Multi-factor Stock Selection Model Based on Machine Learning

Yihua Zhong, *Member, IAENG,* Lan Luo, *Member, IAENG,* Xinyi Wang, *Member, IAENG* and Jinlian Yang, *Member, IAENG*

***Abstract*—On the basis of the analysis of CSI-300 stock, two multi-factor stock picking strategies based on multi-classification logistic regression algorithm and multi-classification XGBoost algorithm are compared with an equal-weighted linear multi-factor stock picking strategy. To control the risk of stock selection strategy and keep the maximum retracement within 10%, a risk model is established in this paper. Moreover, the analysis theory and numerical experiments are reported on the feasibility and efficiency of this two proposed algorithms. The results show that the multi-factor stock selection model based on multi-classification logistic regression algorithm has good earnings prediction and profitability. It has the highest annual sharpe ratio.**

***Index Terms*—multifactorial model; machine learning; sharpe ratio; multi-classification logistic regression; multi-classification XGBoost algorithm .**

I. Introduction

AS the economy grows, the understanding of investment is growing. People are looking for more scientific and reasonable investment methods. Based on the computer technology, quantitative investment provides a meaningful direction for investment decision-making. It combines the traditional investment concepts, uses the reasonable algorithm, and provides the better benefit for the investor. It can obtain higher excess returns, which is gradually accepted by investors. Multi-factor model is the most commonly used in quantitative investment. A growing number of scholars have studied multi-factor stock selection models [1], [2], [3], [4], [5].

In the early stages, according to risk-return characteristics, Ross [6] proposed the arbitrage pricing model (APT) based on the arbitrage behavior of investors. The APT model shows that security asset returns are influenced by a variety of different types of factors rather than a single factor. Fama and French [7] proposed a three-factor model named the Fama-French three-factor model which categorizes the influences of stock returns into three factors, i.e. Rm-Rf, SMB and HML. Based on the APT model and the Fama-French three-factor model, many financial analysts and investors had studied various factors and construction methods. In 2009, Richard Tortoriello [8] enumerated the specific indicators of the factors that can be used to quantify stock selection, and theoretically explained the role of each factor on the stock market. On this basis, the effective factors of U.S. stock market were screened out, and the single-factor, double-factor and multi-factor stock selection models were summarized. Most of them are linear models [9], [10], [11], [12], [13]. Because stock markets are chaotic, complex, and dynamic, the linear models formulating the stock markets may be unreasonable. Comparing with traditional linear multi-factor models, the models by machine learning algorithms are able to capture finer market signals and obtain more robust excess returns through the non-linear expression of factors.

In recent years, the rapid development of machine learning algorithms has provided some new ideas for the research of quantitative investment, such as support vector machine [14], logistic regression algorithm [15] and neural network algorithm [16], [17]. Esmaeil Hadavandi et al. [18] proposed a combination of genetic fuzzy systems and artificial neural networks to construct stock price prediction models. Taking the top 200 stocks in Taiwan stock Market from 1996 to 2011 as research objects, Huang [19] developed a genetic algorithm-improved support vector regression stock-picking model to predict stock returns. Recently, Chen [20] presented the XGBoost algorithm for factor selection using the scoring model based on equal weighted ratings. Kim and Won [21] established a hybrid model integrating long short-term memory (LSTM) with multiple GARCH-type models for stock price index volatility prediction and achieved good results. Wenxing Li and Wenjun Li [22] applied a kind of semi-supervised K-means kernel function clustering algorithm with gravitational influence factors into the multi-factor stock selection model, and showed that the method has a stronger generalization ability than traditional clustering models such that select better stock combinations.

Many examples show that investors are able to effectively reduce risk, mitigate losses, and achieve higher excess returns by studying the stock market and using computer technology. They construct their investment strategies through combining different methods such as big data and cloud computing. Quantitative investing is getting more and more attention, and the numbers of multi-factor stock selection models are proliferating.

The remainder of this paper is organized as the following sections. Sect. 2 provides data analysis. Sect. 3 introduces the constructing process of the proposed multiple classification models by using equal-weighted linear model, multi-classification logistic regression algorithm and multi-classification XGBoost algorithm, respectively. Then three different algorithm of stock picking strategies are established. Sect. 4 shows the comparing of three algorithms in annual sharpe ratio through an empirical analysis. Sect. 5 establishes

Manuscript received June 2, 2020; revised Dec 3, 2020.

Yihua Zhong is a professor in School of Science, Southwest Petroleum University, Chengdu 610500, China. Email: zhongyihua@swpu.edu.cn.

Lan Luo is a master in School of Science, Southwest Petroleum University, Chengdu 610500, China. Email: lanluo3366@163.com. (Corresponding Author)

Xinyi Wang is a master in School of Science, Southwest Petroleum University, Chengdu 610500, China. Email: 1364241552@qq.com.

Jinlian Yang is a master in School of Science, Southwest Petroleum University, Chengdu 610500, China. Email: 2806234881@qq.com.

TABLE I
CANDIDATE FACTOR

Large class factor	Concrete factor
Quality factor	ROE
	ROA
	Current Ratio
	Total Debt Ratio
Growth factor	Net Assets Growth Rate
Value factor	Circulation Market Value
	The Total Market Capitalization
	P/B Ratio
	P/E Ratio
Commonly used technology factor	Decline Indicator
	20-day Moving Average
	5-day Moving Average
Momentum class factor	CMO
	PVT
Emotional factor	20-day Average Turnover Rate
	5-day Average Turnover Rate

a risk model to control the risk of stock selection strategy, and lists the experiment setup and the experimental results. Finally, the paper ends with a conclusion in Sect. 6.

II. DATA ANALYSIS

A. Selection of candidate factors

The choice of candidate factors depends mainly on economic logic and market experience. Obviously, the highly effective factor is undoubtedly one of the key factors to enhance the ability of capturing model information and increase revenue. According to the common 12 factor classifications, six major factors are selected randomly: quality factor, growth factor, value class factor, common technology class factor, momentum class factor, and emotion class factor [23]. (The above factor classification data can be found in dot wide network data dictionary BP factor, and A share market stock factor data is downloaded in the DigQuant quantization community). The specific factors are shown in Table 1.

B. Data pretreatment

(1) Eliminating the extremum

Let the exposure sequence of a factor in the first period on all stocks $\tilde{x}_i$, x_M is the median of this sequence $\tilde{x}_i$, D_{MAD} is the median of the sequence $\mid x_i - x_M \mid$.

$$\tilde{x}_i = \begin{cases} x_M + n * D_{MAD}, & if x_i > x_M + n * D_{MAD}; \\ x_M - n * D_{MAD}, & if x_i < x_M + n * D_{MAD}; \\ x_i, & else. \end{cases} \tag{1}$$

(2) Missing value processing

After obtaining the new factor exposure sequence, the missing value is set as the average value of the same stock.

(3) Standardization

The factor exposure sequence is subtracted from its current mean and divided by its standard deviation to obtain a new sequence $N(0,1)$ which is an approximate distribution.

III. THEORETICAL MODEL

A. Multi-factor stock selection strategy based on equal weight linear model

By scoring the factors, the weights of the factors are given equally, and the way of adding up the scores screened from the stock pool. The specific operation is as follows:

(1) Choosing factor j, ranking the stocks according to the size of factor j, and using the serial number as the score. Among them, the positive index ranks in descending order, the reverse index ranks in ascending order, and the weight of each factor is expressed by k.

(2) Taking the ordinal number as the final score of the stock:

$$Q_i = k \sum_j s_{ij} \tag{2}$$

where s_{ij} is the score of the i-th stock under factor j.

(3) According to the size of the score, the stocks in the stock pool are sorted from large to small and divided into 15 groups. The group with the highest score is selected as our portfolio.

(4) Using "Auto-Trader Strategy Research Return Engine" to back-test the strategy.

B. Multi-Classification Logistic Regression Algorithm

The traditional logistic regression model is always used binary classification tasks, such as a persons gender (male or female). In this section, we introduce the way to adapt the traditional logistic regression model into multiple classification task.

(1) Logistic distribution definition

Let X be a continuous random variable and X obeys logistic distribution, which means that X has the following distribution function and density function:

$$F(x) = P(X \leq x) = \frac{1}{1 + e^{-(x-\mu)/\gamma}} \tag{3}$$

$$f(x) = F'(x) = \frac{e^{-(x-\mu)/\gamma}}{\gamma(1 + e^{-(x-\mu)/\gamma})^2} \tag{4}$$

where μ is the position parameter and $\gamma > 0$ is the shape parameter.

(2) Logistic regression of multiple classifications

Assuming that the set of values of discrete random variables Y is $\{1, 2, \cdots, K\}$. Then multi-classification logical regression model can be described as:

$$P(Y = k \mid x) = \frac{exp(\omega_k \cdot x)}{1 + \sum_{k=1}^{K-1} \exp(\omega_k \cdot x)}, \tag{5}$$
$$k = 1, 2, \cdots, K-1$$

$$P(Y = k \mid x) = \frac{1}{1 + \sum_{k=1}^{K-1} \exp(\omega_k \cdot x)} \tag{6}$$

Where $x \in R^{n+1}, w_k \in R^{n+1}$.

C. Multi-classification XGBoost algorithm

Chen et al. [20] developed the XGBoost algorithm, which is an extension of gradient boosting decision tree (GBDT) algorithm. GBDT algorithm is an implementation of boosting algorithm. Its idea is to continuously reduce the residual of the previous model and make the residual of the previous model decrease in the gradient direction, consequently obtain a new model. Compared with GBDT which only uses the first derivative information, XGBoost expands the loss function in the second order Taylor expansion, makes full use of the first and second derivatives and finds the optimal solution for the regular term outside the loss function. The steps of multi-classification XGBoost algorithm are as follows:

(1) Optimizing objectives

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{7}$$

where x_i is the i-th input sample, $\hat{y}_i$ is the predicted value calculated by mapping relation f, F is a set of all mapping relationships.

(2) Optimizing objective and loss function

$$L(t) = \sum_{i=1}^{n} l(y_i,\ \hat{y}_i{}^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{8}$$

$L(t)$ is the objective function of the t-th iteration, n is the total number of samples, l is a loss function, $\hat{y}_i{}^{(t-1)}$ is the predicted value of the model at $(t-1)$-th iteration. $f_t(x_i)$ is a newly added function, $\Omega(f_t)$ is regularization term.

(3) Formula (8) can be obtained by second-order Taylor expansion and removal of constant terms

$$\bar{L}(t) \cong \sum_{j=1}^{T} \left[(\sum_{i \in I_j} g_i)\omega_j + \frac{1}{2}(\sum_{i \in I_j} h_i + \lambda)\omega_j^2 \right] + \gamma T \tag{9}$$

where I_j is the set of samples on each leaf node j, g_i is the first derivative of l to $\hat{y}_i{}^{(t-1)}$, h_i is the second derivative of l to $\hat{y}_i{}^{(t-1)}$, T is the number of leaf nodes, λ and γ are specific gravity coefficients to prevent over-fitting.

(4) Computing the optimal weight ω_j^* of leaf j and the corresponding optimal function value, i.e.

$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{10}$$

$$\bar{L}(t) = -\frac{1}{2}\frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{11}$$

D. Multi-factor stock selection strategy based on two machine learning algorithms

As shown in Fig. 1, the construction method of multi-factor stock selection model based on multi-classification logistic regression algorithms and multi-classification XGBoost algorithm includes the following steps:

(1) Collecting data

1) Selecting stock pool: the CSI-300 stocks are selected as the stock pool, with the exception of ST stocks and stocks that are suspended from trading on the next trading day of each cross-sectional session. The initial capital is 10 million, and the handling fee is 3 thousandths of bilateral. Positions are adjusted at the end of each month.

2) Determing sample interval: section periods from 2016-01-01 to 2018-09-30.

(2)Training sample

A collection of factor data for the last trading day of each month prior to this month.

(3) Establishing sample labels

Establishing sample labels which are shown in Table 2.

(4) Predicting sample

Factor data from the last trading day of the month was used as the prediction sample for the multi-classification logistic regression model. Through the forecasting results, we can find that the stocks with a rise margin greater than 1 for logistic regression model and greater than 5 for XGBoost model can be purchased as warehousing.

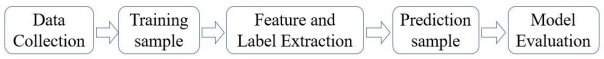

Fig. 1. The process of stock selection strategy execution

TABLE II
LARGE SAMPLE LABEL

Closing price difference (x)	Label(y)
$x < 0$	0
$0 < x \leq 1$	1
$1 < x \leq 5$	2
$0 < x \leq 10$	3
$0 < x \leq 15$	4
$0 < x \leq 20$	5
$x > 20$	6

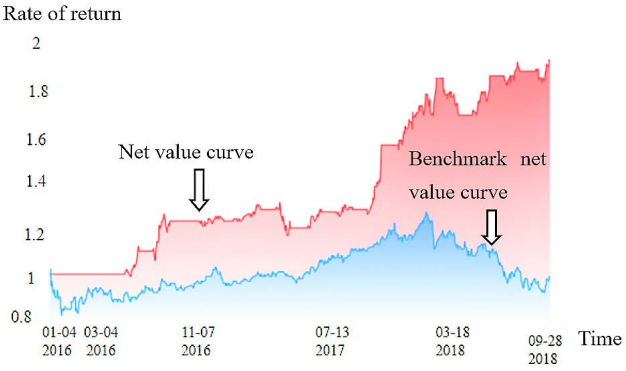

Fig. 2. Contrasts of strategy/indexs net value base on Logical Regression Algorithm

IV. Empirical Analysis

Using "Auto-Trader Strategy Research Retrospective Engine" to back-test the strategy, we can obtain the results which are shown in Fig. 2, Fig. 3 and Fig. 4 according to the multi-factor stock selection strategy of three dfferent algorithms. According to the performance ratio of stock selection strategies of the above three methods, we can also obtain the result which is shown in Table 3.

From the three figures, we can see the net value of the strategy is greater than the baseline net value. From

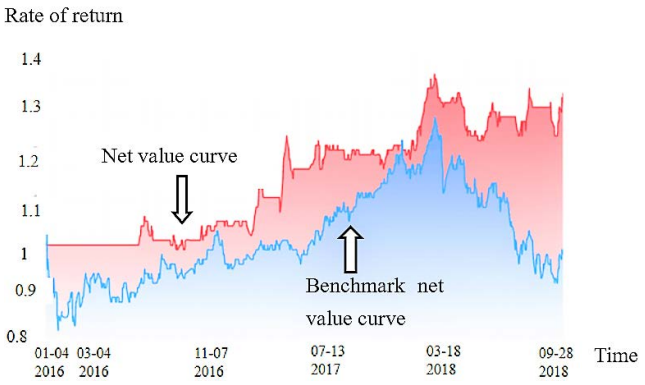

Fig. 3. Contrasts of strategy/indexs net value base on XGBoost Algorithm

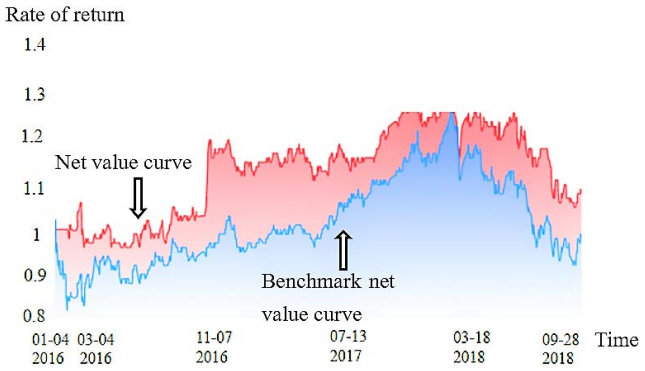

Fig. 4. Contrasts of strategy/indexs net value base on equal weight linear model

TABLE III
LARGE SAMPLE LABEL

Policy name	Logistic	XGBoost	Linear
Cumulative income	99.33%	31.62%	8.7%
Annualized rate of return	28.09%	10.87%	3.18%
Benchmark rate of return	-0.87%	-0.87%	-0.87%
Calmar ratio	2.46	0.84	0.06
Sortino ratio	3.11	1.21	0.11
Sharpe ratio	1.79	0.76	0.08
Information ratio	1.55	0.79	0.42
Maximum retracement	10.62%	10.50%	21.10%
Alpha	0.25	0.10	0.04
Beta	0.21	0.24	0.51

Table 3, we can see logistic regression algorithm has the highest cumulative income, and its annualized rate of return is 28.09%. The annualized rate of return for the equal weight linear model is only 3.18%.

The sharpe ratios are compared from logistic algorithm, XGBoost algorithm and equal weight linearity. As we can see, sharpe ratio of logistic regression algorithm is the highest. The sharpe ratio of these three strategies are all positive, which show that the average yield of products in the measurement period exceeds the risk-free rate. At the same time, the stock selection strategy of multi-classification logistic regression algorithm maximizes the unit risk return of products.

V. Risk Management

In recent years, one of the most important developments in the field of quantitative investment has been the focus on risk management [24], [25], [26]. Risk control means that risk managers take various measures and methods to eliminate or reduce the various possibilities of risk events. Thus, in order to make the stock selection strategy obtain more robust excess return, a structured risk model is established to control the risk [3]. It is required to control the maximum withdrawal within 10% and re-compare stocks selection strategies of two different machine learning algorithms.

A. General form of structured risk model

The core idea of the structured risk model is to explain the stock excess return rate with several common factors and a specific factor that only relates to an individual stock.

$$r_i(t) = \sum_k x_{ik}(t) f_k(t) + \mu_i(t) \tag{12}$$

$r_i(t)$ is the excess return rate of stocks from time t to time $t+1$, $x_{ik}(t)$ is the factor exposure of factor k at time t, $f_k(t)$ is the rate of return of factor k from time t to time $t+1$, $\mu_i(t)$ is the rate of return of specific factor k from time t to time $t+1$.

The excess return rate of the portfolio Q consisting of n stocks:

$$R_Q = \sum_{i=1}^{n} \omega_i \left(\sum_{j=1}^{k} x_{ij} f_{ij} + \mu_i \right) \tag{13}$$

where the weight of stock is $\omega = (\omega_1, \omega_2, ..., \omega_n)^T$, R_Q is the overall rate of return of Q , x_{ij} is the factor exposure of stock i on factor j, f_{ij} is the rate of return of stock i on factor j, μ_i is the rate of return of the i-th stock with a specific factor.

The risk structure of portfolio Q is:

$$\sigma_Q = \sqrt{\omega^\top (XFX^\top + \delta) + \omega} \tag{14}$$

where X is a factor exposure matrix of n stocks on a factor k.

Hence, F is the covariance matrix of k factors.

$$F = \begin{bmatrix} \mathrm{var}(f_1) & \mathrm{cov}(f_1, f_2) & \cdots & \mathrm{cov}(f_1, f_k) \\ \mathrm{cov}(f_2, f_1) & \mathrm{var}(f_2) & \cdots & \mathrm{cov}(f_2, f_k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(f_k, f_1) & \mathrm{cov}(f_k, f_2) & \cdots & \mathrm{var}(f_k) \end{bmatrix} \tag{15}$$

Here, Δ is a diagonal matrix of non-factor return variance.

$$\Delta = \begin{bmatrix} \mathrm{var}(\mu_1) & 0 & \cdots & 0 \\ 0 & \mathrm{var}(\mu_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathrm{var}(\mu_k) \end{bmatrix} \tag{16}$$

B. Model establishment

In order to control the risk of stock selection strategy, it is required to control the maximum retracement within 10%. So we can establish an optimization model as following:

$$\max \frac{R_Q - TC(\omega)}{\sigma_Q} \tag{17}$$

$$s.t. \begin{cases} \frac{\max(D_i - D_j)}{D_i} \leq 0.1 \\ \sum_{i=1}^{n} \omega_i = 1 \end{cases} \tag{18}$$

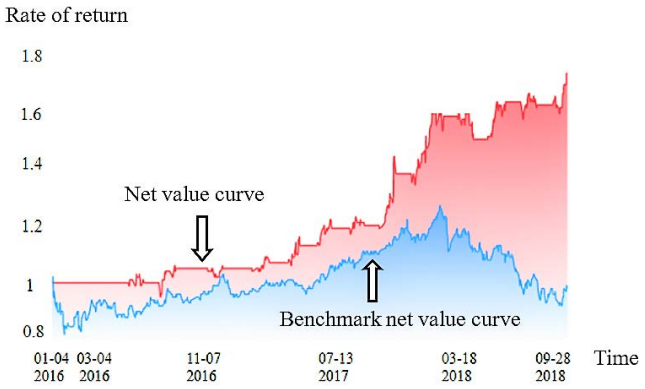

Fig. 5. Logistic–net value curve and accumulated return rate after risk control

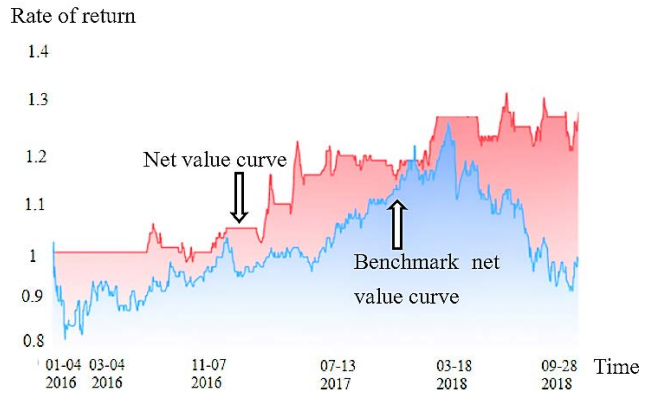

Fig. 6. XGBoost–net value curve and accumulated return rate after risk control

TABLE IV
LARGE SAMPLE LABEL

Policy name	Logistic	XGBoost
Cumulative income	73.3%	28.7%
Annualized rate of return	22.94%	9.94%
Benchmark rate of return	-0.87%	-0.87%
Calmar ratio	2.83	0.77
Sortino ratio	2.48	1.06
Sharpe ratio	1.42	0.67
Information ratio	1.28	0.73
Maximum retracement	7.40%	9.96%
Alpha	0.20	0.09
Beta	0.20	0.23

C. Model Solution

We solve the above single-objective programming model through Python software, control the risk of stock selection strategy, and re-test the two different machine learning algorithms. The results are shown in Fig. 5, Fig. 6 and Table 4.

Table 4 shows that the maximum retracement rates of these two machine learning algorithms decreased to 7.40% and 9.96%, respectively; the sharpe ratio in the stock picking strategy of the multi-classification logistic regression algorithm is still the largest with 1.42, while the sharpe ratio of the XGBoost algorithm is 0.67.

VI. Conclusion

The essence of quantitative investing is to quantify the logic of traditional investing by using a computer, which is characterized with discipline and programmability. In this paper, we select the most commonly used multi-factor models in quantitative investing and establish a multi-factor stock selection model based on machine learning. Firstly, taking CSI-300 stock as the research objective, the theoretical framework of machine learning multi-factor stock selection model is established. Within this framework, we present equal-weighted linear model, multi-classification logistic regression algorithm and multi-classification XGBoost algorithm to obtain the initial portfolio annualized sharpe ratio, and find that the annualized sharpe ratio obtained by multi-classification logistic regression algorithm is the highest. In order to control the risk of the stock picking strategy and keep the maximum reversion within 10%, a risk model is developed in this paper. Through theoretical analysis and numerical experiments, the maximum retreats of multi-classification logistic regression algorithm and multi-classification XGBoost algorithm are reduced to 7.40% and 9.96%, respectively. The multi-classification logistic regression algorithm still has the highest sharpe ratio with an annualized sharpe ratio of 1.42, which shows that the multi-factor stock selection model based on logistic regression algorithm has good earnings prediction ability and profitability.

The further refinement and deepening of this paper will be conducted in the future in the following areas: 1) we shall establish more stock selection models to enrich stock selection strategies; 2) this paper only analyzes the sample as a whole, not the transition of different markets styles, thus subsequent studies will discuss this situation in detail to test whether this method is a good choice.

References

[1] B. G. Malkiel and E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
[2] R. R. Chen N and R. S., "Economic forces and the stock market," *Journal of Business*, vol. 59, pp. 383–403, 1986.
[3] T. YM and T. S, "Adaptive use of technical indibcators for the prediction of intra-day stock prices," *Physica A*, vol. 383, no. 1, pp. 125–133, 2007.
[4] C. C. Chen TL and T. HJ, "High-order fuzzy time series based on multi-period 8 adaptation model for forecasting stock markets," *Phys A*, vol. 387, no. 4, pp. 876–888, 2008.
[5] T. Y. P. Liang O, "Application of quantified investment in futures market. innermongolia coal economy," *Computer Simulation*, pp. 81–82, 2018.
[6] S. A. Ross, "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, vol. 13, no. 3, pp. 341–360, 1976.
[7] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.
[8] Tortoriello and Richard, "Quantitative strategies for achieving alpha," 2009.
[9] Asness and C. S, "The interaction of value and momentum strategies," *Financial Analysts Journal*, vol. 53, no. 2, pp. 29–36, 1997.
[10] R. La Porta, F. Lopezdesilanes, A. Shleifer, and R. W. Vishny, "Law and finance," *Journal of Political Economy*, vol. 106, no. 6, pp. 1113–1155, 1998.
[11] P. M. Dechow and R. G. Sloan, "Returns to contrarian investment strategies: Tests of naive expectations hypotheses," *Journal of Financial Economics*, vol. 43, no. 1, pp. 3–27, 1997.
[12] Piotroski and D. Joseph, "Value investing: The use of historical financial statement information to separate winners from losers," *Journal of Accounting Research*, vol. 38, p. 1, 2000.

[13] P. S. Mohanram, "Separating winners from losers among lowbook-to-market stocks using financial statement analysis," *Review of Accounting Studies*, vol. 10, no. 2-3, pp. 133–170, 2005.
[14] J. Lakonishok, A. Shleifer, and R. W. Vishny, "Contrarian investment, extrapolation and risk," *Journal of Finance*, vol. 49, pp. 1541–1578, 1994.
[15] Y. Yamaguchi, Y. Miyachi, and Y. Shimoda, "Stock modelling of hvac systems in japanese commercial building sector using logistic regression," *Energy and Buildings*, vol. 152, no. oct., pp. 458–471, 2017.
[16] Z. S. Hongxing Y, "Research on wavelet neural network method in stock market forecasting," *Journal of Industrial Engineering and Engineering Management*, vol. 16, no. 2, pp. 32–37, 2002.
[17] Yeh, I-Cheng, Liu, and Yi-Cheng, "Using mixture design and neural networks to build stock selection decision support systems," *Neural Computing and Applications*, vol. 28, no. 3, pp. 1–15, 2017.
[18] E. Hadavandi, H. Shavandi, and A. Ghanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting," *Knowledge Based Systems*, vol. 23, no. 8, pp. 800–808, 2010.
[19] C. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Applied Soft Computing*, vol. 12, no. 2, pp. 807–818, 2012.
[20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," 2016.
[21] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models," *Expert Systems With Applications*, vol. 103, pp. 25–37, 2018.
[22] L. Wenxing and L. Junqi, "Improvement of semi-supervised kernel clustering algorithm based on multi-factor stock selection," *Statistics & Information Forum*, vol. 33, no. 03, pp. 30–36, 2018.
[23] C. Tsai, Y. Lin, D. C. Yen, and Y. Chen, "Predicting stock returns by classifier ensembles," *Appl Soft Comput*, vol. 11, no. 2, pp. 2452–2459, 2011.
[24] Y. W and F. G, "Stock market prediction based on support vector machines," *Computer Simulation*, vol. 23, no. 11, pp. 256–258, 2006.
[25] R. TH, "Forecasting the volatility of stock price index," *Expert Syst Appl*, vol. 33, no. 4, pp. 916–922, 2007.
[26] X. W, C. Y, C. C, and C. T, "Moment matching machine learning methods for risk management of large variable annuity portfolios," *Journal of Economic Dynamics and Control*, vol. 87, pp. 1–20, 2018.

Yihua Zhong was born in Jianyang, Sichuan, P.R.China in 1965. She is a professor and doctor. She is the author of four books, more than 60 articles. Her research interests include issues related to multi-criteria and multiple constraint-level programming, network optimization, fuzzy mathematic, Bayes network, support vector machine, intelligent prediction, data analysis and processing, risk analysis, petroleum engineering computation, etc.. She is the reviewers of the several international journals.