Comprehensive Evaluation of Public Funds Based on Principal Component Analysis

Zhuo Zheng[2], Xuejun Jiang[3] and Yudan Zou[1]

[2]Department of Wealth Management, Chinalin Securities Co., Ltd., Shenzhen, China.
[3]Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, China.
[4]Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China.

Contributing authors: Xuejun Jiang, jiangxj@sustech.edu.cn;

Abstract

In this paper we conduct comprehensive analysis and evaluation on public funds based on principal component analysis model. In the beginning, eight representative original evaluation indexes are selected to determine the principal components. On this basis, the weight of each evaluation index is calculated according to the estimation of the factor loading matrix. For individual index, different scores are assigned according to the percentile segment, then the comprehensive scores are calculated by weighted average. Based on the comprehensive scores, star rating is carried out, which directly reflects the quality of the public fund. Through empirical analysis, it is found that only few funds are rated as 5-star, a small number as 4-star and 3-star, and most of the funds are 1-star or 2-star, which is consistent with the actual market. To verify the effectiveness of the proposed comprehensive evaluation method, we first compare the performance of different star-rated funds with the China Securities Index (CSI) 300 index and similar funds in terms of the average excess return, Sharp index, maximum drawdown, and cumulative abnormal return. Then we test whether the fund comprehensive score plays a significant role in predicting the future return of the fund by establishing a fixed effect model.

Keywords: Public Funds, Principal Component Analysis, Weight Allocation, Comprehensive Evaluation

Mathematics Subject Classification: 62J20

1 Introduction

It has been 21 years since the establishment of China's public funds industry. Now there are nearly 6000 public funds in the whole Chinese financial market, with a total scale of over 1.37 trillion RMB. The characteristics of low starting point, strong liquidity, rich product lines, strict supervision and information transparency have made public funds attract the attention of investors and become one major asset allocation tool for the public.

With the continuous development and opening of China's securities market, the investment scope of public funds is increasingly rich. At the early stage, there were only closed-end funds which invested in the A-share market. Now there are all kinds of opened-end varieties which covers the A-share investment, bond investment, bulk commodity investment, gold investment, Hong Kong stock, and US stock investment. By September 26, 2019, the number of the top three most common funds reached a total of 5221. The number of hybrid funds was 2478, ranked first, the number of bond funds was 1745 in total, ranked second, and the number of stock funds was 998 in total, ranked third. Facing such a large number of funds in the market, the public investors often do not know how to choose excellent target products, so it is particularly important to build a reasonable fund performance evaluation system.

The traditional fund performance evaluation only measures the fund's cumulative net return without considering the risk that the fund bears. This kind of fund evaluation method is relatively simple and intuitive, and is effective for funds with small risk, such as monetary fund. However, it has some defects for evaluating stock funds due to large risk. The adjusted index based on the mean-variance model makes up for this defect, so that funds can be compared at the same risk level.

Sharp in [1] proposed to use the ratio of fund return to risk, i.e. Sharp index, as the measurement index. If the Sharp index of a fund is greater than 0, it is better to use money to purchase the fund than to deposit into the bank. If the Sharp index of a fund is greater than the average Sharp index of similar funds, it means that the performance of the fund is better than the market average. Simply speaking, the larger the Sharp index, the better the fund performance. In [2], Treynor proposed that the fund manager could eliminate the market non-systematic risk by constructing securities portfolio, then all the risks related to the change of fund performance should be systematic risk. The Treynor index is the excess return per unit of system risk assets. Jensen (1968)[3] revealed that the performance difference between the fund and the market comes from the excess return brought by the fund undertaking non-systematic risk, which is the so-called Jensen Index and depends on the investment ability of the fund manager. If the Jensen index of the fund is positive, it indicates that the fund manager has extraordinary investment ability and can outperform the market level. If it is negative, it indicates that the fund manager's investment ability is poor and lags behind the market. The Calmar ratio represents risk by the maximum drawdown rather than the volatility of the fund, which is defined as the ratio of the portion of the funds return that exceeds the risk-free return to the maximum drawdown. Therefore, Calmar ratio is of great reference significance for investors who are very concerned about the maximum drawdown of the fund.

Stock selection and timing ability are also important indicators of fund performance evaluation. Treynor and Mazuy (1966)[4] proposed the famous quadratic regression model, T-M model, which firstly introduced the evaluation of fund's stock selection and timing ability. However, the T-M model only uses one β coefficient to measure the timing ability of the fund and does not distinguish different market trends. Henriksson and Merton (1981)[5] proposed the H-M model, which takes two β coefficients to measure the fund's timing ability in bear market and bull market respectively. Henriksson (1984)[6] updated the H-M model and proposed the C-L model, which clearly distinguishes the market into long market and short market, and distinguishes the market from different cycle risks, and has a stronger applicability.

Fama and French (1993)[7] first proposed the famous three-factor model, which effectively performs attribution analysis on fund performance and finds out the key factors that bring about fund performance. So far, the three-factor model is still widely used in the market. Carhart (1997)[8] further observed the performance of mutual funds and found that the continuity of the performance of the fund manager added a momentum factor, thus constructed a four-factor model, which can reflect the investment strategy of the fund from four aspects and is widely accepted by the research community up to now. Fama and French (2015)[9] studied five-factor model and found that it performed better than the three-factor model of Fama and French (1993)[7] at capturing the size, value, profitability, and investment patterns in average stock returns. However, none of the three-factor model, four-factor model and five-factor model are comprehensive enough to explain the difference in stock returns caused by the company's profit and investment mode. Therefore, we need a comprehensive evaluation method that can incorporate more evaluation factors.

Principal component analysis method and factor analysis method can choose a set of indicators to reflect the characteristics of the research object and objectively assign corresponding weights according to the contribution rates, so it is widely used in the research of performance evaluation. Lin and Chiu (2013)[10] combined principal component analysis and network data envelopment analysis to improve banking performance evaluation by using data on Taiwanese banks. Li et al. (2004) [11] used principal component analysis to select five factors of profitability, asset management capability, debt repayment capability, capital structure and operation development capability from the 16 financial indicators of listed companies, then used the combination of subjective empowerment and objective empowerment based on factor analysis to evaluate the company's business performance. Sha and Gao (2019)[12] compared the commonly used factor models in empirical asset pricing studies and found that Fama and French (2015)[9] five-factor model outperforms other models in the Chinese mutual fund industry and in most fund segments. Zhang (2016)[13] used factor analysis to conduct an empirical research on the performance of hybrid funds, general index funds and themed index funds. Lee, Poon, and Song (2007)[14] studied Bayesian analysis of the factor model with financial applications. For more references on principal component analysis and factor analysis with application in economics and finance, see, for example, Gao and Hu (2009)[15], Mende and Proao (2017)[16], Pan et. al., (2017)[17], Bai et. al., (2018)[18], and Haugh and Lacedelli (2020)[19], among others.

This paper proposes a comprehensive evaluation method for fund performance with principal component analysis method. We select the most representative eight evaluation factors from the market as the original indexes and use principal component analysis method to find the common factors that affect the performance of the fund. On this basis, the corresponding weight of each evaluation index is calculated. For individual index, different scores are assigned according to the percentile segment, then the comprehensive scores are calculated by weighted averages. Based on the scores, star rating is carried out, which directly reflect the quality of the public funds. To verify the effectiveness of the comprehensive scoring and star rating, on one hand, we can compare the performance of different star funds with CSI 300 index and similar funds in terms of the average excess return rate and the cumulative abnormal return rate. On the other hand, we can construct a fixed-effect model to test whether the fund's comprehensive evaluation score has a significant effect on predicting the future excess return of the fund. The above findings all support the effectiveness of our comprehensive scoring and star rating.

The rest of this paper is organized as follows. In Section 2, we introduce theoretical analysis of the performance evaluation of public funds. It introduces some existing index evaluation methods and principal component analysis methods, then analyzes their advantages and disadvantages. In Section 3, the comprehensive scoring system of public funds is proposed. The most representative 8 indexes are selected based on the existing evaluation indexes of the market. According to the historical data, the weight of each index is calculated by principal component analysis method, and the comprehensive scores and stars rating of the target funds are obtained, and the results are verified. In Section 4, we give a summary of this article on fund management.

2 Analysis of Performance Evaluation Index of Public Funds

2.1 Traditional Performance Evaluation Indexes

Traditional performance evaluation mainly starts from the most intuitive angle, only uses the amount of returns to measure the quality of the fund without paying attention to other indicators such as risk, volatility, etc. This type of evaluation indexes includes the growth rate of fund unit net worth (unw):

$$growth\ rate\ of\ unw = \frac{unw\ on\ the\ last\ day - unw\ on\ the\ first\ day}{unw\ on\ the\ first\ day}$$

where unw refers to the fund's unit net worth. This index is very easy to understand, and the data is very easy to obtain. However, this method does not consider the factors such as dividends, the impact of huge redemption of funds on the net value, etc.

2.2 Evaluation of Fund Risk Level

Fund risk evaluation system is divided into qualitative and quantitative indicators. The quantitative indexes include standard deviation, downward standard deviation and maximum drawdown. The risk and yield of a fund are like two sides of a coin. When investors choose a fund, in addition to having expectation for return, fund risk is also an important indicator in their fund investment decision-making. The following are some indicators and meanings mainly used for risk assessment:

(1) Standard Deviation: It is a commonly used index to measure fund risk. The formula of standard deviation is

$$\sigma = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (R_i - \bar{R})^2},$$

where R_i represents the yield of the i-th security, $\bar{R}$ represents the average yield of N secuities. The larger the standard deviation, the worse the performance stability of the fund.

(2) β Coefficient: It is widely used in various models, such as T-M and H-M models, the specific formula of its is:

$$\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)}$$

where R_m represents the market return, β_i reflects the volatility of fund i relative to the overall market. β coefficient is a relative index, the higher the beta coefficient, the greater the volatility of the fund relative to the performance evaluation benchmark. In other words, the higher the β, the greater the risk.

(3) Maximum Drawdown Rate: the ratio of the difference between the lowest net value and the highest net value to the highest net value in the selected period. The maximum drawdown rate is generally a negative number, which is mainly used to describe the worst situation in the trend. It is an indicator that investors pay close attention to when evaluating the risk of the target fund, which is also intuitive and easy to understand.

(4) Annualized Volatility: the observation time interval is divided into several sample intervals according to the specified period, and then the standard deviation of the yield is calculated and annualized to obtain annualized volatility. Annualized volatility is also an important indicator of fund risk. The higher the general volatility, the weaker the risk control ability of the target fund.

2.3 Risk Adjusted Return Evaluation

The core of the evaluation is not just looking at the return of the fund, but to incorporate risk factors, and analyze the performance of the fund by comparing the return with different risks.

As mentioned earlier, the traditional growth rate of fund unit net worth is relatively intuitive, but it can only be used in products with low risk, such as monetary fund and short-term financial products, but it is improper for risky varieties. The vertical comparison of different types of funds can be achieved through risk adjusted return evaluation indicators, which generally include Sharp index, Calmar ratio and Jensen index.

(1) Sharp Index: an index proposed earlier to reflect the comparison of return and risk, and its calculation formula is

$$S_i = \frac{R_i - R_f}{\sigma_i} \tag{1}$$

where the sharp index S_i denotes the excess return that the fund i can obtain for each unit of total risk; R_i represents the average growth rate of unit net worth of the fund i, R_f is the risk-free rate of return, and σ_i represents the standard deviation of the fund i. For stock funds and hybrid funds, the calculation period is generally selected to be years, but for high-frequency trading with $T + 0$, the calculation period can be days or weeks.

The sharp index is very intuitive and easy to compare. Generally, if the sharp index of a single fund in the same time interval is greater than the average value of similar funds, it means that the profitability of this fund is better than the market average. The Sharp index can also be used for comparision among different types of funds. In general, the larger the Sharp index, the stronger the profitability of the fund. The difference between sharp index and other indicators is that the risk selected by sharp index is the total risk including both systematic risk and non-systematic risk, and it is one of the most widely used indicators reflecting fund performance.

(2) Calmar Ratio: an index to describe the relationship between yield and maximum drawdown. Its definition is

$$Calmar\ ratio = \frac{annualized\ rate\ of\ return}{|maximum\ drawdown|}, \tag{2}$$

where the annualized rate of return is calculated in a statistical interval. It is noted that Calmar ratio is different from sharp index in the choice of risk measurement, which uses the maximum withdrawal as the risk measurement value, the larger the better. Compared with sharp index (1), Calmar ratio (2) is more suitable for investors with conservative risk preference.

(3) Jensen Index: an index which studies the difference between the real return rate of the fund and the expected return rate calculated by the CAPM model. The Jensen index of fund i is defined as

$$J_i = R_i - [R_f + \beta_i \times (R_m - R_f)], \tag{3}$$

where R_i is the historical average rate of return of fund i during the study period, R_m is the historical average rate of return for the market, R_f is the risk-free rate of return, and β_i is the systematic risk that the fund i bears during the study period. The Jensen index is the excess return obtained by the fund taking on the non systematic risk. It is widely used to evaluate the benefit of index enhanced fund, such as smart ETF fund. If the Jensen index is larger than 0, the performance of the fund is better than that of the market benchmark portfolio, and the larger the index, the better the performance.

Comparing the above three indexes, we can find that Sharpe index (1) and Calmar ratio mainly (2) measure the relative performance of the fund, while Jensen index (3) measures the absolute performance. In other words, Jensen index ignores the number of stocks in the target fund's investment scope which is called as fund investment breadth, and only considers how much excess return can be obtained which is called as fund portfolio depth. However, the breadth and depth should both be considered when measuring fund performance, single risk adjusted return evaluation index such as Sharpe index, Calmar ratio and Jensen index has its own limitations.

2.4 Evaluation of Fund Managers Ability

Fund evaluation cannot be separated from the measurement of fund manager's ability, which mainly refers to the timing and stock selection ability. In short, timing is the fund manager's control on fund positions in different market conditions, while stock selection is the investment manager's control over stock valuation. Therefore, these two capabilities are also the main indicators to evaluate the ability of fund managers.

(1) T-M Model:

Treynor and Mazuy (1966)[4] proposed to measure the ability of fund managers through stock selection ability and timing ability, and introduced the classical model, T-M model with quadratic term:

$$R_p - R_f = \alpha + \beta_1 \times (R_m - R_f) + \beta_2 \times (R_m - R_f)^2 + \varepsilon_p, \tag{4}$$

where R_p represents the portfolio return, the constant term α represents the stock selection ability of fund managers, the larger its value, the higher the excess return of the fund; β_1 represents the systematic risk that the fund bears, β_2 represents the market timing ability of fund managers, if β_2 is greater than 0, and the larger it is, the stronger the fund manager's market timing ability is, ε_p is random error. T-M model is the most basic model, which only uses one value of β to judge the timing ability of fund managers without distinguishing different market states.

(2) H-M Model:

Henriksson and Merton (1981)[5] improved the T-M model to obtain the H-M model. They thought that the single β value in the T-M model (4) was not comprehensive enough, so they introduced double β value to measure the fund manager's timing ability:

$$R_p - R_f = \alpha + \beta_1 \times (R_m - R_f) + \beta_2 \times (R_m - R_f) \times D$$

where D is a dummy variable, when $R_m > R_f$, $D = 1$, otherwise, $D = 0$. In other words, when the market economic cycle is in a rising state, $\beta = \beta_1 + \beta_2$; when the market economic cycle is in a declining state, $\beta = \beta_1$. Since the studied samples in this paper are hybrid funds, T-M model is used to evaluate the fund's stock selection and timing ability.

2.5 Principal Component Analysis

(1) Introduction to Principal Component Analysis

Principal Component Analysis is a statistical dimension reduction method, and its basic idea is to use a few mutually independent components to reflect the vast majority of information of original variables. Suppose that $\mathbf{x} = (x_1, x_2, ..., x_p)'$ is a $p-$dimensional random vector with mean $\mu = (\mu_1, \mu_2, ..., \mu_p)'$ and covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})$, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the eigenvalues of $\mathbf{\Sigma}$, and $\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_p$ be the corresponding eigenvectors of $\mathbf{\Sigma}$ $(j = 1, 2, \ldots, p)$, then the j-th population principal components of $\mathbf{x}$, denoted by f_j, can be formulated as

$$f_j = \mathbf{t}_j'\mathbf{x},$$

where $Var(f_j) = \lambda_j$ for $j = 1, 2, \ldots, p$. In the above definition, principal components f_j's are mutually uncorrelated. It is common practice to use the cumulative contribution rate $\sum_{j=1}^{m} \lambda_j / \sum_{j=1}^{p} \lambda_j$ to measure the degree to which the total variance is explained, where m represents the number of principal components. For a positive definite matrix, the size of its trace is often determined by a few large eigenvalues. Thus, only a few principal components are selected to make the cumulative contribution rate reach a relatively high level (e.g. 85%), so as to achieve the purpose of dimensionality reduction. Let $\mathbf{f} = (f_1, f_2, \ldots, f_p)'$ be the principal vector composed of all principal components. Then we can express it as

$$\mathbf{f} = \mathbf{T}'\mathbf{x} \tag{5}$$

in matrix form, where the coefficient matrix $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_p)$ is a positive definite matrix.

(2) Data Processing

Suppose that there are n funds to be evaluated, each fund has p indexes, and the observed indexes of the i-th fund are $x_{i1}, x_{i2}, \cdots, x_{ip}, i = 1, 2, \cdots, n$, then the observation data matrix, $\mathbf{X} = (x_{ij})_{p\times n}$, is expressed as

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}$$

Let $\bar{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$, $j = 1, \cdots, p$ be the sample mean of the j-th index of these n funds, $s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x_j})^2$ be the sample variance. In order to ensure the unity of data magnitude, we standardize the original data by $z_{ij} = (x_{ij} - \bar{x_j})/\sqrt{s_j}$. It is obvious that the covariance matrix of the standardized sample is exactly the sample correlation matrix of original data:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix},$$

where $r_{kl} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ik} - \bar{x_k})(x_{il} - \bar{x_l})/s_k s_l$ $(k = 1, \cdots, p,\ l = 1, \cdots, p)$ represents the correlation between the k-th and l-th samples with $r_{kk} = 1$.

(3) Estimation of Coefficient Matrix. Starting from the correlation matrix $\mathbf{R}$, we solve the characteristic equation $|\lambda\mathbf{I} - \mathbf{R}| = 0$ to obtain the eigenvalues and their corresponding eigenvectors, denoted as $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p, \hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2, \cdots, \hat{\mathbf{t}}_p$ respectively. Then, the j-th sample principal components $\hat{f}_j$ has the following expression

$$\hat{f}_j = \hat{\mathbf{t}}_j'\mathbf{z}$$

for $j = 1, 2, \ldots, p$, where $\mathbf{z}$ is the normalized vector of each component.

Similar to (5), the sample principal vector $\hat{\mathbf{f}} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_p)'$ is expressed as

$$\hat{\mathbf{f}} = \hat{\mathbf{T}}'\mathbf{z}, \tag{6}$$

where $\hat{\mathbf{T}} = (\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2, \cdots, \hat{\mathbf{t}}_p)$ is the estimated coefficient matrix.

(4) Comprehensive Evaluation Strategy

Given the predefined threshold α_0, when the cumulative contribution rate $\sum_{j=1}^{m} \hat{\lambda}_j / \sum_{j=1}^{p} \hat{\lambda}_j \geq \alpha_0$, m is selected as the number of principal components. Based on the first m sample principal components, we construct a comprehensive evaluation function as

$$F_z = \sum_{l=1}^{m} (\frac{\lambda_l}{\kappa}) \hat{f}_l = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_p x_p \tag{7}$$

with $\kappa = \lambda_1 + \lambda_2 + \cdots + \lambda_m$, where $\frac{\lambda_l}{\kappa}$ represents the contribution rate of the l-th principal component $\hat{f}_l$ and ω_j is the original weight of index x_j. Note we need to further normalize the weights to satisfy $\sum_{j=1}^{p} \omega_j = 1$. Finally, the comprehensive score (C-Score) of target fund is then calculated as

$$\text{C-Score} = \sum_{j=1}^{p} \omega_j y_j \tag{8}$$

where y_j is the score of the j-th evaluation index in the target fund according to scoring standards in Table A1, whose value is between 0 and 5.

3 Comprehensive Evaluation and Empirical Analysis of Public Funds

In the last section, we introduced some widely used fund evaluation indexes in the current market, however, single index has certain limitations on the overall evaluation of the fund. For example, if we only examine the return rate of the fund, whether before or after the risk adjustment, we can't effectively evaluate the overall performance of the fund. If the volatility and the maximum drawdown are large, it means that the risk control ability of the fund is poor or the performance of the fund is not sustainable. However, both the three-factor model and four-factor models are still not comprehensive enough to describe the expected rate of return. Therefore, on the basis of previous studies, this paper selects eight indicators that best represent the fund's performance for principal component analysis. See Table A1 in Appendix for the specific scoring standards of each indicator, and Table A2 in Appendix for the fund's star rating standards.

3.1 Comprehensive scoring of public funds

We randomly select 500 partial-share hybrid funds and flexible allocation funds that have been established for at least three years in the market (excluding the abnormal net value products caused by missing data or too small shares and large redemption) from Wind database of China. Besides, the 500 funds timing ability (β), stock selection ability (α), percentile ranking of similar funds, Sharp index, Calmar ratio, annualized volatility, maximum drawdown and percentile ranking of similar fund in recent two years are obtained from this database. The studied time interval is from January 1, 2017 to January 1, 2019, which is reasonable because the market has not only experienced a big rise, but also a big fall in this period.

After standardizing the original score data of 8 indexes of the selected 500 funds, we conducted principal component analysis and calculated the eigenvalues and the corresponding cumulative contribution rate of the correlation matrix. The results are shown in Table 1:

Table 1: Cumulative Contribution Rate

No.	Eigenvalue	Contribution Rate	Cumulative Contribution Rate
1	4.1392	0.5174	0.5174
2	1.4319	0.1790	0.6964
3	0.9490	0.1186	0.8150
4	0.6800	0.0850	0.9000
5	0.3623	0.0453	0.9453
6	0.2207	0.0276	0.9729
7	0.1206	0.0151	0.9880
8	0.0961	0.0120	1.0000

It can be seen from Table 1 that the cumulative contribution rate of the first four principal components has reached 90%, so we can use the first four principal components to analyze the original eight indicators.

Table 2: Factor loading matrix

Original index	1	2	3	4
Timing ability(β)	-0.0589	-0.6212	0.7619	0.0617
Stock selection ability(α)	0.8212	-0.1020	-0.3886	0.1380
ranking in last year	0.8871	-0.2624	-0.0196	0.0784
Sharpe index	0.8585	-0.1420	0.0251	-0.4223
Calmar ratio	0.8642	0.0366	0.1142	-0.4182
Annualized volatility	0.2550	0.8218	0.3459	-0.0926
Maximum drawdown	0.6316	0.4895	0.2893	0.3567
Ranking in recent 2 years	0.8524	-0.1743	-0.0115	-0.4025

Note: These first columns represent Profitability, Risk Control Ability, Management Ability and Performance Stability, respectively.

From Table 2, we can see that the first principal component has a large loading (in absolute value) on stock selection ability(α), ranking percentile of similar funds in last year, Sharp index, Calmar ratio, maximum drawdown and ranking in recent 2 years. According to the meaning of these indexes, we interpret the first principal component as 'profitability'. The second principal component has a large loading on timing ability(β), annualized volatility and maximum drawdown, which are then interpreted as 'risk control ability'. The third principal component has a great loading on timing ability(β), stock selection ability(α) and annualized votality, which mainly reflects the management ability of fund managers. Therefore, we interpret the third principal component as 'management ability'. The fourth principal component has a considerable loading on the Sharp index, Calmar ratio, maximum drawdown and ranking in recent 2 years. Therefore, we interpret the fourth principal component as "performance stability".

Through the comprehensive evaluation strategy (7)-(8), we can get the weight of each index for the comprehensive score of the Fund. See Table 3 for details:

Table 3: Weight Result

Original index	Final Weight
Timing ability(β)	-0.0060
Stock selection ability(α)	0.1104
ranking in last year	0.1321
Sharpe index	0.1078
Calmar ratio	0.1349
Annualized volatility	0.1516
Maximum drawdown	0.2104
Ranking in recent 2 years	0.1588
Sum	1.0000

In order to test the significance of the obtained weight coefficients, we calculate the above weights 500 times through resampling and calculate the $t-$value of each weight coefficient according to the sample mean and standard error, the results are in Table 4:

According to the meaning of $t-$value, the larger the absolute value of $t-$value, the more confident we are to reject the assumption that the weight coefficient of the evaluation index is zero. The $t-$value in Table 4 implies the weights of all evaluation indexes are statistically significant.

Further, the weights of four principal components, that is, the weights of profitability, risk control ability, management ability, performance stability are determined by the relative magnitude of its corresponding eigenvalues. Thus the weights of the four first level indexes are 0.5749, 0.1989, 0.1318 and 0.0944 respectively. The results show that in the comprehensive evaluation of the fund, the profitability of the fund is the most important, followed by the risk control ability, which is in line with investors' expectations of the fund. Funds with strong profitability and good risk control ability will naturally be sought

Table 4: t-test for the weight of individual evaluation

	t-value
W1	-3.6667
W2	21.1718
W3	8.2501
W4	18.8066
W5	97.4963
W6	1.9642
W7	3.6895
W8	10.1714

after by investors. Although the management ability has an impact on the performance of funds, they are not decisive in the comprehensive evaluation.

Finally, the comprehensive score of each fund based on the evaluation index can be calculated by the following formula:

$$C - Score = \sum_{i=1}^{8} \omega_i \times x_i^*$$

where ω_i's values are given in Table 3, and x_i^* is the score of the funds i-th evaluation index, whose value is between 0 and 5. Then, based on the comprehensive scoring and star rating criteria (see Table A2), we get the star rating distribution of 500 funds (see Figure 1).

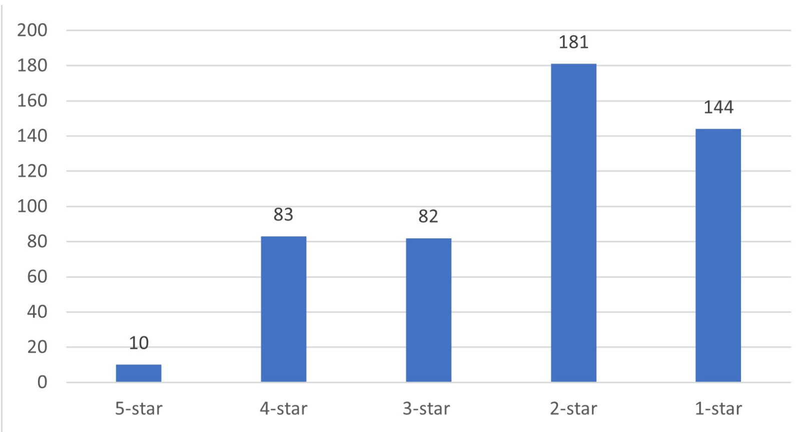

Fig. 1: star rating distribution of 500 funds

The star rating distribution of the 500 funds is basically consistent with the market performance of public funds. There are few good (5-star and 4-star) funds, only 93, accounting for 18.6%; 144 very poor funds (1-star), accounting for 28.8%; and most of the funds are mediocre (2-star and 3-star), accounting for 52.6%. This is in line with the 2-8 principle of market performance of public funds, that is to say, only few of the public funds performed well, most of the funds performed mediocrely, and a few performed poorly.

3.2 Verification of the evaluation result

In order to test the effectiveness of the comprehensive scoring and star rating method we proposed (see (7)-(8)), we selected the average excess return rate, Sharp index and maximum drawdown of each star funds during April 1 to June 30, 2019 to compare with the average of the CSI 300 index and similar funds, among which the similar funds are all the mixed partial stock funds and flexible allocation funds from the market in the same period of time. For the excess return rate, the higher the star level, the better the performance of the fund, where the latest three-month fixed deposit rate of the Bank of China is selected as the market risk-free yield. The performance of CSI 300 index and the average of similar funds are between 3-star and 2-star funds; Sharp index acts similar as excess return rate, except that the Sharp index of 2-star funds is slightly lower than that of 1-star funds and the average of similar funds; maximum drawdown acts similar as Sharp index, except that maximum drawdown of 4-star funds is slightly higher than that of the average of similar funds. See table 5 for details.

Table 5: Comparison Between Different Star Funds

	5-star	4-star	3-star	2-star	1-star	CSI300	Mean of similar funds
Yield rate	1.63%	0.63%	-0.72%	-2.20%	-2.87%	-1.21%	-1.23%
Risk free rate	0.38%	0.38%	0.38%	0.38%	0.38%	0.38%	0.38%
Excess return rate	1.25%	0.25%	-1.10%	-2.58%	-3.25%	-1.58%	-1.60%
Sharp index	0.15	-0.03	-0.08	-0.26%	-0.24%		-0.13%
Maximum drawdown	-4.34%	-10.47%	-11.78%	-14.63%	-14.26%	-13.49%	-9.51%

Finally, considering that the trading rules of the fund are different from the stock, and the short-term holding results in high redemption fee, it is unreasonable to pay attention to the daily return of the fund. Therefore, we divide the inspection period into 13 weeks, and select the cumulative abnormal return rate as the evaluation index to reflect the comparison between each star funds and the CSI 300 index, where the abnormal yield is the difference between the cumulative yield of the fund and the yield of the CSI 300 index. The optimal 5-star funds ranks first in 9 out of 13 weeks, remarkably outperforming other star funds. Note that although the net worth of 1-star and 2-star funds performed well in a few weeks, they are probably not better than other star funds in terms of risk control and profitability. See Figure 2 for details.

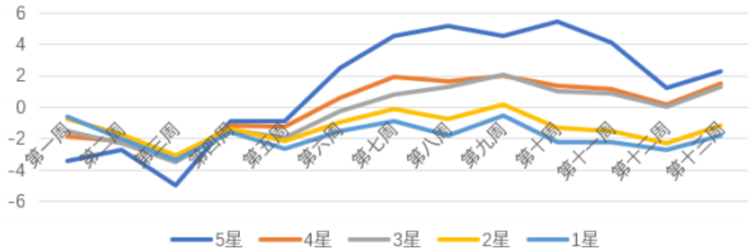

Fig. 2: comparison on cumulative abnormal return rate

3.3 forecast future excess return with comprehensive score

Furthermore, in order to test that the comprehensive score proposed in this paper has a significant effect on the prediction of the excess return rate, we take the excess return rate of the fund as the dependent variable, the comprehensive score of the fund and eight evaluation indexes as the independent variables to establish a linear model. Note that we also introduce classification variables to distinguish different star funds. The validation data is selected from January 3 to April 4, 2019, involving 60 cross-sectional data in the first quarter of 2019. Therefore, we construct the following fixed effect model:

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 X_{1_{it-1}} + \beta_2 X_{2_{it-1}} + \cdots + +\beta_{10} X_{10_{it-1}} \\ +\gamma_1 FS_2 + \gamma_2 FS_3 + \cdots + \gamma_4 FS_5 + \varepsilon_{it} \end{aligned} \tag{9}$$

among which, X_1 to X_9 represent timing ability (β), stock selection ability (α), percentile ranking of similar funds, Sharp index, Calmar ratio, annualized volatility, maximum drawdown and percentile ranking of similar fund in 2017 and 2018, fund comprehensive score, FS_2 to FS_5 respectively represent 2 to 5 star corresponding to star rating results. Using R programming, the fitting results of the model are in Table 6.

It can be seen from Table 6 that the estimated coefficient of the score is positive and its $p-$value is less than 0.05, indicating that the comprehensive score has a significant effect on the prediction of the future excess return rate of the fund. That is, the higher the comprehensive score of the fund, the higher the expectation of the future excess return rate, which further indicating the effectiveness of the proposed evaluation method.

Table 6: Fitting Result

Coefficient	Estimate	Stand error	t-value	p-value(Pr>t)
β_0	-3.589	1.912	-1.878	0.061
β_1	-1.140	0.251	-4.544	7.12×10^{-6}
β_2	-2.659	0.340	-6.649	8.63×10^{-11}
β_3	0.055	0.034	1.636	0.103
β_4	-8.443	1.878	-4.495	8.89×10^{-6}
β_5	15.440	3.408	4.530	7.57×10^{-6}
β_6	4.226	0.753	5.613	3.50×10^{-8}
β_7	2.663	0.837	3.183	0.002
β_8	0.024	0.011	2.153	0.032
β_9	0.060	0.019	3.170	0.002
β_{10}	1.447	0.631	2.295	0.022
γ_1	1.376	1.373	1.002	0.317
γ_2	4.424	1.925	2.299	0.022
γ_3	0.018	2.393	0.007	0.994
γ_4	-2.914	2.895	-1.007	0.315

It should be noticed that the estimated weights vary in different periods. With the rolling of the time (by day), the calculated weight of the evaluation index also has a trend of changing. The starting and ending date of the data selected in this paper is from January 1, 2017 to January 1, 2019. We can also roll the deadline from January 1 to May 31 to get the time series of the weights, and then calculate the comprehensive score and rating of the fund every day in this interval, which would be a dynamic process. Due to the limitation of length, we did not report everyday ranking and star rating of the selected funds. It can be predicted that the star rating of each fund will not change significantly in a short rolling time, but it will fluctuate in the long run, which also explains the scarcity of five-star funds with excellent performance in the market.

4 Conclusion

According to the relevant theories mentioned above, combined with the practical experience of the public funds that have been developed at present, 500 partial-share hybrid and flexible allocation open-end funds that have been established for 3 years are randomly selected, and their performances on January 1, 2017 to January 1, 2019 are scored and rated. The specific conclusions are as follows:

(1) The comprehensive evaluation result shows that the number of the good funds are small, and the proportion of the intermediate rating is the most,

which is in line with the overall situation of the current public fund market (i.e. the $2-8$ phenomenon). This evaluation result has strong objectivity and reference value. It can be seen from the verification that the excess return rate and Sharp index of 3-star, 4-star and 5-star funds are superior to the CSI 300 and the average performance of similar funds in the same period, so it is suggested that investors should select the funds with star rating of 3-star and above.

(2) It is found that many funds have high scores on profitability, risk control, timing and stock selection ability, but more than 80% of them have low scores on performance stability. The reasons for this phenomenon include the replacement of fund manager halfway, the rotation of market style and the rapid change of hot spots in the stock market, etc. For example, growth stocks dominated in 2017, the medical sector and the technology sector tended to be better in 2018, while in 2019, the pharmaceutical sector, the consumer sector, and the 5G sector were better. Therefore, it is very important to maintain the stability of performance, which reflects the macro prediction and policy sensitivity of investment managers. Of course, there are a few excellent funds, such as Zhang Qinghua's fund (Code: 001216) ranked top 30% in two consecutive years' performance, Hu Xinwei's fund(Code: 000083) ranked top 25% in two consecutive years' performance, and the above two funds, ranked 5-star and 4-star respectively in the evaluation system of this article.

(3) The comprehensive scoring and star rating evaluation method proposed in this paper is based on the PCA. Firstly, the principal component analysis is used to determine the 4 common factors for the eight evaluation indexes of the 500 selected fund. On this basis, we obtain the estimation of the factor loading matrix, and the weight of each index is then calculated. Because the calculated weights totally depend on data, this method is objective and scientific, and the results of empirical analysis are consistent with the market performance of public funds. The performances of the selected five-star funds in profitability, risk control and stability are better than that of other star funds, the average of CSI 300 index and the same type of funds in the same period. In addition, we also constructed a fixed effect model and tested that the comprehensive score of the fund plays a significant role in predicting the future excess return, which further proves the proposed method in this paper is effective and has certain reference value for investors.

Acknowledgments. This work was supported by National Natural Science Foundation of China (11871263), and the Shenzhen Sci-Tech Fund No. JCYJ20210324104803010.

Data availability. The data are obtained from Wind database of China with the link: https://www.wind.com.cn/Default.html.

Declarations.

Competing interests. The authors declare that they have no competing interest.

Appendix A

Table A1: Scoring Standards

Index	Weight	Rules Description	Score
timing ability (β)	ω_1	$\beta \geqslant 2$	5
		$1.5 \leq \beta < 2$	4
		$1 \leq \beta < 1.5$	3
		$0.5 \leq \beta < 1$	2
		$0 \leq \beta < 0.5$	1
		$\beta < 0$	0
stock selection ability (α)	ω_2	$\alpha \geqslant 0.0005$	5
		$0.0004 \leq \alpha < 0.0005$	4
		$0.0003 \leq \alpha < 0.0004$	3
		$0.0002 \leq \alpha < 0.0003$	2
		$0.0001 \leq \alpha < 0.0002$	1
		$\alpha < 0.0001$	0
percentile ranking of similar funds	ω_3	$rank \leq 0.1$	5
		$0.1 \leq rank < 0.2$	4
		$0.2 \leq rank < 0.3$	3
		$0.3 \leq rank < 0.4$	2
		$0.4 \leq rank < 0.5$	1
		$rank > 0.5$	0
Sharp index	ω_4	$Sharp\ index \geqslant 3$	5
		$1 \leq Sharp\ index < 3$	4
		$0.5 \leq Sharp\ index < 1$	3
		$0 \leq Sharp\ index < 0.5$	2
		$Sharp\ index < 0$	0

Calmar ratio	ω_5	$Calmar\ ratio \geqslant 3$	5
		$1 \leq Calmar\ ratio < 3$	4
		$0.5 \leq Calmar\ ratio < 1$	3
		$0 \leq Calmar\ ratio < 0.5$	2
		$Calmar\ ratio < 0$	0
annualized volatility	ω_6	$0 < annualized\ volatility \leq 0.1$	5
		$0.1 < annualized\ volatility \leq 0.2$	4
		$0.2 < annualized\ volatility \leq 0.3$	3
		$0.3 < annualized\ volatility \leq 0.4$	2
		$0.4 < annualized\ volatility \leq 0.5$	1
		$0.5 < annualized\ volatility$	0
maximum drawdown	ω_7	$-0.1 \leq maximum\ drawdown < 0$	5
		$-0.2 \leq maximum\ drawdown < -0.1$	4
		$-0.3 \leq maximum\ drawdown < -0.2$	3
		$-0.4 \leq maximum\ drawdown < -0.3$	1
		$maximum\ drawdown < -0.4$	0
percentile ranking of similar fund in recent 2 years	ω_8	stable in the top 30% annually	5
		stable in the top 40% annually	4
		stable in the top 50% annually	3
		stable in the top 60% annually	2
		stable in the top 80% annually	1
		otherwise	0

Table A2: Star Rating Standards

comprehensive score	star rating
$[4,5]$	5-star
$[3,4)$	4-star
$[2,3)$	3-star
$[1,2)$	2-star
$[0,1)$	1-star

References

[1] Sharpe, W.F.: Mutual fund performance. The Journal of business (Chicago, Ill.) **39**(1), 119–138 (1966)

[2] Treynor, J.L.: How to rate management of investment funds. Harvard business review **43**(1), 63–75 (1965)

[3] Jensen, M.C.: The performance of mutual funds in the period 1945-1964. The Journal of finance (New York) **23**(2), 389–416 (1968)

[4] Treynor, J.L., Mazuy, K.K.: Can mutual funds outguess the market? Harvard business review **44**(4), 131–136 (1966)

[5] Henriksson, R.D., Merton, R.C.: On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills. The Journal of business (Chicago, Ill.) **54**(4), 513–533 (1981)

[6] Henriksson, R.D.: Market timing and mutual fund performance: An empirical investigation. The Journal of business (Chicago, Ill.) **57**(1), 73–96 (1984)

[7] Fama, E.F., French, K.R.: Common risk factors in returns on stocks and bonds. Journal of Financial Economics **33**(1), 3–56 (1993)

[8] Carhart, M.M.: On persistence in mutual fund performance. The Journal of finance (New York) **52**(1), 57–82 (1997)

[9] Fama, E.F., French, K.R.: A five-factor asset pricing model. Journal of financial economics **116**(1), 1–22 (2015)

[10] Lin, T.-Y., Chiu, S.-H.: Using independent component analysis and network dea to improve bank performance evaluation. Economic Modelling **32**, 608–616 (2013)

[11] Li: An empirical analysis of the performance of the matching combination of the acquisition company and the target company. Economic Research **6**, 96–104 (2004)

[12] Sha, Y., Gao, R.: Which is the best: A comparison of asset pricing factor models in chinese mutual fund industry. Economic modelling **83**, 8–16 (2019)

[13] Zhang, F.: Research on china's fund performance based on factor analysis. Modern Economic Information (3), 1 (2016)

[14] Lee, S.Y., Poon, W.Y., Song, X.Y.: Bayesian analysis of the factor model with finance applications. Quantitative Finance **7**(3), 343–356 (2007)

[15] Gao, Q., Hu, C.: Dynamic mortality factor model with conditional heteroskedasticity. Insurance, mathematics & economics **45**(3), 410–423 (2009)

[16] Menden, C., Proao, C.R.: Dissecting the financial cycle with dynamic factor models. Quantitative finance **17**(12), 1965–1994 (2017)

[17] Pan, J., Ip, E.H., Dub, L.: An alternative to post hoc model modification in confirmatory factor analysis: The bayesian lasso. PSYCHOL METHODS **22**(4), 687–704 (2017)

[18] Bai, W., Cao, Z.: Research on etf performance evaluation in china. Times Finance (32), 3 (2018)

[19] Haugh, M.B., Ruiz Lacedelli, O.: Scenario analysis for derivative portfolios via dynamic factor models. Quantitative finance **20**(4), 547–571 (2020)