ISSN 2616-5902 Vol. 3, Issue 10: 119-126, DOI: 10.25236/AJBM.2021.031020

The Empirical Analysis and Prediction of China's FinTech Index Based on GARCH Model and BP Neural Network Model

Ming Li[1*], Hefei Wang[2], Hua Fang[3]

[1&3]Business School, University of Shanghai For Science and Technology, Shanghai, 200093, China
[2]School of Public Management, Inner Mongolia University, Hohhot, Inner Mongolia, 010021, China
**Corresponding author: 1143399048@qq.com*

Abstract: *In recent years, the development of FinTech has moved from behind the scenes to the front, greatly promoting the development of innovation and high-quality supply in the financial sector. The development level of China's FinTech is generally leading in the world, and many listed FinTech companies have brought vitality to the financial market. But on the other hand, fintech also contributes to market volatility. Reasonable prediction of its trend and prevention of investment risks are still important issues. In this paper, CNI Xiangmi Lake FinTech Index, which can measure the overall performance of China's FinTech listed companies, is selected as the research object, and GARCH model and BP neural network model are used to conduct empirical analysis and forecast research on the index. The results show that the two time series models can fit the curve and the actual trend well. Compared with the prediction accuracy, the prediction effect of GARCH model is better than BP neural network model.*

Keywords: *FinTech; GARCH model; BP neural network model.; Forecast analysis*

1. Introduction

In recent years, financial technology (FinTech), relying on blockchain, big data, cloud computing and other emerging hot technologies, has moved from the background to the front, and has become an important engine for the development of financial innovation [1]. It is undeniable that fintech can create new business models, personalized and precise products and services, and promote high-quality supply in the financial sector. It can also reduce information asymmetry, reduce market transaction costs, improve the efficiency of resource allocation, and has a strong positive incentive effect on the development of traditional finance.

At present, fintech has become the main direction of China's financial supply-side structural reform and the commanding height of the development of the financial industry [2]. On the whole, China's fintech development level is relatively leading in the world. China has a high level of fintech application, a large number of listed fintech enterprises, and its market innovation and development vitality is constantly improving. However, digital financial infrastructure and fintech regulation need to be further improved [3]. Fintech will be a long-term focus area of the financial landscape for a long time.

But it should also be noted that fintech can also increase volatility and instability in financial markets. If investors ignore its risk for a long time, it is easy to fall into the trap of "boiling frog". Based on GARCH model and BP neural network, this paper conducts empirical analysis and forecast research on the fintech index of China's listed fintech companies, hoping to provide help for rational investment of investors and promote the healthy development of the fintech market.

2. Data selection and preprocessing

2.1 Data selection

The fintech index of listed companies selected in this paper is the CNI Xiangmi Lake FinTech Index (" Fintech "for short, stock code" 399699.SZ "). The index is based on May 26, 2017, with a base point of 3,000 points. High-quality stocks of fintech companies listed on shenzhen and Shanghai Stock Exchanges are selected as sample stocks to reflect the overall performance of fintech listed companies.

ISSN 2616-5902 Vol. 3, Issue 10: 119-126, DOI: 10.25236/AJBM.2021.031020

Eviews10, Rstudio and SPSS software are used to select the relevant data of the index from June 12, 2017 to November 12, 2021 for research. The data source is the official website of Straight flush.

2.2 Preliminary visualization and descriptive statistics of data

Based on the obtained data, the time series chart of the fintech index's daily closing price is drawn (as shown in Figure 1). It can be seen that the index has an outstanding performance in 2020, reaching 4,794.57 points. In 2018, the performance was poor, reaching the lowest point of 2,247.54. On the whole, the fluctuation range is large and it is a nonstationary time series.

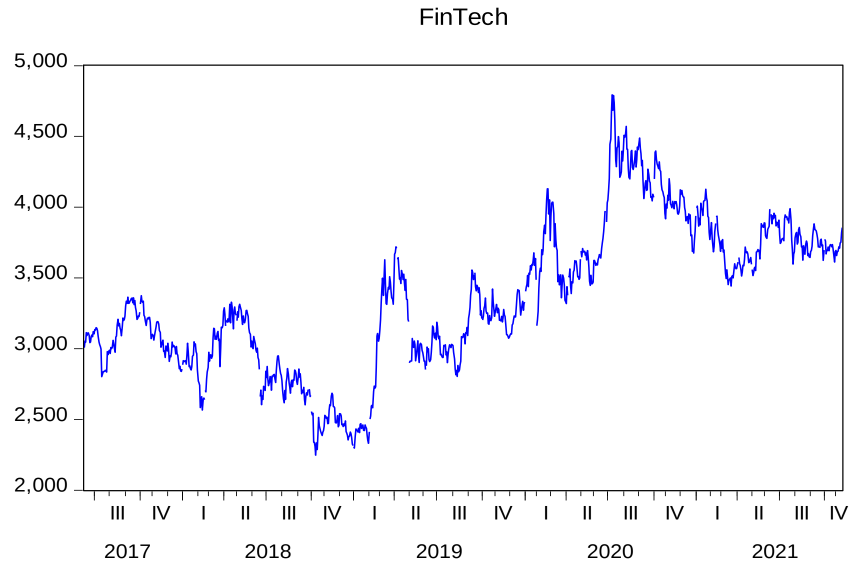


Figure 1 The time series chart of the fintech index's daily closing price

The following constructs the logarithmic return series of the fintech index's daily closing price R_t, that is:

$$R_t = ln(FinTech_t) - ln(FinTech_{t-1}) \tag{1}$$

t stands for time,$FinTech_t$ and $FinTech_{t-1}$ represents the closing price of the Fintech index at time t and the closing price at time t-1 respectively.

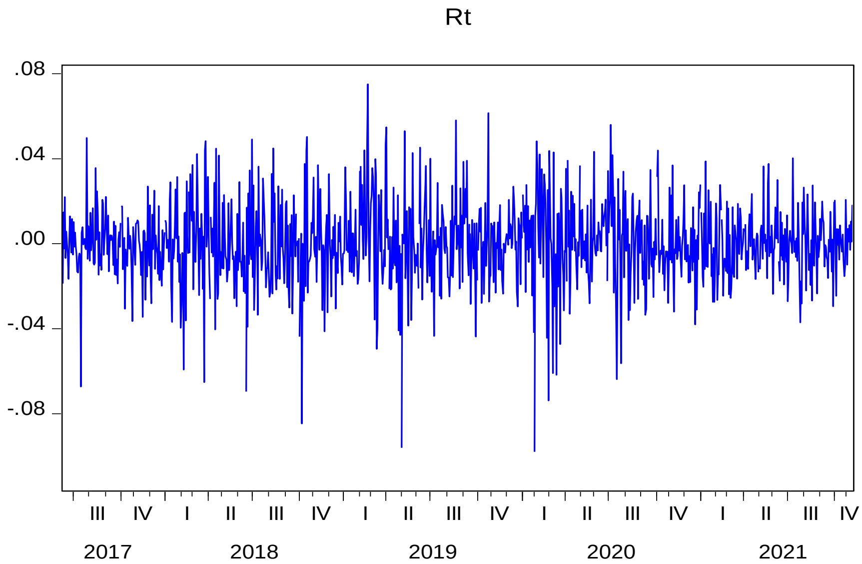


Figure 2 The time series chart of fintech index logarithmic returns

ISSN 2616-5902 Vol. 3, Issue 10: 119-126, DOI: 10.25236/AJBM.2021.031020

Table 1 Basic statistical characteristics of R_t logarithmic return series

N	Mean	Median	Maximum	Minimum	Std.Dev.	Skewness	Kurtosis	Jarque-Bera	p
1078	0.000213	0.000392	0.075069	0.097975	0.0197975	0.258239	5.205867	230.5393	0.0000

R_t logarithmic return series is shown in Figure 2, which shows that there is no obvious trend term and intercept term. Meanwhile, the volatility aggregation effect of the series is prominent. For example, the second quarter of 2019 and the first quarter of 2020 fluctuated greatly; In 2021, the fluctuation is small and heteroscedasticity is suspected. The relevant descriptive statistics are shown in Table 1. It can be seen from Table 1 that the skewness coefficient is -0.258239, which is less than 0 and belongs to the left-skewness distribution. When kurtosis value is greater than 3, it belongs to the thick tail distribution and does not obey the normal distribution, reflecting the data characteristics of "peak thick tail". The statistic under J-B test is 230.5393, which also rejects the null hypothesis of normal distribution at the significance level of 1%. Therefore, the student-t distribution and generalized error distribution (GED) can be used for fitting in subsequent modeling.

2.3 Data stationarity test and autocorrelation test

Since there is no obvious trend term and intercept term in R_t logarithmic return series, ADF test is carried out to verify the stability of data. The test results (see Table 2) shows that the statistics of ADF are less than the critical value and have passed the stationarity test at the significance levels of 10%, 5% and 1%, rejecting the null hypothesis of the existence of unit root. The R_t logarithmic return series can be considered to be stationary.

Table2 The ADF test of R_t logarithmic return series

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-32.90947	0.0000
Test critical values:	**1% level**	-2.567111	
	5% level	-1.941117	
	10% level	-1.616501	

The sequence autocorrelation and partial autocorrelation tests are continued (as shown in Figure 3). It can be seen from the figure that both the autocorrelation and partial autocorrelation coefficients of R_t sequence falls within two times of the estimated standard deviation, and it is preliminarily believed that there is no significant autocorrelation or partial autocorrelation in the sequence. Further Q test also supports this conclusion (see Table 3), p values corresponding to Q statistics are all greater than the confidence level of 0.05, so it is considered that there is no significant correlation in R_t sequence at the significance level of 5%.

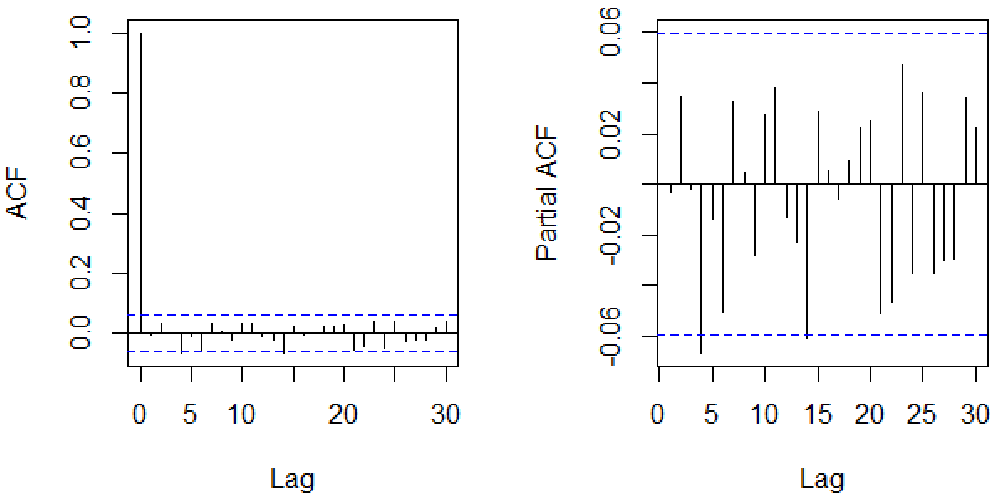


Figure 3 Graphs of autocorrelation and partial autocorrelation of fintech index logarithmic returns

ISSN 2616-5902 Vol. 3, Issue 10: 119-126, DOI: 10.25236/AJBM.2021.031020

Table 3 the Q-test of R_t logarithmic return series

	AC	PAC	Q-Stat	Prob		AC	PAC	Q-Stat	Prob
1	-0.003	-0.003	0.0101	0.920	**19**	0.020	0.022	20.290	0.377
2	0.035	0.035	1.2996	0.522	**20**	0.028	0.025	21.160	0.388
3	-0.002	-0.002	1.3049	0.728	**21**	-0.053	-0.051	24.209	0.283
4	-0.065	-0.067	5.9343	0.204	**22**	-0.042	-0.047	26.143	0.246
5	-0.013	-0.014	6.1238	0.294	**23**	0.041	0.047	28.038	0.214
6	-0.055	-0.050	9.3670	0.154	**24**	-0.051	-0.035	30.876	0.157
7	0.032	0.033	10.506	0.162	**25**	0.041	0.036	32.723	0.138
8	0.005	0.004	10.532	0.230	**26**	-0.027	-0.035	33.540	0.147
9	-0.023	-0.028	11.133	0.267	**27**	-0.023	-0.030	34.122	0.163
10	0.035	0.027	12.440	0.257	**28**	-0.025	-0.029	34.793	0.176
11	0.033	0.038	13.635	0.254	**29**	0.018	0.034	35.163	0.199
12	-0.010	-0.013	13.736	0.318	**30**	0.036	0.023	36.632	0.188
13	-0.020	-0.023	14.182	0.361	**31**	0.020	0.020	37.076	0.209
14	-0.064	-0.061	18.628	0.180	**32**	-0.021	-0.021	37.548	0.230
15	0.025	0.029	19.288	0.201	**33**	-0.044	-0.044	39.687	0.197
16	-0.004	0.006	19.302	0.253	**34**	0.010	0.015	39.805	0.227
17	-0.002	-0.006	19.309	0.311	**35**	0.006	0.009	39.851	0.263
18	0.022	0.009	19.853	0.341	**36**	0.012	0.006	40.015	0.296

3.The construction and prediction of GARCH model

3.1 The white noise mean equation

Since there is no significant autocorrelation in R_t sequence, we cannot establish ARMA mean model for prediction. However, it can be seen from Figure2 that the sequence fluctuation has obvious agglomeration effect and the sequence variance is not constant, so R_t sequence cannot be considered as a white noise process on the whole. However, the mean value equation can be set as white noise. First, set the equation as:

$$R_t = \mu_t + \varepsilon_t \tag{2}$$

Here, μ_t is the mean value changing with t, and ε_t is the residual change with t.

Since the white noise sequence also needs to meet the condition that the mean value is 0, the mean value is de-averaged here. Set the equation:

$$r_t = R_t - 0.000213 \approx \varepsilon_t \tag{3}$$

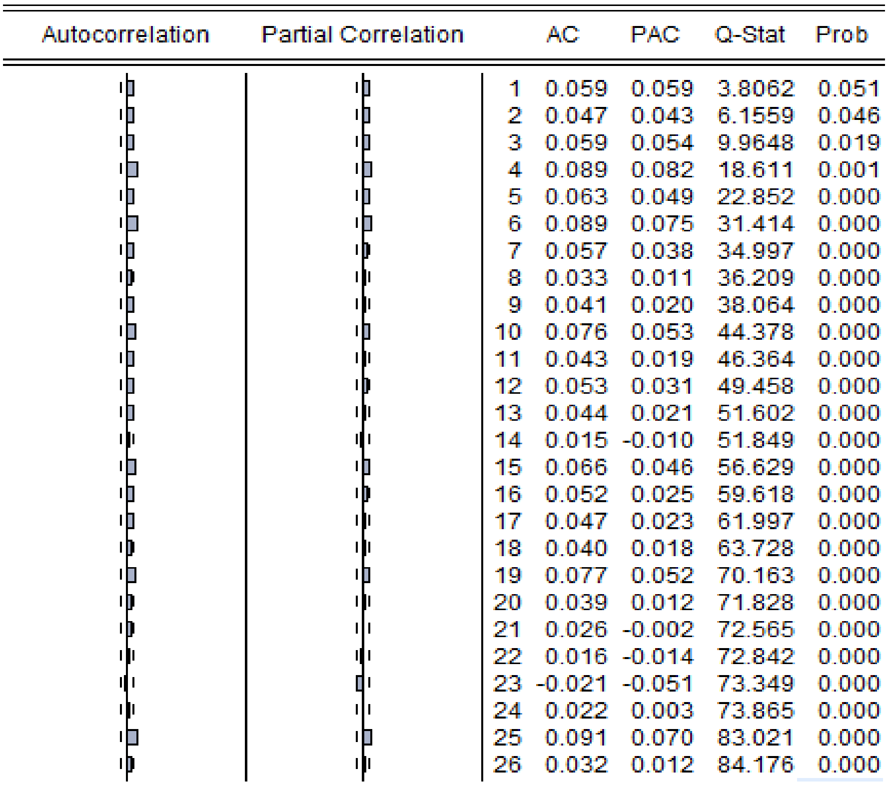
The R_t logarithmic return sequence is uniformly subtracted from its mean value 0.000213 (see Table 1). The sequence r_t is approximately equal to the residual ε_t. Equation (3) is the white noise mean equation.

3.2 The construction of GARCH model

ARCH effect test is carried out on the square term of residual of the white noise mean equation. LM test and residual square correlation graph test are commonly used. Since ARMA modeling is not conducted, the second method is adopted to draw residual square correlation graph (as shown in Figure 4). It can be seen that the Q statistic is relatively significant, and the P value is less than or closes to 0.05, which can be considered as the existence of sequence autocorrelation, so there is ARCH effect.

GARCH model is needed to eliminate ARCH effect. Data from June 12, 2017 to October 15, 2021 are selected as the training set to establish the GARCH model. In the study of financial time series data, the commonly used GARCH models are GARCH(1,1), GARCH(1,2), GARCH(2,1) and GARCH(2,2). At the same time, three distributions of disturbance terms are considered: normal distribution, t distribution and GED distribution. According to AIC, SC and HQ information criteria and whether ARCH effect still exists in the model after modeling, the optimal model is selected.

Academic Journal of Business & Management
ISSN 2616-5902 Vol. 3, Issue 10: 119-126, DOI: 10.25236/AJBM.2021.031020



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.059	0.059	3.8062	0.051
		2	0.047	0.043	6.1559	0.046
		3	0.059	0.054	9.9648	0.019
		4	0.089	0.082	18.611	0.001
		5	0.063	0.049	22.852	0.000
		6	0.089	0.075	31.414	0.000
		7	0.057	0.038	34.997	0.000
		8	0.033	0.011	36.209	0.000
		9	0.041	0.020	38.064	0.000
		10	0.076	0.053	44.378	0.000
		11	0.043	0.019	46.364	0.000
		12	0.053	0.031	49.458	0.000
		13	0.044	0.021	51.602	0.000
		14	0.015	-0.010	51.849	0.000
		15	0.066	0.046	56.629	0.000
		16	0.052	0.025	59.618	0.000
		17	0.047	0.023	61.997	0.000
		18	0.040	0.018	63.728	0.000
		19	0.077	0.052	70.163	0.000
		20	0.039	0.012	71.828	0.000
		21	0.026	-0.002	72.565	0.000
		22	0.016	-0.014	72.842	0.000
		23	-0.021	-0.051	73.349	0.000
		24	0.022	0.003	73.865	0.000
		25	0.091	0.070	83.021	0.000
		26	0.032	0.012	84.176	0.000

Figure 4 Residual square correlation diagram

Table 4 GARCH model information table

	GARCH	AIC	SC	HQ	Whether ARCH effect exists after modeling
Normal	GARCH(1,1)	-5.0861	**-5.2720**	-5.0808	NO(0.8331, 0.8329)
	GARCH(1,2)	-5.0844	-5.0656	-5.0773	NO(0.9277, 0.9276)
	GARCH(2,1)	-5.0842	-5.0655	-5.0771	NO(0.8927, 0.8926)
	GARCH(2,2)	-5.0823	-5.0589	-5.0734	NO(0.7843, 0.7841)
Student' t	GARCH(1,1)	-5.1413	-5.1225	**-5.1342**	NO(0.9396, 0.9395)
	GARCH(1,2)	**-5.1423**	-5.1188	-5.1334	NO(0.6215, 0.6211)
	GARCH(2,1)	-5.1415	-5.1181	-5.1326	NO(0.2108, 0.2105)
	GARCH(2,2)	-5.1409	-5.1128	-5.1303	NO(0.2523, 0.2519)
GED	GARCH(1,1)	-5.1393	-5.1206	-5.1322	NO(**0.9990, 0.9990**)
	GARCH(1,2)	-5.1390	-5.1155	-5.1301	NO(0.6410, 0.6406)
	GARCH(2,1)	-5.1586	-5.1355	-5.1499	NO(0.5514, 0.5509)
	GARCH(2,2)	-5.1371	-5.1089	-5.1264	NO(0.5772, 0.5768)

Note: the figures in brackets in the table are the p-values of the F-statistic of the first-order arch-LM test and the P-values of the Obs* R-Squared statistic respectively

As shown in Table 4, residual errors of all models can pass the ARCH-LM test after modeling, which can eliminate the ARCH effect. Among them, GARCH(1,1) with GED distribution rejects the ARCH effect hypothesis with a probability of 0.999. In terms of information criteria, the GARCH model with student-t distribution has smaller AIC and HQ values. In addition, EGARCH and TGARCH models can also be established on the basis of these models, but the effects are not ideal, so they are removed. After comprehensive comparison, the GARCH(1,1) model under t distribution and the GARCH(1,1) model under GED distribution are better. The variance equation structure of the two models is as follows:

$$\text{student-t distribution:} \quad \sigma_t{}^2 = 8.09 \times 10^{-6} + 0.0458\varepsilon_t{}^2 + 0.9347\sigma_{t-1}{}^2 \tag{4}$$

t value (1.7296) (3.1126) (41.8638)

p value (0.0837) (0.0019) (0.0000)

Logarithmic likelihood =2723.738 AIC=-5.1413 SC=-5.1225 HQ=-5.1342

GED distribution: $$\sigma_t^2 = 1.03\times10^{-5} + 0.0470\varepsilon_t^2 + 0.9261\sigma_{t-1}^2 \tag{5}$$

t value (1.8285) (2.9493) (34.2048)

p value (0.0675) (0.0032) (0.0000)

Logarithmic likelihood =2722.71 AIC=-5.1393 SC=-5.1206 HQ=-5.1322

The coefficients of ARCH term and GARCH term in the two equations under distribution are significant and greater than 0, and the sum of coefficients is 0.9805 and 0.9731 respectively, which are less than 1 and very close to 1, indicating that the impact of conditional variance is persistent and can effectively predict the future trend.

3.3 The application prediction of GARCH model

The following is a forecast of the closing price of the fintech index from October 18, 2021 to November 12, 2021 (20d) based on the two GRACH(1,1) models constructed. The prediction effects of the two models are shown in Figure 5 and Figure 6.

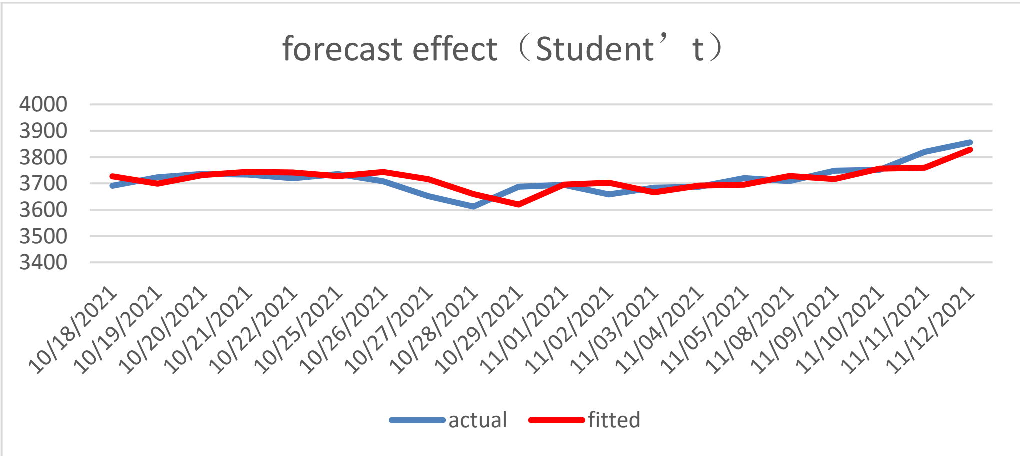


Figure 5 GARCH(1,1)-t distribution forecast effect

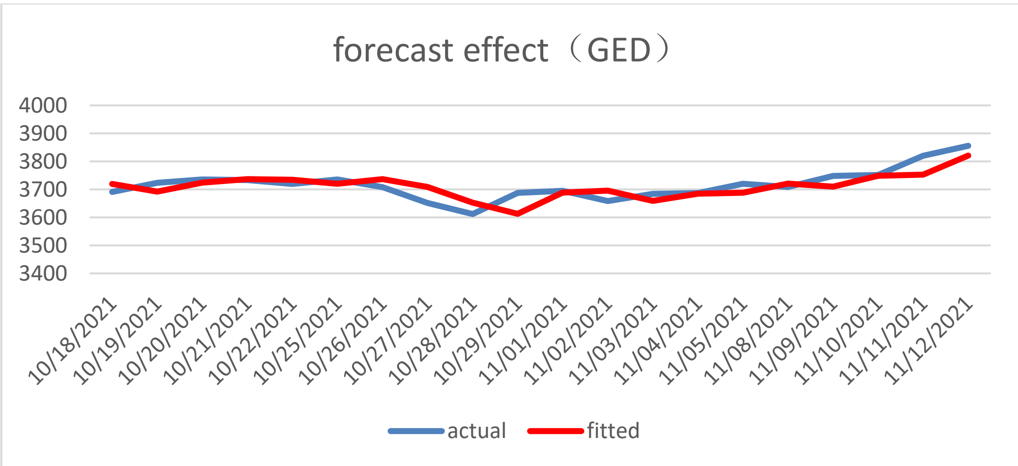


Figure 6 GARCH(1,1)-GED distribution forecast effect

It can be seen that both GARCH models with different distributions can better fit the actual closing trend of fintech index. The average relative error rate is 0.74871% and 0.75929%, respectively. It has good prediction effect. However, it should also be noted that there are still large errors in predicting valleys and peaks. Next, the BP neural network model is constructed for supplementary prediction.

ISSN 2616-5902 Vol. 3, Issue 10: 119-126, DOI: 10.25236/AJBM.2021.031020

4. The construction and prediction of BP neural network model

4.1 Data standardization processing

Before constructing BP neural network model, data standardization is needed. The standardized treatment formula is as follows:

$$\widehat{x_i} = \frac{x_i - minX_i}{maxX_i - minX_i} (i = 1,2...8) \tag{6}$$

x_i is the current actual value of each explanatory variable; $minX_i$ and $maxX_i$ are the minimum and maximum values of each explanatory variable respectively; $\widehat{x_i}$ is the actual value after standardization.

4.2 The application prediction of BP neural network model

BP neural network is a multilayer feedforward neural network trained by error backpropagation algorithm, which includes input layer, hidden layer and output layer. The input layer will input the opening price, low price, high price, close price, total volume and total transaction amount of the previous trading day of the fintech index as variables. Hidden layer Settings are limited to 1-50 layers. The output layer is the closing price of the day. 1058 data from June 12, 2017 are used as the training set, and the last 20 data are used as the test set. The gradient descent optimization algorithm is used, the hidden layer activation function is hyperbolic tangent function, and the output layer activation function is identity.

The model shows that the average relative error rate is 0.76271%. As shown in Figure 7, the BP neural network model can also predict the actual closing trend of fintech index well, but it also faces the problem of inaccurate prediction of peaks and valleys.

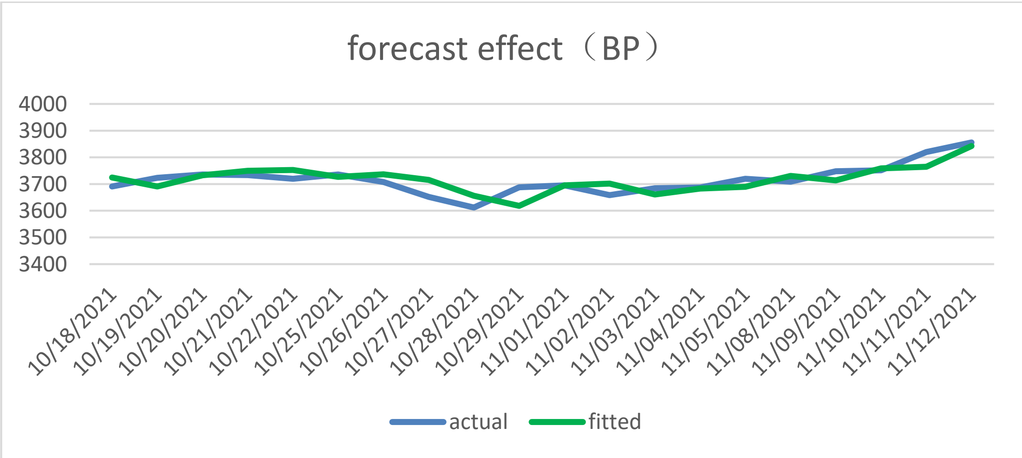


Figure 7 BP neural network forecast effect

5. The comparative analysis of GARCH(1,1) model and BP neural network model

The accuracy of the prediction results of the two models is analyzed in detail below. Common statistical error analysis indexes are introduced: mean error (ME), root mean square error (RMSE), mean absolute percentage error (MAPE) and maximum error(E_{max}).Which, $ME = \frac{1}{N}\sum_{i=1}^{N}|\widehat{y_i} - y_i|$,$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y_i} - y_i)^2}$,$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\widehat{y_i} - y_i}{y_i}\right| \times 100\%$,$E_{max} = max|\widehat{y_i} - y_i|$.$\widehat{y_i}$ is the predicted value of the closing price on the ith trading day, .y_i is the actual value of the closing price on the ith trading day, and N is the forecast number of 20 days.

The comparative analysis results of prediction are shown in table 5. By comparing the four indicators, it is found that the prediction accuracy of GARCH(1,1) model and BP neural network model is not significantly different. The GARCH(1,1) model based on student-t distribution is better, and the three indicators are all minimum, which reflects the strong stability and accuracy of the model.

ISSN 2616-5902 Vol. 3, Issue 10: 119-126, DOI: 10.25236/AJBM.2021.031020

Table 5 Comparison of prediction accuracy of test set(20d)

indicator	**GARCH(1,1)-t**	**GARCH(1,1)-GED**	**BP neural network**
ME	**27.7668**	28.2075	28.2645
RMSE	37.27004	34.77985	**34.32016**
MAPE	**0.7487%**	0.7593%	0.7627%
Emax	**67.918**	74.851	69.41

As for the shorter term prediction, the 5d test set can be used to continue the comparison. As shown in table 6, the four indicators of the GARCH(1,1) model based on GED distribution are all minimum, and the short-term prediction effect is the best.

Table6 Comparison of prediction accuracy of test set(5d)

indicator	**GARCH(1,1)-t**	**GARCH(1,1)-GED**	**BP neural network**
ME	19.36726	**17.9188**	23.512
RMSE	22.28845	**20.86285**	26.52248
MAPE	0.5217%	**0.4824%**	0.6329%
Emax	35.71978	**31.604**	33.31

Through comparison, it can be found that GARCH(1,1) model is better than BP neural network model in predicting the CNI Xiangmi Lake FinTech Index. In terms of the segmentation period, the GARCH(1,1) model based on the student- t distribution of students is better in the long-term prediction (20d), and the GARCH(1,1) model based on the GED distribution is better in the short-term prediction (5d).

6. Conclusion

In this paper, GARCH model and BP neural network model are constructed to carry out empirical analysis and prediction of the CNI Xiangmi Lake FinTech Index. It is found that the index is a "peak and thick tail" distribution, its mean equation can be considered as a white noise process, and its residual has ARCH effect. In terms of application prediction, the two models fit the index trend well, which can provide some reference value for the prediction of stock price performance of Chinese fintech listed companies. Through detailed precision comparison, the prediction effect of GARCH(1,1) model is better than BP neural network model. The GARCH(1,1) model based on student-t distribution is more accurate in medium and long-term prediction (20d), while the GARCH(1,1) model based on GED distribution is more accurate in short-term prediction (5d).

References

[1] Bian Wanli (2021-11-04). The New development pattern of sustainable financial Development and high-quality service driven by fintech innovation. 21st Century Business Herald,007.

[2] Xiong Bin & Zhao Chun.(2021). Risk spillover effect of fintech on financial institutions: Based on GARCH-EVT-Copula-Covar model. Technology and Industry (10),29-37. doi:CNKI:SUN:CYYK.0.2021-10-006.

[3] Xiao xiang, Ding Yangyang & Wang Sichun.(2021). International Comparative study on fintech development Index. Financial Theory and Practice (10),12-21. doi:

[4] Zhang Rumeng & Zhang Huamei.(2021). Comparative analysis of BP neural network and ARMA-GARCH model in stock forecasting. Science Journal of Normal Universities (05) 14-20. doi:CNKI:SUN:GLKX.0.2021-05-004.

[5] Wei Liang, Yaqing Li, Xishou Huang & Hongjiao Li.(2020). Prediction of O_3 concentration in Nanning city based on ARMA-GARCH model. Guangxi science (01), 91-97. doi:10.13656/j.cnki.gxkx.20200311.011.

[6] Song HUA & Hu Taowen.(2021). Risk measurement of interest rate market based on ARMA-GARCH model. Journal of Yibin University (07),39-45. doi:10.19504/j.cnki.issn1671-5365.2021.07.005.

[7] Chen Cang & Zhao Zhiqin. (2021). Study on the volatility of Shanghai and Shenzhen 300 index based on GARCH family model. National circulation economy (15),110-112. doi:10.16834/j.cnki.issn1009-5292.2021.15.035.