networks are a sort of multi-layer feedforward network. The BP algorithm, which it uses, is based on the technique of gradient descent and by applying gradient search technology, decreases the mean square error between the network's predicted and actual output values. Both the processes—signal forward propagation and error back propagation—are involved, which make up fundamental BP algorithm.

Forward propagation is a process where the input signal interacts with the output node via hidden layers to produce output signal after nonlinear changes. Mistakes will be passed on to the back propagation procedure if the actual yield differs from the anticipated yield. Error back propagation is the process by which yield errors are sent from hidden layers to input layers one at a time, with each layer's units receiving an allocation for the error. Each unit's weight is modified based on the error signal that is acquired from each layer. As we move in a gradient, the error decreases through conforming the threshold. After several iterations of learning and training, the settings of network that correlate to the lowest error are established, and training is then completed.

4. Experiment

4.1. The result of ARIMA model

We performed a time series analysis using the ARIMA model, first looking at the results of the ADF as below.

Table 1: ADF test table

ADF Inspection Form							
Variables	Differential orders	t	P	AIC	Threshold values		
					1%	5%	10%
close	0	-2.154	0.223	106.267	-3.467	-2.878	-2.575
	1	-16.025	0.000	98.41	-3.467	-2.878	-2.575
	2	-10.011	0.000	129.089	-3.468	-2.878	-2.576

The outcomes of this series test reveal that the significant P-value is 0.223 at the difference score of order 0 based on the variable CLOSE. Assuming a smooth time series, the significance P-value at order 1 of the difference is 0.0000, which is significant enough to reject the null hypothesis. The significant P-value is 0.000 for an order 2-differential situation, which is significant to the level of rejecting the initial theory that the series is smooth. Contrasting the results of statistical analysis with the results of the ADF Test at different levels of the original theory being rejected at critical values of 1%, 5%, and 10%, with the ADF Test result being less than 1%, 5% and 10% at the same time, shows a strong rejection of the idea. A smaller value of AIC, which is a way to gauge how well the statistical model fits the data, indicates a better fit. It is easy to see that the best differential series plot is a first-order differential.

Based on the information criteria, multiple analysis model comparisons were performed using the AIC and BIC values. The model outputs for the ARIMA model (1,1,0) test table, depending on the variable: near, are determined automatically using the AIC information criteria to discover the best parameters.

Based on the AIC information criteria and the variable CLOSE, the best parameters were discovered automatically, and the model's output was an ARIMA model (1,1,0) test table with one difference in the data with the following model equation: y(t) = -0.007 - 0.175*y(t-1)
Finally, we get the final forecast (Table 2):

Table 2: Predicted Value of ARIMA

Order (time)	Predicted results
1	19.879298944977755
2	19.876772890080023
3	19.869307037917697
4	19.862707258405145
5	19.85595563423288
6	19.849230632308405
7	19.842500962823813

4.2. Result of SVM

Once x has been projected to the feature's new feature space, we indicate by x the initial data point and by $\phi(x)$ the new vector. Then the segmented hyperplane can be denoted as:

$$f(x) = \omega\ \phi(x) + b.$$

The dual problem of our nonlinear SVM becomes a minimization problem:

$$\min_{\lambda} \left\{ \left[\frac{1}{2} \sum_{m=1}^{n} \sum_{n=1}^{n} \lambda_m \lambda_n y_m y_n \left(\phi(x_m) \cdot \phi(x_n) \right) \right] - \sum_{m=1}^{n} \lambda_m \right\}$$

Subjected to:

$$\sum_{i=1}^{n} \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

$$C - \lambda_i - \mu_i = 0$$

Next, the kernel function $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\delta^2}\right)$ is introduced to reduce the calculation. The inner product of x_i and x_j in the eigenspace is equal to what they computed in the original sample space by the kernel function $k(x_i, x_j)$, so the inner product need not be calculated in higher or even infinite dimensions. Then, we get the predicted value after training (Table 5):

Table 3: Predicted Value of SVM

Order (time)	Predicted results
1	19.88
2	19.88
3	19.87
4	19.87
5	19.87
6	19.84
7	19.84

4.3. Result of BPNN

The following diagram depicts the overall layout of the BPNN model, which has three layers: the input layer, the hidden layer, and the output layer. Important information is sent between the input layer and the output layer via the hidden layer.

A hidden layer or layers are always present in BPNN, enabling the network to simulate complicated functions. Forward information propagation and error backpropagation are the two fundamental processes that make up this system. The following is how these three levels relate mathematically to one another:

From the input layer to the hidden layer:

$$y_i = f_h(\mu_i + \sum_{m=t-n}^{t-1} \mu_{im} y_m)$$

From the hidden layer to the output layer:

$$y_t = f_O(\lambda_O + \sum_{i=1}^{h} \lambda_{Oi} y_i)$$

The result is presented below:

Table 4: Predicted Value of BPNN

Order (time)	Predicted results
1	19.91
2	19.91
3	19.90
4	19.88
5	19.88
6	19.87
7	19.86

4.4. Error

The relative errors of prediction are defined by

$$\text{Relative Error} = \frac{|actual - predicted|}{actual}$$

The relative errors are presented:

Table 5: Relative Errors

	ARIMA	SVM	BPNN
1	0.00502	0.005056	0.006572
2	0.001669	0.001507	0
3	0.001542	0.001508	0
4	0.014258	0.013896	0.0134
5	0.028574	0.027886	0.027397
6	0.027475	0.027927	0.026458
7	0.039104	0.039225	0.038257
Average	0.016806	0.016715	0.016012

From Table 7, we get to know that the relative errors between predicted and actual values are small for all three models. Moreover, BPNN model can better match the real data than the other two.

5. Conclusion

In this study, using data from 2022.4.1 to 2022.12.30, we apply the ARIMA, SVM, and BPNN models to anticipate the stock price of Citic Securities. It turns out that the minimum average relative error is from the BPNN model, which means The BPNN model outperforms others in terms of performance. Since the data is from during and after the epidemic, share prices are influenced by more factors than normal, and the data are more non-linear. Due to the better able to capture the non-linear characteristic of data, we draw the conclusion that the BPNN model may be used to forecast stock prices with greater accuracy.

Though this essay contrasts the three models' effectiveness and the BPNN model is best fitted to the actual values, there are a few limitations in this study. For instance, it only takes individual models into consideration. Some hybrid models, like the ARIMA-GARCH model, may forecast more accurately with our data. The combination of the two models may better preserve the linear and nonlinear characteristics. Moreover, different stocks have fared very differently in the wake of the pandemic. So more stocks should be analyzed to have a better insight into the stock price trend after COVID-19.

References

[1] D. Banerjee, "Forecasting of Indian stock market using time-series ARIMA model," 2014 2nd International Conference on Business and Information Management (ICBIM), Durgapur, India, 2014, pp. 131-135, doi: 10.1109/ICBIM.2014.6970973.

[2] A.A. Ariyo, A.O. Adewumi and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2014, pp. 106-112, doi: 10.1109/UKSim.2014.67.

[3] S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," in IEEE Access, vol. 7, pp. 17287-17298, 2019, doi: 10.1109/ACCESS.2019.2895252.

[4] W. Li and J. Liao, "A comparative study on trend forecasting approach for stock price time series," 2017 11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID), Xiamen, China, 2017, pp. 74-78, doi: 10.1109/ICASID.2017.8285747.

[5] Engle, Robert F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." Econometrica 50, no. 4 (1982): 987–1007. https://doi.org/10.2307/1912773.

[6] Tim Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, Volume 31, Issue 3, 1986, Pages 307-327.

DOI: 10.54254/2754-1169/37/20231874

[7] *C. Narendra Babu, B. Eswara Reddy, Prediction of selected Indian stock using a partitioning–interpolation based ARIMA–GARCH model, Applied Computing and Informatics, Volume 11, Issue 2, 2015, Pages 130-143.*
[8] *Zhou, B., He, D., Sun, Z. (2006). Traffic Modeling and Prediction using ARIMA/GARCH Model. In: Nejat Ince, A., Topuz, E. (eds) Modeling and Simulation Tools for Emerging Telecommunication Networks. Springer, Boston, MA.*
[9] *ANN An-Sing Chen, Mark T. Leung, Hazem Daouk, Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index, Computers & Operations Research, Volume 30, Issue 6,2003, Pages 901-923.*
[10] *Ayodele Ariyo Adebiyi, Aderemi Oluyinka Adewumi, Charles Korede Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction", Journal of Applied Mathematics, vol. 2014, Article ID 614342, 7 pages, 2014.*
[11] *Singh, R., Srivastava, S. Stock prediction using deep learning. Multimed Tools Appl 76, 18569–18584 (2017).*
[12] *L. Xiong and Y. Lu, "Hybrid ARIMA-BPNN model for time series prediction of the Chinese stock market," 2017 3rd International Conference on Information Management (ICIM), Chengdu, China, 2017, pp. 93-97, doi: 10.1109/INFOMAN.2017.7950353.*
[13] *Jian-Zhou Wang, Ju-Jie Wang, Zhe-George Zhang, Shu-Po Guo, Forecasting stock indices with back propagation neural network, Expert Systems with Applications, Volume 38, Issue 11, 2011, Pages 14346-14355.*
[14] *Ping-Feng Pai, Chih-Sheng Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, Omega, Volume 33, Issue 6, 2005, Pages 497-505.*
[15] *Wei Huang, Yoshiteru Nakamori, Shou-Yang Wang, Forecasting stock market movement direction with support vector machine, Computers & Operations Research, Volume 32, Issue 10, 2005, Pages 2513-2522.*
[16] *Y. Lin, H. Guo and J. Hu, "An SVM-based approach for stock market trend prediction," The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 2013, pp. 1-7, doi: 10.1109/IJCNN.2013.6706743.*
[17] *G. Mahalakshmi, S. Sridevi and S. Rajaram, "A survey on forecasting of time series data," 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), Kovilpatti, India, 2016, pp. 1-8, doi: 10.1109/ICCTIDE.2016.7725358.*
[18] *Jing Shi, Jinmei Guo, Songtao Zheng, Evaluation of hybrid forecasting approaches for wind speed and power generation time series, Renewable and Sustainable Energy Reviews, Volume 16, Issue 5, 2012, Pages 3471-3480.*