

Construction of Valuation System with Chinese Characteristics Based on Entropy Weight Method and K-means

Chengxin Ye^{1,a}, Weiming Li^{1,b,*}, Yongxuan Ye^{2,c}, Zhaokang He^{3,d}, Wanyu Zheng^{1,e}

¹Guangzhou Maritime University, Guangzhou, Guangdong, China

²Guangdong University of Finance, Guangzhou, Guangdong, China

³Zhuhai College of Science and Technology, Zhuhai, Guangdong, China

^a13427533602@163.com, ^b2393975880@gamall.com, ^c13535585822@163.com,

^d13724101390@163.com, ^e13022098176@163.com

*Corresponding author

Keywords: Valuation System with Chinese Characteristics, Entropy-based TOPSIS Method, K-Means Cluster Analysis, Investment Strategy

Abstract: "Stock valuation" is one of the main reference factors for investment in the securities stock market. With the continuous development of China's securities market, the traditional "price-earnings ratio valuation model" is no longer adapted to the requirements of China's market, the logic of stock valuation in China's securities market needs to be redefined in order to enable investors to make better decisions. In this regard, Yi Huiman, Chairman of the China Securities Regulatory Commission (CSRC), has put forward the idea of constructing a valuation system with Chinese characteristics, so as to enable real quality stocks, i.e., those with high dividends, low valuation and good growth, to obtain more valuation and win the attention of investors. Therefore, this paper is dedicated to constructing a valuation system with Chinese characteristics and redefining the valuation model to help investors better value stocks in the Chinese stock market. After that, this paper categorizes the valuation stocks with Chinese characteristics in Shanghai and Shenzhen A-shares, and obtains two types of stocks: long-term investment type and short-and medium-term investment type, which help investors in stock investment in Chinese stock market.

1. Introduction

The core of portfolio investment focuses on the estimation of the intrinsic value of stocks based on stock market investments. Nowadays, there are many stock valuation models in the capital market, including price-earnings ratio valuation, price-net ratio valuation and discounted cash flow model, etc. However, for the Chinese stock market, it is important to combine the valuation model with the market context, taking into account the policy background under the Chinese market system and the intrinsic growth of the stock.

In 2022, Yi Huiman, Chairman of the China Securities Regulatory Commission (CSRC)[1], first

proposed the concept of "special valuation" or "valuation system with Chinese characteristics". The term "China Special Valuation" refers to a diversified valuation system formed by adding Chinese elements to general market valuation methods and integrating them to adapt to the actual situation of the Chinese market, with distinctive Chinese characteristics.

With the core elements of policy orientation, value investment, capital allocation and risk management, "China Specialized Valuation" is dedicated to improving "low valuation stocks" and "high quality growth stocks", "nationally supported industries and fields" to promote the growth of high-return, high-quality stocks[2].

Therefore, this paper is dedicated to the establishment of "Valuation System with Chinese Characteristics" and the categorization of key stocks under the system, which provides a more targeted program for the evaluation of the intrinsic value of enterprises and the market investment direction, and also provides investors with a more accurate valuation of stocks.

2. Establishment of Valuation System with Chinese Characteristics

2.1 Selection of Characteristic Indicators

Combined with the main factors affecting the valuation system with Chinese characteristics pointed out by Li Xiaorong[3] et al. in their study, this paper selects nine indicators based on the four main core elements of the valuation system with Chinese characteristics, i.e., policy orientation, value investment, risk management, and capital allocation, namely, whether or not it is a state-owned enterprise, ROE (return on equity), whether or not it is a national key project, whether or not it is an innovative science and technology-based enterprise, PE ratio, PCF ratio, PB ratio, PS ratio, and net ratio. This paper draws Characteristic Indicators of Valuation System with Chinese Characteristics Mind Map as in Figure 1.

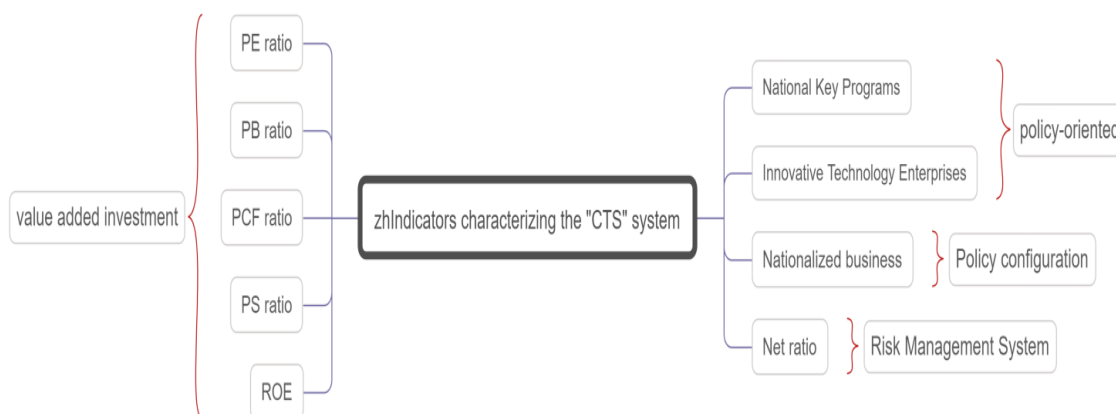


Figure 1: Characteristic Indicators of Valuation System with Chinese Characteristics Mind Map

2.2 Calculation of Characteristic Indicator Weights Using the Entropy Weighting Method

In order to make the model as objective and accurate as possible, this paper chooses to use the entropy weighting method to weight and score the characteristic indexes and hotspot coefficient indexes of the medium and special valuation stocks to derive the scores of each stock, and then use the scores of each stock to react to their short and long-term investment value.

The entropy weighting method is suitable for evaluating a limited number of decisions, comparing the information entropy of different indicators, determining the weights of the indicators,

and finally arriving at the optimal strategy[4]. Generally speaking, if the information entropy of a certain indicator is smaller, the greater the degree of variation of different decisions on the indicator on the surface, the more information it provides, the greater the role it plays, and the greater its weight [5-6].

Step1: Find the information entropy of each indicator:

The information entropy of a set of data according to the definition of information entropy:

$$E_j = -\frac{1}{\ln n} \sum_{j=1}^n P_{ij} \ln P_{ij} \quad (1)$$

Step2: Determine the weight of each indicator:

$$\omega_j = \frac{1-E_i}{k-\sum_i E_i} \quad (2)$$

Step3: The weights of the indicators are given in Table 1

Table 1: Characterization Indicator Weights

Entropy weighting		
Characteristic indicators	Information entropy	Weights (%)
ROE	0.963	4.195
Whether it is a state-owned enterprise	0.989	1.284
Whether it is a national key project	0.375	30.005
Whether it is an innovative technology-based enterprise	0.580	50.860
Net ratio	0.99	1.165
PE ratio	0.994	0.688
PCF ratio	0.924	8.657
PS ratio	0.99	1.135
PB ratio	0.982	2.011

By analyzing the weights of the indicators in Table 1, this paper shows that whether it is an innovative technology company and whether it is a national key project account for more than 80% of the components.

2.3 Exploring Stock Classification under the Valuation System with Chinese Characteristics

Table 2: Collected 75 stocks of Shanghai and Shenzhen A-shares

600007.SH	600415.SH	600938.SH	601390.SH	601808.SH	601985.SH	000065.SZ
600018.SH	600425.SH	600941.SH	601598.SH	601816.SH	603019.SH	001965.SZ
600019.SH	600502.SH	600970.SH	601600.SH	601857.SH	000498.SZ	002051.SZ
600026.SH	600585.SH	601006.SH	601608.SH	601858.SH	000519.SZ	002271.SZ
600028.SH	600685.SH	601018.SH	601611.SH	601866.SH	000617.SZ	002302.SZ
600039.SH	600705.SH	601061.SH	601618.SH	601868.SH	000708.SZ	002423.SZ
600050.SH	600760.SH	601088.SH	601668.SH	601888.SH	000831.SZ	002493.SZ
600072.SH	600886.SH	601117.SH	601669.SH	601898.SH	000923.SZ	002541.SZ
600150.SH	600900.SH	601179.SH	601728.SH	601899.SH	000928.SZ	003816.SZ
600176.SH	600905.SH	601186.SH	601766.SH	601919.SH	000932.SZ	001872.SZ
600339.SH	600916.SH	601298.SH	601800.SH	601949.SH		

This paper collects the data of characteristic indexes of 75 medium-valued stocks in Shanghai and Shenzhen A-shares, then classifies the medium-valued stocks and analyzes the characteristics of the categories, clusters the data of medium-valued stocks by using K-mean clustering analysis and derives the number of categories by using the elbow rule, and finally derives the characteristics of the categories of medium-valued stocks by analyzing the cluster center coordinate and combining with the related domain information.

The 75 stocks of Shanghai and Shenzhen A-shares collected in this paper are shown in Table 2.

2.4 Data Pre-processing

2.4.1 Data Coding

The data, because whether the state holding, whether the national key projects, whether innovation and science and technology categories are fixed type data, in order to convert them into indicators quantitative data, the fixed type data into quantitative data using the form of 0-1 coding, in which "1" represents affirmative, "0" represents negative.

2.4.2 Data Standardization

Because the selected data unit is not uniform, so the data in addition to the influence of the numerical value should be considered in addition to the influence of the scale in order to eliminate the influence of the scale of the data on the standardization process. Different standardization methods should be used for data with different positive and negative orientations to ensure that the processed data not only eliminates the effect of the scale, but also eliminates the effect of positive and negative orientations.

$$x_i = \begin{cases} \frac{x_i - \min x_i}{\max x_i - \min x_i} & \text{(Normalization)} \\ \frac{\max x_i - x_i}{\max x_i - \min x_i} & \text{(Negativization)} \end{cases} \quad (3)$$

2.5 Stock Classification under Valuation System with Chinese Characteristics

2.5.1 K-means Classification Modeling and Solution

In this paper, the collected stocks are classified, and this paper carries out unsupervised K-means clustering analysis on the collected 75 stocks based on the characteristic indexes of the established valuation system with Chinese characteristics.

K-Means clustering algorithm is one of the most common unsupervised learning clustering algorithms, which is a typical distance-based non-hierarchical clustering algorithm. The data are divided into a predetermined number of classes K on the basis of minimizing the error function, and the distance is used as an evaluation index of similarity, i.e., the larger the distance between two objects is considered to be, the more similar they are[7-8].

Step 1: Randomly select k clustering centers of mass then there exist k clusters $C^{(k)}$.

$$C^{(k)}: \mu_1, \mu_2, \dots, \mu_k \in R \quad (4)$$

Step 2: All the data points are associated to the closest center of mass and based on this the clusters are divided. Calculate the distance μ_j between $x^{(i)}$ and each center of mass, then $x^{(i)}$

belongs to the cluster $C^{(j)}$ with his nearest center of mass μ_j .

$$C^{(j)} = \operatorname{argmin}_j \|x^{(i)} - \mu_j\|^2, j \in 1, 2, \dots, k \quad (5)$$

Step 3: For each class $C^{(j)}$, recalculate the value of the cluster's center of mass move the mass to the center of the currently delineated clusters containing all the data points (means)[5]; repeat the second and third steps n times until the sum of the squares of the distances of all the points to the centers of mass of the clusters to which they belong is minimized.

$$\mu_j = \frac{\sum_{i=1}^m I\{C^{(i)}=j\}x^{(i)}}{\sum_{i=1}^m I\{C^{(i)}=j\}} \quad (6)$$

The key formulas in the K-means algorithm involve calculating the distances between the data points and the cluster centers, and updating the cluster centers. Following are the key formulas of K-means algorithm.

Calculating the distance between data points and cluster centers: K-means usually uses Euclidean distance (Euclidean distance) to measure the distance between data points and cluster centers. The Euclidean distance formula is as follows [9].

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (7)$$

Where i denotes each dimension of the data point, and x and y denote the data point and cluster center, respectively.

In K-means, the above formula is usually used to calculate the distance between each data point and each cluster center to determine the closest cluster to assign the data point to. Updating the cluster centers: once the data points are assigned to the nearest cluster, the center of each cluster needs to be updated. Typically, the center of a cluster is the average of all data points within that cluster. The update formula is as follows:

$$\text{New cluster center: } C_k = \frac{\sum_{i=1}^{x_i}}{N_k} \quad (8)$$

Where k denotes the number of the cluster, N_k denotes the number of data points in that cluster, and $\sum_{i=1}^{x_i}$ denotes the accumulation of all data points in that cluster.

This formula calculates the new center of the cluster and then replaces the old cluster center with it. In this way, in the next iteration, the data points will be compared with the new cluster center to reassign the data points.

The K-means algorithm is an iterative process that iterative process calculates distances, reassigns data points, and updates cluster centers, eventually reaching a state of convergence where the cluster centers no longer change or other stopping conditions are met[10].

2.5.2 Stock Cluster Category Visualization and K-Means Cluster Analysis Effectiveness Analysis

In this paper, through the K-mean clustering based on the "elbow rule"[11-13], the collected stocks are clustered into two categories, and the cluster summary is plotted in Figure. 2, and the cluster scatter plot is plotted in Figure. 3.

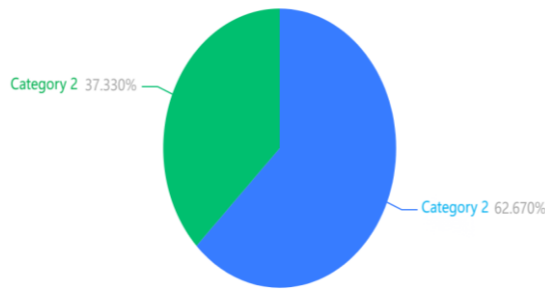


Figure 2: Clustering summary diagram

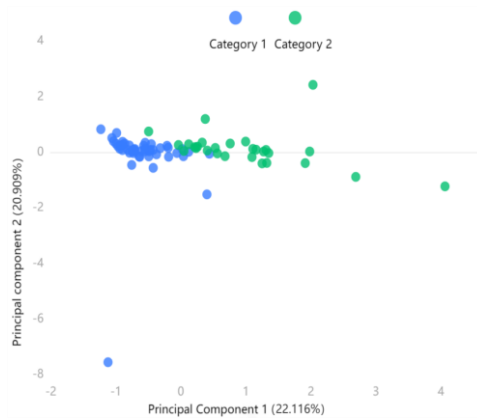


Figure 3: Clustered scatter plot

In this paper, by analyzing Figure. 2 and Figure. 3, we find that there are certain differences between stocks, and the classification effect of the K-Means cluster analysis proposed in this paper is obvious.

2.5.3 Specific Analysis and Definition of Stock Clustering Categories

Table 3: Stock clustering categories

Category 1				Category 2			
600018.SH	600415.SH	601298.SH	601808.SH	600007.SH	600905.SH	601600.SH	603019.SH
600019.SH	600425.SH	601390.SH	601857.SH	600685.SH	600916.SH	601728.SH	000519.SZ
600026.SH	600502.SH	601598.SH	601866.SH	600705.SH	600941.SH	601816.SH	000708.SZ
600028.SH	600585.SH	601608.SH	601868.SH	600760.SH	601006.SH	601858.SH	000831.SZ
600039.SH	600938.SH	601611.SH	601899.SH	600886.SH	601061.SH	601888.SH	000923.SZ
600050.SH	600970.SH	601618.SH	601919.SH	600900.SH	601088.SH	601898.SH	002271.SZ
600072.SH	601018.SH	601668.SH	601949.SH	002423.SZ	001965.SZ	002541.SZ	000932.SZ
600150.SH	601117.SH	601669.SH	601985.SH				
600176.SH	601179.SH	601766.SH	000065.SZ				
600339.SH	601186.SH	000617.SZ	000928.SZ				
001872.SZ	002051.SZ	002302.SZ	002493.SZ				
003816.SZ	000498.SZ	601800.SH					

In this paper, K-Means cluster analysis is used to derive the category 1 stocks and category 2 stocks in Table 3, and extract the two categories of stocks with large differences in the center values

of the Chinese characteristic valuation system characteristic indicators as shown in Table 4.

Table 4: List of center coordinates of clustering centers

Clustering category	Whether it is a state-owned company	Whether it is a national key project	Whether it is an innovative technology enterprise	PS ratio	PB ratio
1	0.957446809	1	0.063829787	0.097020112	0.086061246
2	0.785714286	1.11E-16	0.107142857	0.252027036	0.208220926

By analyzing Tables 3 and 4, the characteristics of the clustering categories are obtained:

Clustering category 1 (long-term investment category):

The composition of state-owned shares is large, consistent with the national key projects and the national strategic direction, the reform of state-owned enterprises and policies are released, and more dividends are expected to obtain more earnings. The price-to-sales ratio and price-to-book ratio of state-owned shares are relatively small, and the valuation is lower than that of cluster 2, which has better growth in the long run and is suitable for long-term investment [14-15].

Cluster category 2 (short to medium term investment category):

Innovation and technology category has a large component, is the national key support industries and fields, release of new products can be a short period of time to get a high investment income, with good growth; by the press release hot spots and industrial adjustment, suitable for short-term investment [16], combined with the results of the clustering can be seen, this category is generally the industry high-tech enterprises.

3. Conclusion

The results of this study find nine characteristic indicators of the valuation system with Chinese characteristics by analyzing the core connotation of the valuation system with Chinese characteristics, which provides help for investors' stock valuation, but there are also some secondary indicators of a valuation system, which need to be considered and analyzed by investors in a comprehensive manner. If more minor influences are to be considered, this can be achieved by further expanding the scope of data collection or introducing more relevant characteristics.

Then, based on the established valuation system with Chinese characteristics, this study selects 75 stocks in the Shanghai and Shenzhen A-share stock markets for K-Means clustering, and obtains two kinds of stocks based on the valuation system with Chinese characteristics in the short- to medium-term investment category and the long-term investment category, which provides assistance to a wide range of investors.

In addition, adding more data and incorporating deeper expertise into the data can improve the accuracy of K-Means clustering, which improves the returns of stock investment strategies.

In conclusion, this study provides decision makers with a novel Chinese stock valuation model for investing in the Chinese stock market. However, to further improve the accuracy and robustness of the model, future research may focus on increasing the amount of data and optimizing the model structure to cope with the volatility and changes in the equity securities market.

References

- [1] Zan Xiuli. Yi Huiman: Strengthening the intrinsic foundation of valuation system with Chinese characteristics. *China Securities Journal*, 2023-06-09(A03).DOI:10.28162/n.cnki.nczjb.2023.003367
- [2] Song Ge; Zhang Hong. Valuation theory system with Chinese characteristics is an incremental contribution to the existing valuation system. *Daily Economic News*, 2023-05-23(002).DOI:10.28571/n.cnki.nmrjj.2023.001919
- [3] Zan Xiuli. Exploring the establishment of a valuation system with Chinese characteristics. *China Securities Journal*,

2023-05-22(A01). DOI:10.28162/n.cnki.nczjb.2023.002441

- [4] Olabhele M E, Jieyun W. Author Correction: Grey hybrid normalization with period based entropy weighting and relational analysis for cities rankings. *Scientific Reports*, 2023, 13(1):16236-16236.
- [5] Zeng W, Fan J, Ren Z, et al. Economic Evaluation Method of Modern Power Transmission System Based on Improved Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Best-Worst Method-Anti-Entropy Weight. *Energies*, 2023, 16(21)
- [6] Yuan J, Yingqing S, Qinkai L, et al. An entropy weight method to integrate big omics and mechanistically evaluate drug-induced liver injury. *Hepatology (Baltimore, Md.)*, 2023
- [7] Olabhele M E, Jieyun W. Grey hybrid normalization with period based entropy weighting and relational analysis for cities rankings. *Scientific Reports*, 2023, 13(1):13797-13797.
- [8] Tai Bin; Huang Yangjue; Shen Kaicheng; Lei Xichun; Liao Siyang. Distributed photovoltaic power fluctuation smoothing control method based on index-weighted K-means++ algorithm. *Journal of Wuhan University (Engineering Edition)*, 2023, 56(11):1413-1424. DOI:10.14188/j.1671-8844.2023-11-012
- [9] Quankun L, Ruixian M, Mingfu L, et al. A novel fault evaluation method based on nonlinear vibration features and Euclidean distance measurement for grid-like structures. *Structural Health Monitoring*, 2023, 22(6):4131-4148.
- [10] Sarah Z, L. C M, Christopher W, et al. Identifying groups of children's social mobility opportunity for public health applications using k-means clustering. *Heliyon*, 2023, 9(9):e20250-e20250.
- [11] Zeng R. M. Improvement of K-means clustering algorithm and its application research. West China Normal University, 2022. DOI:10.27859/d.cnki.gxhsf.2022.000453
- [12] Long W. J.; Zhang X. F.; Zhang C. A business process clustering method based on k-means and elbow rule. *Journal of Jiangnan University (Natural Science Edition)*, 2020, 48(01):81-90. DOI: 10.16389/j.cnki. cn42-1737/n. 2020.01.011
- [13] Nizam M U, Hassan A D M, Norhakim Y, et al. A spatial-temporal clustering for low ocean renewable energy resources using K-means clustering. *Renewable Energy*, 2023, 219(P2)
- [14] Mengyuan L, Guanwen H, Le W, et al. Comprehensive classification assessment of GNSS observation data quality by fusing k-means and KNN algorithms. *GPS Solutions*, 2023, 28(1)
- [15] Russell G, Ioannis K, Perch J N, et al. On optimal constrained investment strategies for long-term savers in stochastic environments and probability hedging. *European Journal of Operational Research*, 2023, 307(2):948-962.
- [16] Islam T M K, Hooi S T, Lee L C, et al. Investment environment, stock market perception and stock investments after stock market crash. *International Journal of Emerging Markets*, 2023, 18(10):3506-3527.