Journal of
Risk and Financial Management

MDPI

Article

Financial Risk Management Early-Warning Model for Chinese Enterprises

Haitong Wei [1,2,*] and Xinghai Wang [2,3]

[1] HongHao Data Intelligence Technology Co., Ltd., Beijing 102699, China
[2] Data Intelligence Branch, Enterprise Financial Management Association of China, Beijing 100195, China; wangxhdr@foxmail.com
[3] Big Data Application Research and Service Center, Modern Education Institute of China Academy of Management Sciences, Beijing 100036, China
* Correspondence: loveweihaitong@foxmail.com; Tel.: +86-185-1364-0920

Abstract: As enterprises face increasing competitive pressures, financial crises can significantly impact on their capital operations, potentially leading to operational difficulties and, ultimately, market exclusion. Consequently, many enterprises have begun to utilize financial early-warning systems to guide and control risks. Currently, there is neither a universal nor comprehensive enterprise financial risk management model in China, nor a unified classification standard for enterprise financial risk management levels. This article takes financial data on A-share listed companies in 2020 as the data sample, including those with special treatment (represented by ST) or non-ST status. We establish an independent indicator system within the framework of profitability, solvency, operational capability, development potential, shareholders' retained earnings, cash flow, and asset growth. The model is constructed employing the factor–logistic fusion algorithm. The factor part addresses the issue of collinearity among risk indicators, and the logistic part presents the results in probabilistic form, enhancing the interpretability of the model. The prediction accuracy of this model exceeds 89%. Finally, by applying the principles of interval estimation theory to statistical hypothesis testing, we categorize the risk levels into Grade A, representing significant risk; Grade B, representing moderate risk; Grade C, representing minor risk; and Grade D, representing no risk. This article aims to provide a comprehensive definition of a universal financial risk management early-warning model applicable to all enterprises in China.

Keywords: financial risk; financial risk warning analysis; financial warning level

check for updates

Citation: Wei, Haitong, and Xinghai Wang. 2024. Financial Risk Management Early-Warning Model for Chinese Enterprises. *Journal of Risk and Financial Management* 17: 255. https://doi.org/10.3390/jrfm17070255

Academic Editor: Svetlozar (Zari) Rachev

Received: 7 May 2024
Revised: 27 May 2024
Accepted: 4 June 2024
Published: 21 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

From a macroscopic viewpoint, over the past four decades of reform and opening up in China, the GDP has increased from CNY 367.8 billion in 1978 to CNY 126 trillion in 2023 (Wang 2014). The total market capitalization of around 5200 listed companies on China's three major exchanges also reached approximately CNY 90 trillion in 2023, indicating rapid growth in the stock market in recent decades. However, because of the relatively late start and rapid development of information technology, various regulatory measures implemented in China's securities market are still not ideal (Lv 2023). These include risks such as illegal operations, fraud, manipulation of profits, and poor management.

From the perspective of the enterprise itself, financial risk runs throughout the development process of every enterprise. This process includes the start-up period, maturity period, and decline period. The operation of a company always comes with financial risks; however, if a company maintains a high level of financial risk for a long period of time, this will not only increase its operating burden and result in a loss of investor confidence but also form a vicious cycle and lead to the breakdown of the company's funding chain. The financial stability of a company is crucial not only for the industry in which it operates but also for the financial system and the entire economy.

In addition, an even more severe situation that occurred is the global COVID-19 pandemic, which began in 2020 and has led to a global economic recession. Facing the threats that have emerged with the global pandemic, competition in the big data market has become even more intense. Both domestic and foreign enterprises are placing greater emphasis on stabilizing and enhancing their financial conditions, as well as refining their financial risk management and early-warning systems (Wen and Chernov 2022).

Financial risk management serves as a diagnostic tool for enterprises, playing a preventive role. It can predict and diagnose financial risks, preventing them from escalating into financial crises. By utilizing financial risk management, businesses can assess their risk profile and make informed investment decisions. Creditors can utilize financial risk management to assess an enterprise's risk profile and establish appropriate credit policies. Government regulatory agencies can effectively oversee corporate risks, thereby guiding and regulating the market. Other stakeholders, such as affiliated parties, can also use financial risk management to guide and manage enterprise risks (Yang et al. 2019). In summary, an effective financial risk management model is crucial for all stakeholders of an enterprise.

The effective prevention of financial risks requires a comprehensive analysis of an enterprise's financial situation and not just the individual indicators. Analyzing financial risks not only aids managers in making informed decisions but also helps businesses address their shortcomings. Therefore, it is essential for businesses to promptly identify and respond appropriately to financial risks, which has led to extensive research on corporate financial risk analysis using indicators from financial statements (Shi 2021; Trencheva 2021).

It is particularly important to establish a comprehensive, flexible, scientific, and accurate enterprise financial risk management and early-warning model in China (Li et al. 2022). This is especially true for investors, governments, banks, creditors, managers, employees, and other stakeholders. Such a model can categorize and promptly warn of different levels of financial risk crises in enterprises, reflecting and eliminating risks at their inception.

Currently, research on financial risk warning models in China faces several issues, as follows:

1. Most scholars employ models already developed abroad, substituting Chinese corporate financial data and overlooking the differences in the contexts in which these models are applied (Lyu 2016; Yang et al. 2017). Given the substantial disparities between domestic and international financial accounting standards (Zhang et al. 2020), the accuracy of their findings needs to be validated.
2. Research on financial early warning lacks theoretical guidance, with a majority relying on qualitative methods for selecting research variables (Zanaj et al. 2023).
3. Each early-warning model focuses on different aspects, resulting in an incomplete selection of financial indicators that is inherently biased (Wang 2020). Previous studies on early-warning models in China have not explicitly incorporated factors such as cash flow and asset appreciation, which are closely linked to a company's financial health. Consequently, a one-sided selection of indicators inevitably impairs the predictive power of these models.
4. Currently, a widely accepted and comprehensive enterprise financial risk management model in China does not exist, nor is there a unified standard for classifying levels of enterprise financial risk management (Peng 2015).

Drawing on the background and objectives of this research, this article primarily examines the following three questions:

1. How can a financial risk model be constructed? What are the criteria for selecting key financial indicator variables?
2. For financial risk management models, is it necessary to design distinct models for manufacturing and nonmanufacturing sectors?
3. What are the thresholds for risk level classification in financial risk models?

The aim of this article is to develop a unified enterprise financial risk management warning model using the financial data on special treatment (represented by ST) and non-ST companies listed on China's A-share market in 2020. The model is constructed using a factor–logistic fusion algorithm, categorizing enterprises into the following four distinct levels of financial risk: Grade A indicates significant risk, Grade B indicates moderate risk, Grade C indicates minor risk, and Grade D indicates no risk.

The scope of this article includes the financial statement data of 4254 A-share listed companies in 2020, with the sample design comprising the selection of both the sample group and the matched group. The sample group consists of ST (including *ST (meaning three consecutive years of losses and delisting warning)) companies, which are considered financially distressed. Subsequently, corresponding financially healthy companies were identified on the basis of the principle of one-to-one pairing year by year. Finally, the sample group and the matched group were divided into modeling samples and prediction samples.

Listed companies are emblematic of technology and business practices. Their benign operations significantly impact society and the national economy. Therefore, studying the operational and managerial risk status of Chinese listed companies directly reflects their business conditions through financial statement data. Financial data are the lifeblood of any enterprise (Jiang et al. 2010). Typically, enterprises do not publicly disclose financial data, and the quality of data in annual business reports is compromised because of irregularities, lack of standardization, and nonaudited status, rendering it of limited research value and relevance. Listed companies differ from other enterprises in that their annual reports must be transparent and audited by professional institutions. Any financial fraud carries legal risks. In conclusion, studying the financial risks of listed companies is valuable and meaningful for numerous aspects of society.

2. Related Literature

2.1. Theoretical Foundation

The research on financial distress prediction abroad has evolved from univariate discriminant analysis and multivariate discriminant analysis to regression analysis, followed by recursive partitioning analysis, artificial neural networks, and, ultimately, to hybrid research methods. Because of the differences in listing systems between domestic and foreign markets, foreign scholars have primarily focused their research on the relationship between corporate financial indicators and the likelihood of a company falling into financial distress, specifically on whether financial indicators can successfully predict corporate bankruptcy.

Financial distress prediction originated from FitzPatrick's 1932 article on univariate bankruptcy prediction (Fitzpatrick 1932). He selected traditional financial variables for comparative analysis, dividing the samples into distressed and nondistressed companies. His research showed that two indicators, "net income/shareholders' equity" and "shareholders' equity/debt", could better discriminate whether a company was experiencing financial distress. These two indicators showed significant differences in the three years prior to a company's bankruptcy.

In 1966, Beaver further extended FitzPatrick's approach by introducing cash flow indicators to establish a financial distress prediction model (Beaver 1968). He randomly selected 79 companies that went bankrupt between 1954 and 1964 and set up a control group for comparison. After univariate testing, he concluded that the "cash flow/total debt" indicator had the strongest predictive ability for corporate financial distress, and the probability of successful prediction increased as the bankruptcy date approached.

Subsequently, in 1968, Altman used multivariate linear discrimination to develop a financial risk prediction model, known as the Z-score model (Altman 1968). This model is linear and comprises weighted coefficients of five financial ratios. By multiplying the weights with the ratios and summing them, a score is obtained. This score serves as a threshold to classify companies as bankrupt or nonbankrupt. The model has high accuracy in financial risk prediction and remains relevant today. However, the multivariate linear dis-

crimination model requires that the independent variables of distressed and nondistressed companies follow a normal distribution with consistent variance and covariance.

To overcome these limitations, Ohlson applied the multivariate logistic regression model to financial distress prediction, in 1980, using a sample of 105 bankrupt companies and 2058 healthy companies in the United States (Ohlson 1980). His empirical article demonstrated that the logistic regression model outperformed multivariate linear discrimination analysis in prediction.

In 1985, Frydman utilized recursive partitioning analysis (RPA) to establish a binary classification tree using financial ratios as decision points (Frydman et al. 1985). Companies were classified as bankrupt or nonbankrupt based on the criterion of minimum misclassification cost.

In 1993, Ofek found that the higher a company's financial leverage, the higher the likelihood of escaping financial distress (Ofek 1993).

Because of the multicollinearity among financial indicators, there are limitations in variable selection for models. To address this issue, Ana M. Aguilera et al. combined principal component analysis with logistic regression, in 2006, to predict corporate default (Aguilera et al. 2006). However, the practical interpretation of the principal components remains challenging.

Additionally, machine learning techniques have been widely applied in model development. For instance, Franco Varetto utilized genetic algorithms to article corporate bankruptcy risk, in 1998, while Jae H. Min and Young-Chan Lee employed support vector machines for credit risk prediction of listed companies, in 2005 (Varetto 1998; Min and Lee 2005). With the development of information technology, Odom et al. used neural networks to predict corporate bankruptcy, in 1990 (Odom and Sharda 1990; Altman et al. 1994). Their results showed that financial distress prediction models built with neural networks exhibited better predictive ability. Artificial neural network models overcome the limitations of statistical methods, demonstrating strong fault tolerance and error correction capabilities, thus gaining recognition from numerous scholars. However, these methods still have drawbacks, including computational complexity, the requirement for large training samples, and the risk of overfitting.

In China, the earliest theoretical development was in 1996, when Zhou Shouhua and others borrowed from Altman's Z-score model and added cash flow ratios (Zhou et al. 1996).

Since then, various scholars have investigated the impact of different financial indicators on models. For instance, in 2002, Zhao Xikang studied the relationship between the debt ratio and financial status of listed companies (Zhao 2002). The research found that a higher debt-to-assets ratio led to a worse corporate financial situation. In addition to traditional financial indicators, Jiang Xiuhua et al. included equity concentration as a variable in their early-warning index system (Jiang and Sun 2001). The empirical results indicated that equity concentration significantly enhanced the analysis's effectiveness.

In 2011, Nie Lijie and Zhao Yanfang conducted an article on financial distress prediction by incorporating cash flow indicators in addition to financial metrics (Nie et al. 2011). The empirical results showed that the model's prediction accuracy and misclassification rate improved when cash flow indicators were included, outperforming traditional financial indicator warning models.

Ma Ruowei and Zhang Wei, in 2014, utilized quarterly data from listed companies to explore variables predictive of financial distress, establishing Fisher discriminant models, linear logistic models, and nonlinear logistic models (Ma and Zhang 2014). The empirical results suggested that logistic models outperformed Fisher models, indicating a nonlinear relationship between financial and nonfinancial indicators and corporate financial distress. Furthermore, both logistic models exhibited higher prediction accuracy than the Fisher model.

In the past four years, Ren Hong and Zhang Yuzhi, among others, studied a financial crisis warning model for Chinese enterprises, in 2020 (Ren 2020). They selected 258 ST-listed companies and another 258 comparison companies (including listed and some nonlisted

firms) from 2015 to 2019 as samples. By analyzing the relevant financial data of these two groups of companies, they employed logistic regression as the model.

In 2022, Zhao Xuefeng and Wu Weiwei, along with others, proposed the CFW-Boost enterprise financial risk warning model oriented toward characteristic causal analysis (Zhao et al. 2022). On the basis of financial and nonfinancial indicators, they constructed multiple financial characteristics, integrated multiple CART trees using feature causality to obtain CFW-Boost, and compared its performance with other warning models through empirical analysis. Huang Dezhong and Zhu Chaoqun et al. introduced corporate asset quality indicators into the financial risk warning model (Huang and Zhu 2022). They selected data from 48 listed companies that were first ST-labeled from 2010 to 2013 and paired them with 96 normal companies. They constructed both a financial warning model with commonly used financial indicators and an asset quality model incorporating asset quality indicators. By comparing the results of these two warning models, they found that the introduction of asset quality indicators could enhance the accuracy of the financial risk warning model.

In 2023, Wang Ning proposed a financial risk warning model for copper enterprises using the entropy method and the efficiency coefficient method (Wang 2023). Zeng Shulian and Wang Tao, among others, introduced the SMOTE double-layer XGBoost model for financial risk warning (Zeng et al. 2023). This model improved the effectiveness of the financial risk warning model, addressed the issue of imbalanced sample data, and optimized the dimension reduction process.

2.2. Theoretical Gaps

2.2.1. Issues Regarding Sample Data Processing

The domestic samples primarily come from ST and non-ST listed companies, with financial data sourced from quarterly public releases. Owing to the imperfect regulatory framework for domestic listed companies, there are instances of financial and profit fraud. Most scholars have not employed outlier handling techniques on the collected financial data. This article applies a two-fold standard deviation adjustment to the financial variables in the model, which underpins the accuracy of the model predictions.

2.2.2. Limited Generalizability of the Model

Primary sample collection, in most studies, typically focuses on data from listed companies in a specific industry. Because of the generally large scale of listed companies and the relatively small sample size, it becomes difficult for small- and medium-sized enterprises within the industry to participate. This article, however, encompasses 4254 samples of listed companies in 2020, covering the entire industry.

2.2.3. Lack of Verification of Nonfinancial Variables

The selection of variables for risk warning modeling is primarily based on financial variables, which are readily accessible. Model validation is primarily conducted to determine whether financial variables have a significant impact. However, there is limited literature exploring whether nonfinancial variables, such as industry attributes and company nature, have a significant impact on company risk (He et al. 2009; Da and Peng 2022). This article incorporates 0–1 variables representing nonfinancial variables, like company nature and industry attributes, verifying that nonfinancial variables have no significant impact on the risk model.

2.2.4. Limited Variety of Modeling Methods

Methodologically, most studies rely on multivariate statistical methods, such as discriminant analysis and logistic regression, with limited application of complex methods like neural network models. Discriminant analysis imposes strong assumptions and limitations on the normal distribution of data, making it unsuitable for today's ever-changing and complex business environment. Logistic regression tends to exclude insignificant

financial variables, leaving fewer financial indicators for analyzing company operations. Additionally, correlations among all financial indicators lead to an incomplete financial risk analysis. Neural networks provide only the prediction results of the model, with the intermediate process remaining a black box. Their mathematical processes are complex and uninterpretable, and the weights of financial variables are not apparent, making it difficult to discern which financial variables influence company risk. This article employs a fusion algorithm combining factor analysis and logistic regression, eliminating correlations among financial variables while retaining most of them. The weights of financial variables on risk are clearly visible, greatly assisting in the analysis of companies' financial risks.

2.2.5. Lack of Scientific Risk Classification

Most domestic and international literature focuses on warning analysis of corporate bankruptcy and ST companies, falling under categorical variable prediction, which is typically dichotomous. The prediction outcomes are often binary, as follows: bankruptcy and nonbankruptcy or ST and non-ST. However, a company's bankruptcy or ST risk is a gradual process, not a sudden one; thus, drawing conclusions directly from bankruptcy or ST status can be arbitrary. On the basis of the samples, the probability of risk is classified as follows: no risk below 5%, minor risk between 5% and 10%, moderate risk between 10% and 50%, and significant risk between 50% and 100%. Furthermore, this prediction is made several years ahead, providing entrepreneurs with sufficient time for reflection and improvement.

2.3. Conceptual Model

On the basis of a review of the domestic and foreign literature, this article proposes a conceptual model for financial risk management early warning in Chinese companies, taking into account the actual situation of their operations, as shown in Figure 1.

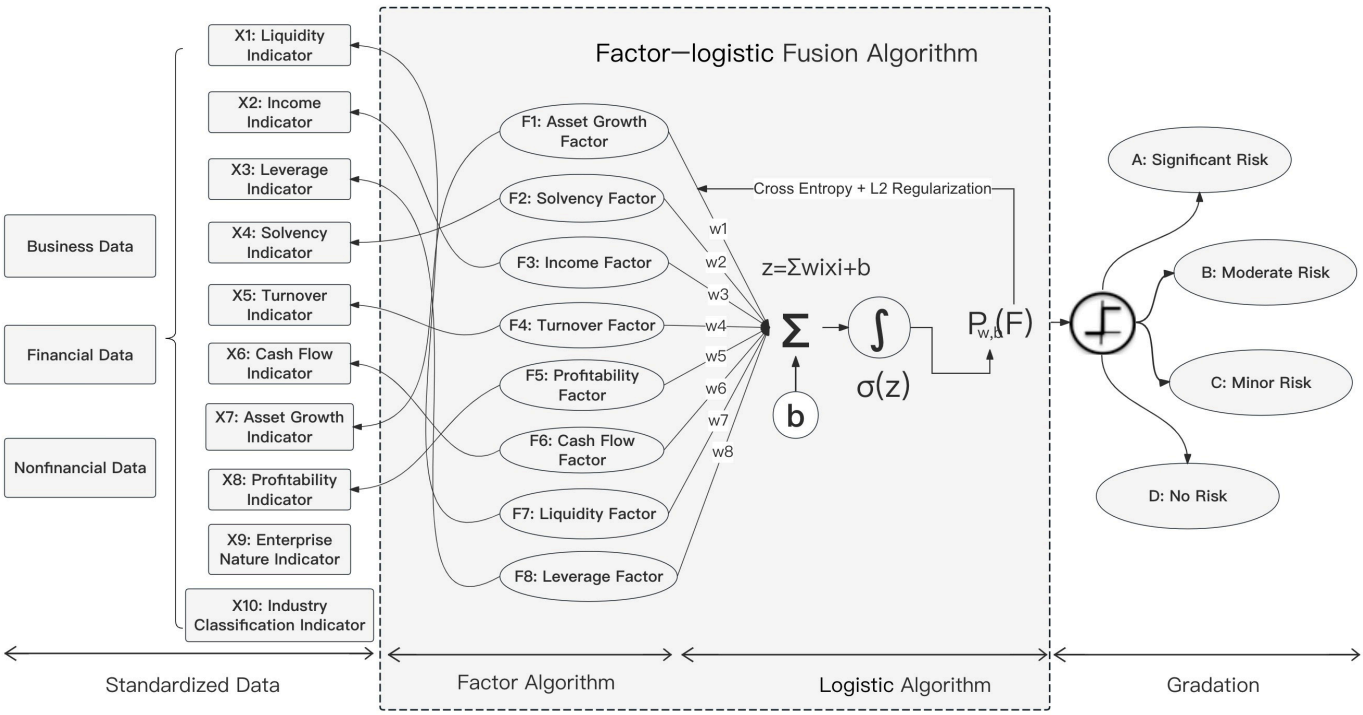


Figure 1. Conceptual model.

The ultimate goal of the conceptual model for risk management is to provide early warning and enable a company to operate smoothly. On the left side of the flowchart lies the data acquisition segment, encompassing business data, such as sales, costs, profits, and financial data, including assets, liabilities, and cash flow, along with categorical nonfinancial

data. These data are utilized to calculate eight financial indicators and two nonfinancial indicators (enterprise nature and industry classification). These indicators are then employed to establish a mathematical model for risk management. The model algorithm adopts the factor–logistic fusion method. Once the model is validated and proven feasible, it is used to assess the company's financial risk management status. The results are classified into the following four levels: Grade A indicates significant risk, Grade B indicates moderate risk, Grade C indicates minor risk, and Grade D indicates no risk. This provides entrepreneurs of risk-prone companies with practical suggestions and improvements, ultimately achieving normal operations.

2.4. Innovation

The primary contributions of this article are highlighted in the following four aspects:

1. **Theoretical Innovation:** The variables and parameters of the financial risk management early-warning model investigated in this article are unprecedented in China. The accuracy of the training samples reaches 89.7%, while the accuracy of the testing samples stands at 95.0%, indicating a high level of precision in model prediction. Based on the Altman Z bankruptcy model, this article incorporates three quantitative indicators, including cash flow, asset growth, and profitability. Additionally, two qualitative indicators are introduced, encompassing company nature and industry characteristics.
2. **Application Innovation:** The proposed model boasts broad applicability, suitable for both listed and nonlisted companies. During the modeling process, various influences stemming from the company's size, nature, and industry characteristics are taken into account. Since all variables are presented as ratio indicators and the original variable data are standardized, the model is not significantly influenced by the company's size. In summary, the model presented in this article can be effectively applied to companies operating normally across various industries.
3. **Innovation in Data Collection and Processing:** This article utilizes 2020 A-share listed companies in China as samples, yielding a total of 4254 samples. After processing the outliers and missing data in the original variable data, 320 qualified samples of financial data are included in the model for calculation. This approach represents an innovative data collection and processing method in China.
4. **Innovation in Risk Classification:** This article categorizes risks into four levels: Grade A indicates significant risk, Grade B indicates moderate risk, Grade C indicates minor risk, and Grade D indicates no risk. Instead of using equal division or effectiveness methods, the classification is based on the predicted probability of ST risk for sample companies. A combination of different statistical significance levels and sigmoid function thresholds is adopted for risk classification. Specifically, Grade D corresponds to an ST risk probability of less than 5%, Grade C ranges from 5% to 10%, Grade B falls between 10% and 50%, and Grade A exceeds 50%. The model employs the Sigmoid function to convert the numerical results into values ranging from 0 to 1, encompassing all potential financial risk outcomes for a company. This risk classification methodology, which integrates statistical significance level classification standards with the characteristics of the sigmoid function for binary classification, is also an innovative approach in China.

3. Model Research

3.1. Sample Source

The sample data for this article were derived from the financial indicator data of 4254 A-share listed companies in 2020, sourced from TongHuaShun Finance (Xu 2024). The sample design encompassed both a sample group and a matched group.

In selecting the sample group, the following factors were considered:

1. Industry of the Sample Data: The majority of ST (including *ST) companies belong to the manufacturing industry. To eliminate the influence of industry factors on the

financial distress prediction model, the article selected an equal number of samples from both manufacturing and nonmanufacturing companies.

2. Research Period of the Sample Data: Most previous studies focused only on the availability of samples, overlooking the issue of the research period. However, because of factors such as macroeconomic trends and economic cycles, enterprises operate in vastly different environments from year to year, leading to significant variations in their financial indicators. If this factor is not considered during sample selection and model construction, the prediction results may be biased by time factors, thereby reducing accuracy. Therefore, this article took this factor into account, using 2020 sample data for modeling, limiting the predictable range of the model to a certain period. With time changes, the optimal model needs to be retrained annually to find the most suitable parameters.
3. Sample Data Size: considering China's unique national conditions, this article deliberately included a balanced number of central- and state-owned enterprises in the training samples, making the model more robust.
4. Integrity of Sample Data: While previous studies often used data integrity as a criterion for sample selection, which disrupts the random sampling requirement for model building, many articles have pointed out that the prediction accuracy of the revised model has not significantly improved. Therefore, this article still uses data integrity as a criterion for sample selection, which facilitates model establishment and ensures the reliability of indicator interpretability.

Taking into account the above factors, the article used A-share listed companies that were subject to ST (including *ST) due to "abnormal financial conditions" as indicators of financial distress (i.e., the research subjects). It was found that there were 200 A-share listed companies with ST and *ST in 2020, of which 182 had valid data (excluding missing values). After outlier treatment, 160 listed companies were selected as the sample group for modeling, including 80 manufacturing enterprises and 80 nonmanufacturing enterprises. Among these 160 ST and *ST companies, 80% were private enterprises, while 20% were state-owned enterprises.

In selecting the matched group, the following factors were considered:

1. Control Factors for Matching Criteria: Because of the influence of reporting seasonality, industry characteristics, and company size, there may be model biases between the financially distressed group and the financially healthy group. Through matching criteria, this bias can be controlled. Control factors for matching criteria typically include accounting year, industry, and asset size. This article selected a certain number of matched samples based on the criteria of accounting year and asset size.
2. Quantity Allocation of Samples Between the Two Groups: Because of the symmetric nature of the logistic function, almost an equal number of financially healthy companies were selected from listed companies with complete data, as compared to the number of ST and *ST companies. Moreover, most empirical studies employ a one-to-one pairing method for sampling, ensuring that the sample group and the control group contain an equal number of research individuals.

On the basis of the above sampling criteria, this article selected 160 financially healthy companies as matched samples from the 4054 non-ST and *ST A-share listed companies. These samples were paired with the sample group one-to-one, with 80 manufacturing enterprises, 80 nonmanufacturing enterprises, 80% private enterprises, and 20% state-owned enterprises, which was entirely consistent with the sample group.

In total, the combined matched sample comprises financial data from 320 companies.

3.2. Indicator Selection

There is a close relationship between financial indicators and knowledge graph technology. Through knowledge graph technology, it is possible to construct a knowledge graph that includes all financial indicators, data, and related information of the enterprise. This makes intelligent financial analysis possible, as knowledge graph technology can

quickly integrate and correlate massive and complex information, providing enterprises with in-depth and comprehensive financial analysis (Wang 2024). This article employed the delphi method to select key financial indicators from the set of financial indicator knowledge graphs as research variables (Zhang et al. 2018). These eight financial indicators, after repeated deliberation by experts, comprehensively cover the core indicators of a company's various aspects of operations, management, and finance, thus forming the financial feature dimensions of this article's financial risk early-warning model.

In addition, enterprise nature (i.e., private or state-owned) and enterprise industry classification (manufacturing or nonmanufacturing) were also included as nonfinancial feature dimensions of our risk early-warning model (Sun 2021; Zhang and Yang 2023). The purpose is to explore whether it is necessary to classify industries and whether industry attributes have a significant impact on company risk. This indicator is added to the feature dimension we selected to explore its impact on the model.

The selected indicators are presented in Table 1 (Trencheva 2021; Zhang et al. 2018), and partial financial indicator data are shown in Table 2 (variable labels will be used in place of indicator names in the following text).

Table 1. Selected indicators.

Variable	Indicator Name	Calculation Method
X1	Liquidity	(Current Assets—Current Liabilities)/Average Total Assets
X2	Income	Retained Earnings/Average Total Assets
X3	Leverage	Earnings Before Interest and Taxes/Average Total Assets
X4	Solvency	Equity Market Value/Total Liabilities
X5	Turnover	Operating Income/Average Total Assets
X6	Cash Flow	Net Cash Flow from Operating Activities/Current Liabilities
X7	Asset Growth	(Total Assets at End of Period—Total Assets at Beginning of Period)/Total Assets at Beginning of Period
X8	Profitability	(Operating Income—Operating Cost)/Operating Income
X9	Enterprise Nature	Private Enterprises = 0, State-Owned Enterprises = 1
X10	Industry Classification	Manufacturing = 0, Nonmanufacturing = 1
Y	Is ST	ST = 0, non-ST = 1

Table 2. Financial indicator data for ST (including *ST) and non-ST A-share listed companies in 2020.

Code	600654	600093	002052	603322		000632	300310	600742	600103
X1	−0.1672	−0.2749	−0.0776	0.0658		0.0384	0.5093	0.0717	0.3632
X2	−0.4564	−0.7002	−1.7612	0.0536		0.032	−0.5171	0.2949	0.0974
X3	−0.0094	−0.9426	−0.1608	0.0508		0.0362	0.0048	0.0628	0.027
X4	0.492	0.6802	1.6943	1.6586		0.261	5.5492	0.6515	2.7525
X5	0.6	0.72	0.35	0.65		0.95	0.87	1.15	0.45
X6	0.0053	−0.0461	−0.0154	0.0549	.....	−0.3073	0.0815	0.3261	0.0611
X7	−0.1454	−0.3738	0.2733	0.0225		0.0752	0.0009	0.1109	−0.0064
X8	0.0624	0.0061	0.1051	0.1976		0.0569	0.1329	0.0983	0.1349
X9	0	1	0	0		1	0	1	1
X10	1	1	0	1		1	1	0	0
Y	0	0	0	0		1	1	1	1

Data sourced from TongHuaShun Finance financial statements. The ellipsis (. . .) indicates that the table has been truncated for brevity. Among them, 600654 represents the stock code of the listed company. The complete table would include all stock codes and corresponding indicator values.

3.3. Data Processing

This article retrieved the necessary financial data for 4254 A-share listed companies. Following data acquisition, missing data were initially addressed. Because of the sufficiently large sample size, listed company samples with missing data were directly eliminated.

Secondly, given that financial data anomalies inevitably exist even in datasets with no missing values, this article adopted the Pauta criterion and Grubbs' test for outlier removal and screening, as follows:

1. Pauta Criterion: It assumes that a set of measurement data only contains random errors. The standard deviation is calculated and processed to determine an interval based on a certain probability. Any error exceeding this interval is deemed not a random error but a gross error, and the data containing such error should be eliminated. This discrimination method is limited to the processing of samples with normal or approximately normal distributions, and it presumes a sufficient number of measurements. When the number of measurements is small, using this criterion to eliminate gross errors is unreliable. Therefore, it is advisable not to apply this criterion in such cases. When the number of repeated tests is significantly greater than 10, the Bessel formula is used to calculate the experimental standard deviation (s). If the absolute value of the difference $(x_a - \overline{x})$ between a suspicious value (x_a) and the mean $(\overline{x})$ of n results is greater than or equal to 3s, x_a is judged as an outlier. After eliminating x_a, the calculation and judgment continue in the same manner until $|x_a - \overline{x}| < 3s$ remains. In practical applications, 2s can also be chosen as the critical value for judging outliers. This article adopts 2s as the critical value, representing an innovative approach in sample data processing.
2. Grubbs' Test: It is a statistical method used to identify and eliminate outliers (or abnormal data) from normally distributed measurement data. This criterion compares the absolute value of the residual (the difference between a measurement value and the mean) with the product of the standard deviation. If the absolute value of a measurement's residual exceeds the corresponding value in the Grubbs' table, that data point is considered an outlier and should be eliminated.

The following section describes the Z-score normalization method applied before the factor analysis in this article.

Z-score normalization is a commonly used method in data processing. It transforms data of different magnitudes into Z-scores on a unified scale for comparison, enhancing the data comparability while reducing their interpretability. The formula is as follows:

$$v' = \frac{v - \overline{x}}{\sigma}$$

where v represents the actual sample, v' represents the standardized sample, $\overline{x}$ is the mean of the sample, and σ is the standard deviation of the sample. Normalization ensures that attributes with larger domains do not overshadow those with smaller domains, making the results more authentic and reliable.

3.4. Indicator Data Analysis and Exploration

3.4.1. Descriptive Analysis

Table 3 presents the descriptive statistics of the data, encompassing the commonly used minimum, maximum, mean, and variance. As evident from the table, among the quantified financial variables, from X1 to X8, the maximum, mean, and variance of X4 significantly exceed those of the other seven variables, while the variance of X2, following closely behind X4, is substantially greater than the variances of the remaining variables. Moreover, descriptive statistical analysis typically focuses on quantitative data rather than qualitative categorical data; thus, the statistical analyses of the company nature variable, X9, and the company industry attribute variable, X10, are not the primary concern.

Generally, when constructing a new model, there tend to be numerous evaluation metrics to comprehensively describe the correlations and causal relationships between model variables. Single-indicator models often exhibit a lower fit and fail to align with the reality, as the key factors influencing a particular issue are often not unique. The existence of multiple influencing factors aligns with practical situations. Similarly, in the risk model

presented in this article, there are 10 independent variables. Because of the differences in the unit sizes and magnitudes among these variables, when there are significant discrepancies in the hierarchical levels of the variables, the impact of the variables with higher numerical values tends to be more prominent in the comprehensive analysis, while the influence of the variables with lower numerical values is relatively diminished. Consequently, to ensure the reliability of the results, it is necessary to standardize the original variable data, eliminating the bias impact of different units and magnitudes on data analysis. After standardization, the mean of each variable is 0, and the variance is 1, facilitating analysis. Additionally, as this article employs factor analysis in the model's analytical approach, it is imperative to standardize the original variable data to eliminate the influence of different variable dimensions and scales.

Table 3. Descriptive statistics of indicator data.

Indicator	Num.	Min	Max	Mean	Variance
X1	320	−2.0559	0.882	0.106496	0.144
X2	320	−6.4172	0.6771	−0.229231	0.757
X3	320	−1.201	0.2227	−0.039178	0.035
X4	320	0.0495	46.2045	5.121727	49.429
X5	320	0	1.59	0.428313	0.094
X6	320	−1.3786	1.1797	0.104675	0.094
X7	320	−0.9561	3.077	0.016794	0.113
X8	320	−0.9251	1	0.235367	0.05
X9	320	0	1	0.203	0.162
X10	320	0	1	0.5	0.251

3.4.2. Correlation Analysis

The correlation coefficient, R, is a measure used to determine the degree of correlation among observed data. Usually, the higher the correlation coefficient, the stronger the correlation. The threshold for the correlation coefficient, R, is between negative 1 and positive 1. The basic evaluation criteria are that when the correlation coefficient R is greater than 0, it belongs to positive correlation, when it is equal to 0, it belongs to uncorrelated, and when it is less than 0, it belongs to negative correlation. Usually, the larger the correlation coefficient, the higher the degree of correlation. Many scholars believe that a correlation coefficient, R, greater than 0.8 indicates a strong positive correlation, moderate correlation in the range of 0.5 to 0.8, weak correlation in the range of 0.3 to 0.5, and noncorrelated in the range of 0 to 0.3. When the correlation coefficient is negative, the classification criteria are the same.

As can be seen from Table 4, there is more or less correlation between financial indicators, and the eight financial variables are generally weakly correlated. The correlation coefficients, R, are all below 0.5. Only between X1 and X3 does R = 0.425, and between X1 and X4 does R = 0.407, both belong to weak correlation. The correlation coefficients among other variables is also less than 0.4, indicating that these financial variables are either weakly correlated or uncorrelated. The conclusion is that there is no strong correlation among variables in the linear model.

Table 4. Correlation coefficients of indicator data.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1									
X2	0.373	1								
X3	0.425	0.316	1							
X4	0.407	0.083	0.169	1						
X5	0.237	0.199	0.304	−0.022	1					

Table 4. *Cont.*

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X6	0.174	0.351	0.28	0.168	0.224	1				
X7	0.339	0.081	0.517	0.18	0.289	0.153	1			
X8	0.307	0.167	0.279	0.262	−0.038	0.265	0.084	1		
X9	−0.108	−0.008	−0.013	−0.112	0.069	−0.055	0.037	−0.151	1	
X10	−0.048	0.006	−0.016	−0.02	−0.107	−0.002	0.05	0.05	−0.023	1

3.4.3. Significance Analysis

Significance testing involves making a hypothesis about the parameters or the overall distribution of a population (i.e., random variable) and then utilizing sample information to determine the plausibility of this hypothesis (i.e., alternative hypothesis), i.e., whether there is a significant difference between the actual situation of the population and the original hypothesis.

The probability level for testing the "null hypothesis" is typically set at 5%, implying that if the same experiment is repeated 100 times and differences in the results occur more than five times due to sampling error, the "null hypothesis" stands, indicating that the difference between the two groups is not significant, often denoted as $p > 0.05$. Conversely, if the differences occur five times or less due to sampling error, the "null hypothesis" is rejected, indicating that the difference between the two groups is significant, often denoted as $p \leq 0.05$. If $p \leq 0.01$, the difference between the two groups is considered highly significant.

Significance testing is a method used to determine whether there is a difference between the experimental treatment group and the control group or between the effects of two different treatments and whether this difference is significant.

The hypothesis to be tested is often denoted as H0, referred to as the null hypothesis, while the hypothesis opposing H0 is denoted as H1, referred to as the alternative hypothesis.

Rejecting the null hypothesis when it is true constitutes a type I error, with its probability of occurrence typically denoted as α.

Failing to reject the null hypothesis when it is false constitutes a type II error, with its probability of occurrence typically denoted as β.

Typically, only the maximum probability of making a type I error (α) is limited, while the probability of making a type II error (β) is not considered. This type of hypothesis testing is also known as significance testing, and the probability α is referred to as the level of significance.

The most commonly used values of α are 0.01, 0.05, 0.10, etc. Depending on the research question, if the loss of rejecting a true hypothesis is substantial, a smaller α value is chosen to reduce such errors and vice versa.

The results of the significance analysis, shown in Table 5, indicate that only four independent variables—X2, X3, X5, and X6—are significant at the 0.05 level, while the significance of the remaining variables—X1, X4, X7, X8, X9, and X10—are all greater than 0.05. This implies that, without data transformation through factor analysis, only these four indicators play a decisive role, while the remaining indicators contribute minimally to the model.

Table 5. Significance analysis.

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
0.232	0	0.005	**0.579**	0	0.008	**0.436**	**0.895**	**0.401**	**0.149**

There are two main reasons for this outcome, as follows:

1. The nonsignificant financial and nonfinancial indicators have minimal impact on the overall financial risk of a company. For instance, the qualitative variables X9 and X10

do not significantly affect financial risk. This conclusion may contradict conventional wisdom. We surveyed numerous entrepreneurs, and their middle and upper management believed that a company's financial risk is related to its nature and industry attributes. They also, generally, agreed that state-owned enterprises are less prone to delisting, while private enterprises facing financial risks are more vulnerable to delisting. Different industries have varying policies, development stages, competitive pressures, and lifecycles, leading to differences in products, markets, profits, and costs. Favorable industries combined with good policies can reduce the likelihood of delisting by the China Securities Regulatory Commission, even in the presence of financial risks. However, our large-sample verification proves that these perceptions are unscientific and merely perceived, highlighting that quantitative analysis is often much more reliable than perceptual analysis. Since these two qualitative indicator variables are not significant in the risk model and have only a weak correlation with the company's financial risk, this means that, from the perspective of the company's nature, there is no practical or obvious differences between state-owned and private enterprises or whether the listed company is ST or non-ST. Similarly, there is no distinction between the manufacturing and nonmanufacturing industries. This analytical conclusion aligns with the goal of this article, as follows: to find a universally applicable early-warning model for financial risk management to avoid financial risks in advance. This finding suggests that the model does not need to distinguish between the nature of the company and the attributes of the industry, making it more universally applicable.

2. There is strong collinearity among the eight financial indicator variables, indicating a linear relationship between the independent variables. Some scholars refer to this as multicollinearity or multiple correlations. Collinearity and correlation are not the same concept. Collinearity involves a strong correlation between multiple independent variables, leading to a linear causal relationship between them, which can interfere with the degree of influence on the dependent variable. In regression or logistic regression analysis, when multicollinearity occurs among independent variables, several issues may arise in the model or parameter testing, such as low prediction accuracy and stability of the model, difficulty in determining the impact of individual variables, increased standard errors in the regression equation, and loss of variable significance. Common methods to address collinearity include directly deleting unimportant collinear variables, increasing the sample size, using ridge regression, partial least squares, or employing principal component or factor dimensionality reduction to eliminate collinearity among independent variables. In our model, the significance of the four financial indicators is greater than 0.05, suggesting strong collinearity among them. Simply deleting the four nonsignificant financial indicator variables would result in a lack of necessary financial information when analyzing the company's financial risks, leading to incomplete and subjective analysis and possibly incomplete interpretation of the remaining four financial variables. Therefore, to avoid losing financial information, factor analysis can be used to eliminate collinearity among the independent variables, representing a methodological innovation in this article.

On the basis of the conclusions drawn from the previous analysis, this article abandons the two nonfinancial indicators of enterprise nature and industry classification and constructs a model solely composed of continuous financial indicators.

3.5. Factor Analysis

Factor analysis is an extension of principal component analysis (PCA), which is more inclined to describe the correlation among the original variables compared to PCA (Kim and Wei 2017; Yang et al. 2018). It studies the problem more deeply, integrating variables with intricate relationships into a few factors to reproduce the relationship between the original variables and factors.

A factor analysis is conducted to find a few random variables that can synthesize the main information of all variables by studying the internal dependencies of the correlation matrices (or covariance matrices) between multiple variables. These random variables cannot be directly measured and are usually called factors. The factors are independent of each other, and all variables can be expressed as a linear combination of common factors. The purpose of a factor analysis is to reduce the number of variables and replace all variables with a few factors to analyze the whole problem.

Factor analysis was first proposed by British psychologist C. E. Spearman. He found that there were certain correlations among students' scores in various subjects. Students with good scores in one subject often had better scores in other subjects, so he could infer whether there were some potential common factors or general intelligence conditions that affected students' academic performance. Factor analysis can reveal hidden representative factors among many variables. Grouping variables of the same nature into one factor reduces the number of variables and tests hypotheses about relationships among variables.

There are n samples and p indicators, in which $X = (X_1, X_2, \ldots, X_p)^T$ are random vectors, and the common factor to be found is $F = (F_1, F_2, \ldots, F_m)^T$. Then, the model,

$$\begin{gathered} X_1 = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \ldots + a_{2m}F_m + \varepsilon_2 \\ \ldots\ldots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \ldots + a_{pm}F_m + \varepsilon_p \end{gathered}$$

is called a factor model. The matrix $A = a_{ij}$ is called the factor loading matrix, where a_{ij} is the factor loading, and its essence is the correlation coefficient between common factor F_i and the variable X_j. ε is a special factor that represents the variable variation caused by influencing factors other than the common factor (which cannot be explained by the common factor), and it is ignored in the actual analysis.

For the obtained common factors, it is necessary to observe which variables have larger loads on them and then explain the actual meaning of the common factors accordingly. However, for the initial factor model obtained by the analysis, the factor loading matrix is often complicated, and it is difficult to provide a reasonable explanation for the factor Fi. In this case, further factor rotation can be performed so that a more reasonable explanation can be obtained after the rotation.

The model obtained by factor analysis has two characteristics. One is that the model is not affected by the dimension; the other is that the factor loading is not unique. By rotating the factor axis, a new factor loading matrix can be obtained, which makes its significance more obvious.

The above is an expression of the formulation of the orthogonal factor model, and the second expression is the estimation method. The estimation of the factor loading matrix is one of the main problems of factor analysis, but the most widely used is the factor rotation method (i.e., maximum variance rotation method), which is an orthogonal rotation method; the goal is to maximize the variance of the squared loads.

Variance-maximizing rotation is a method used in factor analysis to maximize the sum of variances of individual factor loadings through coordinate transformations. In layman's terms, it is as follows:

1. Any variable only has a high contribution rate on one factor, and the loading on other factors is almost 0;
2. Any one factor only has high loading on a few variables, while the loading on other variables are almost zero. A factor loading matrix that satisfies this condition is said to have a "simple structure".

Variance-maximizing rotation is the method used to rotate the loading matrix as close to the simple structure as possible. In terms of the samples represented by this set of variables, the variance-maximizing rotation finds the simplest way to represent the samples, i.e., each sample can be represented by a linear combination of functions of a few variables.

A mathematical expression for variance maximization is as follows:

$$R_{VARIMAX} = \arg\max_{R} \left(\sum_{j=1}^{k} \sum_{i=1}^{p} \left(\bigwedge R \right)_{ij}^{4} - \frac{\gamma}{p} \sum_{j=1}^{k} \left(\sum_{i=1}^{p} \left(\bigwedge R \right)_{ij}^{2} \right)^{2} \right)$$

This method was proposed by Henry Felix Kaiser, in 1958, and is a commonly used orthogonal rotation method (the factors remain linearly uncorrelated after rotation).

Factor analysis has the following three applicable conditions:

1. The Sample Size Should Not Be Too Small: The task of factor analysis is to analyze the internal correlation structure among variables, so the sample size is required to be sufficient, otherwise the results will be unreliable. Generally speaking, the sample size should be more than 10 times the number of variables. To obtain the ideal results, it is better to have more than 25 times the number of variables. The total number of samples cannot be too small and, theoretically, requires more than 100 cases. However, with practical economic and social problems, often the sample size does not meet this requirement. In these cases, factor analysis is not impossible, but it is necessary to realize that because of an insufficient sample size, the model may be unstable, and caution should be exercised when interpreting the results.
2. There Should Be Correlations among the Variables: If the variables are independent of each other, the common factors cannot be extracted from them, and factor analysis cannot be applied. This can be determined by Bartlett's sphericity test. If the correlation matrix is a unit matrix, then the variables are independent and the factor analysis method is invalid. A better correlation indicator is the KMO test statistic, which has a value between 0 and 1. The closer the KMO test statistic is to 1, the stronger the partial correlation among the variables and the better the effect of the factor analysis. In the actual analysis, if the KMO test statistic is greater than 0.7, the effect of the factor analysis is generally better, while if the KMO test statistic is less than 0.5, the factor analysis method is not suitable, and redesign of the variable structure or the use of other statistical analysis methods should be considered.
3. Each Common Factor in the Factor Analysis Should Have Practical Significance: In principal component analysis, each principal component is actually the result of matrix transformation, so it does not matter whether the meaning is clear. However, in factor analysis, it is very important that the extracted factors have practical significance. If a reasonable professional explanation for each factor cannot be found, the analysis should be re-analyzed.

The factor analysis method in IBM SPSS Statistics 23 (represented by SPSS) software was used for the calculations.

First, a KMO measurement and Bartlett's test were conducted on the eight financial indicators. The results are shown in Table 6.

Table 6. KMO measurement and Bartlett's test.

KMO Measurement of Sampling Adequacy		0.684
Bartlett's Test of Sphericity	Approximate Chi-Square Value	486.096
	Degrees of Freedom	28
	Significance	0.000

From the results, we can see that the Bartlett statistic is 486.096, and its corresponding significance probability is 0.000, which is less than the significance level of 0.05. This indicates that the correlation matrix is not an identity matrix; therefore, it is suitable for factor analysis. The KMO value is greater than 0.6, suggesting that the factor analysis results are satisfactory.

Next, using SPSS software, we automatically calculated the eigenvalues and contribution values of each principal component, as detailed in Table 7.

Table 7. Eigenvalues and contribution values of principal components.

	Initial Eigenvalue			Extracted Sums of Squared Loadings			Rotated Sums of Squared Loadings		
	Total	**Variance%**	**Cumulative%**	**Total**	**Variance%**	**Cumulative%**	**Total**	**Variance%**	**Cumulative%**
1.	2.719	33.986	33.986	2.719	33.986	33.986	1.019	12.735	12.735
2.	1.244	15.552	49.537	1.244	15.552	49.537	1.016	12.698	25.433
3.	1.072	13.402	62.939	1.072	13.402	62.939	1.013	12.658	38.091
4.	0.799	9.984	72.923	0.799	9.984	72.923	1.012	12.646	50.737
5.	0.735	9.182	82.106	0.735	9.182	82.106	1.011	12.641	63.378
6.	0.633	7.916	90.022	0.633	7.916	90.022	1.004	12.544	75.922
7.	0.417	5.216	95.238	0.417	5.216	95.238	0.964	12.052	87.974
8.	0.381	4.762	100	0.381	4.762	100	0.962	12.026	100

Extraction Method: Principal Component Analysis.

Taking into account the amount of information represented by the actual indicators and the comprehensiveness of the indicators, we still specify the retention of eight factors. It is believed that these eight common factors reflect the comprehensive information of the original variables. Therefore, the factor analysis in this article only serves the purpose of eliminating collinearity.

Additionally, as can be seen from the scree plot of eigenvalues (Figure 2), the eigenvalue for Factor 8 is not particularly small. Moreover, the differences between Factors 2 to 8 are similar, making it difficult to justify discarding any one of them. Therefore, it is concluded that retaining all eight factors will not result in any loss of information.

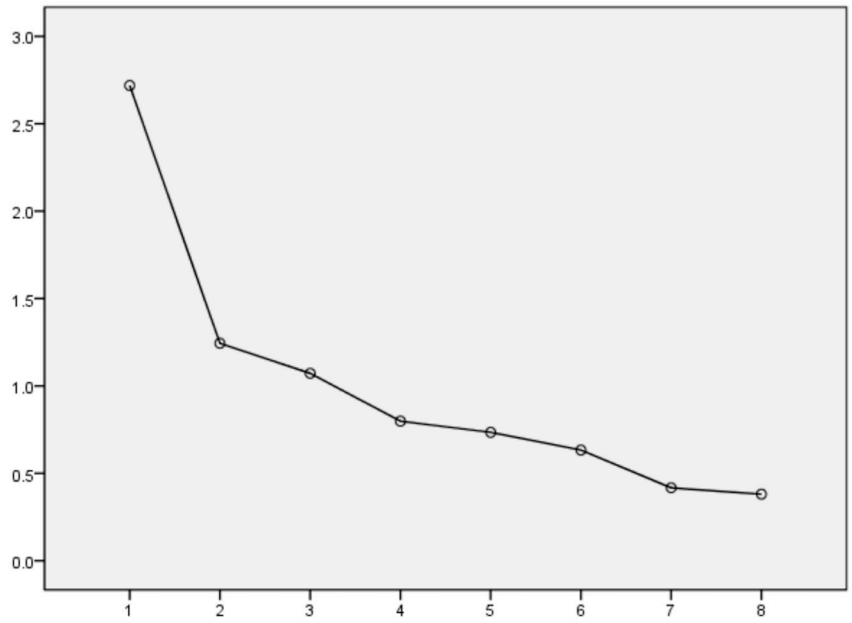


Figure 2. Scree plot.

In order to clearly reflect the relationship between the principal component factors and the original variables, we output the rotated factor loadings, as shown in Table 8.

From Table 8, it can be observed that the asset growth indicator has a relatively large loading on Factor 1; hence, it is named Asset Growth Factor (F1). The solvency indicator has a significant loading on Factor 2; thus, it is designated as Solvency Factor (F2). The profitability indicator exhibits a strong loading on Factor 3, leading to its denomination as Profitability Factor (F3). Similarly, the turnover indicator has a prominent loading on

Factor 4, making it Turnover Factor (F4). The income indicator is heavily loaded on Factor 5, and it is named Income Factor (F5). The cash flow indicator displays a significant loading on Factor 6, resulting in its designation as Cash Flow Factor (F6). The liquidity indicator has a strong loading on Factor 7, and it is named Liquidity Factor (F7). Finally, the leverage indicator is loaded on Factor 8, making it Leverage Factor (F8). The results of the factor analysis firmly validate the strategy of selecting these eight factors.

Table 8. Rotated factor loadings.

Com.	1	2	3	4	5	6	7	8
X1	0.15	0.213	0.185	0.11	0.143	0.042	0.914	0.168
X2	0.005	0.017	0.96	0.081	0.06	0.165	0.16	0.122
X3	0.265	0.056	0.142	0.138	0.133	0.116	0.17	0.91
X4	0.072	0.97	0.017	−0.03	0.115	0.071	0.178	0.048
X5	0.125	−0.029	0.079	0.971	−0.04	0.101	0.093	0.116
X6	0.054	0.072	0.163	0.103	0.122	0.964	0.038	0.099
X7	0.95	0.077	0.004	0.132	0.016	0.054	0.133	0.231
X8	0.017	0.115	0.059	−0.04	0.969	0.12	0.122	0.111

Extraction Method: Principal Component Analysis. Rotation Method: Kaiser Normalization with Varimax Rotation. The rotation converged after 6 iterations.

To establish an accurate relationship between the common factors and the indicators, it is necessary to express the common factors as linear combinations of the individual variables. Using the regression method within the factor analysis function of SPSS software, a factor score coefficient matrix can be generated, as shown in Table 9. This matrix allows us to calculate the factor scores based on the factor score coefficients and the standardized values of the original variables. With these factor scores, further analysis of the financial indicators can be conducted.

Table 9. Component score coefficient matrix.

Com.	1	2	3	4	5	6	7	8
X1	−0.113	−0.203	−0.179	−0.089	−0.113	0.041	1.247	−0.134
X2	0.07	0.032	1.125	−0.048	−0.014	−0.168	−0.207	−0.129
X3	−0.269	0	−0.11	−0.091	−0.112	−0.066	−0.133	1.258
X4	−0.059	1.097	0.033	0.065	−0.09	−0.071	−0.244	0.002
X5	−0.111	0.065	−0.049	1.087	0.081	−0.101	−0.106	0.107
X6	−0.029	−0.07	−0.171	−0.099	−0.119	1.105	0.049	−0.078
X7	1.166	−0.055	0.066	−0.104	0.033	−0.029	−0.126	−0.309
X8	0.036	−0.089	−0.014	0.08	1.093	−0.121	−0.132	−0.135

3.6. Logistic Regression

Logistic regression, also translated as "logarithmic probability regression", is a common algorithm for binary classification tasks. It is a supervised statistical learning method and is mainly used to classify samples. Unlike linear regression, the linear regression model is used to find a function, $f(x) = \omega x + b$, such that f(x) is as close as possible to the true value. Logistic regression, on the other hand, produces and outputs a marker, y (which has a Boolean value of 0, 1), on a binary classification task. The logistic regression is generalized linear regression, which means that for a regression task, if the output of the mark changes on an exponential scale, we can use $\ln(y)$ as the target for the model to predict the output approximation. Similarly, logistic regression is to find a function, g(z), to convert the real value of the predicted output of the linear regression into a Boolean value. Usually we use the sigmoid function, which takes the following form:

$$g(z) = 1/(1 + e^{-z})$$

The function is an S-shaped function. When the independent variable, z, approaches positive infinity, the dependent variable g(z) approaches 1, and when z approaches negative infinity, g(z) approaches 0. Any real number maps to the (0, 1) interval, making it useful to transform an arbitrary-valued function into a function more suitable for binary classification, which is a good fit for our classification probability model. Because of this property, the sigmoid function is also regarded as a method of normalization. Similar to normalization, it is a "scaling" function in data preprocessing, which can compress data into [0, 1]. The difference is that after normalization, 0 and 1 can be taken (the maximum value is 1 after normalization, and the minimum value is 0 after normalization), but the sigmoid function is only infinitely close to 0 and 1. Also, it has a nice derivative property, as follows:

$$g\prime(z) = g(z)(1 - g(z))$$

This is easily obtained by derivation of the function with respect to g(z). If we let z in g(z) be $z = \theta x$, then we obtain the general form of a binary logistic regression model, as follows:

$$h_\theta(x) = 1/(1 + e^{-\theta x})$$

where x is the sample input, and the model output is $h_\theta(x)$, which can be understood as the probability of a certain classification. And θ is the required model parameter of the classification model. For the model output, $h_\theta(x)$, we let it correspond to our binary sample output y (let them be 0 and 1); if $h_\theta(x) > 0.5$ and $\theta x > 0$, y is 1. If $h_\theta(x) < 0.5$ and $\theta x < 0$, then y is 0; y = 0.5 is a critical case, at $\theta x = 0$, the classification of which cannot be determined from the logistic regression model itself.

The smaller the value of $h_\theta(x)$, the higher the probability of being classified as 0, and, conversely, the larger the value, the higher the probability of being classified as 1. Upon nearing the critical point, the classification accuracy will drop. Here, we can also write the model in the following matrix mode:

$$h_\theta(X) = 1/(1 + e^{-\theta X})\Big)$$

After understanding the model of binary classification and regression, we have to look at the loss function of the model. Our goal is to minimize the loss function to obtain the corresponding model coefficient, θ.

Since linear regression is continuous, the loss function can be defined using the sum of the squares of the model errors. However, logistic regression is not continuous, and the experience of defining the loss function in natural linear regression cannot be used. However, we can use the maximum likelihood method to derive our loss function.

According to the definition of binary logistic regression, it is assumed that our sample output is 0 or 1. Then, we have the following:

$$P(y = 1|x, \theta) = h_\theta(x)$$
$$P(y = 0|x, \theta) = 1 - h_\theta(x)$$

These two formulas are then combined into the following single formula:

$$P(y|x, \theta) = h_\theta(x)^y(1 - h_\theta(x))^{1-y}$$

The value of y can only be 0 or 1.

With the probability distribution function expression of y, we can use the likelihood function maximization to solve for the model coefficient θ that is needed.

Here, we used the maximum likelihood estimation method. This method was first proposed by the German mathematician C. F. Gauss, in 1821, but this method is usually attributed to the British statistician R. A. Fisher. Maximum likelihood estimation is just an application of probability theory in statistics, and it is a method of parameter estimation. This means that a random sample is known to satisfy a certain probability distribution, but

the specific parameters are not clear. Parameter estimation involves conducting several experiments, observing the results, and using the results to deduce the approximate value of the parameters. Maximum likelihood estimation establishes that a certain parameter can make the probability of a sample appear to be the largest. Of course, we will not choose other samples with small probabilities, so we simply use this parameter as the real value of the estimate.

In order to facilitate the solution, here we maximize the log-likelihood function, and the inverse of the log-likelihood function is our loss function, J(θ). The algebraic expression of the likelihood function is as follows:

$$L(\theta) = \prod_{i=1}^{m} (h_\theta(x^{(i)}))^{y^{(i)}} (1 - h_\theta(x^{(i)}))^{1-y^{(i)}}$$

where m is the number of samples.

The expression for the inversion of the logarithm of the likelihood function, that is, the loss function expression is as follows:

$$J(\theta) = -\ln L(\theta) = -\sum_{i=1}^{m} (y^{(i)} \log(h_\theta(x^{(i)}))) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))$$

The loss function can be expressed more concisely in the following matrix method:

$$J(\theta) = -Y^T \log(h_\theta(X)) - (E - Y)^T \log(E - h_\theta(X))$$

where E is an all 1 vector.

To minimize the loss function in binary logistic regression, there are many methods, and the most common are the gradient descent method, coordinate axis descent method, quasi-Newton method, etc., with which we derive the formula for each iteration of θ using gradient descent. Because of the tedious derivation of the algebraic method, the matrix method is used to optimize the loss function. Here, the process of deriving the gradient in binary logistic regression by the matrix method is provided.

For the above loss function expressed by the matrix method, we can use the derivation of the $J(\theta)$ for the θ vector to obtain the following:

$$\frac{\partial}{\partial \theta} J(\theta) = X^T \left[\frac{1}{h_\theta(X)} \odot h_\theta(X) \odot (E - h_\theta(X)) \odot (-Y) \right] + X^T \left[\frac{1}{E - h_\theta(X)} \odot h_\theta(X) \odot (E - h_\theta(X)) \odot (E - Y) \right]$$

In this step, we used the chain rule of vector derivation, and the matrix form of the following three basic derivation formulas (g(z) is a sigmoid function):

$$\frac{\partial}{\partial x} \log(x) = \frac{1}{x}$$

$$\frac{\partial}{\partial x} g(z) = g(z)(1 - g(z))$$

$$\frac{\partial x\theta}{\partial \theta} = x$$

For the above derivation formulas, we can simplify them to obtain:

$$\frac{\partial}{\partial \theta} J(\theta) = X^T (h_\theta(X) - Y)$$

Therefore, the iterative formula for each step vector θ in the gradient descent method is as follows:

$$\theta = \theta - \alpha X^T (h_\theta(X) - Y)$$

where α is the step size of the gradient descent method.

In this article, ST listed companies were coded as 0, and non-ST listed companies were coded as 1, serving as the dependent variable. Using the eight influencing factors identified through factor analysis as independent variables, a logistic regression analysis was conducted with the assistance of SPSS software. The regression results are presented in Table 10.

Table 10. Logistic regression coefficients.

Com.	B	S.E.	Wald	df	Significance	Exp(B)
F1	0.724	0.183	15.684	1	0	2.062
F2	0.342	0.201	2.902	1	**0.088**	1.408
F3	2.919	0.509	32.857	1	0	18.521
F4	1.499	0.245	37.334	1	0	4.477
F5	0.456	0.239	3.649	1	**0.056**	1.578
F6	1.354	0.232	34.007	1	0	3.875
F7	1.135	0.241	22.231	1	0	3.11
F8	1.784	0.395	20.442	1	0	5.953
Constant	−0.613	0.248	6.104	1	0.013	0.542

The significance results in Table 10 indicate that the significance levels of F2 and F5 are greater than 0.05 but less than 0.1. We recommend adopting a significance level of 0.1 for modeling on the basis of the following two reasons:

1. Commonly used significance levels are 0.05, 0.01, and 0.001, each with their own merits and drawbacks. A smaller level provides stronger evidence for determining significance, but it also increases the risk of not rejecting a false null hypothesis, thereby heightening the likelihood of a type II error and reducing statistical power. The choice of a significance level inevitably involves a balance between type I and type II errors. In cases in which the consequences of type I errors are not severe, such as in exploratory studies akin to our financial risk early-warning model, we can set the significance level to a higher value, such as 0.05 or 0.1. Furthermore, given our relatively small research sample, increasing the significance level in the article may help enhance the statistical power.
2. Based on the factor variance contribution rates after factor rotation in the previous section, the variance percentages of the eight factor variables are nearly equal, indicating a high degree of balance. Each factor explains a financial variable well and exclusively. The absence of any factor variable signifies a loss of financial risk information represented by that variable. Under such data validation, eliminating or replacing any factor would result in the model missing crucial financial risk information pertaining to the company.

Therefore, the model in this article is established with a significance level of 0.1, thus fully retaining the eight influencing factors.

Using a cut-off threshold of 0.5, observations of the model's performance on the sample data are presented in Table 11.

Table 11. Evaluation of the logistic regression model.

Observations		Estimate		
		ST = 0; non-ST = 1		Percentage Correct
		0.0	1.0	
ST = 0; non-ST = 1	0.0	143	17	89.4
	1.0	16	144	90.0
Overall Percentage				89.7

Table 11 indicates that the logistic regression model achieved an overall prediction accuracy of 89.7% for the sample data. The model incorporates a comprehensive set of dimensional features and exhibits strong explanatory power, suggesting that its predictive capability is reliable and well supported.

3.7. Risk Level Classification

In China, there is no unified standard for classifying financial risk warning levels. Methodologically, an analytical approach using efficiency coefficients is applied (Gao and Wang 2005). Specifically, all comprehensive evaluation scores of sample companies are subject to extreme value processing, and each comprehensive score is mapped to a range between 0 and 100. Generally, the risk levels are divided into the following five categories:

1. High risk: score less than or equal to 60.
2. Significant risk: score greater than 60 but less than or equal to 70.
3. Moderate risk: score greater than 70 but less than or equal to 80.
4. Minor risk: score greater than 80 but less than or equal to 90.
5. No risk: score greater than 90.

While this risk classification method is commonly used, the existing financial risk management warning model with five risk levels is not practical, as it is difficult for most business operators and managers to memorize and distinguish among the risks. Therefore, this article proposes a more scientifically feasible approach by dividing the risk levels into four categories based on the following three significance levels: 0.01, 0.05, and 0.1. This method has been described in previous sections but will be re-explained here for clarity.

The selected financial indicators have a negative correlation with the company's financial risk, indicating that a higher model score corresponds to a lower financial risk for the company. As evident from the sigmoid function, the model score is positively correlated with the sigmoid probability value, suggesting that higher model scores and probabilities correspond to lower financial risks (trending toward no risk). Different significance levels reflect the risk probability of a company being classified as an ST stock, thus requiring a conversion formula, as follows: probability of the company not being an ST stock (NRP) = (1 − ST risk probability (RP)), as shown in Table 12.

Table 12. Corresponding probability levels.

Level	ST Risk Probability (RP)	Corresponding to No Risk (NRP)
A: Significant Risk	RP $\geq$ 50%	NRP < 50%
B: Moderate Risk	10% $\leq$ RP < 50%	50% $\leq$ NRP < 90%
C: Minor Risk	5% $\leq$ RP < 10%	90% $\leq$ NRP < 95%
D: No Risk	RP < 5%	NRP $\geq$ 95%

This classification method aligns with the common sense and scientific understanding of most business management personnel in companies. Utilizing the sigmoid function, the comprehensive score range of the company is mapped to a probability value between 0 and 1, forming a risk probability that is highly suitable for financial risk warning in domestic operating companies. This model of classifying financial risks based on significance levels is also an innovation of this article.

We scientifically categorized the probability values and intervals for the four risk levels. However, the question remains concerning how to establish the mathematical relationships between probability values and model scores. In practical applications, we need to first calculate the company's risk value, convert it into a risk probability value, and determine the risk level based on the probability interval. After identifying the interval, we can ascertain the company's risk level.

In logistic regression analysis, the sigmoid function is used as an activation function to convert linear data into probability data. An intermediate transformation variable, Z, is

required. The linear value of the original risk model is first converted into a Z-score, which is then substituted into the activation function for the probability calculation. The equation is shown in the Figure 3.

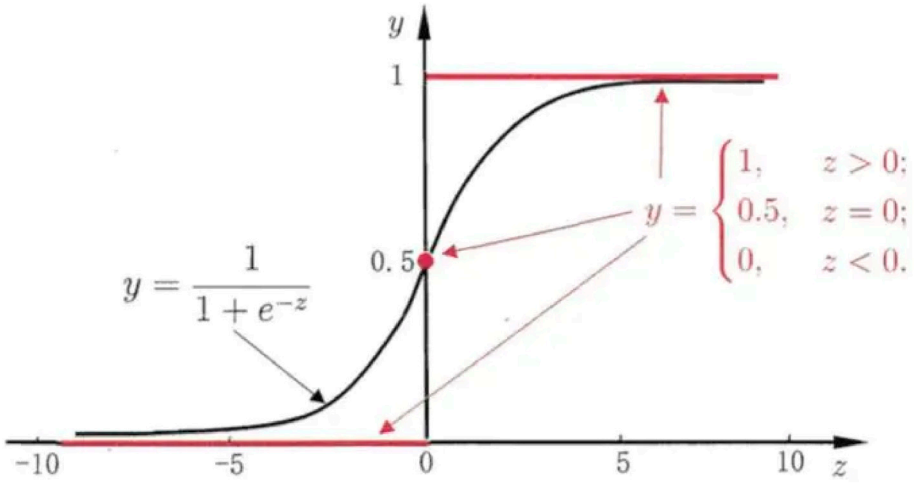


Figure 3. Sigmoid function.

Utilizing the probability values for the four risk levels and the sigmoid function, we derive corresponding Z-values. The calculation results for the four risk levels are providing in Table 13.

Table 13. Corresponding Z-value levels.

Grade A: Significant Risk	Grade B: Moderate Risk	Grade C: Minor Risk	Grade D: No Risk
$Z \leq 0$	$0 < Z \leq 2.2$	$2.2 < Z \leq 2.94$	$Z > 2.94$

Enterprises with economic strength can transform the model into a dynamic monitoring and insight product, enabling real-time data capture and continuous monitoring of their financial risk status. This enhances an enterprise's resilience and adaptability to macro- and micro-environmental risks.

4. General Model

Mathematical Definition

On the basis of the experimental results discussed earlier, this article defines the functional relationships for the general model of financial risk management and warning. Firstly, it is necessary to define four constant matrices for the solution, which are derived from the modeling experiments outlined previously.

The component score coefficient matrix, $C_{8\times8}$, obtained from the factor analysis of the modeling samples, is presented below. Each row and column represents X1 to X8.

$$C_{8\times8} = \begin{bmatrix} -0.113 & -0.203 & -0.179 & -0.089 & -0.113 & 0.041 & 1.247 & -0.134 \\ 0.07 & 0.032 & 1.125 & -0.048 & -0.014 & -0.168 & -0.207 & -0.129 \\ -0.269 & 0.0 & -0.11 & -0.091 & -0.112 & -0.066 & -0.133 & 1.258 \\ -0.059 & 1.097 & 0.033 & 0.065 & -0.09 & -0.071 & -0.244 & 0.002 \\ -0.111 & 0.065 & -0.049 & 1.087 & 0.081 & -0.101 & -0.106 & 0.107 \\ -0.029 & -0.07 & -0.171 & -0.099 & -0.119 & 1.105 & 0.049 & -0.078 \\ 1.166 & -0.055 & 0.066 & -0.104 & 0.033 & -0.029 & -0.126 & -0.309 \\ 0.036 & -0.089 & -0.014 & 0.08 & 1.093 & -0.121 & -0.132 & -0.135 \end{bmatrix}$$

The mean vector, $M_{8\times1}$, and the standard deviation vector, $S_{8\times1}$, of the feature dimensions for the modeling samples are presented below. Each element in these vectors corresponds to X1 to X8. Additionally, the linear weighted weights, $W_{9\times1}$, for the logistic regression are also provided, with the first eight representing the weights of F1 to F8 after the factor analysis, and the last one representing the bias constant.

$$M_{8\times1} = \begin{bmatrix} 0.106496 \\ -0.229231 \\ -0.039178 \\ 5.121727 \\ 0.4283 \\ 0.104675 \\ 0.016794 \\ 0.235367 \end{bmatrix} \quad S_{8\times1} = \begin{bmatrix} 0.378862 \\ 0.8700613 \\ 0.1866928 \\ 7.0305917 \\ 0.30587 \\ 0.3066517 \\ 0.3365726 \\ 0.2228461 \end{bmatrix} \quad W_{9\times1} = \begin{bmatrix} 0.724 \\ 0.342 \\ 2.919 \\ 1.499 \\ 0.456 \\ 1.354 \\ 1.135 \\ 1.784 \\ -0.613 \end{bmatrix}$$

There are N groups of predicted sample feature data matrices, $D_{N\times8}$. The rows and columns are arranged in the characteristic order of X1–X8. The calculation principles and steps are as follows:

Step 1: Transpose the column vectors of the feature dimension mean vector, $M_{8\times1}$, and the feature dimension standard deviation vector, $S_{8\times1}$, to construct N×8 matrices as $M_{N\times8}$ and $S_{N\times8}$, where each row of the new matrices is the original vector.

Step 2: Calculate the $D_{N\times8}$ matrix Z-score standardization to obtain $D'_{N\times8}$; the calculation formula is as follows:

$$D'_{N\times8} = \frac{D_{N\times8} - M_{N\times8}}{S_{N\times8}} \tag{1}$$

Step 3: Multiply the matrix, $D'_{N\times8}$, and the component score coefficient matrix, $C_{8\times8}$, to construct the factor score matrix, $D''_{N\times8}$. The calculation formula is as follows:

$$D''_{N\times8} = D'_{N\times8} C_{8\times8} \tag{2}$$

Step 4: Concatenate the last column of "a" with $D''_{N\times8}$, and set of all 1 vectors of $E_{N\times1}$ to construct a new set of $D'''_{N\times9}$ matrices.

Step 5: Transpose the column vector of the linear weighted weight, $W_{9\times1}$, of the logistic regression to construct an N×9 matrix as $W_{N\times9}$; calculate the matrices $D'''_{N\times9}$ and $W_{N\times9}$; perform the Hadamard product; and then sum the rows; that is, multiply the corresponding elements and calculate the rows. Obtain the linear weighted vector, $Z_{N\times1}$; the calculation formula is as follows:

$$Z_{N\times1} = \sum (W_{N\times9} \circ D'''_{N\times9}) \tag{3}$$

Step 6: Calculate the sigmoid function mapping the $NRP_{N\times1}$ value of the linear weighted vector, $Z_{N\times1}$. The calculation formula is as follows:

$$NRP_{N\times1} = \frac{1}{1 + e^{-Z_{N\times1}}} \tag{4}$$

Step 7: Classify $RP_{N\times1}$ according to the standards of the risk classification provided in Section 3.7. The core formula is RP = 1 − NRP.

Following the general model defined above, formulas can easily be used to determine the probability and level of a company's financial risk. Next, we use the 2019 financial data of two companies, ST Shenji (000410) and ST Culture (300089), for formula verification, showing the calculation process and verification reliability.

First, calculate the metric data needed for our model based on the financial statements of the two companies, as follows:

$$D_{2\times 8} = \begin{bmatrix} 0.0279 & -0.3649 & -0.19 & 1.8751 & 0.07 & -0.0093 & -0.6463 & -0.9778 \\ -0.4912 & -0.1296 & -0.0499 & 0.9885 & 0.13 & -0.0082 & -0.0854 & 0.4627 \end{bmatrix}$$

Step 1: Transform the $M_{2\times 8}$, $S_{2\times 8}$, and $W_{2\times 9}$ matrices, which is not expanded upon here.

Step 2: Find the standardized $D'_{2\times 8}$ matrix, as follows:

$$D'_{2\times 8} = \frac{D_{2\times 8} - M_{2\times 8}}{S_{2\times 8}} = \begin{bmatrix} -0.2075 & -0.1559 & -0.8079 & -0.4618 & -1.1714 & -0.3717 & -1.9701 & -5.444 \\ -1.5776 & 0.1145 & -0.0574 & -0.5879 & -0.97532 & -0.3681 & -0.3036 & 1.0201 \end{bmatrix}$$

Step 3: Find the factor score matrix $D''_{2\times 8}$, as follows:

$$\begin{aligned} D''_{2\times 8} &= D'_{2\times 8} C_{8\times 8} \\ &= \begin{bmatrix} -2.0953 & 0.0733 & 0.0025 & -1.3977 & -5.9083 & 0.5273 & 1.0665 & 0.2781 \\ 0.038 & -0.4327 & 0.4745 & -0.8085 & 1.3058 & -0.4613 & -1.851 & 0.0037 \end{bmatrix} \end{aligned}$$

Step 4: Construct the $D'''_{2\times 9}$ matrix according to the experimental results, as follows:

$$D'''_{2\times 9} = D''_{2\times 8} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Step 5: According to the above requirements, construct the $Z_{2\times 1}$ vector, as follows:

$$Z_{2\times 1} = \sum (W_{2\times 9} \circ D'''_{2\times 9}) = \begin{bmatrix} -4.4665 \\ -2.6837 \end{bmatrix}$$

Step 6: Find the vector of $NRP_{2\times 1}$ after the sigmoid function mapping, as follows:

$$NRP_{2\times 1} = \frac{1}{1 + e^{-Z_{2\times 1}}} = \begin{bmatrix} 0.0114 \\ 0.0639 \end{bmatrix}$$

Step 7: According to the classification standards we proposed and RP = 1 − NRP, these two companies belong to the A-level of significant risk, and the performance is consistent with the actual financial situations of the companies. It is preliminarily verified that our model can be reliably used in practice.

To validate the generalization capability of the model and its parameters, we randomly selected 30 samples from the financial data of listed companies with ST (including *ST) status in 2019, and another 30 samples from financially healthy non-ST listed companies, resulting in a total of 60 validation datasets.

To verify whether the financial risk early-warning model proposed in this article performs equally well on new datasets, we set the following thresholds based on our proposed grading criteria. Specifically, we expected that for the ST-labeled datasets, the predicted risk probability (RP value) would be $\geq$0.5, indicating a significant risk (Grade A). Conversely, for the non-ST datasets, we expected the RP value to be <0.05, indicating a low-risk enterprise (Grade D). This represents the ideal validation outcome.

Following the computational steps of our proposed generalized model and parameters, the prediction accuracies on the 60 validation datasets under the aforementioned thresholds are presented in Table 14.

It can be seen from the above table that only two ST companies were unsuccessful in the prediction, namely, ST Busen (002569) and ST Senyuan (002358), with RP values of 0.36 and 0.38, respectively. They are both Grade B—moderate risk according to our grading criteria.

Table 14. Divide Corresponding Table.

Observations		Estimate		Percentage Correct
		ST = 0 non-ST = 1		
		0.0	1.0	
ST = 0; non-ST = 1	0.0	28	2	93.3
	1.0	0	30	100.0
Overall Percentage				96.7

We also analyzed and reviewed the announcements of these two companies and found the following:

1. The status of ST Busen changed on 8 June 2020 from *ST to ST, indicating that the financial situation of ST Busen improved in 2019. Although its status is ST, the performance for the year according to the financial statement was better, so our prediction of a Grade B—moderate risk is also accurate and reliable.
2. The reason why ST Senyuan was given an ST status, on 29 April 2021, was that it had issued an internal control audit report or assurance report that did not express an opinion or a negative opinion over the last year or failed to disclose the internal control audit report in accordance with regulations. This indicates that, in 2020, ST Senyuan was issued an ST status because of an audit opinion. We checked the company announcements of ST Senyuan and found that it did not have an ST status in 2019, and the financial book situation was relatively good, so we concluded that a prediction of a Grade B—moderate risk is acceptable. This also proves that the risks predicted by our model are instructive for the future.

5. Conclusions

This article proposes a corporate financial risk management early-warning model by examining the operational and management issues existing in Chinese A-share listed companies that were subject to special treatment (ST) status by the China Securities Regulatory Commission (CSRC). Through theoretical and empirical research, 320 ST and non-ST listed companies in 2020 were selected as the research samples (160 for each category, designated as class 0 for ST samples and class 1 for non-ST samples). A risk warning model based on corporate financial risks was established, encompassing two qualitative variables and eight quantitative variables. The following conclusions were drawn through theoretical and empirical research:

1. A comprehensive financial indicator system for the early-warning model was designed, encompassing liquidity, profitability, leverage, solvency, turnover, cash flow, asset growth, profitability, enterprise nature, and industry classification. To address the practical challenge of comparing and assessing risks across companies with varying sizes, the financial variables in this model adopt ratio variables, effectively resolving this issue.
2. The model employed both qualitative and quantitative indicators, including the following two qualitative variables: company nature (X9) and industry classification (X10). Company nature primarily divides companies into two categories—private and state-owned—with the qualitative indicators converted into quantitative indicators. Private enterprises were labeled as 0 and state-owned enterprises as 1. These labels were used as independent variables in the model. During the logistic regression analysis on the SPSS platform, the significance levels of the two qualitative variables fell far below 0.05, indicating that company nature (X9) and industry classification (X10) do not significantly impact corporate financial risks. Therefore, these two qualitative variables were disregarded and excluded from the financial risk management early-warning model.

3. A factor–logistic fusion algorithm was adopted to construct the model. Factor analysis was extensively demonstrated to transform numerous indicators into influencing factors. Logistic regression analysis exhibited strong controllability and interpretable dimensional characteristics. Additionally, the model achieved an accuracy rate of 89.7% in predicting the ST status for 320 A-share listed companies in 2020 using training sample data. Furthermore, it predicted ST risks for 60 A-share listed companies in 2019, two years ahead of time, with a validation sample accuracy rate of 96.7%, thus proving the scientific validity and effectiveness of the corporate financial risk warning model.
4. This article determined an appropriate critical point for the new model and identified thresholds, classification methods, and intervals for different risk levels within the corporate financial risk management early-warning model.
5. This article argues that the financial data of listed companies in China are effective and possess strong predictive capabilities, enabling accurate assessments of listed companies' operational status through scientific induction and analysis.
6. The functional solution relationship of the model was fully quantified, and the calculation process was mathematically formalized. The model's practicality, scientific nature, and accuracy were quantitatively validated by verifying the results of corporate financial risks in 2019.

The risk model proposed in this article was used to conduct theoretical and empirical research solely based on financial data from A-share listed companies in China. However, it lacks financial data from foreign companies and domestic nonlisted companies. Additionally, the design of the model's indicators still lacks qualitative indicators, such as corporate ownership structure, executive equity incentives, board size, age of the company, and geographical distribution. In the future, optimizations and improvements can be made to enhance the model.

Author Contributions: Conceptualization, X.W.; data curation, H.W.; formal analysis, X.W.; investigation, H.W.; methodology, X.W.; validation, H.W.; writing—original draft, H.W.; writing—review and editing, H.W. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this article are available upon request from the corresponding author. The data are not publicly available due to privacy. Please contact the corresponding author before use.

Conflicts of Interest: Author Haitong Wei was employed by HongHao Data Intelligence Technology Co., Ltd., and the Data Intelligence Branch of Enterprise Financial Management Association of China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

Aguilera, Ana M., Manuel Escabias, and Mariano J. Valderrama. 2006. Using Principal Components for Estimating Logistic Regression with High-Dimensional Multicollinear Data. *Computational Statistics & Data Analysis* 50: 1905–24. [CrossRef]

Altman, Edward I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23: 589–609. [CrossRef]

Altman, Edward I., Marco Giancarlo, and Varetto Franco. 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking Finance* 18: 505–29. [CrossRef]

Beaver, William H. 1968. Market Prices, Financial Ratios, and the Prediction of Failure. *Journal of Accounting Research* 6: 179. [CrossRef]

Da, Fang, and Yi Peng. 2022. Non-Financial Indicators for Credit Risk Analysis of Chinese Technology-Oriented Micro and Small Enterprises. *Journal of the Operational Research Society* 74: 1198–10. [CrossRef]

Fitzpatrick, Paul Joseph. 1932. A Comparison of the Ratios of Successful Industrial Enterprises with those of Failed Companies. *Certified Public Accountant* 2: 598–605.

Frydman, Halina, Edward I. Altman, and Duen-Li Kao. 1985. Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *The Journal of Finance* 40: 269–91. [CrossRef]
Gao, Lei, and Sheng Wang. 2005. A Case Study on the Efficiency Coefficient Method of Financial Risk Warning. *Journal of Nanjing University of Finance and Economics* 1: 93–97.
He, Pinglin, Zhongfu Yu, and Jie Tao. 2009. Data Mining of Corporate Financial Risks: Financial Indicators or Non-Financial Indicators. Paper presented at 2009 International Conference on Management and Service Science, Beijing, China, September 20–22.
Huang, Dezhong, and Chaoqun Zhu. 2022. A Financial Risk Early Warning Model Introducing Enterprise Asset Quality Indicators. *Finance and Accounting Monthly (China)* 3: 48–52.
Jiang, Xiuhua, and Zheng Sun. 2001. Weakened Governance and Financial Crisis: A Predictive Model. *Nankai Management Review* 4: 7. [CrossRef]
Jiang, Yingbing, Gang Cui, and Yaowen Jiang. 2010. The Trends and characteristics of financial statement restatement of listed companies: 2004–2008. *Journal of Shanghai Lixin University of Commerce*, 17–27. [CrossRef]
Kim, Kang-Chul, and Hai-tong Wei. 2017. Development of a Face Detection and Recognition System Using a RaspberryPi. *The Journal of the Korea Institute of Electronic Communication Sciences* 12: 859–964. [CrossRef]
Li, Sheng, Hongzhen Liu, and Aihua Zhang. 2022. Financial Management Early Warning Model Based on Two-Layer Kriging Element Model Algorithm. Paper presented at 2022 International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), Bristol, UK, July 29–31.
Lv, Wenjin. 2023. Digital Transformation and Corporate Social Responsibility: Empirical Evidence from Chinese Listed Companies. *BCP Business & Management* 47: 204–12. [CrossRef]
Lyu, Jingwei. 2016. Research on the Establishment and Evaluation of Operational Risk Model of Current Financial İnstitutions in China. *International Journal of Business Administration* 7: 20–25. [CrossRef]
Ma, Ruowei, and Wei Zhang. 2014. Comparative study of financial distress discrimination models for listed companies based on Logistic and Fisher methods. *Journal of Beijing University of Business and Technology: Social Sciences Edition* 29: 8. [CrossRef]
Min, Jae H., and Young-Chan Lee. 2005. Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters. *Expert Systems with Applications* 28: 603–14. [CrossRef]
Nie, Lijie, Yanfang Zhao, and Yifan Gao. 2011. Research on the Construction of a Financial Crisis Warning Indicator System Based on Cash Flow—Based on Data from Listed Manufacturing Companies in China. *Economic Issues* 3: 108–12.
Odom, M. D., and R. Sharda. 1990. A neural network model for bankruptcy prediction. Paper presented at 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, June 17–21; pp. 163–68. [CrossRef]
Ofek, Eli. 1993. Capital structure and firm response to poor performance. *Journal of Financial Economics* 34: 3–30. [CrossRef]
Ohlson, James A. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* 18: 109–31. [CrossRef]
Peng, Fei. 2015. Research on Risk Management of China's Economic Operation under the 'New Normal of Economy'—Overview of the 8th China Lixin Risk Management Forum. *Economic Research* 6: 189–92.
Ren, Hong. 2020. Modeling of Corporate Financial Distress Alarm in China. *Finance* 10: 392–402. [CrossRef]
Shi, Yan. 2021. Exploring the Financial Risks of Listed Real Estate Enterprises—Taking Greenland Holding Group as an Example. Master's thesis, Fujian Jiangxia University, Fuzhou, China.
Sun, Haoyuan. 2021. Application of Financial Risk Early Warning Model of Manufacturing Enterprises Based on Principal Component Analysis. Paper presented at 2021 2nd International Conference on Internet and E-Business, Barcelona, Spain, June 9–11.
Trencheva, M. 2021. Preparation of Financial Statements and Financial Risk Analysis. Paper presented at ICERI Proceedings, Online, November 8–9.
Varetto, Franco. 1998. Genetic Algorithms Applications in the Analysis of Insolvency Risk. *Journal of Banking & Finance* 22: 1421–39. [CrossRef]
Wang, Ning. 2023. Construction and Application of an Early Warning Model Based on Financial Risk Management of Copper Enterprises. *Fulcrum* 7: 86–88.
Wang, Xin. 2020. Research on the Collection of Microcosmic Warning Indicators of Systematic Financial Risk Based on Big Data. Paper presented at 2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Phuket, Thailand, February 28–29.
Wang, Xinghai. 2014. Research on Predicting China's Macroeconomic Development. *The Journal of Statistics & Decision* 18: 117–20.
Wang, Yanjiao. 2024. Research and Application of Movie Recommendation Algorithms Based on Knowledge Graph and Graph Attention Network. Master's thesis, Fudan University, Shanghai, China.
Wen, Riquan, and Alexey Chernov. 2022. The Impact of COVID-19 on China's Small and Medium-Sized Enterprises and Their Response. Paper presented at 4th International Conference on Economic Management and Model Engineering (ICEMME), Nanjing, China, November 18–20.
Xu, Li. 2024. Research on Financial Diagnosis of Listed Companies Based on Data Mining. Master's thesis, Zhejiang University of Finance and Economics, Hangzhou, China.
Yang, Qifeng, Yingying Wang, and Yidong Ren. 2019. Research on Financial Risk Management Model of Internet Supply Chain Based on Data Science. *Cognitive Systems Research* 56: 50–55. [CrossRef]

Yang, Tong, Ibrahim Aliyu, and Chang-Gyoon Lim. 2018. Analysis of Dimensionality Reduction Methods through Epileptic EEG Feature Selection for Machine Learning in BCI. *The Journal of the Korea Institute of Electronic Communication Sciences* 13: 1333–42. [CrossRef]

Yang, Yunyi, Zhonghe Sun, and Jing Lu. 2017. Financial Risk Early Warning Model of Decision Tree Research Based on Z-Score Indicators. *Modern Management* 7: 185–93. [CrossRef]

Zanaj, Blerina, Arianit Peci, and Elma Zanaj. 2023. Reviewing of Different Algorithms Application with Qualitative Variables for the Different Financial Indicators. Paper presented at 2023 10th International Conference on Modern Power Systems (MPS), Cluj-Napoca, Romania, June 21–23.

Zeng, Shulian, Tao Wang, and Yaqiong Duan. 2023. Application of Financial Risk Warning for Listed Companies Based on XGBoost Model. *Business Accounting* 2: 62–66.

Zhang, Chuanqi, Xiang Li, and Gan Luo. 2020. The Differences between Chinese Accounting Standards for Business Enterprises and International Financial Reporting Standards as Well as the Convergence Strategy Method. *Proceedings of Business and Economic Studies* 5: 11–14. [CrossRef]

Zhang, Jianfeng, Wen-sheng Cui, Ming Liu, and Shuo Wang. 2018. Research on Financial Risk Intelligent Early-Warning Method for Colleges and Universities Based on Bayesian Classification and Delphi Method. *DEStech Transactions on Computer Science and Engineering*. [CrossRef]

Zhang, Liangchen, and Guangli Yang. 2023. Reducing Social Security Contribution Rate on the Financial Performance of State-Owned and Non-State-Owned Manufacturing Industries in the Post-Epidemic Era. *PLoS ONE* 18: e0287596. [CrossRef] [PubMed]

Zhao, Xikang. 2002. Empirical Analysis of Asset Liability Ratio and the Quality of Chinese Listed Companies. *Journal of Yan'an University: Social Science Edition* 24: 5.

Zhao, Xuefeng, Weiwei Wu, Delin Wu, Xu Guo, and Huining Shi. 2022. CFW-Boost Enterprise Financial Risk Warning Model for Feature Causal Analysis. *Journal of Systems Management* 31: 317–28.

Zhou, Shouhua, Jihua Yang, and Ping Wang. 1996. On the Early Warning Analysis of Financial Crisis—F-score Model. *Accounting Research* 8: 8–11.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.