Pan, J.S., Rong, Y.T., Lv, Y., Peng, Q.Y., & Cai, X.F. (2024). An enterprise deregistration risk assessment model integrating public credit information and tax Information. In Li, E.Y. *et al.* (Eds.) *Proceedings of The International Conference on Electronic Business, Volume 24* (pp. 662-667). ICEB'24, Zhuhai, China, October 24-28, 2024

An Enterprise Deregistration Risk Assessment Model Integrating Public Credit Information and Tax Information

J.S. Pan [1]
Y.T. Rong [2,*]
Y. Lv [3]
Q.Y. Lin [4]
J.R. Peng [5]
X.F. Cai [6]

*Corresponding author
[1] Shenzhen Public Credit Center, Shenzhen, China, panjs@mail.amr.sz.gov.cn
[2] Xi'an Jiaotong University, Xi'an, China, Ph.D. Candidate, jetryt@163.com
[3] Shenzhen Public Credit Center, Shenzhen, China, Yangl2023@mail.amr.sz.gov.cn
[4] Shenzhen Public Credit Center, Shenzhen, China, linqy1@mail.amr.sz.gov.cn
[5] Shenzhen Public Credit Center, Shenzhen, China, Pengjr2023@mail.amr.sz.gov.cn
[6] Research Assistant Professor, National Center for Applied Mathematics-Shenzhen (NCAMS), Southern University of Science and Technology, Shenzhen, China, caixf@mail.sustech.edu.cn

ABSTRACT

The assessment of firm's credit risk and the prediction of enterprise's default and bankruptcy probability plays an important role in financial organizations and regulatory authorities. Since 1960s, an extensive range of techniques and ratios have been used to predict the default and bankruptcy probability of enterprises. However, default and bakruptcy is not the only form of credit risk. Other forms of firm exit, such as mergers and acquisitions, even malicious cancellation, may happen and lead to credit risk, but have been largely neglected by existing literature. This study proposes an enterprise deregistration risk (including default, bankruptcy, mergers, acquisitions, and malicious cancellation) assessment model integrating public credit information and tax information, considering a broader range of credit risk. The proposed model could help financial organizations and regulatory authorities more comprehensively evaluate the credit risk of enterprises.

Keywords: Credit risk, deregistration, firm exit, public credit information, tax information.

INTRODUCTION

The assessment of firm's credit risk and the prediction of enterprise's default and bankruptcy probability has always received attention from scholars in multiple fileds, including finance, operations research, management science, etc., and practitioners in the financial market (Doumpos et al., 2017; Le et al., 2021; Zhang et al., 2023). On the one hand, once an enterprise defaults, it takes a long time and a high cost to restore normal operation, even leading to bankruptcy liquidation in the worst case, especially for small and medium-sized enterprises (SMEs). On the other hand, an enterprise's default or bankruptcy would affect the interests of all its stakeholders (e.g., investors, creditors, shareholders, employees; Kedia & Mishra, 2024; Sun & Shenoy, 2007). Therefore, it is not only important for financial organizations to make credit grating and investment decisions, but also crucial for regulatory authorities to take action in advance (Doumpos & Zopounidis, 2011; Zhang et al., 2024a).

Since credit risk assessment is of critical importance, an extensive range of methods and algorithms have been used to predict the default or bakruptcy of enterprise. Beaver (1966) was the first to introduce financial ratios (e.g., cash flow to total assets, net income to total debt, quick assets to current liabilities) as predictors of enterprise failure by using univariate analysis. After that, Altman (1968) and Altman et al. (1977) applied multiple discriminant analysis (MDA) to the problem of bankruptcy prediction, which attempts to derive a linear combination of firm charcteristics (e.g., working capital to total assets, sales to total assets) to best discriminate between the bankrupt and non-bankrupt samples. Ohlson (1980) preferred a logistic regression more over MDA as the latter requires less restrictive statistical assumptions and could perform better in bankruptcy discrimination. Though statistical methods remain popular in a long time, they are limited in real application due to strict assumptions and pre-existing functional forms relating response variables. Since 1980s, a variety of artificial intelligence (AI) techniques have been used to offer better prediction performance, such as support vector machines (SVMs; Kim & Sohn, 2010), neural networks (Zhang et al., 1999; Shi et al., 2024), deep learning (Ramesh & Jeyakarthic, 2024; Zhang et al., 2024b).

Besides prediction methods and techniques, prediction ratios play an important role in the performance of credi trisk assessment models. In initial works, Beaver (1966) and Altman (1968) considered only financial ratios in their discriminant models. However, using fanancial ratios to evalute credit risk is always complicated and costly, especiall for SMEs, due to a lack of reliable data source and other factors (Kou et al., 2021). Therefore, more and more studies apply non-financial ratios to predict the default or bankruptcy probability of enterprises. Mirzaei et al. (2016) found that industry related factors, such as the abundance of industry resources, industry volatility, and industry intensity, are also important factors affecting enterprise default. Liang et al. (2016) combined financial ratios with corporate governance indicators in bankruptcy prediction, and found that board structure and ownership structure are important indicators of enterprise bankruptcy.

Though default and bakruptcy risk prediction has been largely analyzed and explored, it is not the only form of credit risk. Other forms of firm exit, such as mergers and acquisitions, even malicious cancellation of the company, always happen, which could also lead to credit risk (Duan et al., 2012). For instance, in May 2021, workers from a catering company in Hongkou District, Shanghai, reported through 12345 that the company had delayed their wages. The human resources and social security department of Hongkou District immediately launched an investigation into the catering company but found that it had been closed due to poor management, though the three shareholders of this company still owe 21 employees a total of more than 110000 yuan in wages. Thus, the human resources and social security department ordered the catering company to make up for the unpaid wages of workers. But the catering company neither submitted an administrative review nor delayed payment of the unpaid wages after more than half a year has passed. In March 2022, when the human resources and social security department decided to apply to the court for compulsory enforcement, it was discovered that the catering company had already been registered for cancellation, which occurred the day before the administrative decision was made. Enterprise deregistration means that it is impossible to enter the execution process, and previous administrative decisions have also been forced to be revoked due to the loss of administrative counterparties, the unpaid wages of workers are still far away. Such cases not only occur in Hongkou District, Shanghai, but also in multiple places such as Ningxia.

Besides unpaid wages, malicious cancellation of the company may lead to more serious social and economic issues. For example, in construction and engineering industry, many homeowners are faced with quality issues of their purchased residences during the warranty period. However, when they go to the developer for repair, they realize that the developer has already cancelled the company, so they don't know who to turn to for warranty responsibility. When this happens, it takes a long time and high cost for homeowners to defend their rights.

However, the majority of current studies of credit risk assessment have focused on predicting the probability of default and bankruptcy, neglecting other forms of credit risk. This study aims at addressing the above issues by proposing an enterprise deregistration risk (including default, bankruptcy, and other forms of firm exit) assessment model integrating public credit information and tax information of enterprises, which are rarely used in the field of credit risk assessment. Zhu and Chen (2024) explored the impact of tax credit on financial credit risk and found that tax credit reduces financial credit risk. But the authors only considered one tax credit ratio, namely, whether the enterprise appears on the list of grade A (the highest grade, annual TC evaluation index score of 90 points or more) taxpayers assessed by China's tax authorities. Yin (2020) combined information found in legal judgments with financial and firm-specific information to evaluate the credit risk of small and medium-sized enterprises (SMEs).

This study intends to make two contributions to the literature. First, though malicious cancellation of the company always happen and lead to serious social and economic issues, it remains underexplored in the literature. To the best of our knowledge, this is the first time that enterprise deregistration risk is used as a prediction target. Second, this study predicts the probability of enterprise deregistration using a combination of public credit information and tax information, but not traditionally financial ratios. As SMEs face inherent problems such as a lack of a complete information disclosure system and the ambiguity in self disclosed information (Chen et al., 2015; D'Aurizio et al., 2015), such a risk assessment index could cover more enterprises.

RESEARCH DESIGN

Variables

Dependent Variable

The task of credit risk assessment is always modelled as a dichotomous classification problem, assigning "1" to default ones and "0" to non-default ones (Rong et al., 2023; Serrano-Cinca & Gutiérrez-Nieto, 2016). Similarly, this study regard enterprise deregistration risk assessment as a dichotomous classification task, where "1" represents deregistration enterprises and "0" represents remaining enterprises. The objective is to estimate the deregistration probabilities of enterprises and distinguish enterprises that are supposed to be remaining from those that are very likely to deregistrate in 6 months.

Independent Variables

This study includes two main categories of independent variables, public credit information and tax information.

Public Credit Information: Public credit information can be roughly divided into three categories, including basic business information, administrative management information, and change information. Basic business information contains information generated in the process of registration. Administrative management information refers to the information generated by the government in the process of managing administrative affairs of enterprises, such as administrative licenses, administrative penalties. Enterprise change information reveals enterprises' changes in shareholders, capital, and other aspects.

Tax Information: China's tax authorities annually assess a tax credit rating to taxpayers based on taxpayers' fulfillment of tax obligations within a certain period of time. If an enterprise scores 90 points or more according to tax authorities' evaluation index, it obtains a Grade A tax credit. Thus, enterprises with a Grade A tax credit is assigned a value 1 and 0 otherwise. Moreover, this study includes: 1) whether enterprises owe taxes to tax authorities and are included in the overdue collection list; and 2) whether enterprises are recognized by tax authorities as abnormal taxpayers.

All dependent and independent variables are listed below in Table 1.

Table 1: Dependent and independent variables

Variables	Description
Dependent variable	
Deregistration	Whether the target enterprise will deregistrate in 6 months.
Independent variables: Public credit information	
Basic business information	Years of establishment.
	Whether an enterprise is publicly listed.
	Industry categories an enterprise belongs to.
	Amount of registered capital.
	Number of people engaged in the target enterprise.
	Whether the target enterprise is rated as a high-tech enterprise.
Administrative information	Whether the target enterprise is on the list of dishonest executors.
	Whether the target enterprise is on the list of seriously dishonest entities.
	The number of administrative penalties imposed on the target enterprise in the past year.
	The amount of administrative penalties imposed on the target enterprise in the past year.
	The number of times the target enterprise appears in the list of abnormal operations by the Market Supervision Administration in the past year.
	Whether the target enterprise has been included in the industry regulatory red list issued by the National Development and Reform Commission.
Enterprise change information	The number of times the target enterprise has changed its name in the past year.
	The number of times the target enterprise has changed its legal representative, chairman, executive director, general manager, and other key responsible personsin the past year.
	The number of times the target enterprise has changed its registered address in one year.
Independent variables: Tax information	
Grade A tax credit	Whether an enterprise is assigned a Grade A tax credit by tax authorities.
Overdue collection list	Whether an enterprise owes taxes to tax authorities and is included in the overdue collection list.
Abnormal taxpayer	Whether an enterprise is recognized by tax authorities as abnormal taxpayer.

Research Methodology and Model Selection

Logistic regression has been the most widely used in the field of credit risk assessment (Guo et al., 2016). Therefore, we first choose logistic regression as our benchmark model. Besides, AI techniques have been widely used in credit risk assessment, among which, decision trees, support vector machines and neural networks remain popular and competitive (Ma et al., 2021; Malekipirbazari & Aksakalli, 2015). Thus the above three AI techniques will be used to further validate the proposed index for enterprise deregistration risk assessment. After that, He et al. (2018) and Xia et al. (2017) pointed out that tree-based ensemble method can achieve promising performance in credit risk assessment tasks. Thus, tree-based ensemble method will be further adopted to improve the model performance. The above models will be compared with each other according to the following evaluation metrics to select the most suitable one for this task.

Evaluation Metrics

Choosing proper evaluation metrics is crucial for credit risk assessment tasks to guide model selection. *Accuracy* is the most popular measure for quantifying the performance of binary classification models (Kou et al., 2021). Therefore, we first choose *accuracy* as our evaluation metric. However, *accuracy* may be inappropriate when class distributions and misclassification costs are rarely uniform. For instance, in the context of enterprise deregistration, there are always much more remaining enterprises than deregistration borrowers, and the cost of misclassifying a deregistration borrower is much larger than that of misclassifying a remaining borrower. Thus, the purpose of enterprise deregistration risk assessment is to identify as more deregistration enterprises as possible, while retain all remaining enterprises. Accordingly, we choose *recall* as our second evaluation metric, which measures how many of the positive samples (i.e., deregistration enterprises) are captured by the positive predictions. *Recall* is often used when it is important to identify all positive samples (Wu et al., 2022).

Data Collection

The first part of dataset used in this study will be crawled from the nationwide credit information sharing platform and "Credit China" platform established by Chinese government (Figure 1), which provide enterprises' public credit information (e.g., administrative penalty information, administrative licensing information, red and black list information) and some pieces of tax information (e.g., whether the enterprise is on the list of grade A taxpayers assessed by China's tax authorities) across China.

Source: https://www.creditchina.gov.cn.

Figure 1: "Credit China" platform.

Then, we will collect enterprise deregistration information from National Enterprise Credit Publicity System (Figure 2) to construct the dependent variable. In China, reasons for enterprise deregistration include adjudication of bankruptcy, mergers, acquisitions, business period expiration without renewal, or the company is dissolved internally, etc., which could all lead to credit risk.

Source: https://www.gsxt.gov.cn/corp-query-homepage.html.

Figure 2: National Enterprise Credit Publicity System.

Public credit information and tax information will be combined to establish the enterprise deregistration risk assessment index and predict the risk of enterprise deregistration. Expectedly, our sample will include all deregistration enterprises from January 2024 to May 2024. In accordance with Wu et al. (2022), the dataset will be randomly divided into two parts (training set and testing set) by the ratio of 8:2, where the training set is used to establish the classification model and the test set is use for model comparison.

DISCUSSION

In summary, this research proposes an enterprise deregistration risk assessment model. On the one hand, the proposed deregistration risk assessment model considers a broader range of credit risk, taking other forms of firm exit into account besides default and bankruptcy, and sets firm deregistration as the prediction target, which has been largely ignored in the literature. On the other hand, the proposed deregistration risk assessment model uses a combination of public credit information and tax information. The research results show that both public credit information and tax information could be used as important indicators of enterprise deregistration risk.

This study may provide feasible practical and managerial implications. First, as financial organizations and regulatory authorities have difficulty in obtaining reliable and complete information of enterprises to predict their credit risk, especially financial information, the proposed deregistration risk assessment index using a combination of public credit information and tax information could help expand credit granting and supervision to larger groups. Second, our classification model takes time horizons into account, namely, whether the enterprise will deregistrate in 6 months. Thus, there will be enough time for financial organizations and regulatory authorities to take actions to avoid the risk brought by enterprise deregistration, even to discover and prevent enterprises' malicious cancellation in advance.

ACKNOWLEDGMENT

The authors would like to acknowledge the partial grant support to the research (Grant ID: 72061127002) and the Young Innovative Talent Program of Guangdong (Grant ID: 2023WQNCX067). This research is also supported by DeFin research center, National Center for Applied Mathematics Shenzhen; Shenzhen Key Research Base in Arts & Social Sciences (Intelligent Management & Innovation Research Center of SusTech), and the National Laboratory of Mechanical Manufacture System, XJTU, China.

REFERENCES

Altman, E. I. (1968). Financial ratios, discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589-609. https://doi.org/10.2307/2978933

Altman, E. L., Edward, I., Haldeman, R., & Narayanan, P. (1977). A new model to identify bankruptcy risk of corporation. *Journal of Banking & Finance*, 1, 29-54. https://doi.org/10.1016/0378-4266(77)90017-6

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111. https://doi.org/10.2307/2490171

Chen, Y., Huang, R. J., Tsai, J., & Tzeng, L. Y. (2015). Soft Information and Small Business Lending. *Journal of Financial Services Research*, 47(1), 115-133. https://doi.org/10.1007/s10693-013-0187-x

D'Aurizio, L., Oliviero, T., & Romano, L. (2015). Family firms, soft information and bank lending in a financial crisis. *Journal of Corporate Finance*, 33, 279-292. https://doi.org/10.1016/j.jcorpfin.2015.01.002

Doumpos, M., Andriosopoulos, K., Galariotis, E., Makridou, G., & Zopounidis, C. (2017). Corporate failure prediction in the European energy sector: A multicriteria approach and the effect of country characteristics. *European Journal of Operational Research*, 262(1), 347-360. https://doi.org/10.1016/j.ejor.2017.04.024

Doumpos, M., & Zopounidis, C. (2011). A multicriteria outranking modeling approach for credit rating. *Decision Sciences*, 42(3), 721-742. https://doi.org/10.1111/j.1540-5915.2011.00328.x

Duan, J. C., Sun, J., & Wang, T. (2012). Multiperiod corporate default prediction-A forward intensity approach. *Journal of Econometrics*, 170(1), 191-209. https://doi.org/10.1016/j.jeconom.2012.05.002

Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, 249(2), 417-426. https://doi.org/10.1016/j.ejor.2015.05.050

He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, 98, 105-117. https://doi.org/10.1016/j.eswa.2018.01.012

Kedia, P., & Mishra, L. (2024). Credit risk management: A systematic literature review and bibliometric analysis. *Journal of Credit Risk*, 20(1), 82. https://doi.org/10.21314/JCR.2024.002

Kim, H. S., & Sohn, S. Y. (2010). Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research*, 201(3), 838-846. https://doi.org/10.1016/j.ejor.2009.03.036

Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems*, 140, 14. https://doi.org/10.1016/j.dss.2020.113429

Le, R., Ku, H., & Jun, D. (2021). Sequence-based clustering applied to long-term credit risk assessment. *Expert Systems with Applications*, 165. https://doi.org/10.1016/j.eswa.2020.113940

Liang, D., Lu, C.-C., Tsai, C. F., & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561-572. https://doi.org/10.1016/j.ejor.2016.01.012

Ma, Z. W., Hou, W. J., & Zhang, D. (2021). A credit risk assessment model of borrowers in P2P lending based on BP neural network. *Plos One*, 16(8), 21. https://doi.org/10.1371/journal.pone.0255216

Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10), 4621-4631. https://doi.org/10.1016/j.eswa.2015.02.001

Mirzaei, M., Ramakrishnan, S., & Bekri, M. (2016). Corporate default prediction with industry effects: Evidence from emerging markets. *International Journal of Economics and Financial Issues*, 6(3), 161-169.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131. https://doi.org/10.2307/2490395

Ramesh, R., & Jeyakarthic, M. (2024). Enhancing credit risk prediction with hybrid deep learning and sand cat swarm feature selection. *Multimedia Tools and Applications*, 21. https://doi.org/10.1007/s11042-023-17974-3

Rong, Y., Liu, S., Yan, S., Huang, W. W., & Chen, Y. (2023). Proposing a new loan recommendation framework for loan allocation strategies in online P2P lending. *Industrial Management & Data Systems*, 123(3), 910-930. https://doi.org/10.1108/IMDS-07-2022-0399

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems*, 89, 113-122. https://doi.org/10.1016/j.dss.2016.06.014

Shi, Y., Qu, Y., Chen, Z., Mi, Y., & Wang, Y. (2024). Improved credit risk prediction based on an integrated graph representation learning approach with graph transformation. *European Journal of Operational Research*, 315(3), 786-801. https://doi.org/10.1016/j.ejor.2023.12.028

Sun, L. L., & Shenoy, P. P. (2007). Using Bayesian networks for bankruptcy prediction: Some methodological issues. *European Journal of Operational Research*, 180(2), 738-753. https://doi.org/10.1016/j.ejor.2006.04.019

Wu, Y., Huang, W., Tian, Y., Zhu, Q., & Yu, L. (2022). An uncertainty-oriented cost-sensitive credit scoring framework with multi-objective feature selection. *Electronic Commerce Research and Applications*, 53, 24. https://doi.org/10.1016/j.elerap.2022.101155

Xia, Y., Liu, C., & Liu, N. (2017). Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electronic Commerce Research and Applications*, 24, 30-49. https://doi.org/10.1016/j.elerap.2017.06.004

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), 16-32. https://doi.org/10.1016/S0377-2217(98)00051-4

Zhang, W., Tong, M., Yin, Y. H., & Shang, J. (2024). How do credit rating agencies and bond investors react to credit guarantees? Evidence from China's municipal corporate bond market. *Journal of Credit Risk*, 20(1), 82. https://doi.org/10.21314/JCR.2023.011

Zhang, W., Yan, S., Li, J., Peng, R., & Tian, X. (2024). Deep reinforcement learning imbalanced credit risk of SMEs in supply chain finance. *Annals of Operations Research*, 1-31. https://doi.org/10.1007/s10479-024-05921-w

Zhang, Y., Ye, S., Liu, J., & Du, L. (2023). Impact of the development of FinTech by commercial banks on bank credit risk. *Finance Research Letters*, 55, 8. https://doi.org/10.1016/j.frl.2023.103857

Zhu, R., & Chen, F. (2024). Tax and financial credit risks-Empirical evidence from Chinese investment enterprises. *Finance Research Letters*, 61, 7. https://doi.org/10.1016/j.frl.2023.104917