Research on Credit Rating of China A-share Listed Companies

Junhao Chen
dept. Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI, USA
cjhhao@umich.edu

***Abstract*—This study develops a tailored credit rating model for 1,513 China A-share listed companies using factor analysis and cluster analysis. Through statistical sampling, companies are categorized based on similar credit risk levels, and corresponding rating standards are established. Empirical testing demonstrates the model's strong effectiveness in assessing and predicting credit risk for A-share listed companies, offering a robust framework for credit evaluation in this context.**

***Index Terms*—Credit Rating Model, A-share Market, Factor Analysis, Cluster Analysis, Credit Risk Assessment**

I. Introduction

The emergence of a market economy and globalization has positioned China within a credit-based economic framework. As a cornerstone of modern finance, credit ratings mitigate information asymmetry through independent assessments of creditworthiness [1]. Despite this, China's credit rating industry is still in its infancy, characterized by small financial intermediary institutions and underdeveloped methodologies. While international agencies such as Standard & Poor's (S&P), Moody's, and Fitch have mature frameworks, their evaluations of Chinese enterprises are largely confined to state-owned or large corporations, leaving a significant gap in the analysis of the broader A-share market.

This study aims to bridge this gap by developing a credit rating model tailored to China A-share listed companies. Utilizing **factor analysis** and **clustering analysis**, the model will assess and predict creditworthiness, providing a robust framework adapted to China's financial landscape.

Credit evaluation techniques can be broadly categorized into traditional statistical models and machine learning models. Statistical approaches, such as Altman's Z-Score [2] and Ohlson's O-Score [3], rely on financial metrics to evaluate credit risk. In contrast, machine learning models integrate market and behavioral data, employing advanced techniques like random forests, gradient boosting (GBDT), and neural networks. For instance, S&P's CreditModel™ leverages index density models and Akaike Information Criterion (AIC), while Moody's RiskCalc adopts boosting algorithms. Enhanced machine learning models, such as the BSVM [4], have demonstrated superior performance in specific credit rating scenarios.

Credit ratings, typically represented on a scale from AAA (highest stability) to D (default risk), provide a standardized measure of financial health. By building on these established methodologies, this study aims to create a comprehensive, data-driven model to address the unique challenges of credit evaluation in China A-share market.

II. Model Construction

A. Model and indicator selection

To develop a credit rating model for China A-share listed companies, this study employs **factor analysis** to extract latent common factors from a set of financial indicators reflecting credit risk. The comprehensive credit risk evaluation function is expressed as a linear combination of these factors, given by:

$$Z = \alpha_1 F_1 + \alpha_2 F_2 + \cdots + \alpha_m F_m$$

where:

- Z is the comprehensive credit risk evaluation function of the evaluated object;
- F_i $(i = 1, 2, \ldots, m)$ are the m factors extracted through factor analysis;
- α_i $(i = 1, 2, \ldots, m)$ are the weights of the factors (i.e., the variance contribution rate of the i-th factor).

Factor analysis calculates factor scores for each evaluated company, thereby reducing variable dimensionality and addressing multicollinearity issues. These scores are subsequently utilized in **K-means clustering** to categorize companies into distinct credit groups, forming the basis of the credit rating model.

The financial indicators system used in this project comprises 14 metrics across 5 categories, covering key aspects of company performance, as detailed in Table 1.

This study utilizes data from 2,524 non-financial China A-share listed companies from 2014 to 2023 to compute the aforementioned indicators. A 60% random sampling method, based on the "Guidelines for the Industry Classification of Listed Companies (2012 Revision)", yielded a final sample of 1,513 companies. The model's predictive performance was further validated using 2024 data from companies outside the modeling sample.

B. The empirical study of factor analysis

To extract representative factors from numerous original variables, this study follows three main steps: data preprocessing, factor extraction, and factor score computation.

TABLE I
FOURTEEN FINANCIAL INDICATORS

Category	Metrics
Profitability	Earnings per share
	Return on asset
	Return on equity
	Net profit to revenue
Operating Efficiency	Accounts receivable turnover
	Inventory turnover
	Current asset turnover
Debt Paying Ability	Current ratio
	Debt to asset ratio
	Interest coverage ratio
	Long-term debt to working capital ratio
Growth Ability	Total asset growth ratio
	Operating revenue growth ratio
Cash Flow Status	Cash flow ratio

TABLE II
SEVEN-LEVEL CREDIT RATING SCALE

Level	Symbol	Score Z	Description
1	AAA	$Z > 0.5884$	Highest quality, excellent debt repayment ability.
2	AA	$0.3445 < Z \leqslant 0.5884$	Very high quality, strong debt repayment ability.
3	A	$0.1488 < Z \leqslant 0.3445$	High quality, reliable but sensitive to changes.
4	BBB	$-0.0300 < Z \leqslant 0.1488$	Good quality, reasonable guarantees, affected by changes.
5	BB	$-0.2225 < Z \leqslant -0.0300$	Speculative, weaker guarantees, significant uncertainties.
6	B	$-0.4660 < Z \leqslant -0.2225$	Highly speculative, thin guarantees, high uncertainty.
7	CCC	$Z \leqslant -0.4660$	Substantial risks, extremely weak or no guarantees.

First, the raw data is preprocessed. Based on the average values of 14 financial indicators, the data undergoes de-extreming, industry-neutralization, and standardization to ensure reliability and comparability.

Second, factors are extracted using principal component analysis (PCA). Six factors are set, with a cumulative variance contribution rate of 71.089%, capturing most of the information in the original 14 indicators.

Third, factor scores are calculated. This study employs the regression method to determine the factor scoring coefficients, enabling the computation of factor scores for the original financial indicators. Subsequently, the comprehensive credit score Z for each company is derived by weighting factor scores based on their variance contribution rates.

C. The empirical study of clustering analysis

The K-means algorithm, an indirect clustering method based on sample similarity, is employed to classify companies using their comprehensive scores Z from 1,513 modeling samples. The clustering process identifies group boundaries, forming the basis for a seven-level credit rating standard.

Based on the clustering results, companies are assigned to credit levels by minimizing the distance to the nearest cluster center. To provide clear and intuitive descriptions of creditworthiness, the credit rating scale is aligned with definitions from domestic and international rating agencies, as detailed in Table 2.

III. MODEL VALIDATION

The credit rating model is validated in two stages: first, by testing it on ST (Special Treatment) companies within the modeling sample, which are identified as financially distressed or at risk of delisting; second, by evaluating its predictive performance using companies outside the modeling sample.

A. Validation with ST Companies in the Modeling Sample

The modeling sample comprises 1,513 listed companies, of which 112 have been designated as ST for more than two years. Under the seven-level credit rating model, the validation results are presented in Table 3.

TABLE III
VERIFICATION RESULTS OF ST COMPANIES IN THE MODEL SAMPLE

Level	Level 1	Level 2	Level 3	Level 4
Sample Size	51	171	251	338
ST Verification	0	0	2	6
Proportion	0.0%	0.0%	0.8%	1.78%

Level	Level 5	Level 6	Level 7
Sample Size	337	232	125
ST Verification	16	27	21
Proportion	4.75%	11.64%	16.8%

The results show a higher proportion of ST companies in lower credit ratings, confirming the model's effectiveness. Theoretically, ST companies should ideally fall into the bottom three levels; however, while ST status indicates an elevated default risk, it does not equate to default. Furthermore, minor inaccuracies due to information loss in factor analysis are reasonable and expected.

B. Evaluation of Predictive Capability

To evaluate the model's predictive capability, companies marked as ST for at least three months in 2024 are selected as the validation sample. Excluding 80 companies already in the modeling sample, 49 remain for testing. Using their 2014–2023 financial data, credit ratings are assigned based on the seven-level standard, with results shown in Figure 1.

Figure 1 shows that among the 49 companies, 4 are rated at Level 7, 23 at Level 6, and 17 at Level 5, with 89.8% (44 out of 49) falling into the bottom three levels. This indicates that the proposed model demonstrates strong predictive capability.

IV. CONCLUSION

This study establishes a credit rating model through factor analysis and clustering analysis based on 1,513 A-share listed companies across various industries in China. The model incorporates 14 financial indicators, capturing the majority of information relevant to the creditworthiness of listed companies. Model validation demonstrates its strong predictive capability for credit ratings.

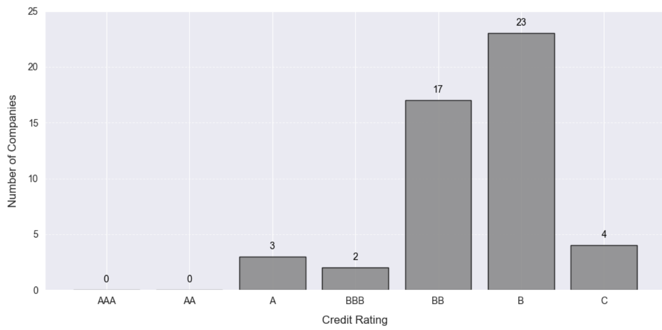

Fig. 1. Credit Rating Results for ST Companies in 2024 Validation Sample

Moreover, the application of factor analysis allows the model to provide not only final credit ratings but also intermediate variable insights, offering a more nuanced understanding of corporate credit conditions. This dual functionality enhances its practical utility for both lending institutions and rated companies, supporting informed decision-making and comprehensive credit assessments.

REFERENCES

[1] M. H. Adelson, "The role of credit ratings in the financial system," *Global Credit Portal*, pp. 1–12, 2012. [Online]. Available: http://www.standardandpoors.com/ratings/articles/en/us/?articleType=HTML&assetID=1245333790527

[2] E. I. Altman, R. G. Haldeman, and P. Narayanan, "Zetatm analysis a new model to identify bankruptcy risk of corporations," *Journal of Banking Finance*, vol. 1, no. 1, pp. 29–54, 1977. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0378426677900176

[3] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131, 1980. [Online]. Available: http://www.jstor.org/stable/2490395

[4] H. H. K. K. Viswanathan, P. Murugesan, S. Rengasamy, and L. Vilvanathan, "A modified boosted support vector machine to rate banks," *Benchmarking*, vol. 28, no. 1, pp. 1–27, 2021.