International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1608-1620

https://doi.org/10.70517/ijhsa463124

Construction of Enterprise Tax Risk Prediction and Prevention and Control System Based on Logistic Regression in Tax Big Data Environment

Xinjuan Wang[1,*]

[1] School of Finance and Economics, Guangdong University of Science and Technology, Dongguan, Guangdong, 523000, China
Corresponding authors: (e-mail: wangxinjuan1989@163.com).

Abstract This paper takes the enterprise tax risk early warning as the core objective, combines with the crisis management theory, and constructs the multidimensional tax risk early warning indicator system. Using multiple logistic regression method to analyze the significant indicators related to tax risk. By empirically analyzing the enterprise financial data, the risk prediction model is established and the model prediction effect is verified. The regression analysis shows that the asset-liability ratio, the proportion of financial personnel, and the return on net assets are significantly related to tax risk at the 95% confidence level. Capital intensity is significantly associated with tax risk at 90% confidence level. The established multivariate logistic regression model has a steeper gain curve and higher prediction accuracy. This study proposes strategies covering capital structure optimization application and other strategies through quantitative analysis.

Index Terms logistic regression, tax risk early warning, indicator system, gain analysis, prediction accuracy

I. Introduction

In the new era, tax is an essential policy on the road of national development, it serves the people, but also from the people, and enterprises are a major source of tax [1]. Enterprise development is often to their own maximum profit as the goal, while the government tax is to regulate income fairness for the purpose and for the mandatory implementation of the enterprise, which inevitably contradicts the development of the enterprise, some enterprises are wandering in this contradiction, which triggered the tax risk of the enterprise [2]-[4]. The existence of tax risk is a major stumbling block on the road of enterprise development, and if enterprises are unable to overcome this difficulty, it will seriously impede their own development, and even be punished by the government, which is not conducive to the long-term development of enterprises [5], [6]. Therefore, enterprises should recognize the important position of tax work in their own development, actively fulfill their obligations, and do a good job of financial management at all levels and stages within the enterprise, effectively prevent tax risks, and promote their long-term good development [7], [8].

With the rise of big data information technology, more and more enterprises' main products are transformed from physical products to intangible assets, which further increases the degree of information asymmetry between tax authorities and taxpayers, and brings great challenges to enterprise tax risk management [9]-[11]. However, with the help of data analysis, data mining, data screening and other technical means, enterprises can realize accurate identification, prediction and assessment of tax risks, so as to respond more effectively to the increasingly complex tax compliance environment [12]-[14]. In the future, enterprises should make full use of the big data platform and intelligent technology in the process of digital transformation to promote the continuous optimization of tax risk management and escort the sound development of enterprises [15]-[17].

In the study of enterprise tax risk, literature [18] emphasizes that tax risk is an important part of the overall risk management system of the enterprise, and a systematic tax risk management program will try to have an important impact on the stable operation of the enterprise, the economic development and the formulation of strategic objectives. Literature [19] discusses the importance of a sound enterprise tax risk management system under the current tax system, and the reduction of enterprise tax risk through internal control and other means not only reduces the enterprise tax cost, but also helps to make the enterprise maintain a stable business model.

At the same time, by assessing, analyzing and identifying the tax risks generated in the process of enterprise operation, and identifying the matters that are prone to risks to form an early warning mechanism, the impact of tax risks on enterprises can be effectively reduced. Literature [20] utilizes the hierarchical analysis method and entropy power method to establish the index system affecting the enterprise tax risk management, and provides decision support for the enterprise's risk control strategy by evaluating the enterprise tax risk situation in the context of big

data. Literature [21] constructed an enterprise tax risk analysis and prediction model based on artificial neural network multilayer perceptron, and compared with the rule-based risk assessment method, the proposed model is more adaptable to the unbalanced data samples and has a higher prediction accuracy. Literature [22] introduced the MALDIVE model into the corporate tax risk assessment process, through data mining and visualization analysis of taxpayer relationship network, so as to accurately identify the set of taxpayers with the risk of tax evasion and tax avoidance behavior. Literature [23] combined feature change technology and machine learning algorithms to design a tax default prediction system for tax management, and by focusing on the data transformation and nonlinear relationship between corporate financial warning indicators and tax default behavior, it can significantly improve the effective detection of corporate financial default information. Literature [24] shows that under the current complex and huge tax system, it is more difficult to carry out personalized prediction analysis for different taxes, and proposes the use of multi-source big data fusion algorithm to integrate analysis and prediction of enterprise financial and tax data, which provides valuable suggestions for enterprise financial risk management.

From the above research on tax risk and tax risk early warning research, it can be seen that scholars at home and abroad pay more attention to the risk of enterprise tax, and since Chinese small and medium-sized enterprises (SMEs) occupy a central position in the country's economic structure, the tax risk of SMEs needs to be paid attention to. Among them, logistic regression model, as a commonly used model for evaluating various types of risks, has practical value in the construction of enterprise tax risk prediction and prevention and control system, which can be used to deal with the linear impact of various factors on tax risk and has important research value.

The increasingly strict tax regulation and the increase in the complexity of enterprise operation make the tax risk affect the sustainable development of enterprises. This paper combines the logistic regression algorithm with enterprise financial indicators to construct a quantifiable tax risk early warning indicator system. Taking the enterprise financial data as the research object, the correlation between specific indicators and tax risk is explored, and the prediction model is constructed. Judge the prediction effect of the constructed tax risk prediction model through gain analysis. Propose targeted tax management strategies to reduce the risk of corporate tax crisis.

II. Indicator system construction and research methodology analysis

This chapter analyzes enterprise tax risk related indicators and constructs enterprise tax risk early warning indicator system. It describes the specific process of the logistic regression method chosen in this paper, and studies the specific relationship between relevant indicators and tax risk.

II. A. Construction of Enterprise Tax Risk Early Warning Indicator System

(1) Solvency indicators

According to the theory of crisis management, solvency is mainly the ability of the enterprise to repay the principal and interest when there are various types of debts. When the enterprise has strong solvency, it means that the enterprise has the ability to bear the pressure of debt, and can still repay the debt normally in the case of tighter financial situation, to avoid debt default and other situations. In this paper, asset-liability ratio and speed ratio are used to measure the solvency of enterprises. The gearing ratio considers the relationship between an enterprise's liabilities and assets. Generally speaking, the value of the enterprise's assets should be higher than the total liabilities, and once the enterprise is at risk, it can pay off its debts by selling its assets, and equation (1) is the formula for its calculation.

$$\text{Asset-liability ratio} = \frac{\text{Total liabilities}}{\text{Total assets}} \tag{1}$$

Generally speaking, the lower the asset-liability ratio of an enterprise, it means that the value of the enterprise's assets will be higher than its liabilities, and when the enterprise pays off its debts, it can utilize its assets to offset its debts, and therefore has a better solvency in the long run. Quick ratio can be used to measure the situation of the enterprise's current assets, reflecting the enterprise's liquidity. The higher the value of the quick ratio, indicating that the enterprise's cash liquidity is better, when the short-term debt expires, the enterprise can quickly obtain a certain amount of cash to repay the debt, to prevent the enterprise from tax risk, formula (2) is its calculation formula.

$$\text{Quick ratio} = \frac{\text{Quick assets}}{\text{Current liabilities}} \tag{2}$$

(2) Profitability indicators

Taxes are mainly derived from the profitability of enterprises. Among them, the return on net assets can reflect the ability of enterprises to utilize the profit of asset activities, when the higher the indicator, it means that the

enterprise can make money the higher the level of net profit, the enterprise has a better profitability, formula (3) is the formula for its calculation.

$$\text{Return on net assets} = \frac{\text{Net profit}}{\text{Net assets}} \tag{3}$$

The percentage of financial personnel represents the cost expenditure efficiency of the enterprise. Generally speaking, in the process of generating profits, enterprises need to invest in various costs and expenses, so when the unit cost of expenses is lower, it indicates that the enterprise's profitability is stronger, and equation (4) is the formula for its calculation.

$$\text{Proportion of financial personnel} = \frac{\text{Employee salary expenditure}}{\text{Total cost and expense}} \tag{4}$$

(3) Operating capacity

Capital intensity is a core indicator of the efficiency of capital input and output of an enterprise, calculated by the ratio of fixed capital investment to total output, reflecting the degree of capital dependence of an enterprise. The lower the degree of capital dependence of an enterprise, the more efficiently it can operate its assets, and at this time the possibility of tax risk is relatively lower, formula (5) is the formula.

$$\text{Capital intensity} = \frac{\text{Fixed capital investment}}{\text{Total output}} \tag{5}$$

The accounts receivable turnover ratio represents the accounts receivable of the firm. From the analysis of the research background, it can be seen that a number of enterprises have tax risk problems because of accounts receivable problems. The higher the turnover rate of accounts receivable, the stricter the control of tax risk in the operation process of the enterprise, and equation (6) is the formula for its calculation.

$$\text{Accounts receivable turnover ratio} = \frac{\text{Net operating income}}{\text{Average balance of accounts receivable}} \tag{6}$$

Current asset turnover reflects the operation of enterprises on various types of current assets, the higher the efficiency of the turnover of current assets, the enterprise has better liquidity, which is conducive to reducing the tax risks faced by the enterprise, formula (7) is the formula for its calculation.

$$\text{Current asset turnover ratio} = \frac{\text{Net operating income}}{\text{Average current assets}} \tag{7}$$

Inventory intensity reflects the cash that a firm can quickly recoup through the sale of inventory when facing debt maturity. In the case of this indicator is good, the enterprise has better cash flow, which is conducive to avoid the occurrence of short-term tax risk, equation (8) is its calculation formula.

$$\text{Inventory intensity} = \frac{\text{Net operating cash flow}}{\text{Average total assets}} \tag{8}$$

Inventory turnover rate reacts to the process of enterprise operation, the flow rate of goods in stock, generally the higher the flow rate, the stronger the enterprise's ability to cope with the tax risk, formula (9) is the formula for its calculation.

$$\text{Inventory turnover rate} = \frac{\text{Operating cost}}{\text{Average inventory balance}} \tag{9}$$

(4) Growth capacity

The situation of cost of goods sold ratio reflects whether the company is in the stage of growth in its business operations, the lower the cost of goods sold ratio, it means that the company's earnings grow faster, equation (10) is its calculation formula.

$$\text{Ratio of sales cost} = \frac{\text{Sales cost amount}}{\text{Sales revenue amount}} \tag{10}$$

Net assets per share experiences the ratio of shareholders' equity to the total number of shares in a company. This indicator reflects the present value of assets owned per share of stock. The higher the net worth per share, the more value of assets per share owned by shareholders; the lower the net worth per share, the less value of assets per share owned by shareholders. The higher the net assets per share, the more room for corporate growth. Equation (11) is the formula for its calculation.

$$\text{Net assets per share} = \frac{\text{Total shareholders' equity}}{\text{Total number of shares}} \tag{11}$$

Regional tax incentives reflect the external support received by the enterprise, which is a direct reflection of the growth potential of the enterprise and demonstrates the expansion capacity of the enterprise, and equation (12) is the formula for its calculation.

$$\text{Regional tax preferences} = \frac{\text{Amount of tax reduction and exemption for the current year}}{\text{Total amount of taxes and fees payable}} \tag{12}$$

Based on the above analysis, this paper constructs the index system of enterprise tax risk early warning. Table 1 is the constructed enterprise tax risk early warning indicator system. As can be seen from Table 1, the enterprise tax risk early warning indicator system constructed in this paper contains 5 dimensions and 12 indicators, which comprehensively summarizes the situation of enterprises that may be related to tax risk.

Table 1: Enterprise tax risk early warning index system

Dimensionality	ID	Index	ID
Solvency	A1	Asset-liability ratio	B1
		Quick ratio	B2
Profitability	A2	Return on equity	B3
		Proportion of financial personnel	B4
Operational capacity	A3	Capital intensity	B5
		Accounts receivable turnover rate	B6
		Turnover of current assets	B7
		Inventory density	B8
		Inventory turnover	B9
Growth ability	A4	Cost of sales ratio	B10
		Net asset value per share	B11
		Regional tax preference	B12

II. B.Logistic regression

Logistic regression is a method commonly used in machine learning a linear regression obtained by some degree of mathematical alteration. It is used to predict the chances of a particular event occurring by fitting the data to a logistic function. The initial steps include the construction of a regression model, followed by the use of great likelihood estimation techniques to determine the parameters of the model and calculate the fitted values. After that, a mathematical strategy is applied to make decisions and solve the classification problem. This classification algorithm is widely used in several fields. For example, in corporate tax research, logistic regression can help identify potential risks in corporate tax and predict the probability of their occurrence. Classification problems can be categorized into binary logistic regression and multivariate logistic regression depending on the number of categories.

II. B. 1) Binary logistic regression

In the binary logistic regression model, the input features are first processed by weighted merging and then transformed by a Sigmoid function to produce a probability value between 0 and 1, which indicates the probability that the sample belongs to a specific classification. In the model training phase, a loss function based on great likelihood estimation is established by constructing a prediction function, and the weighting parameters of the model are iteratively updated using gradient descent methods with a view to achieving the goal of minimizing the loss function.

The prediction function is a Sigmoid function in the form shown in equation (13).

$$g(t)=\frac{1}{1+e^{-t}} \tag{13}$$

The logistic regression model is defined as shown in equation (14).

$$\Pr\left(Y=0\middle|X=x\right)=g\left(x\omega^{T}\right)=\frac{1}{1+e^{-x\omega^{T}}} \tag{14}$$

where Y denotes the category, $x\in R^{f}$ denotes the features, and $\omega\in R^{f}$ denotes the weight vector to be learned during model training.

For logistic regression, the smaller the loss function, the better the model. Using the great likelihood estimation method, the loss function can be obtained as shown in equation (15).

$$J(\omega)=-\frac{1}{n}+\alpha\sum_{i=1}^{n}\left[y_i\log g\left(x_i\omega^{T}\right)+\left(1-y_i\right)\log\left(1-g\left(x_i\omega^{T}\right)\right)\right] \tag{15}$$

II. B. 2) Multiple logistic regression

In practice, we often encounter the need to deal with multi-categorization learning tasks, multi-categorization logistic regression can be used as an extension form of binary logistic regression in multi-categorization tasks. In this paper, we use the idea of “decomposition method” to realize the expansion process. Common decomposition methods include “one-to-one” (OvO) and “one-to-many” (OvR) and “many-to-many” (MvM).

Given the dataset $D=\{(x_1,y_1),(x_1,y_1),\ldots,(x_1,y_1)\}$, $y_i\in\{C_1,C_2,\ldots,C_k\}$. A demonstration of the OvO strategy is shown in Fig. 1.The OvO strategy creates multiple binary classification tasks by pairing all categories two by two. In the final stage of integrating the predicted results, the category that has been predicted the most times is selected as the final categorization result.

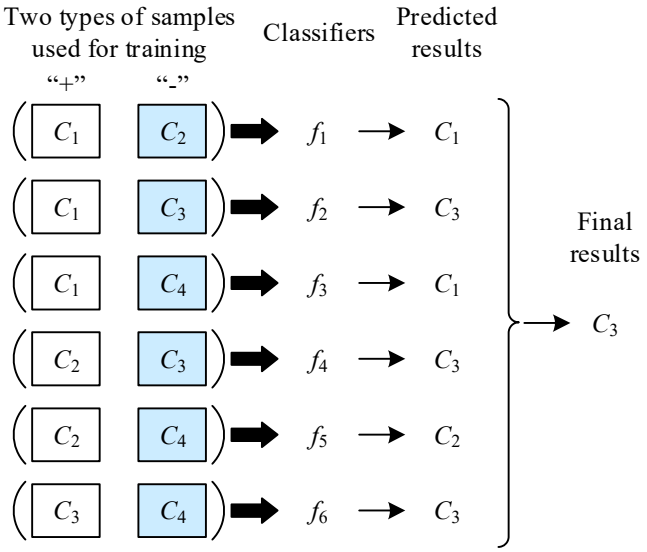

Figure 1: Shows the OvO policy

Figure 2 is a demonstration of the OvR strategy. The OvR strategy is able to produce a relative number of classifiers for new test samples. In the case where the whole test is a positive class, the class is the final split final result, if only one classifier preclassifies the sample labels. If more than one classifier predicts the sample as a positive category, the prediction confidence of each classifier is compared and the category with the highest confidence is selected as the final classification result.

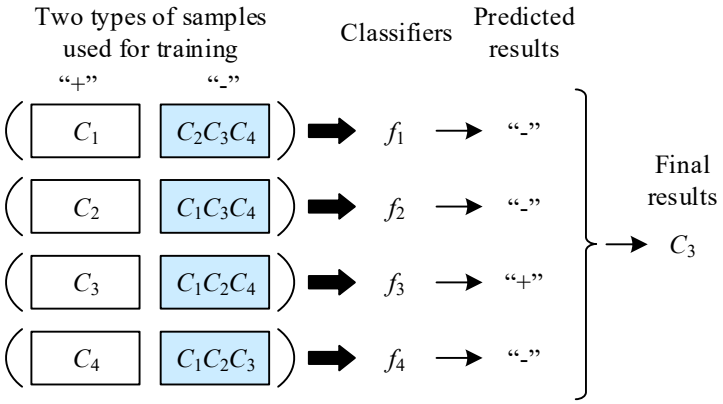

Figure 2: Shows the OvR policy

By comparing Figures 1 and 2, it is found that the OvO strategy tends to have more overhead in storage and testing time than the OvR strategy. In contrast, the OvR strategy has a lower training time overhead when the number of categories is small; therefore, this study chooses to adopt the OvR strategy to extend the application of binary logistic regression to the scenario of multicategorical logistic regression.

The specific steps of the multi-category logistic regression algorithm are as follows: the user initializes the weight matrix ω by providing the dataset matrix X, the label matrix Y and the learning rate α. During the iterative process, for each category, the system will perform a series of steps, including calculating the inner product, performing the Sigmoid function calculation, and updating the weight vectors, and finally obtaining the corresponding weight parameters for each classifier.

III. Practice of corporate tax risk analysis

This chapter takes a specific enterprise as an example, conducts regression analysis on enterprise-related tax index data, mines the indexes affecting tax risk. And the effectiveness of the regression model is evaluated.

III. A. Basic introduction of S Enterprises

S Software Enterprise was founded on April 1, 2005, with a registered capital of 365,599,000 yuan, and a total share capital of 366,555,191 shares as of the end of 2019. S Software Enterprise is the first enterprise in China to be listed on the GEM board based on the "cloud computing" enterprise informationization service, and the first enterprise to be listed on the GEM board in the province of XX, and has been rated as a national key software enterprise from 2011 to 2013. S Software was established as a limited liability company at the beginning, and was changed into a joint stock enterprise on July 30, 2008, and was listed on the Shenzhen Stock Exchange in March 2011.

Enterprise S belongs to the software and information technology service industry, and passed the national "double-soft certification" in 2013, and obtained the qualification of the first batch of high-tech enterprises of City A in 2009 in October 2009, and passed the review in 2016, according to the articles of association of the enterprise, Enterprise S is a national key software enterprise. According to the Articles of Association, Enterprise S is a company that mainly utilizes the SAAS model and independently develops software products and services, mainly for small and medium-sized enterprises in China, and produces software products and services such as enterprise e-mail, e-commerce website development and construction, Internet domain names, customer relationship management software systems and office automation management systems (OA).

At present, the company has 36 software patents, and has successively won the honorary titles of "Five-star Excellent Domain Name Registration Service Agency", "High-tech Enterprise in City A", and "Credit Enterprise with AAA Credit Rating". The enterprise through continuous independent innovation, focusing on market demand, and constantly meet customer needs, increase investment in product research and development, and constantly enhance the strength of research and development, and promote product improvement, in order to promote the improvement of product functions and product iteration process, to improve the competitiveness of the products produced by the enterprise in the market, and constantly integrate the enterprise existing enterprise mailbox, OA, enterprise instant messaging and other mobile office applications, to create a mobile Internet-based enterprise application service platform, to provide customers with efficient, open, innovative, synchronized, easy to manage and cloud office services. Internet-based enterprise application service platform, to provide customers with efficient, open, innovative, synchronized, easy-to-manage cloud office services.

III. B. Financial statements of enterprise S

Financial statements reflect the funds and profits of a business over a certain period of time and can reveal potential tax risk situations. The following are the financial statement data of the enterprise for the four years from 2020-2023. Table 2 lists some abnormal data of the balance sheet of Enterprise S during the four years 2020-2023. It can be seen that from 2020-2023, the enterprise's accounts receivable and inventory decreased substantially, the cash turnover ratio and inventory turnover ratio increased, and goodwill showed an abnormal and substantial decline, and it is known from Enterprise S's annual report in 2022 that a large amount of goodwill impairment loss was accrued in that year, but the reasons for the accrual were not clearly stated in the annual report, and the goodwill value returned to normal in 2023. Taxes payable decreased from year to year, but in combination with the profits of the enterprise, it is known that the income of enterprise S did not fluctuate significantly, which is contrary to the normal change.

Table 2: Company S main balance sheet data (yuan)

Subject	2023	2022	2021	2020
Notes and accounts receivable	30,713,400	25,834,100	40,115,800	52,436,800
Stock	543,800	367,400	68,700	2,503,500
Goodwill	80,522,400	304,524,800	682,005,400	683,425,500
Long-term deferred expenses	16,832,600	15,945,500	5,040,600	2,235,500
Notes payable and accounts payable	37,539,200	50,504,300	57,324,600	136,850,100
Taxes payable	4,723,800	2,534,600	4,055,100	2,348,200
Total of other payables	15,006,300	23,877,200	5,020,500	7,068,600
Among them: interest payable	697,400	750,800	103,500	584,800
Long-term loan	402,400,100	302,500,100	286,500,100	45,250,100
Undistributed profit	-452,064,600	-254,200,400	60,355,800	2,005,600
Total owner's equity attributable to parent company	312,085,400	567,550,920	982,902,200	906,825,700
Minority interest	-245,200	-4,733,400	69,906,800	81,730,500

Table 3 presents the main data of the income statement of Enterprise S for the four-year period 2020-2023. Analyzing Table 3, it can be found that Enterprise S has a large rate of change in its profit from main business during the four-year period of 2020-2023, and the enterprise grows steadily during the two-year period of 2020-2021, and achieves a small increase in the level of profit and revenue. During the 2021-2022 period, revenues decrease but operating costs increase, with interest expense and asset impairment losses decreasing more, and it is important to consider whether Business S is utilizing the tax shield effect of interest and the potential tax risks involved. The increase in costs also leads to a plunge in profits, which are negative in 2022-2023.

Table 3: S Company income statement data (Yuan)

Subject	2023	2022	2021	2020
1 Total business income	71,874,226	20,850,551	32,568,194	23,442,027
2 Total operating costs	50,720,416	69,504,310	28,037,371	11,812,839
Among them: operating costs	26,772,186	20,330,841	8,546,082	3,783,730
Business taxes and surcharges	775,875	83,905	284,714	421,235
Selling expense	7,670,879	7,106,332	18,809,830	8,605,705
Administrative expenses	2,914,477	9,162,457	1,929,630	15,570,697
Research and development expense	4,797,231	7,925,280	829,872	4,283,757
Financial expense	2,836,175	3,570,573	2,134,206	54,760
Among them: interest expense	3,153,978	1,461,602	488,833	4,433,657
Interest income	304,095	573,550	341,324	5,042,810
Asset impairment loss	28,592,427	1,287,321	426,299	3,309,630
3 Operating profit	-36,067,425	-7,525,937	11,733,760	14,704,002
4 Total profit	-73,078,445	-91,346,071	6,226,886	14,441,569
Less: Income tax expense	544,692	57,596	-35,145	-21,362
5 Net profit	5,624,558	-78,482,029	9,534,353	1,041,731

III. C. Enterprise tax risk assessment model

III. C. 1) Data introduction

In order to analyze the potential tax risk of enterprise S, the data of capital intensity and other indicators of this enterprise are studied as sample data. In order to facilitate the experiment, the data containing missing values are deleted, and finally 1500 pieces of relevant index data of S enterprise are organized. SPSS19 software is utilized to statistically process these 1500 index data.

Table 4 shows the specific results. As can be seen from Table 4, there are 1500 experimental data, of which 1500 are valid data, of which the larger fluctuations are the inventory turnover ratio and current assets turnover ratio, accounts receivable turnover ratio, the full distance of the data are all above 110, and the variance terms are all greater than 20. Inventory turnover ratio is too large, which indicates that the enterprise has better sales, but it is too large and also increases the risk of failure of collection. A large accounts receivable turnover ratio indicates that the company has a large proportion of credit sales. The next big changes are current ratio, current asset turnover, cost of goods sold ratio and net income per share. Current ratio, current asset turnover and cost of goods sold ratio may be affected by industry characteristics. Net earnings per share, on the other hand, reflects that the company has some ability to grow.

Next are Capital Intensity, Finance Staff Ratio, Inventory Intensity, and Net Asset Intensity, whose minimum values are shown as 0.00. This may be caused by the small value of this sample indicator, and SPSS19 automatically retains two decimals by default. It does not affect the feasibility of the experiment in terms of the next research. From the overall mean and variance, the overall distribution of the indicators is more reasonable, and the data are more stable. The enterprise tax risk is a triple categorical variable with 1, 0, -1 distribution, from high to bottom indicates that the performance of the enterprise is getting worse, and the tax risk is increasing gradually. The mean value of the tax risk of enterprise S is 0.21, which is greater than 0, which indicates that although there is anomaly on the financial statement of the enterprise, but the enterprise has a tax crisis is still a small probability event.

Table 4: Descriptive statistic

	N	Range	Minimum value	Maximum value	Mean value	Standard deviation	Variance
Capital intensity	1500	0.75	0.00	0.75	0.21	0.24	0.017
Asset-liability ratio	1500	0.89	0.06	0.89	0.43	0.18	0.021
Current ratio	1500	23.04	0.21	27.08	3.31	0.94	4.05
Inventory turnover	1500	100.52	0.05	121.34	7.81	9.34	73.81
Accounts receivable turnover rate	1500	241.82	0.71	273.82	10.54	15.80	192.45
Turnover of current assets	1500	12.35	0.15	9.33	1.87	0.92	19.55
Cost of sales ratio	1500	16.57	1.31	15.57	1.83	0.89	0.91
Proportion of financial personnel	1500	0.55	0.00	0.13	0.08	0.01	0.00
Net asset value per share	1500	14.38	0.35	12.43	5.05	3.72	5.04
Inventory density	1500	0.84	0.00	0.72	0.12	0.13	0.01
Return on equity	1500	0.45	0.00	0.43	0.23	0.06	0.01
Regional tax preference	1500	1.00	0.00	1.00	0.31	5.21	0.21
Tax risk	1500	2.05	-1.00	1.00	0.25	0.64	0.41
Valid N (list state)	1500	-	-	-	-	-	-

III. C. 2) Optimization of indicators

There are three tests for bivariate testing of variables in SPSS19, which are Pearson test, Kendall's tau-b test & Spearman test. Among them, the Pearson test requires that all independent and dependent variables obey a joint normal distribution, and the product difference correlation coefficient is utilized to test whether there is a linear relationship between the independent and dependent variables. When $P<0.5$, it is significantly correlated at the 95% confidence level. The product correlation coefficient is affected by the extremes of the data and can only be applied to all data samples where the independent and dependent variables are continuous variables. Kendall's tau-b test utilizes the rank correlation coefficient and is applicable to data samples where the independent and dependent variables are categorical variables. Spearman's test does not require that the data in the sample follow a normal distribution and is a non-parametric test. It is a non-parametric test. Meanwhile, there is no requirement that the independent variables and dependent variables are categorical or continuous variables. In this paper, the tax risk

degree and “regional tax incentives” in the sample data are categorical discrete variables, and the other variables are continuous variables, so this paper chooses the Spearman to test the correlation of the data.

Table 5 shows the correlation coefficient results. Where C is the correlation coefficient and S is Sig. (two-sided). This test is a bivariate test, where each cell of data reflects the correlation between the variable whose column heading it is and the variable whose row heading it is. Thus, it can be seen that when the column variable and the row variable are the same variable, the correlation coefficient is 1.00. The sig value in the table is the P-value of the test, and when the P-value is greater than 0.5 the original hypothesis is accepted that there is no significant correlation between the variables, and when the P-value is less than 0.5, the original hypothesis is rejected, and the variables are considered to be significantly correlated with one another at the 95% confidence level; Significantly correlated at 99% confidence level when p-value is less than 0.1. This paper focuses on the correlation between corporate tax risk and other variables and thus focuses on the results shown in the first column.

Table 5: Correlation coefficient result

		Tax risk	B5	B1	B2	B9	B6	B7	B10	B4	B11	B8	B3	B12
Tax risk	C	1.00												
	S	-												
B5	C	0.043*	1.00											
	S	0.015	-											
B1	C	-0.060*	0.058	1.00										
	S	0.034	0.111	-										
B2	C	0.045	-0.276**	-0.875**	1.00									
	S	0.191	0.001	0.000	-									
B9	C	0.040	0.235**	0.042	-0.272**	1.00								
	S	0.358	0.000	0.106	0.000	-								
B6	C	-0.002	0.356**	0.036	-0.187**	0.292**	1.00							
	S	0.304	0.000	0.289	0.000	0.000	-							
B7	C	0.003	0.472**	0.215**	-0.405**	0.522**	0.044**	1.00						
	S	0.909	0.000	0.000	0.000	0.00	.000	-						
B10	C	0.015	-0.176**	-0.482**	0.358**	-0.336**	-0.192**	-0.400**	1.00					
	S	0.601	0.000	0.000	0.000	0.000	0.000	0.000	-					
B4	C	-0.047*	-0.306**	0.164*	-0.005	-0.024	0.011	-0.171**	0.064*	1.00				
	S	0.011	0.000	0.038	0.098	0.480	0.746	0.000	0.034	-				
B11	C	0.045	0.008	-0.194*	0.192**	0.016	0.064*	0.033	0.083**	-0.145**	1.00			
	S	0.151	0.801	0.003	0.006	0.676	0.045	0.318	0.002	0.000	-			
B8	C	0.016	0.265**	-0.506	0.108**	-0.109**	-0.063*	-0.121**	-0.007	0.061	0.018	1.00		
	S	0.581	0.000	0.082	0.002	0.001	0.042	0.000	0.984	0.051	0.601	-		
B3	C	0.085**	-0.058	-0.017	0.701	-0.003	-0.008	0.016	0.059	-0.014	-0.003	-0.023	1.00	
	S	0.004	0.097	0.665	0.218	0.830	0.798	0.622	0.065	0.739	0.978	0.478	-	
B12	C	0.038	0.051	-0.011	-0.0123	0.005	-0.011	-0.015	-0.016	-0.001	-0.021	0.012	-0.123**	1.00
	S	0.217	0.100	0.764	0.726	0.898	0.755	0.651	0.638	0.914	0.521	0.674	0.001	-

* The correlation is significant when the confidence (double test) is 0.05.

** The correlation is significant when the confidence (double test) is 0.01.

Through Table 5 we observe that capital intensity (B5) has a p-value of 0.18 which is less than 0.5, gearing ratio (B1)0 has a p-value of 0.24 which is less than 0.5, finance staff ratio (B4) has a p-value of 0.11 which is less than

0.5, and return on equity (B3) has a p-value of 0.005 which is less than 0.1, and the variables are linearly significantly associated with firms' tax riskiness at 95%, 95%, 95%, and 99% levels of The variables are linearly and significantly correlated with the firm's tax riskiness at 95%, 95%, 95%, and 99% confidence levels, respectively. So far, we have come up with four independent variables that are significantly correlated with corporate tax risk: capital intensity, asset-liability ratio, financial personnel ratio, and return on net assets. Among them, capital intensity, asset-liability ratio and return on net assets are related to whether the business departments fully consider the enterprise tax risk in their operation activities, and whether the enterprise managers fully consider the tax risk factors in the investment decision-making process, which shows that the optimized indexes correspond to the results of the enterprise endogenous analysis, and can better reflect the tax risk management information of the S enterprises in various aspects.

III. C. 3) Feasibility analysis

In regression analysis, sometimes there is an exact high correlation, or a highly correlated relationship, between the independent variables, a phenomenon that can distort or make it difficult to accurately estimate the model. Full contributivity occurs less frequently, and is generally an approximate covariance, which has to be synthesized by combining a number of tests to make a comprehensive judgment on the covariance. Ordered logistic regression also takes this effect into account, requiring that there is no multicollinearity in the independent variables within the model. In this paper, the sample data are tested for multicollinearity in SPSS.

Table 6 shows the results of the test. Two covariance statistics are given in Table 6: tolerance and VIF. Tolerance is obtained based on the calculation of regression done on each independent variable as a dependent variable on the other independent variables. The smaller the tolerance, the more accurately the variable is predicted by the other variables and the stronger the covariance. Generally, covariance is considered to be a problem when the tolerance is less than 0.1. In this paper, it can be seen that the independent variables return on net worth, finance staff ratio, capital intensity and gearing ratio are all greater than 0.95, which is much greater than 0.1. VIF, the variance inflation factor, is the inverse of the degree of tolerance, which takes a value between 1 and 10, the smaller the better. According to Table 6, the variance inflation factors of the four significantly correlated independent variables are all close to 1.00, which is much smaller than 10, indicating that there is no multicollinearity among the four significantly correlated independent variables.

Table 6: Test coefficient result

Model	Nonnormalized coefficient		Standard coefficient			Collinear statistics	
	B	Standard error	Trial version	T	Sig.	Allowance	Vif
Constant	0.249	0.081		3.845	0.000		
Return on equity	0.606	0.362	0.053	1.888	0.062	0.997	1.005
Proportion of financial Personnel	-2.382	1.103	-0.072	-2.093	0.029	0.966	1.035
Capital intensity	0.371	0.141	0.081	1.734	0.067	0.966	1.037
Asset-liability ratio	-0.343	0.172	-0.083	-2.480	0.015	0.983	1.014

Dependent variable: tax risk

III. C. 4) Results of regression analysis

The four significantly correlated independent variables were regressed and their respective parameter estimates were calculated. Table 7 shows the parameter estimates. The parameter results of the ordered logistic regression are given in Table 7, from which it can be seen that the significant P-value of capital intensity is 0.058, which is slightly greater than 0.05, insignificant at the 95% confidence level, and significant at the 90% confidence level; the P-value of the gearing ratio is 0.019, which is less than 0.05, and significant at the 95% confidence level; the P-value of the percentage of finance staff is 0.036, which is less than 0.05, significant at 95% confidence level; return on net assets P-value is the same less than 0.05, significant at 95% confidence level. Therefore, it can be judged that the independent variables that have the greatest impact on the enterprise tax risk are the asset-liability ratio, the proportion of financial personnel, the return on net assets, and capital intensity, and accordingly construct the tax risk prediction model of S enterprise.

Table 7: Parameter estimates

		Estimate	Standard error	Wald	Df	Significance	95% confidence interval	
							Floor limit	Upper limit
Threshold value	[Risk=-1.00]	-1.837	0.218	70.019	1	0.000	-2.265	-1.407
	[Risk=0.00]	0.349	0.211	2.753	1	0.098	-0.064	0.758
position	Capital intensity	0.850	0.450	3.551	1	0.058	-0.036	1.735
	Asset-liability ratio	-0.835	0.354	5.608	1	0.019	-1.529	-0.143
	Proportion of financial personnel	-6.114	2.923	4.371	1	0.036	-11.845	-0.382
	Return on equity	2.012	0.965	4.358	1	0.036	0.124	3.901

III. C. 5) Evaluation of Model Gain Plots

In order to verify the effect of the enterprise tax risk prediction model constructed based on multiple logistic regression analysis, this paper takes the financial data from 2020 to 2023 as the training data, and selects four models, namely C&R tree, Bayesian network, neural network and SVM model, as the comparison models to jointly predict the tax risk situation of Enterprise S in 2024 and judge the prediction effect of the model. Figure 3 shows the multi-model prediction result gain chart. When making a multi-model comparison, for a gain plot, if the model is better, the steeper the model's curve on the plot will rise to 100% and then slowly flatten out. The more the curve deviates from the baseline, the better the model and also the higher the prediction accuracy. As can be seen from Fig. 3, the curve generated by the prediction model constructed based on the multivariate logistic regression method rises from 4.5% gain degree to 100% gain degree after only 1 steep change and rises the fastest and deviates from the baseline to the greatest extent. Accordingly, it can be concluded that the enterprise tax risk prediction model constructed based on the multiple logistic regression method in this paper is the most effective.

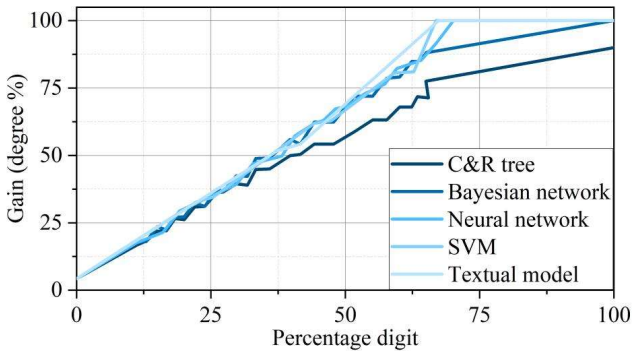

Figure 3: Multiple model gain graph

IV. Enterprise tax risk prevention and control strategies

According to the analysis results of the enterprise tax risk prediction model, it is necessary to focus on four financial indicators, namely, asset-liability ratio, financial personnel ratio, return on net assets, and capital intensity, to formulate corresponding enterprise tax risk prevention and control strategies, and to enhance financial soundness and market competitiveness.

A high gearing ratio indicates that an enterprise is overly reliant on debt financing and may face debt servicing pressure in the long term, which in turn may lead to tax risks. In this regard, it is recommended that:

1) Adjust the financing structure. Expand the proportion of equity financing by introducing strategic investors and issuing additional shares to reduce the scale of debt.

2) Debt maturity matching. Link long-term debt with fixed asset investment projects to avoid liquidity crisis caused by the concentrated maturity of short-term debt. For example, use long-term loans for R&D investment or technology upgrading to enhance asset earning capacity.

Too low a proportion of finance staff reflects the enterprise's insufficient investment in tax management, which may lead to tax accounting errors or insufficient utilization of policies. Specific strategies include:

1) Expand the professional team. Increase the number of personnel in the finance department, especially senior personnel with experience in tax planning, to ensure compliance and efficiency in tax filings and applications for preferential policies.

2) Regular training and assessment. Organize internal training in conjunction with the latest tax policies and enhance the team's ability to deal with complex tax issues through case simulations.

The fluctuation of return on net assets directly affects the stability of corporate profits, which in turn affects the ability to pay taxes. It is recommended to optimize it in the following aspects:

1) Strengthen cost and expense control. Establish a cost accounting refinement system, distinguish the reasonableness of R&D, sales and management expenses, and avoid ineffective expenditures.

2) Expand high value-added business. Relying on the advantages of cloud computing technology, develop enterprise-level value-added services, improve product gross margins, and enhance profit sustainability.

High capital intensity indicates low asset turnover efficiency, which may lead to cash flow constraints due to idle assets. Recommendations:

1) Optimize fixed asset investment. Adopt an asset-light operation model, shift part of the infrastructure to leasing or cloud services, and reduce the proportion of fixed assets.

2) Accelerate inventory and accounts receivable turnover. Monitor the inventory level in real time through digital supply chain management system to avoid backlog; meanwhile, strengthen the credit assessment of customers and shorten the accounts receivable account period.

V. Conclusion

This paper realizes the prediction of enterprise tax risk by constructing the index system and logistic regression model. Through data analysis, it is found that the asset-liability ratio ($P<0.05$), the proportion of financial personnel ($P<0.05$), the return on net assets ($P<0.05$) are significant at the 95% confidence level, and the capital intensity is significant at the 90% confidence level, which become the independent variables with the highest contribution to the enterprise tax risk. Further comparing the prediction effect of multiple models, the slope of the gain curve of the multivariate logistic regression model constructed in this paper is significantly better than that of the control group, which verifies that it has a higher prediction accuracy for enterprise tax risk. Combined with the prediction results and the four independent variables, the corresponding enterprise tax risk prevention and control strategies are given. In the future, deep learning models can be introduced to improve the mining ability of relevant indicators to enhance the quality of the early warning model.

Funding

Project 1: Guangdong University of Science and Technology "Innovation Strong School Project" Collaborative Innovation Center "Institute of Finance, Taxation and Economics" (Project Number: GKY-2020CQJG-4).

Project 2: Guangdong University of Science and Technology Major Research Achievement Cultivation Program "Research on the Layout and Co construction Mechanism of Advanced Manufacturing Innovation Centers in the Guangdong Hong Kong Macao Greater Bay Area" (Project Number: GKT-2020CQPT-4).

Project 3: Guangdong University of Science and Technology First-class Major Program in Accounting (Project Number: GKZLGC2024352).

References

[1] Bellingwout, J. W. (2020). Global Enterprise Taxation (GET). Intertax, 48, 317.

[2] Chen, W., Liang, S., & Zhang, L. (2024). Tax uncertainty and corporate innovation output: evidence from China. The European Journal of Finance, 30(10), 1127-1163.

[3] Chen, L., & He, R. (2024). Does digital tax enforcement drive corporate digitalization? Evidence from the Golden Tax Project III in China: A pre-registered report. Pacific-Basin Finance Journal, 83, 102242.

[4] Ramesh, V. K., & Athira, A. (2024). Geopolitical risk and corporate tax behavior: international evidence. International Journal of Managerial Finance, 20(2), 406-429.

[5] Holkova, B., Malichova, E., Falat, L., & Pancikova, L. (2023). Determinants of tax ethics in society: Statistical and logistic regression approach. Mathematics, 11(10), 2341.

[6] Cai, D. F., Ni, N. N., & Cai, J. (2015). Big enterprise tax risk management: Warning, simulation and application. International Journal of u-and e-Service, Science and Technology, 8(4), 135-146.

[7] Ievsieieva, O. (2025). State Tax Regulation of Enterprises in the Conditions of Digitalization. International Journal of Accounting, Finance and Risk Management, 10(2), 86-93.

[8] Meylis, A. (2019). Analysis and prevention of enterprise financial risk under the new tax policy. Open Journal of Business and Management, 7(4), 1943-1952.

[9] Darono, A., & Pratama, A. (2022). Tax Audit in the Era of Big Data: The Case of Indonesia. Journal of Tax Administration, 7(2), 27-50.

[10] Barik, T. R., & Ranawat, P. (2024). Transformation of Traditional Corporate Tax Planning into AI-Driven Corporate Tax Planning. Involvement International Journal of Business, 1(4), 269-280.

[11] Rahman, R. A., Masrom, S., Omar, N., & Zakaria, M. (2020). An application of machine learning on corporate tax avoidance detection model. IAES International Journal of Artificial Intelligence, 9(4), 721.

[12] Brink, W., & Hansen, V. (2018). Using Big Data to identify tax risk. Tax Adviser, 49, 318.

[13] Li, H., & Yu, X. (2025). Construction of tax risk management evaluation system for listed enterprises from the perspective of big data. Journal of Computational Methods in Sciences and Engineering, 14727978251317308.
[14] Zheng, Q., Xu, Y., Liu, H., Shi, B., Wang, J., & Dong, B. (2024). A survey of tax risk detection using data mining techniques. Engineering, 34, 43-59.
[15] Tang, Y., Li, Y., Qu, M., & Yang, Q. (2021, December). Research on Risk Early Warning Methods of Enterprise Finance and Taxation Management Based on Big Data Mining. In 2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS) (pp. 157-161). IEEE.
[16] Xu, C., & Kong, Y. (2024). Random forest model in tax risk identification of real estate enterprise income tax. Plos one, 19(3), e0300928.
[17] Akinrinola, O., Addy, W. A., Ajayi-Nifise, A. O., Odeyemi, O., & Falaiye, T. (2024). Application of machine learning in tax prediction: A review with practical approaches. Global Journal of Engineering and Technology Advances, 18(02), 102-117.
[18] Lavrenteva, E. (2019). Tax risk management at the enterprise of the energy sector. In International Scientific Conference Energy Management of Municipal Facilities and Sustainable Energy Technologies EMMFT 2018: Volume 2 (pp. 146-157). Springer International Publishing.
[19] Shuping, L., Linmin, X., Shuwen, Y., & Xiaoyun, Z. (2021). Research on Enterprise Internal Control and Tax Risk Prevention. Academic Journal of Business & Management, 3(3), 74-78.
[20] Ouyang, S., & Fang, Y. (2022). Enterprise Financial and Tax Risk Management Methods under the Background of Big Data. Mathematical Problems in Engineering, 2022(1), 5831866.
[21] Chen, Z. (2024). A Study on the Construction of an Intelligent Prediction Model for Corporate Tax Risk. In Innovative Design and Intelligent Manufacturing (pp. 144-150). IOS Press.
[22] Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F., & Pagliuca, D. (2020). Combining network visualization and data mining for tax risk assessment. Ieee Access, 8, 16073-16086.
[23] Abedin, M. Z., Chi, G., Uddin, M. M., Satu, M. S., Khan, M. I., & Hajek, P. (2020). Tax default prediction using feature transformation-based machine learning. IEEE Access, 9, 19864-19881.
[24] Zhu, L. (2022). Methodology and Application of Fiscal and Tax Forecasting Analysis Based on Multi‐Source Big Data Fusion. Mathematical Problems in Engineering, 2022(1), 8028754.