Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230714

The measurement of expected return on assets based on pricing models and machine learning methods

Qingxin Dong

College of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, 250014, China

carlos.paquin@sus.edu

Abstract. Equilibrium in the pricing of capital assets is the basis of modern finance research. In recent decades, as portfolios have become more diversified and information transparency has increased, investors have become increasingly concerned about the expected return on equity assets. The pharmaceutical manufacturing industry is crucial for a country with a considerable market and rigid demand for pharmaceutical consumption while affecting the country's people's health and economic development. Currently, China's pharmaceutical level is among the top pharmaceutical countries in the world, and the system of the pharmaceutical industry and pharmaceutical circulation policy have been quite complete. Hence, it uses pricing models with machine learning methods, including the CAPM model and the Fama-French three-factor model, to measure the expected return on assets of several large listed pharmaceutical companies in China, then using the "Transformer model" based on the machine learning algorithm to predict the expected return and comparing with the well-known pricing models.

Keywords: machine learning, pricing models, neural networks.

1. Introduction

The Chinese securities market has been developed relatively short compared to the securities markets in Europe and the United States, and ordinary stockholders do not have a deep understanding of investment concepts. Most ordinary investors have not mastered the scientific methods of assessing company value and investment risk. The analysis of stock investment value is only at a primary level, while investors' attention to the expected return on equity assets is increasing. Return on assets is an essential indicator for judging investors' investment behavior and is used in other investment programs such as corporate investment and financing and institutional investment. Pharmaceutical manufacturing is a relatively important industry in China's industrial system. In recent years, China's aging population has been increasing, the incidence of chronic diseases and complications is gradually increasing, the pressure of life of young people is increasing, and most people are in a sub-healthy state of life. With the national publicity on the importance of health care, people's awareness of health care is increasing, coupled with medical consumption is a rigid demand, so the investment value of medical stocks is becoming more and more apparent, investors prefer the pharmaceutical sector of the A-share market, so it chooses several large Chinese listed pharmaceutical companies as the research object, using a pricing model based on machine learning methods to measure their expected return research.

© 2023 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/6/20230714

In the securities market, the purpose of investors is to obtain the maximum return with the minimum risk. The pharmaceutical manufacturing industry is also essential for people's well-being, and the pharmaceutical industry's output and stock market value is growing. Under the premise of the continuous development of the pharmaceutical manufacturing industry, the correct estimation of the expected return on assets of the pharmaceutical industry is beneficial for investors to judge whether the listed pharmaceutical enterprises in China can continue to maintain a good development trend and for investors to judge what risks to take according to their ability and for pharmaceutical enterprises to choose the appropriate financing according to their situation.

2. Feasibility analysis

2.1. Basic characteristics of Chinese pharmaceutical industry

As people's standard of living and quality of life is improving, the awareness of healthy living is also improving, but most young people stay up late. Other irregular work and rest, the same to the health of the great hidden danger, and disease can adversely affect health, so people's demand for medicine is rigid, less affected by the macroeconomic, medical demand will not change too much. Medical demand and the population are proportional to each other. China has a large population base, but the aging of the population has increased in recent years, and medical demand is growing. Plus, this year's COVID-19 epidemic has a greater demand for medical treatment and prevention of the COVID-19 epidemic. Overall, the demand for pharmaceuticals and medical devices in China is growing.

The pharmaceutical industry is technically demanding and risky, requiring a significant investment in research and development, a technology-capital-intensive industry with high requirements for new entrants. In addition, medicine is closely related to human health, the quality of medicine is the object of strict national supervision, the industry and technology need to be constantly developed and improved, and pharmaceutical companies who want to get high returns to needing to pay much investment in the research and development of new products, including human and material resources and financial resources, for complex or chronic diseases, the success of the research and development and testing of effective drugs and will promote the development of the entire enterprise, the stock price rose, gain, However, the degree of exclusivity of the new drug for the market will decline over time, as products with similar efficacy enter the market, the company's share price will decline, and at the same time the market may appear more cost-effective drugs. Side effects and government regulation will also impact the new drug, all of which will impact the company's profits. The process of research and development of new drugs also requires significant investment in human and material resources to develop new production technologies and drugs to improve competitiveness in the market. In contrast, small companies that fail in research and development will be forced to withdraw from the competition, and market concentration will continue to increase.

This characteristic of the pharmaceutical industry leads investors to have different expectations of return on assets. This paper will analyze the factors that affect the E(R) of the pharmaceutical industry and conduct research and analysis through the pricing models.

2.2. Feasibility analysis

The previous section analyzed the characteristics of China's pharmaceutical industry. It concluded that the returns of China's listed pharmaceutical enterprises are affected by various aspects such as policies and R&D. The enterprises have more sub-industries with extensive internal differentiation, effectively dispersing the non-systematic factors. It is clear from previous studies that the impact of macro factors on China's pharmaceutical industry is not significant, which is also in line with the current situation of China's pharmaceutical industry. When considering the impact of macro factors on the E(R) of China's listed pharmaceutical enterprises, the stock and securities market risk factors are strongly related to it. The market risk alone cannot fully measure the main risks faced by China's listed pharmaceutical enterprises. If only the CAPM model is used for analysis, it cannot comprehensively portray the main risk factors faced by Chinese pharmaceutical companies. So the CAPM and multi-factor models with

DOI: 10.54254/2755-2721/6/20230714

machine learning methods will be used, and the Fama-French model will be optimized according to the actual situation.

3. Related works

History average model. Duarte, Rosa used historical averaging to find the E(R) of the U.S. market over the previous 53 years (1960-2013) of 10.5% by investigating the information of previous S&P500 returns [1]. Risk-reward framework. Damodaran, Duarte, and Rosa measured asset returns on the U.S. stock market for different time periods [2,3]. The risk-reward model is more operational. The risk-return model has attracted many scholars to empirically study and improve the model from several aspects since its creation in 1964, and its applicability is widely recognized. Research on the Chinese securities market. The Chinese securities market is late in development compared to the European and American securities markets, and Chinese scholars started their research on this issue late. However, it also allows Chinese scholars to gain more research experience from European and American scholars' research. After the classification of measurement models by European and American scholars was determined, Chinese scholars Yaxin Li and Yonggang Li sorted out and analyzed the measurement models, respectively, and used the historical average model to measure the expected return of specific assets in different periods to study [4,5]. However, because of its intense subjectivity, the assumptions do not match the actual conditions, making the measurement results have errors. Based on the actual situation of China's stock market, Jinyan Song compared the studies of European and American scholars, added three explanatory factors, PE, dividend/share price ratio, and market price ratio, and used the extended model multi-factor model of CAPM to calculate E(R) of 11.89% for China from 1995 to 2009 [6]. Yang Cao found that the stock liquidity of pharmaceutical-listed companies is primarily influenced by R&D investment, and stock liquidity also impacts expected stock returns [7]. However, the disclosure of R&D expenses is not mandatory in the country, and only a few companies voluntarily expose it, so the sample selection leads to the non-universality of the results. Kai Liang used statistical models to analyze the stock returns of China's listed pharmaceutical companies for the period 2009-2015 [8]. It found that macro factors have a particular influence on the return on assets of China's listed pharmaceutical companies. Macro factors such as GDP and CPI can have a significant influence on the return on assets of China's listed pharmaceutical companies.

4. Methods

4.1. CAPM model and its extensions

Mean-variance model. "H.M. Markowitz" proposed the mean-variance model in 1952 to measure the risk of a portfolio, in which risk is defined by the volatility of the expected return [9]. In this way, the analysis and judgment of the portfolios are completed in a mathematical-statistical way, and the results are more convincing. H.M. Markowitz's article "Portfolio Selection" proposes a risk metric model in which risk is defined by the volatility of the expected return. In this way, the judgment and analysis of the portfolio are completed mathematically and statistically so that the best equilibrium effect of the mutually constrained objectives is achieved. H.M. Markowitz made the following assumptions: 1) Investors with the same return. 2) returns are assumed to follow a normal distribution. 3) there are no transaction costs. 4) investors' decisions are based only on security returns and risks. A mean-variance mathematical model is developed:

$$\boldsymbol{min\,\sigma^2 \Sigma x_i x_j\, cov\left(R_i - R_j\right)} \tag{1}$$

Equation (1) is the objective function

$$\boldsymbol{R_P = \Sigma x_i R_i} \tag{2}$$

The limitation is:

$$\boldsymbol{1 = \Sigma \chi_i} \tag{3}$$

DOI: 10.54254/2755-2721/6/20230714

Short selling is allowed for equation (3). In the above equation, Rp denotes the portfolio return, Ri denotes the return of the ith stock, $x_i x_j$ denotes the investment ratio of the two securities, σ^2 (R_P) denotes the total risk of the portfolio, and $cov(R_i - R_j)$ denotes the covariance, which can be solved by the Lagrangian objective function for the minimum value of the portfolio risk. The minimum set of covariances can be solved by calculating the set of minimum variances that can be achieved by minimizing the investment risk and the expected return of investors.

In practice, the number of asset portfolios increases, and the calculation of the mean-variance model becomes very large. The model only considers the case of risky assets and does not consider the impact of risk-free assets on the calculation. In 1964, American scholars John Lintner, William Sharpe and others proposed the famous Capital Asset Pricing Model (CAPM model).

Capital Asset Pricing Model. The CAPM model assumes that the expected return on an asset with investment risk is positively correlated with the unit of measurement of that asset's risk (the β of the stock), which can be specified as follows:

$$r_P = r_f + \beta(r_m - r_f) \tag{4}$$

In formula (4), r_P is the expected rate of return for an individual stock or portfolio. r_f represents Risk-free return. r_m is stock market expected return. β risk index to measure the price volatility of individual stocks relative to the overall stock market. $\beta = Cov(r_p, r_m)/\sigma_m^2$ And the σ_m^2 means the variance of market returns. $\beta = 1$ means that the risk profile of the asset is the same as the risk profile of the market portfolio due to the year-over-year change in the average risk-return of the asset and the market portfolio $\beta > 1$ means that the return on the asset is higher than the average risk-return of the market portfolio and the risk profile is greater than the risk profile of the market portfolio. $\beta < 1$ means that the return on the asset is lower than the average risky return of the market portfolio, and the risk profile is less than the risk of the market portfolio. The CAPM model helps investors to determine the value of capital assets, i.e., the linear relationship between expected return and market risk (systematic risk) at market equilibrium.

Fama-French three-factor model. In the "Fama-French three-factor model [10]", the excess return on a portfolio can be explained by its exposure to three factors: the market asset mix $(r_m - r_f)$, the market capitalization factor (SMB), and the book-to-market ratio factor (HML).

$$R_{it} - R_{f_t} = \beta_i(R_{mt} - R_{ft}) + s_i SMB_t + h_i(HML_t) \tag{5}$$

In regression model (5), R_{it} represents the return rate of assets at time t. R_{f_t}: the risk-free rate of return at time t. market rate of return at time t. R_{mt} is market rate of return at time t. SMB_t: simulated portfolio return on market capitalization factor at time t. HML_t : simulated portfolio return of the book-to-market ratio factor at time t, the difference between the average returns of under-valued and over-valued firms.

4.2. Machine learning methods

The "Recurrent Neural Network (RNN)" and "Long short-term memory (LSTM)" can be used to predict the expected return on stock or portfolio during a specific period [11, 12]. For example, those methods can use 200 days of daily quote data to predict the value of the following seven days. However, RNN and LSTM algorithms can only compute sequentially from left to right or from right to left. This mechanism poses the following problems: 1) The computation of time slice t relies on the computation results now t-1, which limits the parallelism capability of the model. 2) Although the structure of gate mechanisms like LSTM alleviates the problem of long-term dependence to some extent, LSTM is still powerless for long-term dependence. So, the "Transformer model" has been developed, including many optimization methods [13, 14]. A practical method is suggested, which is the "self-attention mechanism." It can optimize the distance between any two positions in the sequence as a fixed value of "1" and solve the problem that the RNN algorithm processing long sequences will cause missing information.

DOI: 10.54254/2755-2721/6/20230714

Moreover, the "multi-head attention layer" can make multi groups be computed in parallel, and second, different groups can capture information from different subspaces. Since self-attention has no circular structure, Position Embedding (PE) has been proposed and was used to represent the relative or absolute position relationships of elements in the sequence.

5. Specific analysis steps

Research data from the Shanghai Stock Exchange selected a few more well-known companies, such as Harbin Pharmaceutical, Tongrentang, etc.

Specific steps of pricing models. For the CAPM model, the American economist "Michael Jensen" introduced the α parameter. The CAPM model after the introduction of the α parameter is as follows [15]:

$$R_{it} - R_{f_t} = \alpha_i + \beta\left(R_{mt}/R_{ft}\right) + \epsilon_{it} \tag{6}$$

The subscript "it" indicates that this is serial data, here α indicates that the return is higher than the part of the securities market, the trend is significantly stronger than the securities stock, α value will be greater than 0, if the trend is not as strong as the securities stock, then α will be less than 0. Here if the target return rate has been determined, when α is greater, then β is smaller, that indicates the risk could be smaller.

Specific steps: 1. Assuming a risk-free fixed return of 3.2%, calculate the average daily risk-free return. 2.Then calculate the respective risk premium reward of the selected stock and the CSI 300 index.3. Determine whether there is a significant linear relationship between the selected stock and the CSI 300 index. 4.Use the CAPM model with the introduction of α parameter for modeling calculation.

For the Fama-French model, introducing market asset mix $\left(r_m - r_f\right)$, market capitalization factor (SMB), and book-to-market ratio factor (HML) in the three-factor model. 1.According to the circulating market value, the stocks are divided into large market value and small market value stocks with a ratio of 1:1.2. According to the data, the stocks are divided into 3:4:3 high, medium and low groups respectively, and six portfolios can be obtained by combining the first step. 3. Use the market capitalization weighted average to find the return of each group. 4. find the SMBt (market capitalization portfolio return series) and HMLt (portfolio return rate series)

The x1, x2 and x3 in the following results correspond to the "market risk premium factor", "market capitalization factor" and "book-to-market ratio factor", respectively. Excess return is α.

Specific steps of machine learning methods. The generated sine and cosine data with different frequencies are added as position codes to the input sequence for position encoding (PE). Instead of using the original Encoder-Decoder architecture, the Decoder is replaced with a fully connected layer for outputting the predicted values, where the mask will be used to avoid introducing future information. The first 70% of the data is used for model training, and the remaining 30% for model testing. By using Mean Squared Error (MSE) as the loss function, "adaptive moment estimation" (adam) as the optimizer, which is better than confessional stochastic gradient descent, and the scheduler that sets the learning rate, the final run is 300 epochs to bring the loss down to a stable and low level [14, 15].

6. Results and analysis

Results of two pricing models. Firstly, the least squares method is chosen to conduct a linear regression analysis of the expected return on assets. Since the regression analysis is first conducted on a single factor (market risk), no autocorrelation test is required, and the empirical results of the CAPM model using the introduction of the alpha factor are shown below. From the CAPM model's results, it can be obtained that α for "xiuzheng Pharmaceutical" is almost 0 (i.e. -0.0005), β = 0.7735 and R^2 is 0.159; From the results of CAPM, it can be obtained that α for "Beijing Tongrentang" Pharmaceuticals is almost 0 (i.e. -0.0005), β = 0.6766 and R^2 is 0.254; α for "HPGC" Pharmaceuticals is still almost 0 (i.e. -0.0009), β = 0.635 and R^2 is 0.114.

DOI: 10.54254/2755-2721/6/20230714

Analysis of two pricing models results. The analysis shows that the value of the regression result R-squared is around 0.159, which indicates that the linear fitting of the stock is average and the market risk alone does not explain the expected return on assets well. The performance of HPGC in recent years is worse than one year, the intrinsic value is shrinking, there is no long-term investment value in the next two years, and the performance of the return is obviously lower than the market return, which is basically consistent with the actual situation.

Based on the regression results, it believes that only relying on the market risk factor and the explanation of expected return on assets is not good enough, the following is the analysis of Fama-French three-factor model for those pharmaceutical companies return rate. The R^2 values in all three results are not high and the fit of the regression model is not good, indicating that the the $Rm - R_F$, SMB , HML factors are not significant for explaining the expected return on assets of listed pharmaceutical companies in China. Here is the comparison of CAPM model and Fama-French model. Firstly, the α factor indicates the expected return rate, the following figure visualizes the prediction results of the CAPM model and Fama-French model for the α factor.

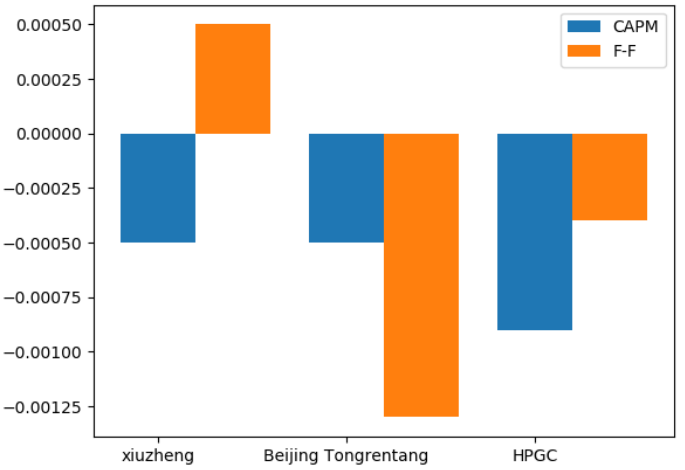
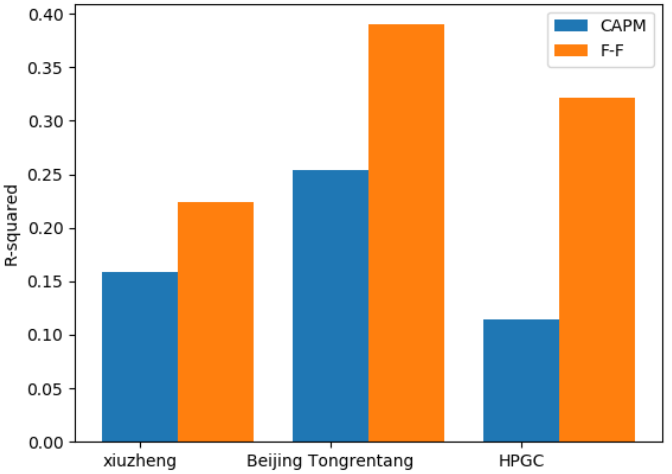


Figure 1. α result of two pricing models.

The “figure 1” shows visually expresses the forecast for the expected returns of several traditional Chinese pharmaceutical industries, with "α" close to 0 and in some cases even less than 0, indicating a possible downward trend in future returns (y-axis means the value of "α")



DOI: 10.54254/2755-2721/6/20230714

Figure 2. R^2 result of two pricing models.

In “figure 2”, the "R^2" result intuitively showed by the bar chart which indicates that the fitting of Fama-French model is obviously better than the CAPM model, but the "R^2" is still small. That means the fit of two models are not so good.

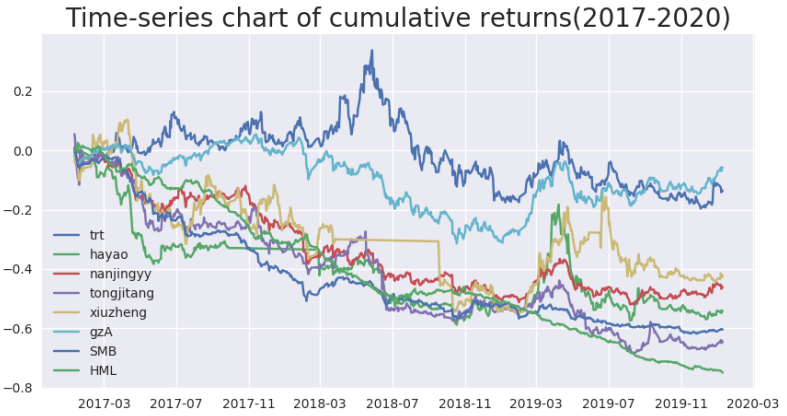


Figure 3. Cumulative returns.

Results of machine learning method. The experiment was run for 300 epoches, and the Figure 3 to Figure 5 are the results of the 300th epoch.

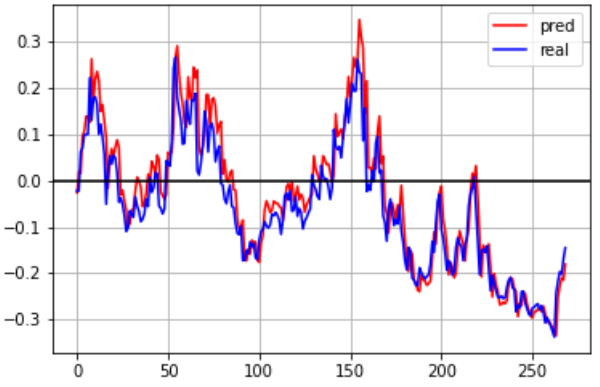


Figure 4. Beijing tongrentang result.

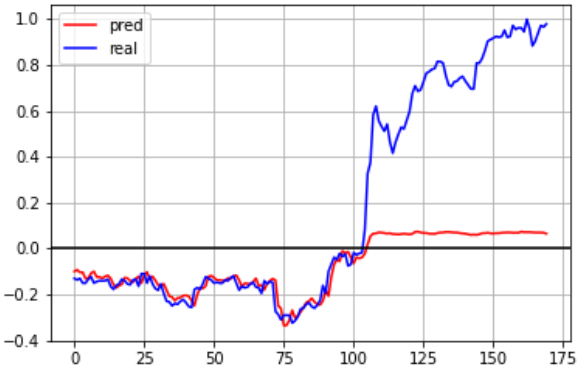


Figure 5. HPGC result.

DOI: 10.54254/2755-2721/6/20230714

The blue line is the real situation and the blue one is the prediction by the “Transformer model”

Analysis of machine learning method. From the above results, although the applied "transformer model" is better than RNN and LSTM, there are still inaccuracies in predicting some data. For example, the prediction of "HPGC" in the later stage has a sizeable directional error. However, the prediction of "Beijing Tongrentang" is more relevant to reality, and the error is small.

"Figure 5" shows that all Chinese traditional pharmaceutical companies mentioned above have seen their revenues decline over this period, and even though some rebounded, the overall return was declining. Although the R-squared is not high in the OLS regression result, the direction of the prediction is as expected.

The multi-factor Fama-French model fits relatively better than the single-factor CAPM model, but the overall fit of both is not very favorable. Market risk alone does not explain the expected return well. The results of the multi-factor model results analysis show that these several high-profile listed pharmaceutical companies are not developing rapidly. Macro factors are not the most critical factors affecting Chinese-listed pharmaceutical companies. The parent and subsidiary companies of pharmaceutical companies have mediocre performance, lack the ability of scientific and technological innovation, cannot develop vaccines, coupled with most of the medical resources used to fight the new coronavirus, the lack integration of resources by pharmaceutical companies, and the failure to discover effective drugs harms performance is also adversely affected.

7. Conclusion

The machine learning method performed exceedingly well, and the prediction value data is close to the actual situation. However, the predicted direction of some companies contrasts with the actual situation. Moreover, the fit of the pricing model with the data is not so good. However, the alpha factor accurately predicted the expected return rate of assets. Generally, according to data forecasts through 2020, the expected return on assets of the Chinese traditional pharmaceutical companies selected in the article could not grow rapidly and even on a downward trend.

References

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D. S. Touretzky and T.K. Leen (eds.), Advances in Neural Information Processing Systems 7, pp. 609-616. Cambridge, MA: MIT Press.

[2] Duarte F, C Rosa. The equity risk premium: A review of models [R]. New York: Federal Reserve Bank of New York Staff Reports,2015.

[3] Damodaran A. Equity Risk Premiums (ERP): Determinants, Estimation and Implications [N]. Unpublished paper Stem School of Business,2013-11-25(2).

[4] Duarte F, C Rosa. The equity risk premium: A review of models [R]. New York: Federal Reserve Bank of New York Staff Reports,2015.

[5] Li Yaxin. Research on the method of estimating the risk premium of Chinese market [D].2013.

[6] Li Yonggang. Analysis on the determination method of market risk premium ERP [J]. China asset evaluation,2015 (1): 30-34

[7] Song Jinyan. Research on Equity Premium: An Empirical Study Based on China's stock market[D].2011.

[8] Cao Yang. Whether stock liquidity promotes R & D innovation - A Study Based on China's pharmaceutical sector [J]. Financial review, 2015 (6)

[9] Liang Kai. Empirical Study on the influencing factors of stock return in China's pharmaceutical industry [J]. business, 2016 (1): 187-188

[10] Markowitz, H.M. (1952) Portfolio selection. Journal of Finance, 7, 77-91.

[11] Fama, E.F. and French. K.R. (1992) The Cross-Section of Expected Stock Returns. Journal of Finance, 47, 427-465.

[12] Wojciech Zaremba, Ilya Sutskever, Oriol Vinyals: “Recurrent Neural Network Regularization”, 2014; arXiv:1409.2329.

DOI: 10.54254/2755-2721/6/20230714

[13] Sepp Hochreiter and J urgen Schmidhuber. Long short-term memory. Neural computation, 9(8): 1735 1780, 1997.
[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: “Attention Is All You Need”, 2017; arXiv:1706.03762.
[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: “Attention Is All You Need”, 2017; arXiv:1706.03762.
[16] Michael Jensen The Performance of Mutual Funds during 1945-1964