

Mining Factors Affecting Stock Prices from the Perspective of Asset Pricing Based on ANN-RBF Algorithm

Shaoyu Yang^{1,#}, Yuzhe Wang^{2,#}, Zejiang Ni^{3,*,#}

¹ Business School, Renmin University, 100872, Beijing, China

² School of Management and Economics, Beijing Institute of Technology, 102401, Beijing, China

³ School of international, Jiangxi Finance and Economic University, 330013, Jiangxi, China

*Corresponding author: lzj18228916968@163.com

Equally contributed to this works

Abstract. Pricing of assets through machine learning has been given more attention. This article attempts to study the factors affecting the stock value. In addition to the Fama French factor, this article selects the stocks in the A-share market and adds seven other factors affecting stock value to construct a stock pricing model. The sum of squares error (SSE) of the RBF neural network's prediction results was 0.4, and the relative error was 0.955. Among the 12 factors, the economic prosperity index (HJ), consumer expectations index (CEI), and an inflation index (CPI) were significantly crucial for the growth of the A-share market value. This study is conducive to exploring the factors affecting stock prices, helping investors and other stakeholders identify significant influencing factors, and making correct responses to changes in factors to obtain additional returns.

Keywords: Machine learning; ANN-RBF; Stock pricing models; A-share.

1. Introduction

In the current environment of the worsening global epidemic, the Chinese stock market has experienced several large pullbacks, and there is a high risk of volatility in the stock market. In the two sessions held in China in early March 2022, the government emphasized strengthening the policy measures to stabilize growth, which means that the long-term trend of an attractive Chinese stock market still exists. At the same time, the advent of the big data era has made it easier for people to get the data results it needs through machine learning. It has important theoretical and practical significance in such an era to materialize various risk factors and construct a stock pricing model with high explanatory strength through Machine Learning (ML). At present, there is much research in these fields.

The American scholar William Sharp et al. developed the CPAM asset pricing model based on the Markowitz model as early as 1964. Later, Fama et al. [1] found in 1993 that two factors, stock market value and book-to-market ratio, could explain the vast majority of stock price movements, and constructed the Fama-French Three-factor Model, which could explain the most of stock price movements in the context of the time. In later empirical studies, some stocks could not be well explained by the three-factor model. Therefore, Fama et al. [2] added investment and profitability factors to construct a better explanatory Fama-French Five-factor Model, and confirmed the validity of their model by using more than 50 years of market data in the United States. Xia et al. [3] conducted an empirical study on the effect of investment heat on stock excess return in 2019. They pointed out that the pricing model explained significantly stronger after adding the heat factor measured by the market capitalization growth rate. In addition, national macroeconomic indicators such as GDP, FDI, etc. [4] have also been studied as stock price factors in previous studies.

Meanwhile, machine learning has been widely applied in asset pricing, and many experts and scholars also tried to optimize pricing models through machine learning. Gourav et al. [5] constructed a hybrid evolutionary intelligence system and hybrid time series econometric model for stock price prediction in 2021. Zheng et al. [6] used Bat Optimization Algorithm for efficient stock price prediction. The application of neural networks for asset prices has also been frequently proposed.

Patel et al. [7] applied ANN, and made an empirical study on the stock price forecast of Indian stock market by inputting technical characteristic indicators. In 2017, Yan et al. [8] used Bayesian-Regularised-Artificial Neural Networks (BR-ANN), inputting historical price data, to reliably predict the future one-day closing price of the Shanghai Stock Exchange Index. Bai et al. [9] predicted the risk of multi-project resource conflicts based on artificial neural networks, and the final goodness-of-fit could reach 0.98, i.e., the goodness-of-fit was very high. The above studies play an important role in this paper, as they are crucial theoretical guidance for studying stock pricing.

However, the number of studies on China's A-share pricing models is insufficient. Most of them are applicability studies of previously proposed models, and the selection of impact factors is not comprehensive enough in terms of impact factor identification. In summary, the article proposes using the ANN-RNF method to confirm the magnitude of the influence of each macro factor on stock prices to construct a highly explanatory stock pricing model suitable for A-shares. The marginal contribution of the model is that investors and other stakeholders of listed companies can effectively identify which factors have a significant impact on stock prices, so that they could have a more accurate judgment on the trend of stock price changes when the influencing factors change, ultimately, to make reasonable decisions.

2. Identification of influencing factors

Table 1. Identification of influencing factors

Name of indicator	Indicator code	Reasons for selection and data sources	Reference
Market premium factor	X1	The difference between the monthly market return of cash dividend reinvestment and the monthly risk-free interest rate.	[10]
Size factor	X2	The difference between the monthly returns of small-cap stocks and large-cap stocks.	[10]
Book-to-market ratio factor	X3	The difference in monthly returns between a high book-to-market ratio portfolio and a low book-to-market ratio portfolio.	[11]
Profitability factor	X4	The difference in monthly returns between high-profit stock portfolios and low-profit portfolios.	[11]
Turnover rate factor	X5	Trading volume / total equity = turnover rate, using the average turnover rate of all stocks in the A-share market this month as the turnover rate index.	[11]
m2 year-on-year growth rate	X6	The year-on-year growth rate was last year.	[11]
Volatility factor	X7	Pairwise Value of Weighted Average of Market Value for Monthly Volatility of Stocks	[11]
CPI	X8	Monthly CPI data released by the National Bureau of Statistics	[12]
Economic prosperity index HJ	X9	Data published by the National Bureau of Statistics	[11], [12]
Consumer expectations index CEI	X10	Data published by the National Bureau of Statistics	[11], [12]
Total quantity consumed C_t	X11	The total retail sales of social consumer goods announced monthly by the National Bureau of Statistics	[13], [14]
Consumption growth speed ΔC_t	X12	Growth in consumption	[13], [14]

For China's A-share market, the Fama-French five-factor model is used for pricing, which has a strong explanatory power. Meanwhile, using the turnover factor to replace the investment factor can make the pricing model more explanatory [10]. Therefore, we also use the volume factor instead of the investment factor when constructing the model. In addition, to build a more explanatory model, we add the economic environment factor, volatility factor, and heat factor to the model. We selected variables including market premium factor (X1), market value factor (X2), book market value ratio factor (X3), profitability factor (X4), turnover factor (X5), money supply growth rate (m2), year-on-

year growth rate (X6), volatility factor (X7), CPI (inflation factor) (X8), economic prosperity index (X9), consumer expectations index (X10), total consumption (X11), consumption growth rate (X12) and foreign direct investment (X13). The specific reasons for selection and references are shown in table 1. Furthermore, in the following discussion, the coronavirus is considered a black swan event instead of a financial crisis. It is due to the fact that the Chinese stock market did not experience a significant plummet from 2020 to 2021. Even though the volatility increased, there was still a rising trend during the period because of the successful epidemic control and rapid resumption of production, making the economy back to life within a short time. Hence, there is no need to analyze the Chinese stock market separately according to the pre-Covid19 and Covid19 eras.

3. ANN Method Introduction

In 1985 Powell proposed a radial basis function method for multivariate interpolation, a real value function whose value depends only on the distance from the origin. By 1988, Moody and Darken presented a new neural network, namely RBF neural network. RBF neural network belongs to the forwarding neural network, and its structure is similar to a multi-layer forward network. It is a three-layer feed-forward neural network composed of the input layer, hidden layer, and output layer, as shown in Figure 1. The input layer is mainly composed of signal source nodes. The number of hidden units in the second hidden layer is determined by the problem described. The transformation function of hidden units is the RBF radial basis function, a radially symmetric and attenuated non-negative nonlinear function for the center point. The third input layer responds to the effect of the output mode. The transformation from the input space to the hidden layer space is nonlinear, and the transformation from the hidden layer space to the output layer space is linear. Compared with the traditional BP neural network, the structure of the RBF neural network is simpler and has local solid approximation characteristics. It can approximate any continuous function with arbitrary accuracy. The approximation ability and convergence speed are generally better than BP neural network.

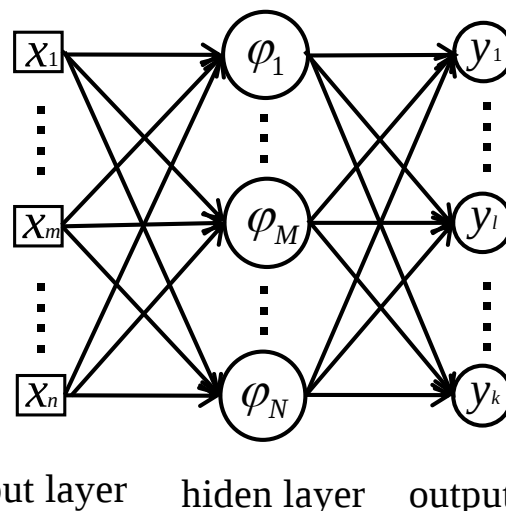


Figure 1. Block diagram of traditional RBF neural network

$$f(x) = \sum_{i=1}^n w_i \varphi_i(\|x - c_i\|) \quad (1)$$

Suppose the input vector is $x = [x_1 \cdots x_m \cdots x_n]$, the number of hidden layer nodes is N, the output of the RBF network can be expressed as : $(1) w_i (i = 1, 2, \cdots, n)$: represents the weight between the first hidden layer node and the output layer: $\varphi_i(\|x - c_i\|)$ is the hidden layer activation function.

$$\varphi_i(\|x - c_i\|) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right) \quad (2)$$

In formula (2): x is an n -dimensional input term, σ_i^2 represents the normalizing constant of the width of the i center point of the basis function, c_i represents the center value of the i basis function, and $\|x_i - c_i\|$ represents the Euclidean distance from the sample point x to the hidden layer center. The learning speed and diagnostic accuracy of the parameter c_i , σ_i , w_i , RBF neural network, have a great influence. To improve the accuracy of the model as much as possible, the error function E is defined as the mean square error, and the goal is to minimize the error function:

$$E = \frac{1}{2m} \sum_{i=1}^m e_i^2 = \frac{1}{2m} \sum_{i=1}^m (f(x) - y)^2 \quad (3)$$

$$\sum_{i=1}^m (f(x) - y)^2 = \sum_{i=1}^m \left(\sum_{j=1}^q w_j \varphi(x, c_j) - y \right)^2 \quad (4)$$

4. Case Study

4.1 Data Analysis & Collection

Connor (1995) describes three different types of factor models: macroeconomic, fundamental, and statistical factor models. Each model mostly only uses the variables that the name implies. However, in this case, different types of variables are integrated as features in the understanding of machine learning to shed light on how important they can be for the dependent variable. Specifically, the dependent variable is the monthly market value growth rate of A-share in both the Shanghai and Shenzhen Stock Exchange, and the independent variables include macroeconomic features (X6 and X8 to X12), fundamental features (X1 to X4), and statistical features (X5 and X7). The data are collected from the China Stock Market & Accounting Research Database and the Chinese National Bureau of Statistics of China, and the time span is from March 2018 to March 2021. As the pre-coronavirus and coronavirus era are included, the data are more diverse (for example, the fluctuation range is larger), so the importance of each factor could be better revealed.

As Figure.2 shows, 12 selected factors are used as the input layer of the RBF neural network, and the monthly market value growth rate is used as the output layer to explore the influence of selected factors on the monthly market value growth rate A shares and their respective importance. The training and test sets account for 76.7 % and 23.3 % of the sample. SPSS software is used to obtain the neural network results of a single hidden layer composed of four neurons. The softmax function classifies the hidden layer, and the activation function of the output layer is an identity that does not change the output. The neural network is shown in Fig. 1. At this time, the prediction results' square sum error (SSE) is 0.4, and the relative error is 0.955. The error of the test set is low enough, and the model confidence is high.

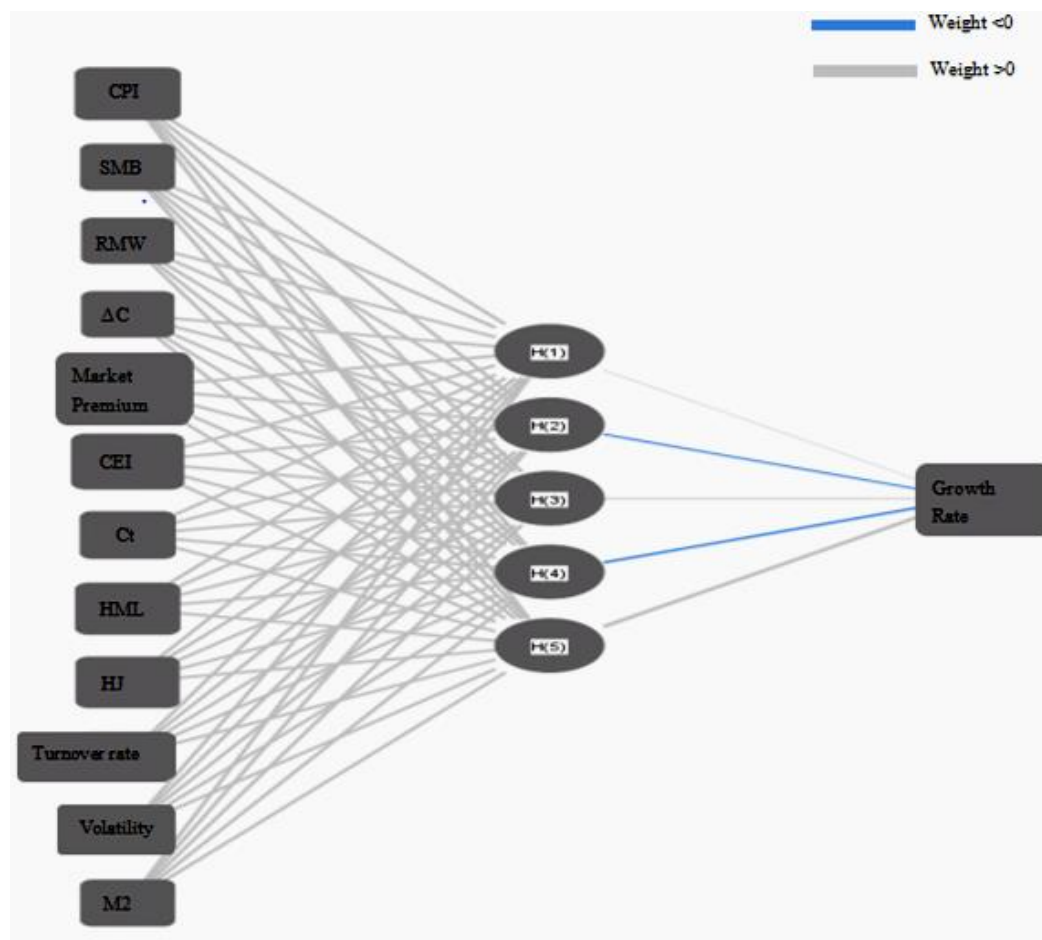


Figure 2. The Synapses weights of each layer

4.2 Result and Suggestion

Based on the training results, after calculating the neural network's weights, each index's influence on the monthly growth rate of A stock market value is quantified to illustrate the importance and normalized importance of independent variables in Figure 3. The results of the RBF model show that the market premium has the most significant impact, with 100% normalized importance to the growth rate. Nevertheless, the other fundamental features are less robust, especially the size factor (SMB) and profitability factor (RWM) are both less than 15%. SMB, RWM, and HML are considered substitutions of fundamental features according to Fama & French (2014). In addition, this result shows this assumption could be biased owing to the profound significance of market premium, which can explain the most variation in growth rate, leading to an implication that those substitutions might not be representative and irreplaceable. On the other hand, the difference between RBF and the linear model could also change the result, so further study is necessary.

As for macro factors, the economic prosperity index (HJ), consumer expectation index (CEI), and an inflation index (CPI) have a noticeable impact on the growth of the A-share market value with 36.7%, 57.6%, and 41.7% normalized importance respectively. Meanwhile, another macroeconomic factor (ΔC , C_t , and M2) seems less important in terms of predicting market growth. This further verifies Jiang Fuwei's research on the economic prosperity index and consumer expectation index, which can be used as a reliable macro prediction index for future research on market value growth. Although some scholars, such as Gultekin (1983), indicated that the stock market lacked a reliable correlation with inflation, the experiment in this model showed that the A-share market might be sensitive to inflation. This could be owing to the amplification of the correlation between inflation and stock market growth caused by China's easy monetary and fiscal policies from 2020 to 2021 during the coronavirus period. However, the current result illustrates, according to the arbitrage pricing theory (APT) of Ross (1976), that the Chinese stock market could provide a positive premium

in inflation. By comparison, the statistical factors (volatility and turnover rate) applied in this model could not well reflect the variation of market growth, both of whose normalized importance is no more than 20%.

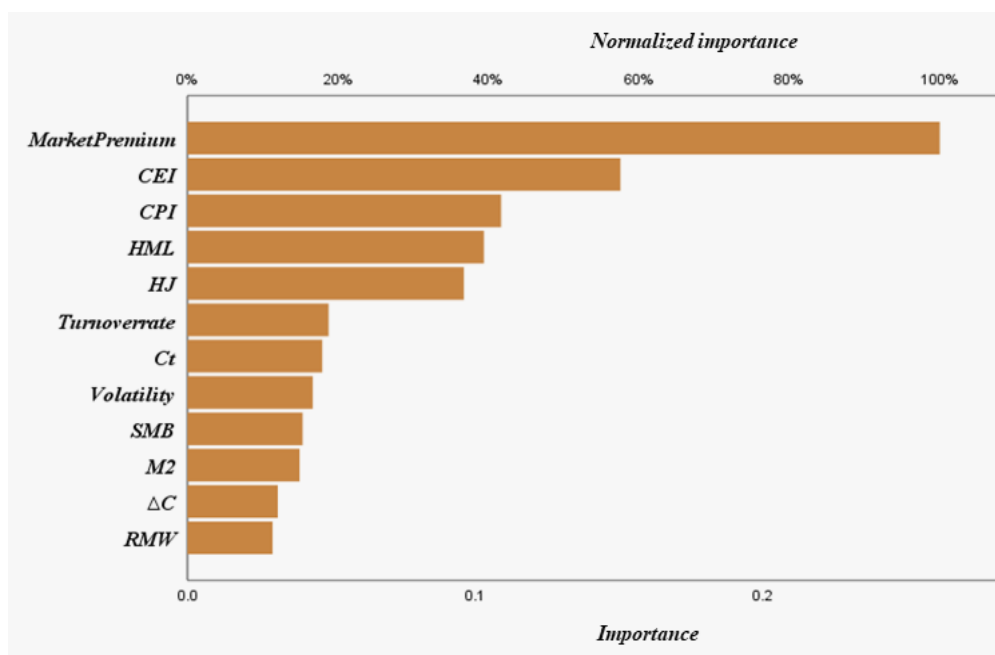


Figure 3. The importance and normalized importance of each factor

5. Conclusion

1) RBF neural network is a kind of feed-forward neural network with excellent performance and high efficiency, with better generalization ability and higher accuracy than the BP neural network. Based on the advantages of the RBF neural network, this paper innovatively combines RBF neural network with the CAPM asset pricing model, effectively improves the accuracy of the model, and can predict the asset price of China's A-share market more accurately.

2) This paper selects 12 macro factors that impact China's A-share stock market and finds that among the 12 selected factors, the market premium factor plays the most important role in asset price prediction. In contrast, the profit factor (RMW) has little effect. The square sum error (SSE) of the prediction results of the model is 0.4, and the relative error is 0.955, which has high confidence and provides a foundation for the research of asset valuation, capital cost budget, and resource allocation.

3) This paper only studies the macro factors for data processing and neural network modeling analysis and has achieved good results, but there is still a problem that the research scope is not comprehensive enough. Future research focuses on establishing a more optimized model and algorithm to solve the asset pricing model prediction data under the combined action of macro and micro factors.

Acknowledgments

The author's name is in alphabetical order, and the workload of each author is equivalent.

References

- [1] EUGENE, F., FAMA, KENNETH, R., & FRENCH. (1992). The cross-section of expected stock returns. The Journal of Finance, 47(2), 427-465.
- [2] Fama, E. F., & French, K. R.. (2014). A five-factor asset pricing model. Journal of Financial Economics, 116(1).

- [3] Xia, S.L., & Kang, M.. (2019). A study on the impact of investment heat and financial leverage on stock excess return--an empirical analysis based on a multi-factor asset pricing model for the a-share market. *Price Theory and Practice* (12), 5.
- [4] Verma, R. K., & Bansal, R.. (2021). Impact of macroeconomic variables on the performance of stock exchange: a systematic review. *International Journal of Emerging Markets*, 16(7), 1291-1329.
- [5] Kumar, G., Singh, U. P., & Jain, S.. (2021). Hybrid evolutionary intelligent system and hybrid time series econometric model for stock price forecasting. *International Journal of Intelligent Systems*.
- [6] Chen, H.. (2021). The stock index prediction is based on the SVR model with the bat optimization algorithm. *Algorithms*, 14.
- [7] Patel, J., Shah, S., Thakkar, P., & Kotecha, K.. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259-268.
- [8] Yan, Dingqi, Zhou, Qi, Wang, & Jianzhou, et al. (2017). Bayesianregularisation neural network based on artificial intelligence optimization. *International Journal of Production Research*.
- [9] Bai, L., Wang, Z., Wang, H., Huang, N., & Shi, H.. (2020). Prediction of multi-project resource conflict risk via an artificial neural network. *Engineering Construction & Architectural Management*, ahead-of-print(ahead-of-print).
- [10] Li Z-Bing, Yang G-Y, Feng Y-C, & Jing L. (2017). An empirical test of the Fama-french five-factor model in the Chinese stock market. *Financial Research* (6), 16.
- [11] Zhang, X.Y., & Wang, Z.S.. (2022). A test of the applicability of the Fama-french five-factor model in a-share market. *China Price* (1), 4.
- [12] Deng, Kebin, Guan, Zihuan, & Chen, Bin. (2018). Macroeconomic policies and systematic risk in the stock market - A proposal and test of hybrid macro and micro beta estimation method. *Economic Research*, 53(8), 16.
- [13] Jiang Fuwei, Ma Tian, & Zhang Hongwei. High-risk low return? An explanation of the dynamic CAPM model based on machine learning. *Journal of Management Science*.
- [14] Xu Yue, Qiu Zhigang, & Wang Hefei. (2019). The impact of macroeconomic states on asset pricing factors. *Economic Theory and Economic Management* (9).
- [15] Connor, G.. (1995)" The Three Types of Factor Models: A Comparison of Their Explanatory Power. *Financial Analysts Journal* 51(3) 42–57.
- [16] Fama, E. F., & French, K. R.. (2014). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1).
- [17] Ross, S. A.. (2015). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341-360.
- [18] Gultekin, N.B.. (1983). Stock market returns and inflation: evidence from other countries. *Journal of Finance*, 38(1), 49-65.