mathematics

Article

A Study on Early Warnings of Financial Crisis of Chinese Listed Companies Based on DEA–SVM Model

Zhishuo Zhang [1], Yao Xiao [1], Zitian Fu [2], Kaiyang Zhong [3,*] and Huayong Niu [1,*]

1. International Business School, Beijing Foreign Studies University, Beijing 100089, China; zhangzhishuo@bfsu.edu.cn (Z.Z.); xiaoyaoibs@bfsu.edu.cn (Y.X.)
2. School of Economics, Sichuan Agricultural University, Chengdu 611134, China; 2021208028@stu.sicau.edu.cn
3. School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China
* Correspondence: zhongky@smail.swufe.edu.cn (K.Z.); niuhuayong@bfsu.edu.cn (H.N.)

Abstract: In the era of big data, investor sentiment will have an impact on personal decision making and asset pricing in the securities market. This paper uses the Easteconomy stock forum and Sina stock forum as the carrier of investor sentiment to measure the positive sentiment index based on stockholders' comments and to construct an evaluation index system for the public opinion dimension. In addition, the evaluation index system is constructed from four dimensions, which include operation, innovation, finance and financing, to evaluate the overall condition of listed companies from multiple perspectives. In this paper, the SBM model in the data envelopment analysis method is used to measure the efficiency values of each dimension of the multidimensional efficiency evaluation index system, and the efficiency values of each dimension are the multidimensional efficiency indicators. Subsequently, two sets of input feature indicators of the SVM model were established: one set contains traditional financial indicators and multidimensional efficiency indicators, and another set has only traditional financial indicators. The early warning accuracy of the two sets of input feature indicators was empirically analyzed based on the support vector machine early warning model. The results show that the early warning model incorporating multidimensional efficiency indicators has improved the accuracy compared with the early warning model based on traditional financial indicators. Then, the model was optimized by the particle swarm intelligent optimization algorithm, and the robustness of the results was tested. Moreover, six mainstream machine learning methods, including Logistic Regression, GBDT, CatBoost, AdaBoost, Random Forest and Bagging, were used to compare with the early warning effect of the DEA–SVM model, and the empirical results show that DEA–SVM has high early warning accuracy, which proves the superiority of the proposed model. The findings of this study have a positive effect on further preventing and controlling the financial crisis risk of Chinese-listed companies and promoting as well as facilitating the healthy growth of Chinese-listed companies.

Keywords: multidimensional efficiency indicators; public opinion index; financial crisis warning; data envelopment analysis; support vector machine

MSC: 91G50

1. Introduction

Since the Shenzhen Stock Exchange and Shanghai Stock Exchange officially opened in December 1990, China's stock market has gone through 32 years of development. On 4 September 2020, the total number of Chinese-listed companies in Shanghai and Shenzhen exceeded 4000. These Chinese-listed companies have raised funds for enterprise operation and development, deepened enterprise reform, and realized the conversion of operation mechanisms through the stock market. At the same time, these Chinese-listed companies cover various industries and are the leading enterprises in each field, and their financial

status is of great concern to stakeholders such as government departments, internal and external shareholders, business operators and creditors.

At present, China is facing challenges brought by structural and cyclical problems, economic slowdown, world economic downturn, trade protectionist measures and trade disputes provoked by some countries, as well as the impact of the new corona pneumonia epidemic on the supply chain, industrial chain and value chain, all of which have an impact on Chinese-listed companies. Once the financial crisis occurs in Chinese-listed companies, the harm is multifaceted; for the internal aspects of the enterprise, it will hinder normal production and operation of the enterprise, reduce the competitiveness of the enterprise, make it passive in the market competition and affect the production motivation of the employees. At the same time, as the foundation of the national economy, the financial crisis of Chinese-listed companies will affect national financial security and economic development processes, especially Chinese-listed companies as public interest entities, as their financial status and operating results are related to the interests of the majority of individual investors and are thus closely related to social stability. Therefore, by constructing an effective financial crisis early warning model, it helps managers and operators of Chinese-listed companies to anticipate risks in advance, to further analyze risks, control them, take timely countermeasures to resolve them, and protect the legitimate rights and interests of stakeholders.

The occurrence of financial crises is generally not a one-time occurrence. They can be predicted before they occur, and crisis warnings can be issued to relevant stakeholders, thus enabling the relevant managers to take targeted measures against the company. It is not the difference in financial indicators that generates a crisis but the numerous internal and external factors that are reflected in the financial indicators that make them a sign [1]. Therefore, this paper composes the theoretical basis of financial crisis early warning from multiple perspectives to provide the theoretical basis for the subsequent selection of early warning variables and model construction, mainly with the theories as follows. (1) Economic cycle theory: economic cycle refers to the cyclical occurrence of economic expansion and economic contraction in economic operation [2], which brings about the financing, investment, and operation risks of enterprises, and being in economic cycle will result in corresponding changes in the financial situation of enterprises. Economic cycle theory is one of the important theories to study the financial crisis of enterprises and its early warning. (2) Principal-agent theory: principal-agent theory is mainly developed and generated with the separation of ownership and operation of an enterprise, and the principal-agent problem is generated [3]. Through financial crisis warning, the principal can indirectly grasp the overall financial situation of the enterprise and its business development and evaluate the agent's operation and management ability, and the agent can also indirectly prevent and control the various risks encountered in its operation and management through financial crisis warning. (3) Enterprise crisis management theory: enterprise crisis management can generally be divided into three stages. The first stage is crisis prevention stage, which focuses on promoting the crisis early warning system, making the management and all employees aware of the crisis; the second stage is the crisis disposal stage, where once people discover the crisis, they immediately take corresponding measures to minimize their losses; the third stage is the crisis summary stage, which mainly analyzes and summarizes the experience and lessons learned in the second stage and further establishes and improves the crisis management mechanism [4]. The financial crisis warning is an important stage of enterprise crisis management, which can help managers and employees to identify risks and take countermeasures as early as possible to reduce losses; thus, it is of certain practical significance to apply this theory to guide the financial crisis warning.

This paper investigates the early warning model for the financial crisis of listed companies based on data envelopment analysis and support vector machines. Data envelopment analysis, as a nonparametric analysis method, extends the concept of single-input, single-output engineering efficiency to the evaluation of the efficiency of multiple-input, multiple-output decision units, and has become a common and important analytical tool.

In 1978, Charnes, Cooper, and Rhodes [5] created the first theoretical approach to DEA, the CCR model, which is composed of the initials of the last names of the three individuals. Moreover, the DEA method has been widely used in environmental assessment, bank efficiency evaluation and other important fields in recent years [6–8]. Support vector machine (SVM) is a binary classification model, that is, a linear classifier that finds the partitioned hyper-plane with the largest interval, and its learning strategy is interval maximization, which can eventually be translated into the solution of a convex quadratic programming problem. According to literature studies, SVM has higher warning accuracy compared to other early warning models of machine learning and statistical methods. Support vector machine, first proposed by Cortes and Vapnik in 1995, has shown many unique advantages in solving small-sample, nonlinear and high-dimensional pattern recognition, and their excellent performance in text classification has become a major technique in machine learning, and it is extended to other machine learning problems such as function fitting. In recent years, classification algorithms have been a research hot spot, and the earliest classification algorithm used is a neural network, which can classify and predict unknown data. However, as a heuristic learning machine, it inherently has a large empirical component. To avoid the problems of neural network structure selection and local mini-ma, Vapnik proposed a support vector machine, and with the rise of statistical learning, the addition of kernel functions formally laid the theoretical foundation of SVM [9–11].

The innovative contributions of this paper are mainly as follows: (1) Previous financial crisis early warning studies mostly focus on traditional financial indicators without considering multidimensional efficiency indicators. This paper constructs a multidimensional efficiency evaluation index system from operational, innovation, financial and financing dimensions to evaluate the overall situation of listed companies more comprehensively. (2) Based on the background of big data, this paper introduces the positive sentiment index of stockholders' comments from the Easteconomy stock forum and Sina stock forum. Then, the construction of an evaluation index system of public opinion dimension was completed to explore the early warning effect of investor sentiment on the financial crises of listed companies through the perspective of public opinion. (3) The SBM model in data envelopment analysis is used to measure the efficiency value of the multidimensional efficiency evaluation index system. The efficiency value obtained after the measurement is combined with traditional financial indicators to form the index system of financial crisis early warning, which enriches the research of the SVM early warning index system. (4) This paper applies the hybrid DEA–SVM model to the study of financial crisis early warning of listed companies, which effectively improves the accuracy of the early warning model and is verified by the data of Chinese-listed companies. (5) This paper uses the particle swarm optimization algorithm for parameter seeking to improve the early warning effect of the model and to verify the robustness of the early warning effect of the DEA–SVM model. (6) This paper compares the early warning effect of six mainstream machine learning methods, including Logistic Regression, GBDT, CatBoost, AdaBoost, Random Forest and Bagging, with that of the DEA–SVM model, which further verifies the superiority of the proposed model and provides a new feasible method for listed companies to conduct financial crisis early warning.

The structure of this paper is as follows. Section 2 is the literature review, Section 3 introduces the basic models and methods involved in this study, Section 4 designs the financial crisis early warning indicators, mainly including the construction of a multidimensional efficiency evaluation index system and the selection of traditional financial indicators, and Section 5 conducts an empirical analysis and discusses the results based on the mentioned index design and model selection. First, the SBM model is used to measure the efficiency values of each dimension. Subsequently, an empirical study is conducted based on a support vector machine model for financial crisis early warning. Through the optimization of the original model by particle swarm optimization algorithm, the effectiveness of the DEA–SVM model is further evaluated by comparing with the early warning

effect of various mainstream machine learning models. Section 6 is a summary and an outlook on future research directions.

In conclusion, this paper constructs an evaluation index system from five aspects of multiple dimensions—operation, innovation, finance, financing, and public opinion. It uses data envelopment analysis to measure the efficiency value of the constructed evaluation index system to obtain multidimensional efficiency indicators, and the SBM model is used for the selection of the data envelopment analysis method, which effectively improves the DEA–CCR and BCC models that cannot measure all slack variables, which is beneficial to the accuracy of the efficiency assessment. The early warning model is constructed based on this efficiency value and traditional financial indicators and by comparing the early warning effect of different kernel functions of nonlinear support vector machine finally. Then, a financial crisis early warning model with more early warning accuracy is derived, and the stability of the results is tested using the SVM model under the particle swarm optimization algorithm. This paper also adopts a variety of mainstream machine learning models for comparative study, which further proves the research results of this paper.

2. Literature Review

Financial crisis early warning generally starts with selecting some predictive indicators that are indicative of whether a financial crisis will occur in the future and then with applying statistical analysis or machine learning methods to predict the financial crisis based on these predictive indicators. According to this process, the effect of crisis early warning depends on the selection of forecasting methods and techniques and on whether the forecasting indicators can make a scientific and reasonable comprehensive evaluation of the overall situation of Chinese-listed companies.

In terms of forecasting method techniques, relevant research has gradually evolved from single-variable models to multivariable and conditional probability analysis models and then to the application of artificial intelligence. As the research progressed, modern researchers focused more on the application of machine learning to early warning models. The first research on financial crisis early warning was conducted by Fitzpatrick [12] who discerned that the ratios of net income/shareholders' equity and shareholders' equity/debt showed significant differences three years before the company's bankruptcy through the univariate bankruptcy early warning model he made, which was limited to descriptive analysis due to the lack of advanced tools at that time. Since then, Altman [13] applied multivariate statistical analysis to the field of financial crisis early warning, pioneering a multivariate early warning model that is more accurate than the previous approach of using a particular financial ratio to predict a company's bankruptcy crisis. Altman et al. [14] proposed the ZETA model for predicting financial crisis early warning, which extended the original Z-score model and overcame the shortcomings of it, which was only applicable to short-term forecasting. Ohlson [15] used the logistic regression method to construct a financial early warning model, which has no requirement for the distribution of independent variables and well solves the situation of difficult-to-satisfy preconditions faced in discriminant analysis; thus, it gradually replaces the discriminant analysis method and occupies a mainstream position in the field of financial early warning models. With the development of information technology, machine learning methods are gradually introduced into the field of financial crisis warnings. Vapnik [9] proposed the support vector machine method, which can better solve the small sample, nonlinear and high-dimensional support vector machine. Fan and Palaniswami [16] were the first to use support vector machine to solve the financial crisis early warning problem. Min and Lee [17] used a support vector machine model to predict 944 samples in the manufacturing industry and concluded that the support vector machine model outperformed BP neural networks, logistic regression analysis, multivariate linear warning models, and univariate warning models for early warning. Yan et al. [18] and Song [19] used the support vector machine method to construct an early warning model for the financial crisis of Chinese-listed companies, and both concluded that the prediction accuracy of the support vector

machine model was higher than other methods. Boyacioglu et al. [20] used various methods to conduct financial crisis early warning research, and the results showed that the support vector machine method was more effective than other methods.

In terms of forecasting indicators, there is no universally accepted standard for selecting indicators, and most studies have selected financial indicators and unidimensional efficiency indicators from the financial statements of Chinese-listed companies as forecasting indicators, and the relevant literature studies are as follows.

Beaver [21] constructed a univariate model based on a single financial ratio indicator based on statistical methods and conducted an empirical study on 158 firms. By applying 30 financial ratio indicators to a dichotomous test to find out the crisis cut-off point, the results showed that its early warning accuracy rate was higher than 70% five years before the enterprise's bankruptcy. Wu et al. [22] selected twenty-one financial ratios from five aspects, including profitability, solvency, operating capacity, growth capacity and enterprise size, and constructed an early warning model about the financial crisis. Wang et al. [23] selected eight financial crisis indicators including corporate per-share indicators, profitability, solvency, growth capacity, operating capacity and capital structure. Barboza et al. [24] used financial indicators such as operating profit margin, sales, number of employees, and change in return on equity as predictors. Xiao et al. [25] selected 15 financial indicators for financial crisis prediction from five aspects such as solvency, profitability, development capacity, operational capacity, and cash flow indicators of enterprises. As scholars in various countries have deepened their understanding of the causes of financial crises, some scholars have incorporated the efficiency evaluation indicators of Chinese-listed companies into the prediction indicators of early warning models. Ran et al. [26] proposed the application of relative operating efficiency as a predictor and combined it with the SVM model for the financial crisis early warning aspect. Xu et al. [27] introduced the enterprise operational efficiency into the early warning model to improve the model's early warning accuracy. Li et al. [28] considered the efficiency indicators of enterprises and constructed a financial early warning model accordingly.

In summary, the existing literature on financial crisis early warning shows more research on traditional financial indicators and unidimensional efficiency indicators and lacks the integration of multidimensional efficiency indicators. The construction of multidimensional efficiency indicators is beneficial to comprehensively evaluate the overall situation of Chinese-listed companies and to improve the early warning accuracy of the model.

3. Description of the Methodology

3.1. SBM Model

Since the efficiency values measured by the CCR and BCC models only reflect the part of radial improvement, i.e., the part of an equal proportional improvement to inefficiency and on the part of slack improvement, there are certain shortcomings in their efficiency assessment. For the consideration of improvement, Tone [29], i.e., Equation (1), takes into account the input–output relaxation problem and makes the efficiency measurement results more accurate.

$$\min \rho = \frac{1 - \frac{1}{m} \sum\limits_{i=1}^{m} \frac{s_i^-}{x_{ik}}}{1 + \frac{1}{q} \sum\limits_{r=1}^{q} \frac{s_r^+}{y_{rk}^g}} \quad \text{s.t.} \quad \begin{array}{l} X\lambda + s^- = x_k \\ Y\lambda - s^+ = y_k^g \end{array} \quad s^-, s^+, \lambda \geq 0 \tag{1}$$

In this model, ρ represents the efficiency value of DMU (x_0, y_0). s_i^- represents the redundancy of the ith input, s_r^+ represents the deficiency of the rth output. λ is the adjustment matrix, $X\lambda$ represents the amount of inputs on the frontier and $Y\lambda$ represents the amount of outputs on the frontier. $\frac{1}{m} \sum\limits_{i=1}^{m} \frac{s_i^-}{x_{ik}}$ is the average of the ratio of the redundancy of m inputs to the respective actual amount of inputs, i.e., the average inefficiency level

of m inputs. $\frac{1}{q}\sum_{r=1}^{q}\frac{s_r^+}{y_{rk}^g}$ is the ratio of the deficiency of q outputs to the respective actual amount of outputs, which is the average of the inefficiency level of output q, and $\frac{1}{1+\frac{1}{q}\sum_{r=1}^{q}\frac{s_r^+}{y_{rk}^g}}$ represents the efficiency level of output. It can be seen that in the model, the efficiency value of each DMU is the product of the average efficiency level of each input and the average efficiency level of each output.

Equation (2) is the non-oriented type of the SBM model, which considers both input-oriented and output-oriented efficiency. If considered from the management perspective of a Chinese-listed company, it is more concerned with the degree of input reduction required to achieve technological efficiency without reducing output. Therefore, the numerator is taken in the objective function of the non-oriented SBM model, which constitutes the input-oriented SBM model used in this paper.

$$\min \rho = 1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{ik}} \quad \text{s.t.} \quad \begin{array}{l} X\lambda + s^- = x_k \\ Y\lambda \geq y_k^g \end{array} \quad s^-, \lambda \geq 0 \tag{2}$$

3.2. Support Vector Machine (SVM)

Support vector machine is a binary classification model. The SVM algorithm works by finding an optimal segmentation hyper-plane. Its algorithm not only correctly separates the two classes of data, but also maximizes the classification margin between the two classes of data, which is represented graphically in Figure 1.
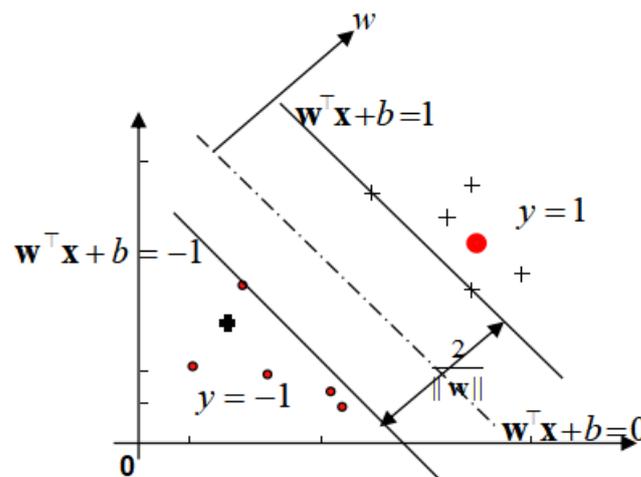
Figure 1. Graphical representation of support vector machine with linear separability.

In the case where hyperplane $\mathbf{w}^\top \mathbf{x} + b = 0$ is determined, $\left|\mathbf{w}^\top \mathbf{x} + b\right|$ can represent the distance from point $\mathbf{x}$ to the hyperplane, and the correctness of the classification depends on whether the sign of $\mathbf{w}^\top \mathbf{x} + b$ coincides with the sign of the marker y. Therefore, the correctness and certainty of the classification can be judged according to the positive or negative of $y(\mathbf{w}^\top \mathbf{x}) + b$. When the value of $y(\mathbf{w}^\top \mathbf{x}) + b$ is larger, the classification reliability is larger. By the principle of the maximum interval, the optimization problem of the linear separable problem is obtained as:

$$\begin{array}{l} \min\limits_{w,b} \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t.} \quad y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1, \; i = 1, 2, \ldots, m \end{array} \tag{3}$$

$\|\mathbf{w}\|$ denotes the Euclidean norm of the vector $\mathbf{w}$. Its geometric meaning is the length of the vector.

The optimization problem is solved by constructing a Lagrangian function and introducing a Lagrangian multiplier $\alpha_i \geq 0$. The Lagrangian function of Equation (4) is:

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \tfrac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{m} \alpha_i\left(1 - y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right)\right)$$
$$i = 1, 2, \ldots, m \tag{4}$$

Its corresponding pairwise problem is:

$$\max_{\boldsymbol{\alpha}}\min_{\mathbf{w},b}\frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{m} \alpha_i\left(1 - y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right)\right) \text{s.t. } \alpha_i \geqslant 0,\ i = 1, 2, \ldots, m \tag{5}$$

The optimization of $(\mathbf{w}, b)$ in the inner layer of Equation (5) is an unconstrained optimization problem, and then the bias derivative is made equal to zero.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i$$
$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^{m} \alpha_i y_i = 0$$
$$i = 1, 2, \ldots, m \tag{6}$$

Bringing Equation (7) into Equation (6) gives

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j$$
$$\sum_{i=1}^{m} \alpha_i y_i = 0$$
$$\text{s.t. } \alpha_i \geqslant 0,\ i = 1, 2, \ldots, m;\ j = 1, 2, \cdots, n \tag{7}$$

According to the Kuhn–Tucker theorem, the optimal solution satisfies $\alpha_i\left(1 - y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right)\right) = 0$. Thus, $\mathbf{w}$ can be found according to $\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i$. For any support vector $(\mathbf{x}_i, y_s)$, there is $y_i\left(\mathbf{w}^\top \mathbf{x}_s + b\right) = 1$. Finally, the classification decision function of the linear SVM can be obtained in Equation (8).

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i \in SV} \alpha_i y_i \mathbf{x}_i^\top \mathbf{x} + b\right)$$
$$i = 1, 2, \ldots m \tag{8}$$

Linear divisibility is an ideal situation. In a real situation, the sample data are mostly nonlinearly divisible. When dealing with nonlinear problems, they are generally transformed into linear problems and handled by the already constructed linear support vector machine. The main idea is to map the original sample data from a two-dimensional plane of nonlinear classifiable points into a three-dimensional space so that a hyperplane can be constructed to perform linear partitioning of the sample points. Thus, this optimization problem is shown in Equation (9).

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \quad \tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m} \xi_i$$
$$\text{s.t.} \quad y_i\left(\mathbf{w}^\top \phi(\mathbf{x}_i) + b\right) \geq 1 - \xi_i,\ i = 1, 2, \ldots, m$$
$$\xi_i \geq 0,\ i = 1, 2, \ldots m \tag{9}$$

where ξ_i denotes the relaxation variable at the ith training point. $C > 0$ is the regularization parameter, and ϕ represents the mapping relationship.

Mapping the sample data to a high-dimensional feature space and then classifying it with a linear support vector machine will make the whole model more complicated. Therefore, by introducing a suitable kernel function, an optimal classification hyperplane can be

constructed to achieve fast processing of high-dimensional inputs. First, the kernel function $\mathcal{K}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \phi\left(\mathbf{x}_i\right)^\top \phi\left(\mathbf{x}_j\right)$ [30,31], at which time the final decision function introduced into the kernel function is:

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i \in SV} \alpha_i y_i \phi\left(\mathbf{x}_i\right)^\top \mathbf{x} + b\right)$$
$$i = 1, 2, \ldots m$$

(10)

The following kernel functions are currently in common use.

1. The linear kernel function, whose expression is shown in Equation (11)

$$\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{x}_i^\top \mathbf{x}_j + c$$

(11)

where c represents an optional constant.

2. The polynomial kernel function, whose expression is shown in Equation (12).

$$\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\lambda \mathbf{x}_i^\top \mathbf{x}_j + \eta\right)^d$$

(12)

where λ represents the slope, η is a constant term, and d is a power of the polynomial.

3. The radial basis kernel function, which is currently the most dominant kernel function, has the following expressions in Equation (13).

$$\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}\right)$$

(13)

where $\left\|\mathbf{x}_i - \mathbf{x}_j\right\|$ represents the open root sign of the sum of squares of the components after the vector is differenced, and σ is the parameter of the Gaussian kernel.

4. Sigmoid function, whose expression is shown in Equation (14).

$$\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) = \tanh(\mu\left(\mathbf{x}_i^\top \mathbf{x}_j\right) - r)$$

(14)

where μ and r are the parameters of kernel function.

4. Financial Crisis Early Warning Indicator Design

4.1. Multidimensional Efficiency Evaluation Index System Design

Although there are some single-factor input–output indicators such as gearing ratio and return on assets in the current financial crisis early warning index system of Chinese-listed companies, these are difficult to assess in the overall input–output situation of all aspects of Chinese-listed companies. Therefore, this paper proposes to construct a multidimensional efficiency evaluation index system, based on the efficiency value measured by each of its dimensions, as multidimensional efficiency indicators. The index system contains four dimensions, namely, operation dimension, innovation dimension, financial dimension and financing dimension, which are four important dimensions to ensure the normal operation of Chinese-listed companies. In addition, the public sentiment dimension as the fifth dimension can be measured by collecting the information of stakeholders' comments on enterprises on the Internet and measuring the emotional tendency based on the comments, and this dimension can reflect the overall situation of enterprises through the perspective of big data.

First, it reflects the ability of Chinese-listed companies to generate economic benefits from their assets from the operational dimension. From the input–output perspective, it can be described as the optimal state of input costs and output returns in the process of operating activities of Chinese-listed companies. If the operational efficiency of Chinese-listed companies is low, it will make Chinese-listed companies face the situation that the products are sold but the initial input capital cannot be obtained back, which will lead to the financial crisis. Based on this, this paper selects operating costs and total expenses as

input indicators from the cost perspective and selects revenue and total profit as output indicators from the revenue perspective, as shown in Table 1.

Table 1. Operation dimension evaluation index system.

Dimensions	Input Indicators	Output Indicators
Operation dimension	Operating costs Total expense	Revenue Total profit

Second, innovation is an important factor for Chinese-listed companies to be able to expand their business scale and to achieve sustainable and rapid development. The lack of core innovation technology may lead to low production efficiency and high product cost, thus making Chinese-listed companies face financial crises. The innovation dimension is mainly use to study the input and output of innovation factors. In this paper, R&D, the number of R&D personnel, and government subsidies are used as input indicators to evaluate the efficiency of the innovation dimension, and ROE and rate of return on total assets are used as output indicators to evaluate the innovation dimension, as shown in Table 2.

Table 2. Innovation dimension evaluation index system.

Dimensions	Input Indicators	Output Indicators
Innovation dimension	R&D Number of R&D personnel Government subsidy	ROE Rate of return on total assets

Third, in terms of the finance dimension, the financial dimension evaluation index system can fully comprehensively measure and evaluate the financial status and operating results of Chinese-listed companies. The financial efficiency of Chinese-listed companies determines whether they have sufficient funds for debt repayment and reproduction. The financial situation of Chinese-listed companies that is not optimistic will possibly lead to the emergence of a financial crisis. Therefore, this paper selects oper-cash into current debt, debt assets ratio, total asset rate, and current assets turnover as the input indicators for evaluating financial efficiency. ROE, the ratio of profits to cost, and return on assets are used as output indicators for evaluating financial efficiency, as shown in Table 3.

Table 3. Finance dimension evaluation index system.

Dimensions	Input Indicators	Output Indicators
Finance dimension	Oper-cash into current debt Debt assets ratio Total asset rate Current assets turnover	ROE Ratio of profits to cost Return on assets

Fourth, the financing dimension is used to evaluate the efficiency of Chinese-listed companies in mobilizing resources to obtain needed funds in financing activities. For the financing dimension, the lower the value, the better. Since financing activities are dynamic and continuous, the net cash flow from financing activities is chosen to measure the financing cost of Chinese-listed companies at first. Then, the current total liabilities and total assets reflect the current financial leverage of Chinese-listed companies, which indicates the financing stock of Chinese-listed companies and can be regarded as an important indicator reflecting the financing scale of Chinese-listed companies. Operating costs reflect the efficiency of Chinese-listed companies for the use of capital, which largely determines the efficiency of financing allocation. In the case of established income, the lower the cost, the higher the efficiency. Therefore, operating costs are also used as one of the input indicators to evaluate the financing efficiency of Chinese-listed companies. In addition, the

total asset rate is the efficiency of the flow of assets from input to output in the business activities of listed companies in China, which can be used as one of the output indicators. The ROE indicator reflects the level of return on owner's equity, which can be used to measure the efficiency of owned capital of Chinese-listed companies. The value of the indicator shows a positive correlation with the return from investment. The final revenue reflects the operating efficiency of the Chinese-listed companies in terms of integrating funds. Therefore, total asset turnover, ROE and revenue are selected as output indicators for evaluating the financing dimension, as shown in Table 4.

Table 4. Financing dimension evaluation index system.

Dimensions	Input Indicators	Output Indicators
Financing dimension	Cash flow from financing activities Total liabilities Total asset Operating costs	Total asset rate ROE Revenue

Fifth, the public opinion dimension introduces network comments into the evaluation index system of Chinese-listed companies. In this paper, the input index is selected as the number of events announced by Chinese-listed companies, which will have a direct impact on online comments and is closely related to the generation of online comments. The output index is introduced to the shareholder comments of the Eastmoney stock forum and Sina stock forum. The positive sentiment index is calculated from the quantification of the texts of stockholders' comments on the Eastmoney Stock Forum and Sina Stock Forum. The article determines the sentiment attitude of each text through machine learning methods. Then, the positive sentiment index is derived by the tone calculation formula proposed by Xie and Lin [32], and the formula is (number of bullish posts-number of bearish posts)/(number of bullish posts + number of bearish posts). The evaluation index system is shown in Table 5.

Table 5. Public opinion dimension evaluation index system.

Dimensions	Input Indicators	Output Indicators
Public opinion dimension	Number of events announced by listed companies in China	Shareholder comments positive sentiment index

4.2. Principles of Selecting Traditional Financial Indicators and Calculation Caliber

There is no accepted standard for financial indicators used in financial early warning models in relevant studies. Based on the principles of selecting traditional financial indicators and based on relevant literature [33–35], five aspects of traditional financial indicators are selected based on the combination of modern financial management theories [36] and the Interim Measures for the Management of Integrated Performance Evaluation of Central Enterprises issued by the State-owned Assets Supervision and Administration Commission of China [37]. The financial status of Chinese-listed companies also depends mainly on the five aspects of profitability, short-term solvency, long-term solvency, operational efficiency of assets and cash flow quality of Chinese-listed companies. These aspects contain a total of 17 specific indicators, and the specific calculation caliber considers whether to use point-in-time or period values, excludes the influence of episodic values, and remains consistent across different Chinese-listed companies in the calculation of the same indicators. The specific indicators and algorithms are shown in Tables 6–10.

Table 6. Indicators and algorithms of the profitability index.

Indicators	Algorithms
Gross income ratio (%)	(Operating income-Operating costs)/Operating income × 100%
Expense ratio (%)	Selling expenses as a percentage + management expenses as a percentage + financial expenses as a percentage + R&D expenses as a percentage (all calculated as XX expenses/operating income × 100%)
Operating profit ratio (%)	Operating profit/Operating revenue × 100%
Net profit margin on sales (%)	(Net income attributable to shareholders of the parent company + minority interests)/operating income × 100%
ROE (%)	Net income/average net assets × 100%
EPS (Yuan)	Net income (also known as profits or earnings) divided by available shares.

Table 7. Indicators and algorithms of the short-term solvency index.

Indicators	Algorithms
Current ratio(times)	Current assets/current liabilities
Quick ratio (times)	(Current Assets-Inventories-Prepaid Expenses)/Current Liabilities
Super quick ratio (times)	Super quick assets divide by the current liabilities of a business (super quick assets = cash + trading financial asset + notes receivable + accounts receivable + other receivables)
Oper-cash into current debt (%)	Cash flow from operating activities/current liabilities × 100%

Table 8. Indicators and algorithms of the long-term solvency index.

Indicators	Algorithms
Debt assets ratio (%)	Total liabilities/total assets × 100%
Interest cover (times)	EBIT/interest expense (EBIT = net income + interest expense + income tax expense, the numerator "interest expense" is taken from the interest expense of financial expenses in the income statement, the denominator "interest expense" is taken from the interest expense of financial expenses in the income statement + capitalized interest included in the balance sheet fixed assets and other costs)

Table 9. Indicators and algorithms of operational efficiency of assets index.

Indicators	Algorithms
ART rate (frequency)	Operating income/notes and accounts receivable (using operating income as the numerator here would overestimate the turnover rate of accounts receivable, and it is more reasonable to use credit sales, as credit sales data are not available; thus, operating income is used. In addition, there are seasonal factors affecting notes and accounts receivable; thus, it is more reasonable to take the average of multiple time points)
Inventory turning rate (frequency)	Operating costs/inventory (the assessment here is the performance of inventory management, not short-term solvency analysis; thus, operating costs are chosen and inventory is still using the average of the beginning and end of the year)
Total asset rate (frequency)	Operating income/total assets (where total assets are averaged at the beginning of the year-end)

Table 10. Indicators and algorithms of Cash flow quality index.

Indicators	Algorithms
Operation cash per share (Yuan)	Cash flow from operating activities/Total common share capital at the end of the year
Cash rate of sales (times)	Cash received from sales of goods and services/operating income

5. Empirical Analysis

5.1. Design of DEA–SVM Financial Crisis Early Warning Method

As shown in Figure 2, the main steps of the DEA–SVM method-based financial crisis early warning are as follows.
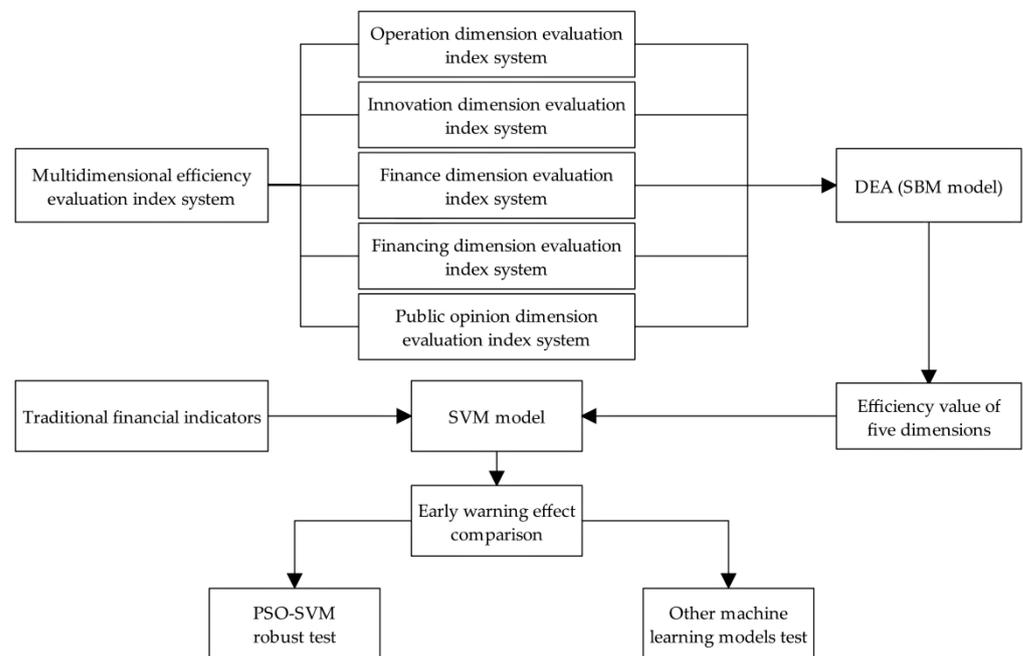
Figure 2. Schematic diagram of the process of the DEA–SVM financial crisis warning method.

In the first step, a multi-dimensional efficiency index system is constructed, which includes the evaluation index system of operation, innovation, financial and financing dimensions. At the same time, considering the perceptive role of investor comments on the financial status of listed companies, this paper introduces the positive sentiment index of stockholders' comments of the Easteconomy stock forum and Sina stock forum to complete the construction of the evaluation index system of the public opinion dimension.

In the second step, the input and output indexes of the above five dimensions are measured using the SBM model of DEA method in Section 3.1, and the efficiency values of the five dimensions are calculated.

In the third step, the efficiency values calculated by the DEA method are introduced into the SVM model as part of the predictors.

The fourth step is to use the nine selected traditional financial indicators and the efficiency value indicators measured by DEA as one group and those with only nine traditional financial indicators as another group. Then, the training and test sets are distinguished for 156 samples of the Chinese-listed companies, and the training set is modeled and solved using the SVM model to determine the optimal hyperplane equation and to derive the classifier of the support vector machine.

In the fifth step, the training SVM classifier is used to calculate the classification of the test set, and the prediction results are obtained.

In the sixth step, a robustness test is performed using the PSO–SVM model to further validate the early warning results of the model.

In the seventh step, six mainstream machine learning methods besides the SVM model are selected to verify whether the proposed model in this paper has better prediction results.

5.2. Data Source and Sample

The data are obtained from Eastmoney Choice Financial Data Terminal, Resset Financial Research Database and CSMAR Database. In this paper, special treatment (ST) of the Chinese-listed companies due to abnormal financial status is taken as the identifier of Chinese-listed companies in financial crisis, and the year in which a financial crisis occurs in Chinese-listed companies is defined as T-0 year. The sample is selected from ST stocks corresponding to the number of non-ST stocks. Since the number of ST stocks is limited, this paper constructs a financial crisis warning model without distinguishing industries; thus, no sampling is performed. All ST stocks in the current year are selected, totaling 83 stocks. The *ST stocks are not included because they represent three consecutive years of losses for Chinese-listed companies, and *ST stocks can have more obvious data abnormalities from financial statements in T-3 years. To improve the effectiveness of early warning, only ST stocks, i.e., stocks of Chinese-listed companies with only two consecutive years of losses, are selected here. The non-ST stocks are selected by taking non-ST stocks with the same tertiary industry (industry according to Shenwan industry category) and similar asset size as ST stocks while taking into account the similarity of the main business scope as much as possible. The ST and the corresponding non-ST samples belong to the same three-tier industry segment, and both have similar asset size and similar scope of main business as much as possible. Some ST stocks do not have a corresponding non-ST stock in the tertiary industry segment (consistent with both having similar asset size and similar scope of main business as possible); thus, non-ST stocks under the upper-level industry segment are selected up to the primary industry segment. Ultimately, there are five ST stocks that belong to non-ST stocks that do not correspond to a similar size, similar business scope and industry segmentation. In order to maintain one-to-one correspondence of the data and to follow the principle of scientificity, these five ST stocks are not selected when the sample data are collated. The five ST stocks are ST Shuanghuan (000707.SZ), ST Jintai (600385.SH), ST Zhongpu (600084.SH), ST Electric Energy (600877.SH) and ST Hongsheng (600817.SH).

In addition, ST stocks are those that have been given special treatment for two consecutive years of losses; thus, the financial data indicators for T-2 years can already show the financial crisis. To reflect the effect of early warning, the annual indicator data of T-3, i.e., the financial statement data of 2017, are used.

In this paper, based on the above selection principles, 78 ST stocks in Shanghai and Shenzhen were selected, and 78 non-ST stocks were selected according to the constraints, for a total of 156 stocks. Please read Appendix A for details.

5.3. Data Processing

First, the collected data are cleaned, mainly to deal with some cases of lack of relevant values, the presence of singular and extreme values of indicators and non-positive values of indicators.

For the lacked data, the relevant values were filled in according to the notes to the annual report or relevant quarterly reports at first. If also in the absence of such values, the relevant data of listed companies in China for the relevant year are used to fill in. In the case of a lack of relevant values even in the close years, the data of similar Chinese-listed companies are used to fill in the values.

In this paper, the singular and extreme values in the sample data are treated according to the quartile formula [38]. When the sample data processed are an even number, the

formula is shown in Equation (15), where n represents the total number of sample data and Q represents the position of each quantile:

$$\begin{aligned} Q_2 &= (n+1)/2 \\ Q_1 &= (Q_2+1)/2 \\ Q_3 &= (Q_2-1)+Q_1 \end{aligned}$$ (15)

When an odd number of sample data are processed, the formula is shown in Equation (16).

$$\begin{aligned} Q_2 &= (n+1)/2 \\ Q_1 &= (Q_2+1)/2 \\ Q_3 &= (Q_2-1)+Q_1 \end{aligned}$$ (16)

In practice, the values that are in the region of singular and extreme values are adjusted to the values of $Q_1 - 1.5QR$ or the critical value at $Q_1 + 1.5QR$ in this paper.

For the case of non-positive values of decision units, Shen [39] and Kerstens et al. [40] summarized the relevant treatments and generally used the dimensionless approach to map the data to the dimensionless interval of (0,1].

5.4. Efficiency Value Measurement Based on SBM Model

A box plot is drawn based on the efficiency values of the multidimensional evaluation index system measured by the SBM model to compare the efficiency values of each dimension for ST and non-ST Chinese-listed companies, as shown in Figure 3. From the figure, it can be seen that there are some differences in all dimensions, and the above five-dimensional efficiency values will be used as feature values in the next stage of training and prediction of the support vector machine early warning model.
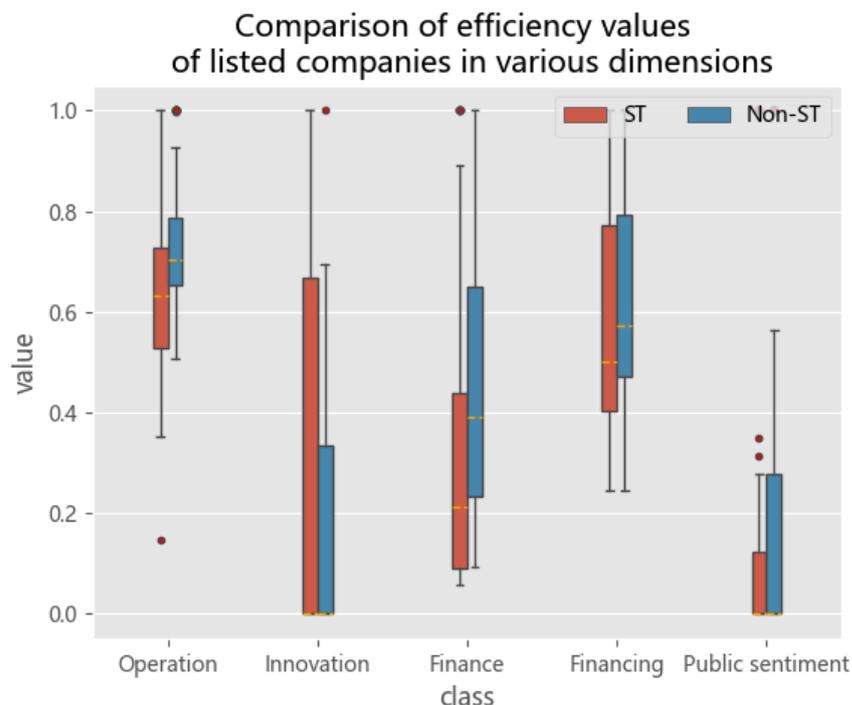
Figure 3. Comparison of efficiency values of listed companies in various dimensions.

5.5. Development and Experimentation of a Support Vector Machine-Based Early Warning Model

This paper focuses on whether the early warning effect is improved after incorporating the efficiency value of multidimensional efficiency indicators on the basis of traditional financial indicators. However, considering that the support vector machine is taking the hyper-segmentation surface, the adoption of the eigenvalues with strong correlation is

beneficial to enhance the early warning effect of the model; thus, the traditional financial indicators are pre-processed. First, the violin plots of 17 traditional financial indicators data of ST Chinese-listed companies and non-ST Chinese-listed companies are constructed, as shown in Figure 4. Then, the violin plots are compared in turn, and the traditional financial indicators with insignificant eigenvalues and traditional financial indicator data with similar eigenvalues are eliminated by the median and density distribution of the violin plots. This also ensures that at least one indicator remains in each of the five areas of profitability, short-term solvency, long-term solvency, asset operating efficiency, and cash flow quality. The excluded indicators are gross profit margin, expense, net profit margin on sales, ROE, quick ratio, super quick ratio, inventory turnover ratio, and total asset ratio. Finally, nine traditional financial indicators are retained for the next step of early warning model construction.
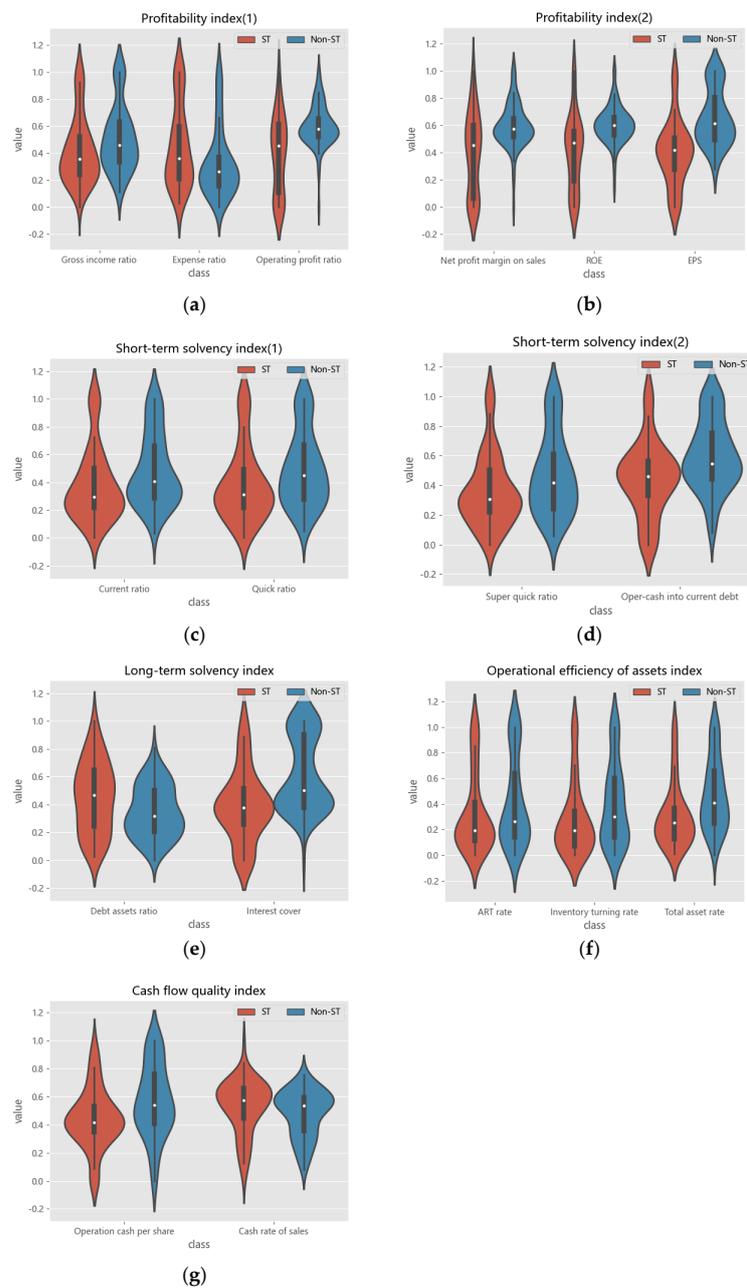
Figure 4. Violin chart based on the company's financial indexes. (**a**) Profitability index (1); (**b**) Profitability index (2); (**c**) Short-term solvency index (1); (**d**) Short-term solvency index (2); (**e**) Long-term solvency index; (**f**) Operational efficiency of assets index; (**g**) Cash flow quality index.

The multidimensional efficiency index data of 156 Chinese-listed companies were organized into Excel with the traditional financial index data. Among them, 78 ST Chinese-listed stocks are identified as 1, and 78 non-ST Chinese-listed stocks are identified as 2. By comparing the correct classification scores of linear kernels function, radial basis kernel function, polynomial kernel function and sigmoid kernel function, the highest score is a radial basis kernel function; thus, the support vector machine based on radial basis kernel function is used. Then, the model algorithm is used to calculate the average warning accuracy, the warning accuracy of crisis-identified Chinese-listed companies and the warning accuracy of non-crisis-identified Chinese-listed companies. To avoid the existence of overfitting, the K-fold cross-validation method with K = 10, equilibrium parameter C = 1, and regularization parameter alpha = 0.001. This experiment is all conducted through Python 3.8.

5.6. Empirical Analysis Results

According to the model test scores, the average prediction accuracy rate of the experimental group containing multidimensional efficiency indicators, the prediction accuracy rate of crisis-labeled Chinese-listed companies, and the prediction accuracy rate of non-crisis-labeled Chinese-listed companies are all higher than those of the experimental group without them. The average forecast accuracy increased by 3.75%. This result verifies that the early warning model containing multidimensional efficiency indicators is more effective in crisis warning, and the specific prediction accuracy results are shown in Table 11.

Table 11. The results of SVM prediction accuracy.

Experimental Group	Average Prediction Accuracy	Forecast Accuracy of Crisis Marker for Chinese-Listed Companies	Forecast Accuracy of non-Crisis Marker for Chinese-Listed Companies
Contains multi-dimensional efficiency metrics	75.09%	85.89%	64.29%
Without multidimensional efficiency indicators	71.34%	83.39%	59.29%

Based on the above experimental results, this paper uses the particle swarm optimization algorithm (PSO) to optimize the penalty parameter c and the parameter g of the kernel function in the SVM. To ensure the robustness of the experimental results, the average of the warning accuracy of the test set is sought after 10 tests. The results show that the average prediction accuracy of the experimental set containing multidimensional efficiency indicators, the prediction accuracy of crisis-labeled Chinese-listed companies, and the prediction accuracy of non-crisis-labeled Chinese-listed companies are all higher than those of the experimental set without efficiency values, and the average prediction accuracy is improved by 9.81%. The prediction results are shown in Table 12, and the data classification effect of the test set containing multidimensional efficiency indicators is detailed in Figure 5.

Table 12. The results of PSO–SVM prediction accuracy.

Experimental Group	Average Prediction Accuracy	Forecast Accuracy of Crisis Marker for Chinese-Listed Companies	Forecast Accuracy of Non-Crisis Marker for Chinese-Listed Companies
Contains multi-dimensional efficiency metrics	83.83%	89.74%	77.92%
Without multidimensional efficiency indicators	74.02%	87.01%	61.03%
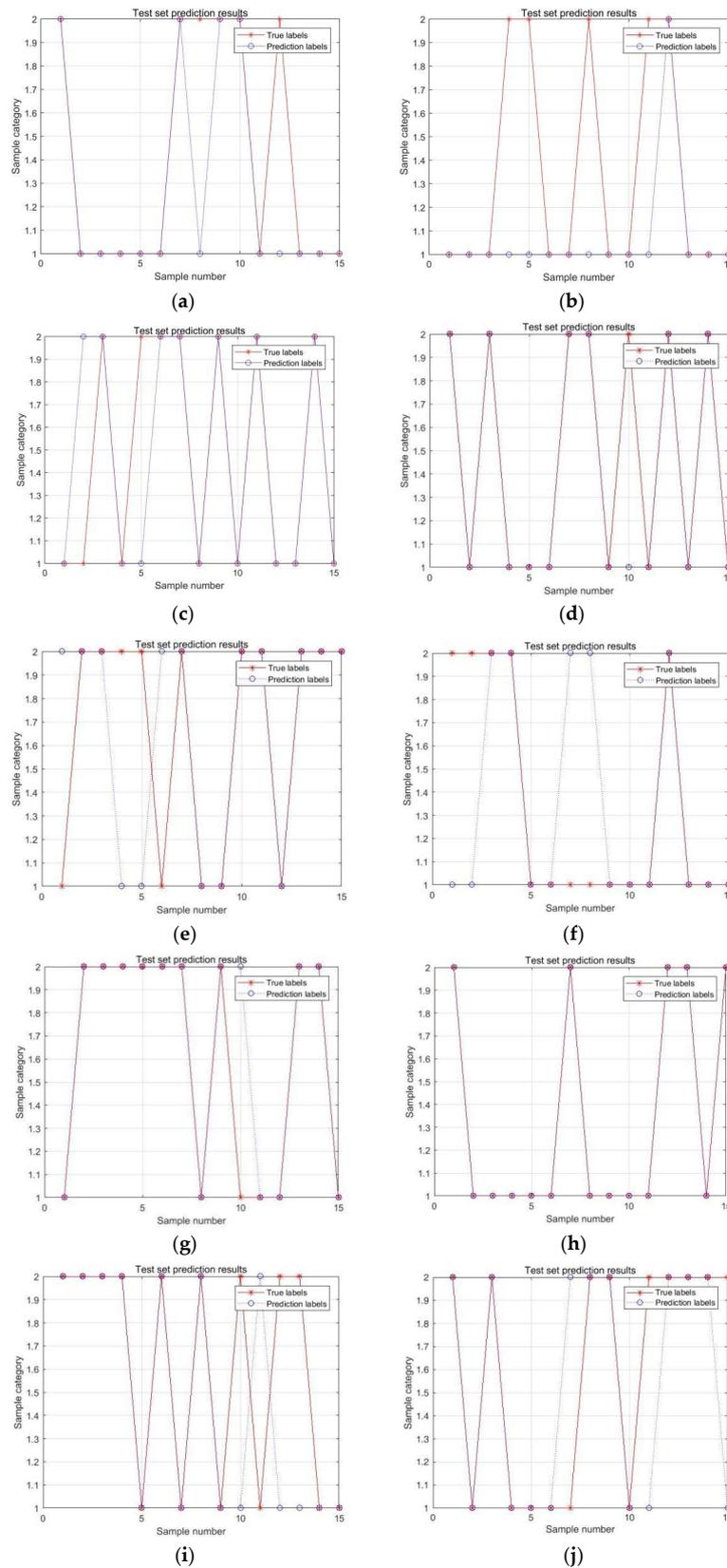
Figure 5. Classification effect of the test set data containing multidimensional efficiency index graphs. (**a**) The first test; (**b**) The second test; (**c**) The third test; (**d**) The fourth test; (**e**) The fifth test; (**f**) The sixth test; (**g**) The seventh test; (**h**) The eighth test; (**i**) The ninth test; (**j**) The tenth test.

The above findings show that the early warning models incorporating multidimensional efficiency indicators have improved early warning accuracy compared with those based on traditional financial indicators. Among them, the SVM based on radial basis kernel function has the best early warning effect, and the results are still robust and show higher early warning accuracy after retesting by PSO–SVM. The model effects are compared to construct a more effective early warning model for financial crises of Chinese-listed companies. This provides methodological support to further prevent and control the financial crisis risk of Chinese-listed companies and has a positive effect on promoting and facilitating the healthy growth of Chinese-listed companies.

5.7. Empirical Results of DEA and Other Machine Learning Models

Based on the aforementioned study, this paper further verifies the forecasting effects of DEA and other machine learning models. Six mainstream machine learning algorithms are selected, and the results show that the forecasting effects are improved by adding multidimensional efficiency indicators. Both forecast accuracy of crisis marker for Chinese-listed companies and forecast accuracy of non-crisis markers for Chinese-listed companies and average prediction accuracy have been improved, which validates the idea proposed in this paper that input feature indicators incorporating multidimensional efficiency indicators have better prediction results, as detailed in Table 13.

Table 13. Results on the prediction accuracy of other machine learning models.

	Experimental Group	Average Prediction Accuracy	Forecast Accuracy of Crisis Marker for Chinese-Listed Companies	Forecast Accuracy of non-Crisis Marker for Chinese-Listed Companies
Logistic Regression	Contains multi-dimensional efficiency metrics	70.12%	68.06%	72.18%
	Without multidimensional efficiency indicators	66.20%	64.77%	67.63%
	Predicted improvement after incorporating multidimensional efficiency indicators	3.92%	3.29%	4.55%
GBDT	Contains multi-dimensional efficiency metrics	64.38%	60.34%	68.42%
	Without multidimensional efficiency indicators	61.08%	57.73%	64.42%
	Predicted improvement after incorporating multidimensional efficiency indicators	3.30%	2.61%	4.00%
CatBoost	Contains multi-dimensional efficiency metrics	67.88%	64.41%	71.35%
	Without multidimensional efficiency indicators	63.92%	56.66%	71.17%
	Predicted improvement after incorporating multidimensional efficiency indicators	3.96%	7.75%	0.18%
AdaBoost	Contains multi-dimensional efficiency metrics	68.97%	63.20%	74.74%
	Without multidimensional efficiency indicators	65.26%	57.02%	73.49%
	Predicted improvement after incorporating multidimensional efficiency indicators	3.71%	6.18%	1.25%
Random Forest	Contains multi-dimensional efficiency metrics	68.60%	64.27%	72.92%
	Without multidimensional efficiency indicators	65.13%	60.41%	69.85%
	Predicted improvement after incorporating multidimensional efficiency indicators	3.46%	3.86%	3.07%
Bagging	Contains multi-dimensional efficiency metrics	66.85%	60.77%	72.92%
	Without multidimensional efficiency indicators	65.72%	59.34%	72.10%
	Predicted improvement after incorporating multidimensional efficiency indicators	1.13%	1.43%	0.82%

Meanwhile, comparing with the prediction effect of multiple models, the forecast accuracy of crisis marker for Chinese-listed companies and the average prediction accuracy of the DEA–SVM model are the best among all models, which can effectively prove the

superiority of the proposed model in this paper, and it can better identify the enterprise financial crisis. The specific effect comparison is shown in Figure 6.
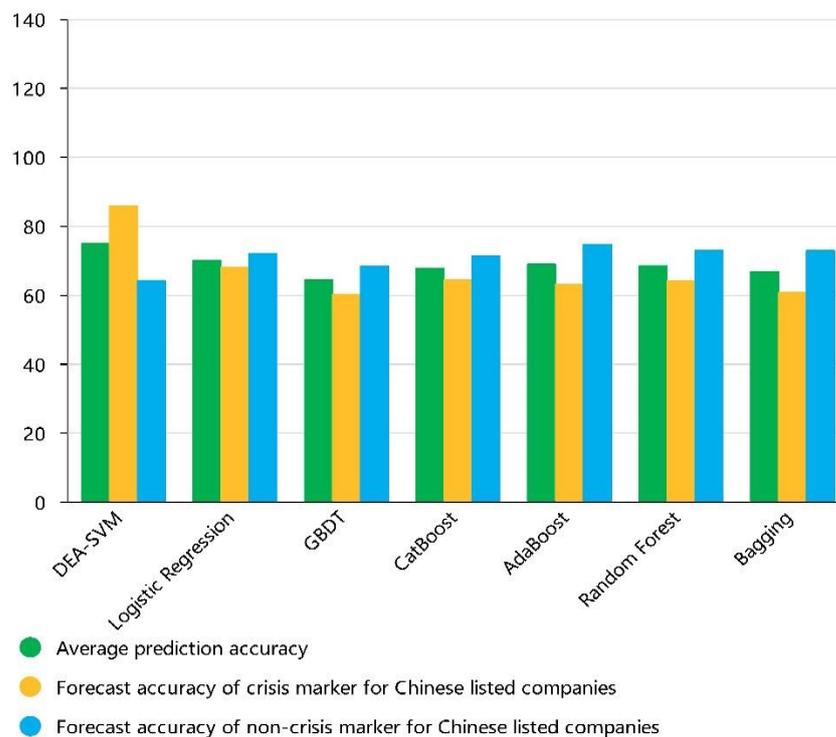
Figure 6. Comparison of early warning effects between six machine learning models and DEA–SVM model.

6. Conclusions

This paper constructs a multidimensional efficiency evaluation index system based on five dimensions: operation, innovation, finance, financing, and public opinion, and evaluates the efficiency values of each dimension using the SBM model in the data envelopment analysis method. In this paper, the measured efficiency values of the five dimensions are used as multidimensional efficiency indicators and are combined with traditional financial indicators to apply the unique advantages of the DEA method to the study of financial crisis early warning problems. In this paper, the financial crisis early warning research of Chinese-listed companies is conducted by the DEA–SVM early warning model, and the following four main research results are achieved.

1. In this paper, based on the principles of traditional financial indicators selection and based on the relevant literature, a total of 17 traditional financial indicators in four aspects of profitability, solvency, asset operating efficiency, and cash flow quality are screened based on the combination of modern financial management theory and the Interim Measures for Integrated Performance Evaluation Management of Central Enterprises (2016) issued by the State-owned Assets Supervision and Administration Commission of China. This is a useful exploration of indicator selection and expands the related research.

2. This paper makes a more comprehensive evaluation of the operation condition of listed companies through the DEA method, introduces the efficiency indicators obtained from the evaluation into the index system of early warning model, and proposes a prediction index system containing multidimensional efficiency indicators evaluated by DEA and traditional financial indicators.

3. This paper considers the construction of a multidimensional efficiency evaluation index system from the input–output perspective and uses big data elements in the construction of the evaluation index system. Then, this paper introduces the public opinion dimension into the index system of the financial crisis early warning model by introducing the

stockholders' comments of the Easteconomy stock forum and Sina stock forum and by measuring the positive sentiment index based on the stockholders' comments.

4. This article constructs a DEA–SVM model together with the SVM model based on the advantages and characteristics of the DEA method and conducts a financial crisis early warning study on 156 listed companies in China. This article proves the effectiveness of the method through empirical research, improves the early warning accuracy of the model, and further optimizes the model using an intelligent optimization algorithm.

However, for the listed companies, falling into financial crisis is mostly a gradual process that does not occur overnight. According to the research of this paper, it is possible to identify and predict the financial crisis. Early warning through the establishment of a financial crisis early warning model can enable stakeholders to detect the signal of deteriorating financial situation of listed companies at an early stage and to take corresponding measures, which has the significance from micro and macro perspectives, mainly reflected in the following aspects.

1. It can help to facilitate the management of listed companies to locate risks and formulate policies.

Through financial crisis, the early warning can make the management of listed companies recognize the problems in the operation of enterprises and take corresponding measures to stop the further deterioration of the financial situation of listed companies, which will also help to improve the risk early warning ability of listed companies.

2. It can help to facilitate creditors and other investors to make correct decisions.

Financial crisis early warning models can objectively forecast the financial situation of enterprises, and such forecast results can provide a basis for investment decisions of creditors and other investors and reduce investment risks.

3. It can help the government understand the overall risk profile of listed companies.

Through financial crisis early warning, the government can have an overall grasp of the risk situation of listed companies, which can better supervise listed companies and help the government make monetary policy or fiscal policy according to the actual situation.

4. It can help to assist other information users in making judgments.

Auditors can use the financial crisis early warning model to assess the financial situation of the audited listed company, as to formulate reasonable audit procedures, which helps auditors accurately judge the listed company's ability to continue as an ongoing concern.

5. It can promote the healthy development of China's capital market.

The operation of listed companies is related to the healthy development of the capital market. If abnormalities can be detected in time through financial crisis warning and the problems behind the abnormalities can be found, strategies and policies will be adjusted in time. This will be conducive to the normal operation of listed companies and thus promote the healthy development of the capital market.

6. It helps to optimize the allocation of social resources.

Predicting the development prospects of industries and enterprises through financial crisis early warning models will help realize the rational allocation of people, money and materials, which helps to improve the efficiency of capital utilization of the whole society.

This paper has achieved desirable results in the empirical study, but there are still some limitations of this paper that need to be further researched and explored in the future.

1. The study is based on the Chinese stock market, and whether it can be extended to the global capital market to form a more universal early warning model is a direction for further research in this paper.

2. This study used stock bar commentary texts to the model. This will be a direction for future exploration as to whether other texts, e.g., industry research reports, news media reports, etc., also have a role in identifying early warnings of financial crises.

3. This paper uses the positive sentiment index in text analysis, but it is to be further explored whether word frequency, readability, and tense can improve the warning accuracy of the warning model.

Appendix A

Table A1. The stock codes of 156 Chinese-listed companies.

Stock Codes			
002766.SZ	002333.SZ	002232.SZ	002082.SZ
002005.SZ	600238.SH	002745.SZ	002646.SZ
600702.SH	300325.SZ	000799.SZ	000973.SZ
002420.SZ	601258.SH	300475.SZ	600297.SH
002306.SZ	002592.SZ	000721.SZ	603239.SH
002336.SZ	002684.SZ	605188.SH	603730.SH
000572.SZ	600698.SH	601127.SH	601689.SH
600423.SH	002700.SZ	002556.SZ	002911.SZ
002427.SZ	002770.SZ	002998.SZ	002946.SZ
002650.SZ	300367.SZ	600305.SH	300352.SZ
601113.SH	600654.SH	002419.SZ	600602.SH
600696.SH	000868.SZ	600830.SH	600686.SH
002021.SZ	000504.SZ	600843.SH	002693.SZ
603188.SH	600080.SH	603980.SH	000661.SZ
300029.SZ	600275.SH	603628.SH	600257.SH
000972.SZ	600399.SH	600251.SH	600507.SH
002102.SZ	002089.SZ	300267.SZ	600105.SH
600421.SH	600119.SH	300838.SZ	300240.SZ
002290.SZ	002289.SZ	603519.SH	688299.SH
600408.SH	600666.SH	603113.SH	002387.SZ
600725.SH	002692.SZ	600792.SH	002533.SZ
600265.SH	002473.SZ	002679.SZ	002959.SZ
000422.SZ	600721.SH	002386.SZ	688222.SH
600091.SH	600767.SH	000510.SZ	600763.SH
600319.SH	002872.SZ	000635.SZ	002462.SZ
600301.SH	600652.SH	600753.SH	002425.SZ
600608.SH	300038.SZ	000151.SZ	000607.SZ
600870.SH	002175.SZ	600822.SH	601599.SH
600157.SH	002629.SZ	601088.SH	300164.SZ
300064.SZ	002200.SZ	600172.SH	603316.SH
002569.SZ	600610.SH	300840.SZ	002374.SZ
002656.SZ	600112.SH	002029.SZ	300423.SZ
600365.SH	600518.SH	000869.SZ	600252.SH
603779.SH	601399.SH	600543.SH	600320.SH
000611.SZ	600193.SH	600156.SH	002713.SZ
000737.SZ	600209.SH	002709.SZ	603030.SH
300446.SZ	000409.SZ	300847.SZ	600784.SH
600228.SH	600149.SH	002909.SZ	000632.SZ
600539.SH	600234.SH	603408.SH	600051.SH

References

1. Wu, X.Z. Financial Crisis Early Warning Research: Problems and Framework Reconstruction. *Account. Res.* **2011**, *2*, 59–65.
2. Zhang, Y.; Jiang, T.K. Hayek and Schumpeter—Introduction, comparison and evaluation of the two Austrian schools of economic cycle theory. *Economist* **2018**, *7*, 96–104.
3. Ross, S.A. The Economic Theory of Agency: ThePrincipal's Problem. *Am. Econ. Rev.* **1973**, *63*, 134–139.
4. Wang, L.; Xing, H.B. A Review of Foreign Corporate Crisis Management Research. *Yunnan Financ. Econ. Univ. J. Econ. Manag.* **2011**, *26*, 49–53.
5. Charnes, A.; Cooper, W.; Rhodes, E. Measuring the efficiency of decision-making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
6. Zhang, M.; Li, L.; Cheng, Z. Research on carbon emission efficiency in the Chinese construction industry based on a three-stage DEA-Tobit model. *Environ. Sci. Pollut. Res.* **2021**, *28*, 51120–51136. [CrossRef]
7. Zhong, K.; Li, C.; Wang, Q. Evaluation of Bank Innovation Efficiency with Data Envelopment Analysis: From the Perspective of Uncovering the Black Box between Input and Output. *Mathematics* **2021**, *9*, 3318. [CrossRef]
8. Zhong, K.Y.; Wang, Y.F.; Pei, J.M.; Tang, S.M.; Han, Z.L. Super efficiency SBM-DEA and neural network for performance evaluation. *Inf. Process. Manag.* **2021**, *58*, 102728. [CrossRef]
9. Vapnik, V.N. The nature of statistical learning theory. *Nat. Stat. Learn. Theory* **1995**, *20*, 273–297.
10. Vapnik, V.N. Statistical learning theory. *Encycl. Sci. Learn.* **1998**, *41*, 3185.
11. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [CrossRef]
12. Fitzpatrick, F. A comparison of ratios of successful industrial enterprises with those of failed firm. *Certif. Public Account.* **1932**, *10*, 598–605.
13. Altman, E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *J. Financ.* **1968**, *23*, 589–609. [CrossRef]
14. Altman, E.I.; Haldeman, R.G.; Narayanan, P. Zetatm analysis a new model to identify bankruptcy risk of corporations. *J. Bank. Financ.* **1977**, *1*, 29–54. [CrossRef]
15. Ohlson, J.A. Financial ratios and the probabilistic prediction of bankruptcy. *J. Account. Res.* **1980**, *3*, 109–131. [CrossRef]
16. Fan, A.; Palaniswami, M. Support vector machines symposium proceedings. In Proceedings of the IEEE Computer Society, Washington, DC, USA, 4–6 November 2002.
17. Min, J.H.; Lee, Y.C. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Syst. Appl.* **2005**, *28*, 603–614. [CrossRef]
18. Yan, J.J.; Liu, J.H. Support vector machine early warning model for financial crisis of listed companies. *Stat. Decis.* **2006**, *12*, 158–159.
19. Song, J. An Empirical Study on Early Warning of Financial Crisis of Listed Companies Based on Support Vector Machine. *Mark. Mod.* **2007**, *25*, 391–392.
20. Boyacioglua, M.A.; Kara, Y.; Baykan, K. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Syst. Appl.* **2009**, *36*, 3355–3366. [CrossRef]
21. Beaver, W.H. Financial rations as predictors of failures. *J. Account. Res.* **1966**, *4*, 71–111. [CrossRef]
22. Wu, S.N.; Lu, X.Y. A study on the prediction model of financial distress of listed companies in China. *Econ. Res. J.* **2001**, *6*, 46–55+96.
23. Wang, J.T.; Sun, T.T. Research on financial early warning of energy-consuming enterprises based on genetic BP neural network. *J. Harbin Univ. Commer.* **2012**, *5*, 110–116.
24. Barboza, F.; Kimura, H.; Altman, E. Machine learning models and bankruptcy prediction. *Expert Syst. Appl.* **2017**, *83*, 405–417. [CrossRef]
25. Xiao, Y.; Xiong, K.L.; Zhang, X. Research on early warning model of enterprise financial risk based on TEI@I methodology. *Manag. Rev.* **2020**, *32*, 226–235.
26. Ran, M.S.; Zhou, S.; Huang, L.Y. Application of SVM model based on DEA indicators in financial warning. *Stat. Decis.* **2009**, *20*, 143–145.
27. Xu, X.; Wang, Y. Financial failure prediction using efficiency as a predictor. *Expert Syst. Appl.* **2009**, *36*, 366–373. [CrossRef]
28. Li, Z.; Crook, J.N.; Andreeva, G. Dynamic prediction of financial distress using malmquist DEA. *Soc. Sci. Electron. Publ.* **2017**, *80*, 94–106. [CrossRef]
29. Tone, K. A slacks-based measure of efficiency in data envelopment analysis. *Eur. J. Oper. Res.* **2001**, *130*, 498–509. [CrossRef]
30. Cherkassky, V.; Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* **2004**, *17*, 113–126. [CrossRef]
31. Sun, H.W.; Yu, Z.X. Extension of Mercer's theorem. *J. Univ. Jinan* **2004**, *3*, 280–282.
32. Xie, D.R.; Lin, L. Can the tone of management predict the future performance of the company?—A textual analysis based on annual earnings presentations of listed companies in China. *Account. Res.* **2015**, *2*, 20–27.
33. Billio, M.; Casarin, R.; Costola, M.; Pasqualini, A. An entropy-based early warning indicator for systemic risk. *J. Int. Financ. Mark. Inst. Money* **2016**, *45*, 42–59. [CrossRef]

34. Ma, Y.; Chen, Y. Financial imbalance index as a new early warning indicator: Methods and applications in the Chinese economy. *China World Econ.* **2014**, *22*, 64–86. [CrossRef]

35. Zhang, Y.L.; You, X.L.; Qiang, W. LASSO-based early warning of corporate financial crisis and selection of key indicators. *J. Henan Norm. Univ.* **2016**, *44*, 160–165.

36. Jensen, M.C.; Meckling, W.H. Theory of the firm: Managerial behavior, agency costs and ownership structure. *J. Financ. Econ.* **1976**, *3*, 305–360. [CrossRef]

37. China State-owned Assets Supervision and Administration Commission. Notice on the Issuance of the Implementation Rules for the Comprehensive Performance Evaluation of Central Enterprises. *Announc. State Own. Assets Superv. Adm. Comm. State Counc. China* **2006**, *9*, 5–16.

38. Zhao, X.J.; Zhao, X.F.; Shao, L.J. Box plotting based on a new formula for calculating quartiles. *J. Tangshan Norm. Univ.* **2020**, *42*, 13–15.

39. Shen, J.J.; Long, W. Treatment of negative output in DEA models-based on the use of software DEAP. In Proceedings of the Annual China Management Conference, Hefei, China, 6–8 December 2015.

40. Kerstens, K.; Van, D. Negative data in DEA: A simple proportional distance function approach. *J. Oper. Res. Soc.* **2011**, *62*, 1413–1419. [CrossRef]