

## Research Article

# The Multifactor Quantitative Investment Model Based on Association Rule Mining and Machine Learning

Kefu Yi 

*University of International Business & Economics, School of International Trade & Economics, Beijing 100029, China*

Correspondence should be addressed to Kefu Yi; 201800150012@uibe.edu.cn

Received 17 January 2022; Revised 14 February 2022; Accepted 15 February 2022; Published 18 March 2022

Academic Editor: Xin Ning

Copyright © 2022 Kefu Yi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The security information database has accumulated a large amount of historical data due to the continuing development of the securities market. People are concerned about how to fully utilize these data to investigate the securities market's law. In the financial field, financial asset pricing is a major issue. To some extent, the size of the return is determined by the difference between asset prices and their intrinsic value. The total global investment scale of quantitative funds will surpass 20 billion yuan by the end of 2021. Global asset management firms have turned to quantitative funds as their most important investment tool. Quantitative investment applies a specific investment idea to a specific model by creating specific indicators and parameters and then executes the investment strategy, greatly increasing the breadth and depth of investment. The goal of investors is to understand risk and maximize returns on investment. Researchers and investors alike value quantitative investment because of its scientific and efficient operation. In quantitative stock selection, a multifactor stock selection model is a critical tool for building a portfolio. This paper builds a multifactor investment strategy based on the relevant factors of corporate finance and valuation, selects the portfolio, and calculates the excess return using a machine learning classification algorithm.

## 1. Introduction

Looking back at the development of quantitative investment, it has been more than 40 years in foreign countries [1]. China's development in the field of quantitative investment is relatively short. The first quantitative fund was born in China in 2004. After more than ten years of tempering, it was not gradually recognized by domestic investors until 2015. The rapid expansion of data has produced the problem and phenomenon of "data explosion and lack of knowledge" [2]. By the end of 2016, the total global investment scale of quantitative fund has exceeded 3 trillion US dollars, and its proportion in global fund scale is close to one third. Quantitative fund has become the most important investment tool of global asset management companies. However, due to the backward information technology at that time, the huge and complicated data could not be processed in time and effectively, which led to the slow development of quantitative investment theory [3]. Therefore, how to quickly and effectively choose profitable stocks or companies from many

stocks becomes a very practical problem, or from the point of view of higher requirements, how to choose stocks that exceed the market benchmark returns from many stocks [4], to understand the data mining in the securities industry and its system structure research [5].

Starting with massive data in the securities industry, it examines data mining technology that is appropriate for each type of data and systematically investigates the basic process and functional components of the securities data mining system [6, 7]. The concept of quantitative investment is first introduced through the research background, and then the development process of quantitative investment theory is explained, with a focus on the theoretical basis of the multi-factor model [8]. Quantitative investment, in comparison to traditional investment, has the advantages of relatively accurate risk control and relatively stable excess return, attracting increasing interest from the fund industry [9]. At the same time, the quantitative stock selection model and quantitative timing strategy are combined to effectively control some factors of human intervention in each link,

from selecting individual stocks to optimizing position control, from judging portfolio risk to executing trading, which can not only reduce investment risk but also reduce trading costs. It can not only provide relatively stable investment income but also help the entire stock investment market run smoothly and efficiently [10]. The multifactor classification model based on decision tree algorithm improves the traditional factor score model based on a priori experience, and the classification rules obtained by the decision tree classification model can be used as the if rules conditions for investors to choose stocks, with a wide structure, easy to understand, and rational basis [11].

The main innovations of this paper are as follows: from the aspect of data selection, there are 45 factors used for analysis in this paper. The selected data are the data of recent years, including the data of 2021, which has good timeliness and uses the data of 2015, which makes the results of this paper have better adaptability and better explanation for today's capital market. In the aspect of architecture design of securities data mining system, the system architecture covers three main application aspects of data mining technology in securities industry. The system architecture includes commonly used data mining algorithms and adopts advanced artificial intelligence technologies such as intelligent selection control technology and metaknowledge base. From the aspect of regression model, this paper adopts the elastic net regression model to screen the factors. Combining the elastic net regression with the multifactor stock selection model is a new and efficient way. Compared with the ordinary OLS regression, it has a better screening effect, and the formed investment strategy has advantages in stability and profitability, which provides some new ideas for the future application and development of quantitative investment.

## 2. Association Rule Mining Technology

*2.1. Association Rule Mining.* Association rule mining is an important branch of data mining research. Association rules are the most typical of many knowledge types of data mining [12, 13]. The composition of each part is shown in Figure 1 below.

Because the market mechanism of Chinese stock market is not perfect enough to reflect all the information in the market, and its effectiveness is weak, there are relatively many stocks that are wrongly priced, and the purpose of studying quantitative investment is to capture as many investment opportunities as possible in the capital market. Therefore, quantitative investment research in China's capital market is profitable [14]. At the same time, because China's stock market often has problems such as disorderly rise and fall and strong randomness, combining quantitative stock selection model with quantitative timing strategy can effectively control some factors of human intervention in every link from selecting individual stocks to optimizing position control, from judging the risk of portfolio to executing trading, which can not only reduce investment risks and obtain relatively stable investment returns but also promote

the stable, healthy, and sustainable development of the stock market [15].

In a transaction database, association rule mining can uncover interesting connections between items or attributes. These connections are not known ahead of time and cannot be found using logical database operations like table connections or statistical methods [16]. Nonlinear science has emerged as a new research method, with fruitful results in financial practice, as a result of continuous iteration of technology and continuous innovation of financial theory [17]. To inject vitality into the development of financial research, new methods such as legacy algorithms, decision trees, and neural networks are used in model construction. Figure 2 depicts the historical context of quantitative investment theory.

In addition, from the development experience of foreign mature financial markets, using quantitative methods for investment can effectively improve the liquidity of financial markets [18]. Association rules are characterized by concise form, easy explanation, and understanding and can effectively capture the important relationships between data. In factor selection, most models give priority to fundamental factors. Domestic scholars mostly do empirical research on multifactor models. By adopting various quantitative investment strategies, it is proved that the stock market has not reached an effective state, and stable excess returns can be achieved. Stock price fluctuation is influenced by many factors, and there may be a strong correlation between different factors [19]. In 2007-2015, a multifactor quantitative stock selection model based on regression method was established for the constituent stocks of Shanghai Stock Exchange 180 Index, respectively. According to the idea of value investment, the investors' portfolio was constantly adjusted through the quantitative stock selection model built year by year, starting from the 2007 annual report and ending with the 2015 annual report, and there were nine positions adjustment. The unified multifactor model has three main strategies.

The selection of factors and the construction method of multifactor models [20] are the main differences between different models. The investment logic is that the data of existing characterization factors in the stock market can be related to the stock's future earnings. As a result, the effective characterization factors can be considered comprehensively, many factors affecting the company's stock price and company value can be selected, the quantitative model for these factors can be established, the results are scored and sorted, and whether to hold some stocks can be selected based on the sorting results to form a stock portfolio with a higher return than the market [21]. In reality, there is no such thing as a highly efficient market. We can divide the effectiveness of the market into three types according to the degree of access to information, as shown in Figure 3.

The application of association rules is not limited to market basket analysis but has a wide range of applications, such as business and finance, census data analysis, engineering data analysis, medical care, finance, macrodecision support, e-commerce, website design, and Internet. In the empirical research of the multifactor quantitative stock

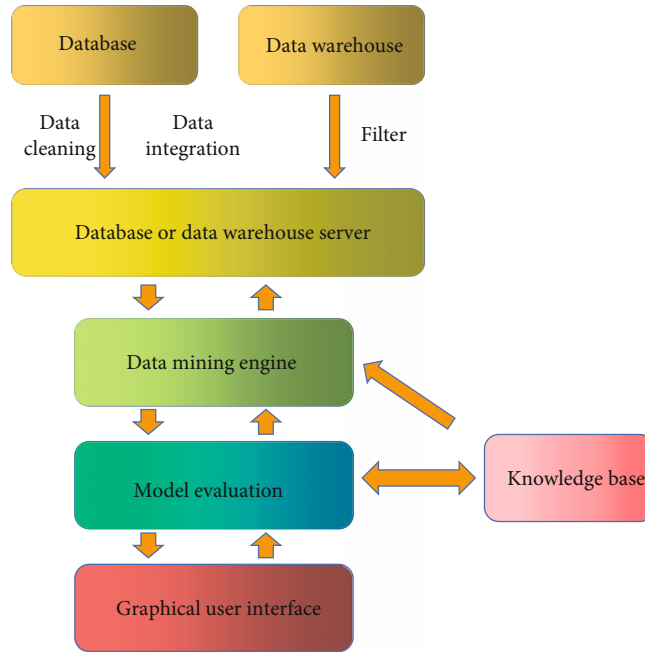


FIGURE 1: Typical data mining system structure.

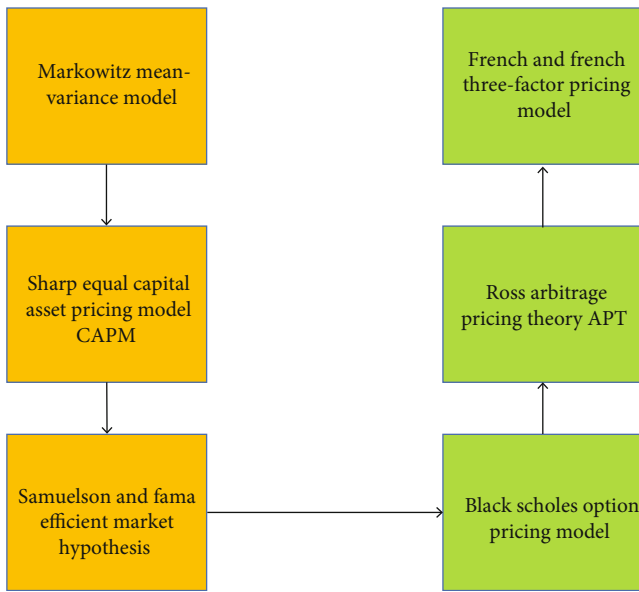


FIGURE 2: Development of quantitative investment theory.

selection model, the final results may be different due to different candidate factors selected by different scholars when establishing the model. Confidence is a measure of the accuracy of association rules, indicating that the strength support of rules is a measure of the importance of association rules and the frequency of rules. The support degree of the rule indicates how representative it is in all transactions. The larger its value, the more important the association rule is. If the confidence level of the association rule is high but the support level is low, it means that the association rule has little practical chance. If the support level is high and

the confidence level is low, it means that the rule is unreliable.

**2.2. Data Mining.** With the improvement of data acquisition and storage technology, a large number of large databases have been produced in various fields of human life [22]. The general idea of establishing the multifactor quantitative stock selection model is as follows: (1) selection of factors and (2) how to use the selected factors to build a model with relatively higher yield and more stable results. The establishment of quantitative stock selection model is mainly based on CAPM model, APT model, and Fama-French three-factor model. The multifactor model building process is shown in Figure 4 below.

The definition of data mining has always been controversial in academic circles, and there is no completely unified and accurate definition. For the application research of this paper, data mining quotes the following definition [23].

The mean-variance model employs mathematical tools to investigate financial issues, resulting in a highly data-driven and structured approach to securities research [24]. Long-term, medium-term, and short-term stock returns are all divided into three categories in various studies. When a candidate factor has a high ability to explain the stock return, it is added to the list of effective factors for further study, and the effectiveness test framework for candidate factors is established, as shown in Figure 5.

Among them, the long-term rate of return is represented by the 5-year or 3-year cumulative rate of return. The medium-term rate of return mainly refers to the annual rate of return and semiannual rate of return, while the short-term rate of return includes quarterly rate of return, monthly rate of return, and weekly rate of return. The CAPM model is the most important basis of security investment theory.

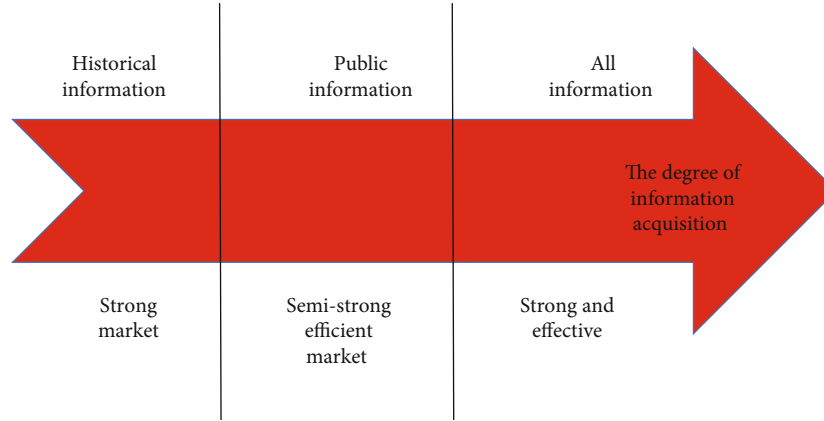


FIGURE 3: Efficient market hypothesis.

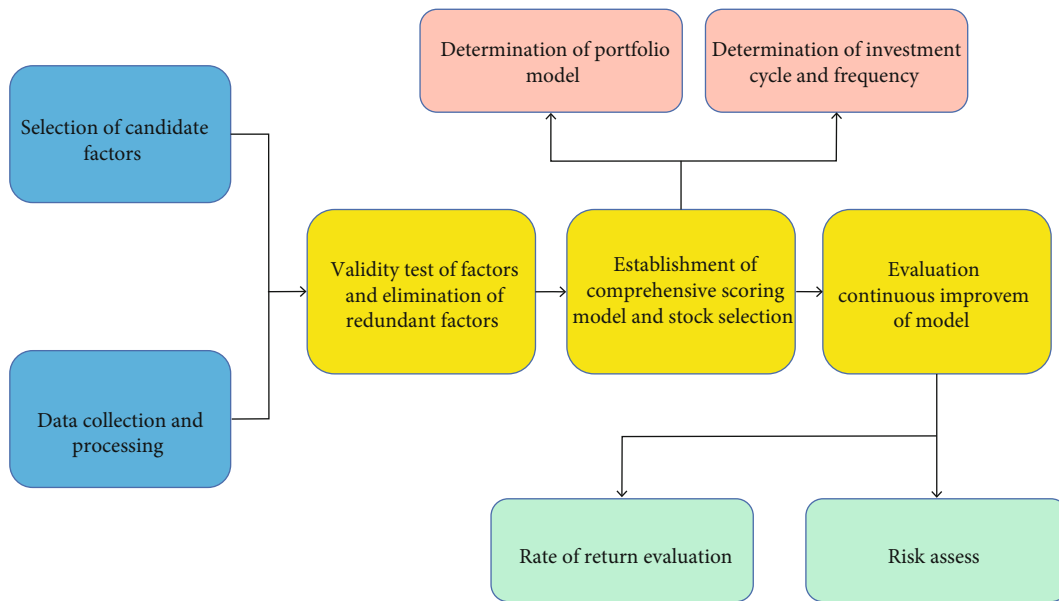


FIGURE 4: Flow chart of the multifactor stock selection model.

According to the very strict assumptions of the CAPM model, all investors obtain the same unit risk return due to the change of investment portfolio; that is, nonmarket risk will not affect the expected return of investment portfolio. If the rate of return obtained by the model is different from that of other investors, the difference of these returns can only be caused by system risk [25].

When it comes to describing data mining from the standpoint of commercial application, it is discovered that it differs from other research fields such as machine learning [26] and has had a strong commercial application purpose since its inception. The discovery of universal truth is not required for data mining. All newly discovered knowledge is relative, and it serves as a guide for specific business operations [27]. The present value of the stock price is related to the future forecast, but the evolution mode of the original variables and historical data is not [28]. Due to the long investment process of a quantitative fund, it is necessary to

select the annual rate of return in the medium-term rate of return for empirical research in order to control the risks generated in the investment process as much as possible, fully tap the information published by various listed companies, and timely and effectively find stocks with growth potential.

### 3. The Establishment and Empirical Analysis of the Multifactor Stock Selection Model

*3.1. Data Correlation Analysis.* After the standardization of data processing, it is necessary to analyze the correlation of data. Ensure that the data to be selected must be true and accurate original data. In this sample research range, it basically contains all the possible market trends of the stock market every year. It is found that the selection of models and variable factors in various multifactor models is often based on simple methods, such as regression analysis and

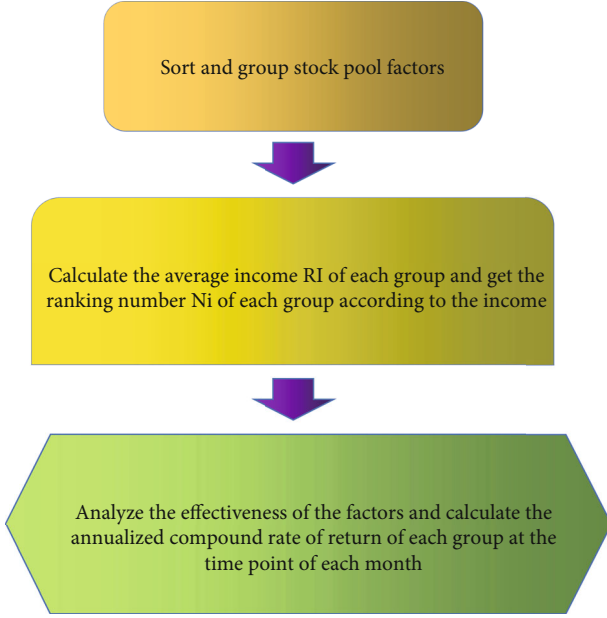


FIGURE 5: Candidate factor validity checking framework.

correlation analysis. The entire market is frictionless. The capital market is a complete information market. All valuable information in the market has been reflected, and there are no taxes or transaction costs. Based on the above assumptions, the calculation formula of CAPM is

$$E(R_i) = R_f + b_i(E(R_M) - R_f). \quad (1)$$

$E(R_i)$  is the expected return on securities  $I$ ,  $R_f$  is the return on risk-free assets, and  $E(R_M)$  is the market expected rate of return is the annualized rate of return, which converts the current rate of return into the annual rate of return.

Common quantitative factors such as index per share refer to the related financial indicators per share, such as EPS and free cash flow of enterprises per share. While scanning the database, all frequent itemsets are directly generated by operation, and the most computationally intensive part of the frequent itemset generation step is completed. Many researchers regard data mining as a synonym for knowledge discovery in another commonly used term database, while others just regard data mining as a basic step in. The coefficient value of MTM and EPS under different factor coefficient values is shown in Figure 6 below.

This is the theoretical rate of return, not the actual rate of return. The formula is as follows:

$$R = \left(1 + \frac{N_2 - N_1}{N_1} \times 100\%\right)^{\frac{365}{T_2 - T_1}}. \quad (2)$$

$T_2 - T_1$  is the investment period,  $365/T_2 - T_1$  is the convert the investment period into years, and  $N_2 - N_1/N_1$  is the total rate of return during the investment period.

The multifactor stock selection model extracts a large amount of data and enough sample stocks for analysis, ensuring that the results are more applicable, and that the

model's effectiveness is ensured. The factor coefficient of the current asset ratio in the quality factor is -0.003, which is the lowest absolute value of the factor coefficient in the table, according to the regression coefficient table. As a result of the regression results, it can be deduced that the current asset ratio has a minor impact on the monthly return of listed companies' stocks. Figure 7 shows a comparison of the fix asset ratio and the gross income ratio for various factor coefficients.

$B$  represents  $\beta$  coefficient of  $I$ , that is, the size of system risk, and its calculation formula is

$$\beta_i = \frac{\text{Cov}(R_i, R_M)}{\text{Var}(R_M)} = \frac{\sigma_{iM}}{\sigma_M^2}. \quad (3)$$

Figure 8 below shows the three PB curves under different market net coefficients.

Combining the variable selection algorithm, decision tree algorithm, and traditional multifactor quantitative stock selection model, innovatively construct a multifactor quantitative classification model. In the process of studying the relationship between multifactor variables and stock returns and quantitatively predicting stock returns, a multiple linear regression model is used, and its calculation formula is as follows:

$$Y = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_j X_{ij} + \varepsilon. \quad (4)$$

The coefficient of the momentum factor five-day moving average (Ma5) in the technical factor is about -7.71, which is the largest absolute value in the coefficient table, which shows that the 20-day moving average has the greatest impact on the monthly return of stocks. The comparison between Ma5 and mA20 is shown in Figure 9 below.

The association rule states that "when the stock prices of IBM and the sun rise within a certain time period  $T$ , in 95% of cases, Microsoft's stock will rise at the same time." Because of the  $T$ -time constraints, the model is referred to as a one-dimensional relational model. When using the multifactor stock selection model to better explain excess returns, we must consider a variety of factors in order to find a more effective portfolio, and we must also determine whether there is a large causal relationship between the selected factors and future stock returns. The mutual verification of the test cycles inside and outside the sample proves the validity of the multifactor stock selection model. Empirical research shows that multiple variable indicators have a strong correlation, which is related to the calculation method of indicators and the economic meaning they represent. By querying different transaction sets, the difference between discovering a two-dimensional relational model and a one-dimensional model is similar to "when the stock prices of IBM and SUN rise in a certain time range  $T$ , in 80% of cases, the stock will rise after a certain interval."

**3.2. Optimization Algorithm Performance Analysis.** The time cost of the algorithm mainly depends on the time of transaction database scanning and generating frequent itemsets. 30 factors are selected to establish the model. These factors

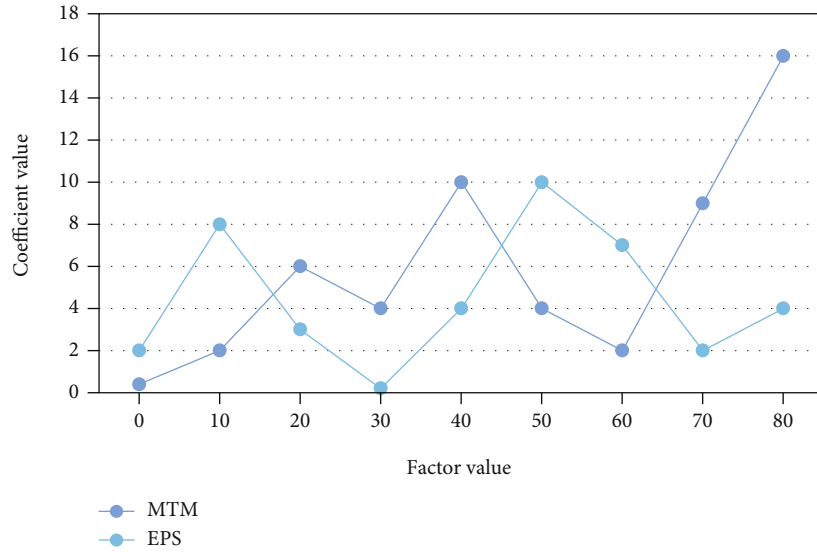


FIGURE 6: Coefficient values of MTM and EPS under different factor coefficient values.

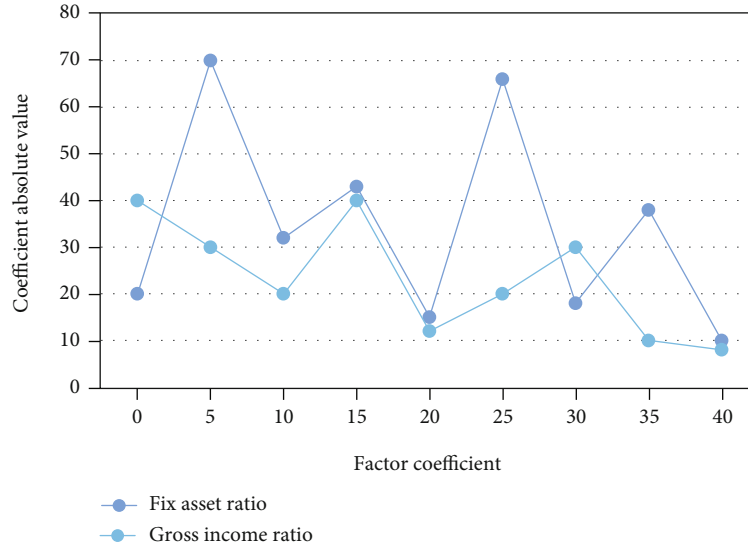


FIGURE 7: Comparison of fix asset ratio and gross income ratio under different factor coefficients.

include  $P/E$  ratio, sales gross profit margin, and  $P/B$  ratio in the value factor; shareholders' equity ratio, fixed assets ratio, return on assets, earnings per share, logarithm of current market value, return on assets, and return on equity in the quality factor; volume ratio, turnover rate, moving average, Hearst index, and momentum index in technical factors; and the current stock price in the liquidity factor is in the stock price position of the past year, historical beta value, trading volume ratio, historical fluctuation, and stock price skewness. As a classical frequent itemset generation algorithm, it plays a milestone role in the research of association rules. Among the technical factors, the momentum factors, including five-day moving average and sixty-day average turnover rate, have coefficients of 2.28 and 3.77, respectively, which are larger values in the coefficient matrix, which indicates that momentum factors have strong explanatory power

for the monthly stock returns, and can greatly affect the stock returns of listed companies. For example, the comparison between ROA and VOL20 under different characteristic values is shown in Figure 10 below.

The scoring method is used to calculate the comprehensive score of the stock, but the meaning and unit of each factor are different. If you want to compare and analyze factor data of different dimensions at the same time, you need to standardize them. The value of stock  $i$  factor  $j$  in all samples is as follows:

$$Z_{i,j} = \frac{X_{i,j} - \mu_j}{\sigma_j}. \quad (5)$$

$\mu_j$  is the mean of the  $j$  factor, and  $\sigma_j$  is the standard deviation of the  $j$  factor.

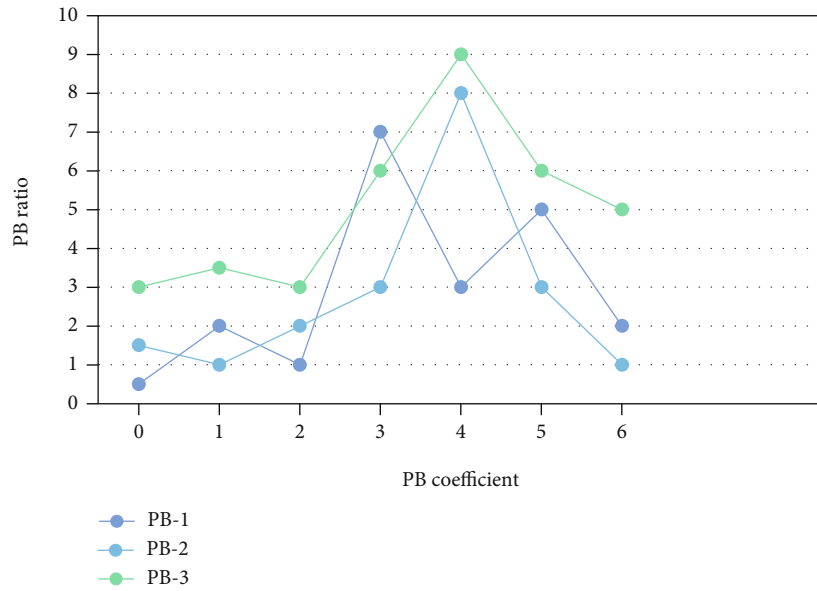


FIGURE 8: Three Pb curves under different  $P/b$  coefficients.

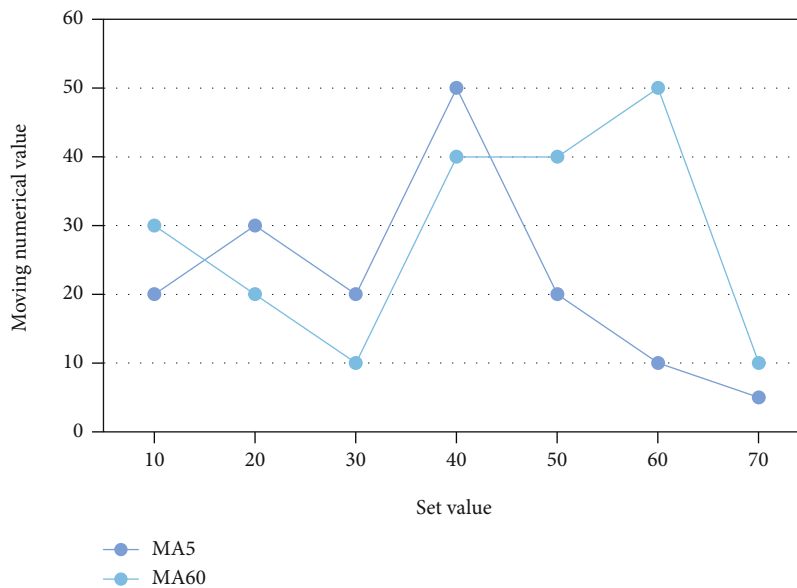


FIGURE 9: Moving values of MA5 and MA60 under different settings.

The number of transaction sets in each cycle will also affect the efficiency of the whole algorithm. The problem is particularly prominent when the amount of data is large. How to shorten the transaction set in each cycle also needs to be further improved. Among the quality factors, the coefficient of book leverage is about -0.5, which is also a factor that has a greater impact on the stock return. Figure 11 shows the comparison between Hurst curve and skewness curve under the changing set value.

Although the current application of this model has changed a lot, its ideas have always had a great influence on many scholars and quantitative investors. The basic expressions of

Fama and French models are as follows:

$$E(R_i) - R_f = b_i [E(R_m - R_f)] + s_i E(\text{SMB}) + h_i E(\text{HML}). \tag{6}$$

$R_i$  is the return on asset  $i$ ,  $R_f$  is the risk-free rate of return,

$R_m$  is the revenue from the market, SMB is the market value factor, which represents the difference between the returns of large-cap stocks and small-cap stocks, and HML

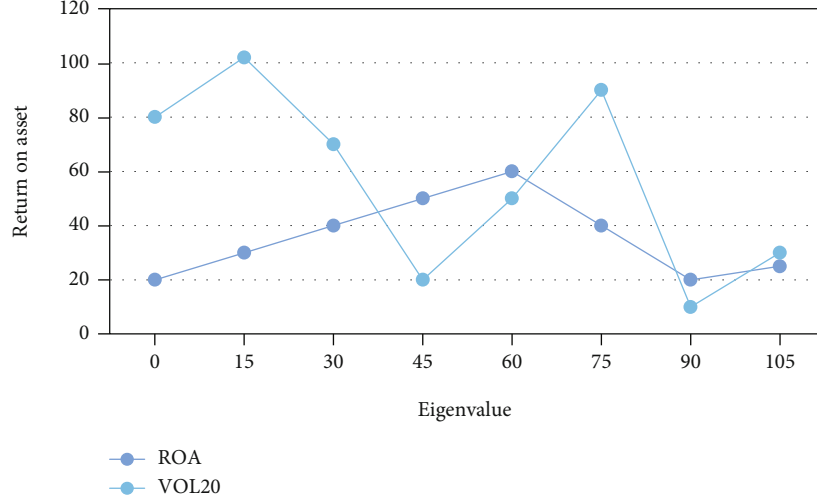


FIGURE 10: Comparison between ROA and VOL20 under different characteristic values.

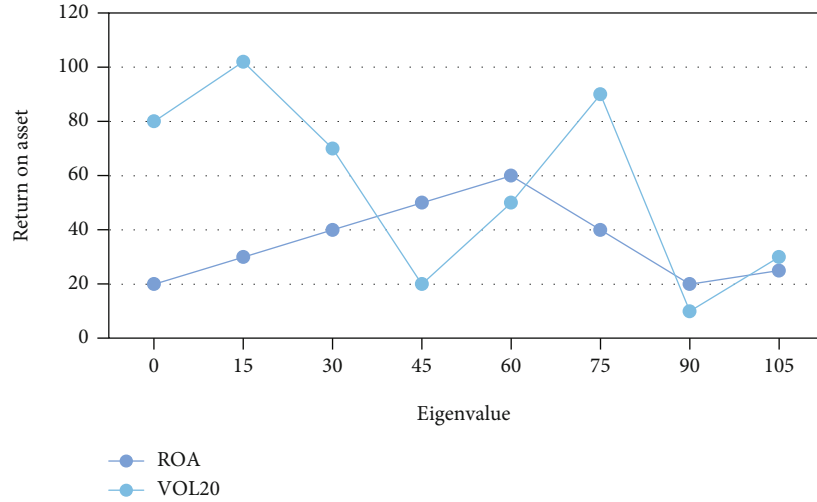


FIGURE 11: Comparison of Hurst curve and skewness curve.

is the book-to-market ratio, a measure of the difference in earnings between value stocks and growth stocks.

The rate of return on net assets can also be divided into net sales interest rate, asset turnover rate, and equity multiplier, which is a comprehensive index reflecting the efficiency of shareholders' capital use. The effective valuation factors of GEM are as follows:  $P/E$  ratio,  $P/E$  ratio,  $P/B$  ratio, net assets per share, total liabilities, and earnings per share. The expression of lasso method can be written as follows:

$$\bar{\beta} = \arg \min \left\{ \|Y - X\beta\|^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (7)$$

The main board's and GEM's effective factors are combined to create a set of valuation effective factors. We can see that the selection sensitivity of the underlying stocks is not strong for the evaluation factors; that is, the only difference between the effective factors of the main board and the

GEM is the enterprise value multiple and the total liabilities, and the enterprise value multiple in the main board is effective, partly because most of the listed companies on the main board are public companies. Figure 12 shows the comparison of the FTWJH and BLEV curves at various setting values.

Ridge penalty has great advantages in solving the multicollinearity problem; that is, it adds an  $l_2$  penalty on the basis of least square regression, which is expressed as follows:

$$\bar{\beta} = \arg \min \left\{ \|Y - X\beta\|^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}. \quad (8)$$

Reflected in the above algorithm, it is mainly reflected in the low efficiency when using generation, and in order to calculate the support of each candidate element, all transactions in the transaction database have to be compared once



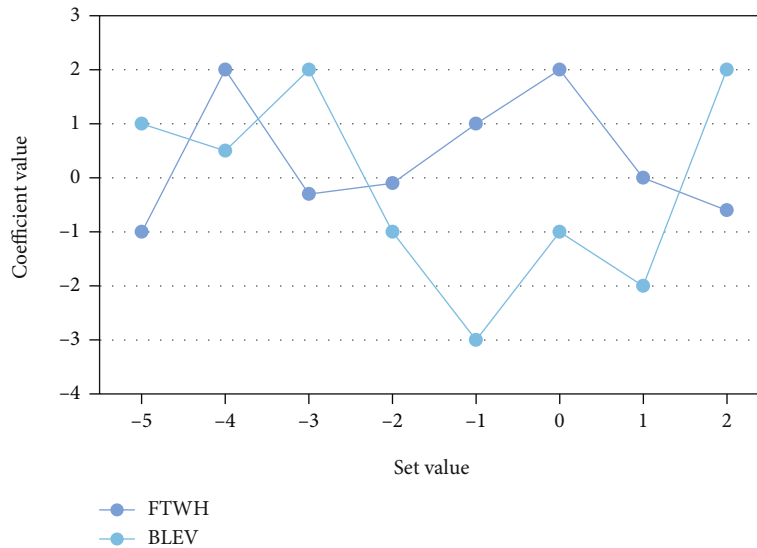


FIGURE 12: Comparison of FTWH and BLEV curves.

and compared again every cycle, so that the time complexity of this algorithm increases and the efficiency decreases. Association rules were first started by shopping basket analysis. However, with the expansion and deepening of research, the application scope of association rules is constantly expanding; so, there are various forms of association rules research. It should be adjusted in accordance with the quarterly and annual report release dates: we can only use the third quarterly report of the previous year from January to April because the most recent available financial data is from the previous year's third quarterly report. From May to August, the most recent financial data is derived from either the current year's first quarterly report or the previous year's annual report, the latter being more authoritative. The data from the previous year's semiannual report was used for analysis in the months of September and October. Use the data from the third quarterly report of that year during the months of November and December.

#### 4. Conclusions

The multifactor stock selection strategy begins with an examination of the possible causes of stock excess returns, followed by the development of a quantitative investment strategy based on the index data obtained in order to construct an investment portfolio that outperforms the market benchmark return rate. The preprocessing module of stock data is designed, and the stock data suitable for association rule mining is generated, based on the research and analysis of association rule theory, combined with the characteristics of stock original data. Statistics provides a practical and useful framework for examining problem solutions. Modern statistics is based on models. Model selection and calculation are frequently regarded as an afterthought and a subset of model creation. This paper constructs a multifactor stock selection model with equal weight based on the selected effective factors and conducts an empirical test during the out-of-sample test period. The results show that the multi-

factor model performs well in terms of yield, stability, and risk measurement, demonstrating its effectiveness and superiority in China's stock market. The model's portfolio consistently outperforms the market benchmark in various stock markets year after year, and when combined with a quantitative timing strategy, it can help investors achieve positive returns and higher investment returns. The decision tree classification algorithm model was introduced in an innovative way to determine the combination selection rules, based on the traditional multifactor model quantification and drawing inspiration from the factor classification effectiveness test method. The corresponding index rules of the stocks of different groups of returns were learned during the selection process. The degree of control algorithm learning can help the model achieve a good balance of learning and generalization capabilities.

#### Data Availability

The data used to support the findings of this study are included within the article.

#### Conflicts of Interest

The author does not have any possible conflicts of interest.

#### References

- [1] B. J. C. David Goliath, "Who wins the quantitative Battle?," *The Journal of Portfolio Management*, vol. 43, no. 3, pp. 34–37, 2017.
- [2] C. R. Harvey, S. Rattray, A. Sinclair, and O. Van Hemert, "Man vs. machine: comparing discretionary and systematic hedge fund performance," *The Journal of Portfolio Management*, vol. 44, no. 7, pp. 11–15, 2018.
- [3] J. Gibson and F. Rioja, "The welfare effects of infrastructure investment in a heterogeneous agents economy," *The BE Journal of Macroeconomics*, vol. 20, no. 1, pp. 155–172, 2020.

- [4] J. Lopez, A. Sakhel, and T. Busch, "Corporate investments and environmental regulation: the role of regulatory uncertainty, regulation-induced uncertainty, and investment history," *European Management Journal*, vol. 35, no. 1, pp. 91–101, 2017.
- [5] R. Feng and B. Yi, "Quantitative modeling of risk management strategies: stochastic reserving and hedging of variable annuity guaranteed benefits," *Insurance: Mathematics and Economics*, vol. 85, no. 3, pp. 60–73, 2019.
- [6] E. Berman, J. Bound, and S. Machin, "Implications of skill-biased technological change: international evidence," *Papers*, vol. 113, no. 4, pp. 1245–1279, 2018.
- [7] X.-B. Jin, W.-T. Gong, J.-L. Kong, Y.-T. Bai, and T.-L. Su, "A variational Bayesian deep network with data self-screening layer for massive time-series data forecasting," *Entropy*, vol. 24, no. 3, p. 335, 2022.
- [8] M. B. Mathur and D. B. Reichling, "Navigating a social world with robot partners: a quantitative cartography of the uncanny valley," *Cognition*, vol. 146, pp. 22–32, 2016.
- [9] F. Caravelli, L. Sindoni, F. Caccioli, and C. Ududec, "Optimal leverage trajectories in presence of market impact," *Papers*, vol. 158, no. 3, pp. 437–442, 2016.
- [10] X. Deng, W. Ying, and G. T. Yeo, "Enterprise perspective-based evaluation of free trade port areas in China," *Maritime Economics & Logistics*, vol. 19, no. 3, pp. 451–473, 2017.
- [11] D. K. Burns, A. P. Jones, Y. Goryakin, and M. Suhrcke, "Is foreign direct investment good for health in low and middle income countries? An instrumental variable approach," *Social Science & Medicine*, vol. 181, pp. 74–82, 2017.
- [12] P. A. Haslam and N. A. Tanimoune, "The determinants of social conflict in the Latin American mining sector: new evidence with quantitative data," *World Development*, vol. 78, pp. 401–419, 2016.
- [13] S. Alessandro, F. Andrea, and P. Fabio, "Investment communities: behavioral attitudes and economic dynamics," *Social Networks*, vol. 55, pp. 170–188, 2018.
- [14] E. Aca and A. Mozumdar, "Investment–cash flow sensitivity: fact or fiction?," *Journal of Financial and Quantitative Analysis*, vol. 52, no. 3, pp. 1111–1141, 2017.
- [15] T. J. Chemmanur, E. Mine, and K. Karthik, "Is it the investment Bank or the investment banker? A study of the role of investment banker human Capital in Acquisitions," *Journal of Financial and Quantitative Analysis*, vol. 54, no. 2, pp. 587–627, 2019.
- [16] J. Lewellen and K. Lewellen, "Investment and cash flow: new evidence," *Journal of Financial & Quantitative Analysis*, vol. 51, no. 4, pp. 1135–1164, 2016.
- [17] P. Zhang, K. Yu, J. Yu, and S. U. Khan, "QuantCloud: big data infrastructure for quantitative finance on the cloud," *IEEE Transactions on Big Data*, vol. 4, no. 3, pp. 368–380, 2018.
- [18] D. Bui-Thi, P. Meysman, and K. Laukens, "MoMAC: multi-objective optimization to combine multiple association rules into an interpretable classification," *Applied Intelligence*, vol. 52, no. 3, pp. 3090–3102, 2022.
- [19] D. Martín, M. Martínez-Ballesteros, D. García-Gil, J. Alcalá-Fdez, F. Herrera, and J. C. Riquelme-Santos, "MRQAR: a generic mapreduce framework to discover quantitative association rules in big data problems," *Knowledge-Based Systems*, vol. 153, pp. 176–192, 2018.
- [20] D. Martín, J. Alcalá-Fdez, A. Rosete, and F. Herrera, "NIC-GAR: a niching genetic algorithm to mine a diverse set of interesting quantitative association rules," *Information Sciences*, vol. 355, pp. 208–228, 2016.
- [21] G. S. Navale and S. N. Mali, "A multi-analysis on privacy preservation of association rules using hybridized approach," *Evolutionary Intelligence*, vol. 3, pp. 1–15, 2019.
- [22] G. Fan, W. Shi, L. Guo, J. Zeng, K. Zhang, and G. Gui, "Machine learning based quantitative association rule mining method for evaluating cellular network performance," *IEEE Access*, vol. 7, pp. 166815–166822, 2019.
- [23] Y. Liu, H. Gan, and K. Karim, "The effectiveness of chief financial officer board membership in improving corporate investment efficiency," *Review of Quantitative Finance and Accounting*, vol. 57, no. 2, pp. 487–521, 2021.
- [24] G. J. Power, N. Rani, and A. Mandal, "Corporate control and the choice of investment financing: the case of corporate acquisitions in India," *Review of Quantitative Finance and Accounting*, vol. 58, no. 1, pp. 41–68, 2022.
- [25] C. Jiang, K. John, and D. Larsen, "R&D investment intensity and jump volatility of stock price," *Review of Quantitative Finance and Accounting*, vol. 57, no. 1, pp. 235–277, 2021.
- [26] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Graph sample and aggregate-attention network for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [27] S. Cederburg, "Pricing intertemporal risk when investment opportunities are unobservable," *Journal of Financial & Quantitative Analysis*, vol. 54, no. 4, pp. 1759–1789, 2019.
- [28] C. Su, H. Zhang, K. Bangassa, and N. L. Joseph, "On the investment value of sell-side analyst recommendation revisions in the UK," *Review of Quantitative Finance and Accounting*, vol. 53, no. 1, pp. 257–293, 2019.