

RESEARCH

Open Access



Clues from networks: quantifying relational risk for credit risk evaluation of SMEs

Jingjing Long¹, Cuiqing Jiang^{1,2*} , Stanko Dimitrov³ and Zhao Wang¹

*Correspondence:
jiangcuiq@163.com;
jiangcuiq2017@163.com

¹ School of Management,
Hefei University of Technology,
Hefei 230009, Anhui, People's
Republic of China

² The Philosophy and Social
Sciences Laboratory of Data
Science and Intelligent
Society Governance, Ministry
of Education, Hefei, Anhui,
People's Republic of China

³ Department of Management
Sciences, University of Waterloo,
200 University Avenue West,
Waterloo, ON N2L 3G1, Canada

Abstract

Owing to information asymmetry, evaluating the credit risk of small- and medium-sized enterprises (SMEs) is difficult. While previous studies evaluating the credit risk of SMEs have mostly focused on intrinsic risk generated by SMEs, our study considers both intrinsic and relational risks generated by neighbor firms' publicly available risk events. We propose a framework for quantifying relational risk based on publicly available risk events for SMEs' credit risk evaluation. Our proposed framework quantifies relational risk by weighting the impact of publicly available risk events of each firm in an interfirm network—considering the impact of interfirm network type, risk event type, and time dependence of risk events—and combines the relational risk score with financial and demographic features to evaluate SMEs credit risk. Our results reveal that relational risk score significantly improves both discrimination and granting performances of credit risk evaluation of SMEs, providing valuable managerial and practical implications for financial institutions.

Keywords: SMEs, Credit risk evaluation, Interfirm network, Risk event, Relational risk

Introduction

Small- and medium-sized enterprises (SMEs) critically support national economies, being main contributors to employment and economic development in many countries (OECD 2020). However, SMEs have difficulty obtaining the credit loans needed for their growth and development. Most SMEs are private firms, and their financial information is less transparent and reliable than that of large firms, which are, in general, publicly traded. Hence, financial institutions have difficulty conducting effective evaluations of SMEs' credit risk. Consequently, financial institutions are prone to *credit rationing* (i.e., limiting the supply of additional credit to SMEs requiring funds) even if the latter are willing to pay higher interest rates (Stiglitz and Weiss 1981; Murro and Peruzzi 2019). To solve SMEs' financing difficulties, helping financial institutions effectively evaluate the credit risk of SMEs is crucial.

To this end, many studies focus on extracting features from various factors to improve credit risk evaluation performance (e.g., Altman 2010; Yin et al. 2020). Commonly

used features include demographic, financial, and credit history features (e.g., Abdou et al. 2016; Djeundje et al. 2021), which predominantly depict the intrinsic risk profile of SMEs. However, some studies demonstrate that a firm's risk can be affected by that of its neighbor firms (e.g., Giesecke and Weber 2004; Beaver et al. 2019). Simultaneously, SMEs have weak anti-risk capability and are vulnerable to external factors; hence, SMEs are more susceptible to the risks of their neighbor firms (OECD 2016). Accordingly, to effectively evaluate SMEs' credit risk, both intrinsic risk arising from SMEs and relational risk arising from SMEs' neighbor firms must be considered. We refer to the impact of a firm's neighbor firms' risk on the risk of that firm as *relational risk*.

Considering that relational risk is essential in the credit risk evaluation of SMEs, and the network approach greatly contributes to the understanding of the interdependency between firms (Zha et al. 2020), several studies use network analysis to explore relational risk for credit risk evaluation of SMEs. Some studies extract statistical features related to relational risk (e.g., number of neighbor firms and defaulted neighbor firms) from interfirm networks and incorporate these features in credit risk evaluation (e.g., Vinciotti et al. 2019; Letizia and Lillo 2019). These features consider only the distribution of neighbor firms, but they rarely capture the impact of a firm's neighbor firms' risk on the risk of that firm. Other studies use historical defaults of each firm in an interfirm network to quantify relational risk for credit risk evaluation of SMEs (i.e., quantifying the impact of credit history information of a firm's neighbor firms on the credit risk of that firm) (e.g., Calabrese et al. 2019; Óskarsdóttir and Bravo 2021). In our study, quantifying relational risk based on each firm's credit history information in an interfirm network for credit risk evaluation refers to quantifying relational risk based on *homogeneous information*. In turn, this implies that the credit history information of each firm in an interfirm network is known. However, obtaining the credit history information of each firm in an interfirm network faces cost and access issues, since credit history information is either purchased (e.g., in the United States or Germany)¹ or not publicly available owing to the lack of credit information sharing systems (e.g., in some developing countries) (Fosu et al. 2020; Saruni and Koori 2020). As existing methods of quantifying relational risk based on homogeneous information suffer from both access and cost issues, a novel methodology for quantifying relational risk for credit risk evaluation of SMEs is required.

In this study, we present a method for quantifying relational risk based on *heterogeneous information* for credit risk evaluation of SMEs. Specifically, we quantify relational risk based on the publicly available risk events (instead of credit history information) of each firm in an interfirm network for credit risk evaluation. With the development of the government information disclosure system, some risk events of firms are disclosed by government agencies and freely accessible to everyone (e.g., legal judgments, administrative penalties) (Henninger 2013; Medvedeva et al. 2020). We refer to such risk events disclosed by government agencies as publicly accessible risk events. Thus, these publicly available events can be considered reliable while freely and easily accessible. When quantifying relational risk based on heterogeneous

¹ In the U.S., everyone can purchase a credit report of a firm via three business credit bureaus—Equifax, Experian, and Dun & Bradstreet (Berger and Frame, 2007). In Germany, financial institutions can purchase a credit report of a firm via a credit bureau Creditreform (Dierkes et al. 2013).

information for credit risk evaluation of SMEs, we aim to answer the following three research questions. (1) Does the relational risk quantified based on heterogeneous information improve the performance of credit risk evaluation of SMEs? (2) Does network type affect the effectiveness of relational risk in credit risk evaluation of SMEs? (3) Does network type and risk event type simultaneously affect the effectiveness of relational risk in the credit risk evaluation of SMEs?

Aiming to quantify relational risk to improve the performance of credit risk evaluation of SMEs and to map to the logic of the preceding three questions, we propose a framework for quantifying relational risk based on heterogeneous information for credit risk evaluation of SMEs. Our framework considers key elements of relational risk quantified based on heterogeneous information (i.e., network type, risk event type, and time dependence of risk events). In the proposed framework, we assign a time-dependent risk score to each firm in an interfirm network based on that firm's involvement in publicly available risk events. We then use a smoothed version of the weighted-vote relational neighbor (wvRN) algorithm to generate a score for each firm quantifying relational risk by weighting time-dependent risk scores of neighbor firms in an interfirm network. Finally, we combine the relational risk score with financial and demographical features to evaluate the credit risk of SMEs.

We evaluate our proposed framework using two datasets, namely, the manufacturing SMEs dataset and National Equities Exchange and Quotations (NEEQ) SMEs dataset (Wang et al. 2020a, b). We evaluate the utility of relational risk quantified based on heterogeneous information in credit risk evaluation of SMEs in terms of discrimination performance (i.e., the ability to distinguish bad loans from good loans) and granting performance (i.e., the ability to grant loans at a low default rate). Our results indicate that incorporating relational risk based on heterogeneous information can significantly improve discrimination performance. Moreover, we found that granting performance of credit risk evaluation of SMEs and network type, risk event type, and time dependence of risk events affect the effectiveness of the relational risk in credit risk evaluation of SMEs.

Our study provides the following contributions to research and practice. First, our study presents a novel methodology for quantifying relational risk. While existing studies quantify relational risk using the credit history information of each firm in the network (e.g., Calabrese et al. 2019; Óskarsdóttir and Bravo 2021), our study quantifies relational risk using the publicly available risk events of each firm in the network. Second, we propose a framework for quantifying relational risk based on heterogeneous information for credit risk evaluation of SMEs. Our framework introduces the impact of network type, risk event type, and time dependence of risk events, which are not simultaneously introduced in the framework of quantifying relational risk based on homogeneous information for credit risk evaluation (e.g., Van Vlasselaer et al. 2016; Tobback et al. 2017). Third, we provide a comprehensive analysis of whether relational risk based on heterogeneous information can improve the performance of credit risk evaluation of SMEs. Hence, we provide guidance for financial institutions.

Literature review

Two main research streams on credit risk evaluation (i.e., loan default prediction) are available. The first focuses on credit risk evaluation methods, which classify loan applications as default or nondefault (e.g., Angilella and Mazzù 2015; Figini et al. 2017; Maldonado et al. 2020), and the second focuses on exploring effective features related to credit risk. Meanwhile, the features can be divided into two types, wherein the first comprises basic features depicting the intrinsic risk of borrowers (e.g., Yin et al. 2020; Stevenson et al. 2021) and the second comprises relational risk features depicting the relational risk arising from borrowers' neighbors (e.g., Vinciotti et al. 2019; Kou et al. 2021). Since we aim to quantify relational risk using the network approach and incorporate quantified relational risk to improve the performance of credit risk evaluation of SMEs, we review related literature as follows: (1) credit risk evaluation methods, (2) basic features in credit risk evaluation, and (3) interfirm networks and relational risk.

Credit risk evaluation methods

Several studies on credit risk evaluation methods have been made to effectively evaluate the credit risk of firms in both industry and academia. For example, in the industry, credit rating agencies (e.g., S&P, Moody, and Fitch) are key in the credit market as they develop credit rating models to assess firms' creditworthiness and provide these credit ratings for market participants in decision-making (e.g., Wojewodzki et al. 2018; Wu et al. 2022). In academia, Altman (1968) developed the Z-score model based on multivariate discriminant analysis for bankruptcy prediction, which was then applied to credit risk evaluation. Based on Altman's work, researchers designed several representative credit risk evaluation models based on statistical methods, including linear discriminant analysis (LDA) and logistic regression (LR) (Abdou et al. 2016; do Prado et al. 2016). Currently, focus has shifted toward machine learning models, including support vector machines (SVM), neural networks, random forest (RF), and eXtreme gradient boosting (XGB) (Abellán and Castellano 2017).

Recently, the development of large neural networks with many layers of neurons (i.e., deep learning), has become popular in credit risk evaluation owing to its competitive computing and discrimination performance (Gunnarsson et al. 2021). However, deep learning models have low interpretability owing to their black box nature. In this study, we use LR, RF, and XGB for credit risk evaluation owing to their common usage (Lessmann et al. 2015).

Basic features in credit risk evaluation

To evaluate credit risk, researchers typically gather from firms' own information and extract basic features, which are mainly divided into three groups: demographic, financial, and credit history. Demographic features mainly describe a firm's essential attributes, including age, size, industry type, geographical location, character of management team, and percentage of insider ownership (Altman et al. 2010; Cassar et al. 2015). Financial features mainly describe a firm's profitability, leverage, liquidity, efficiency, and growth opportunity (e.g., current ratio, return on assets ratio, quick ratio, and assets turnover ratio) (Angilella and Mazzù 2015; Liang et al. 2016). Credit history features

mainly describe a firm's debt and debt repayment over a period. History of trustworthiness and expectations of continued performance demonstrate a borrower's ability to pay, including number of banks that lent money to the firm, number of short-term loans, and historical defaults (Bai et al. 2019; Djeundje et al. 2021).

The studies mentioned demonstrate the effectiveness of the basic features in SMEs' credit risk evaluation. However, some studies highlight that SMEs' information, especially financial information, is not always reliable (Altman et al. 2010; Cassar et al. 2015). Additionally, some studies find that firm risk not only depends on its own information but is also affected by information of its neighbor firms (Orton et al. 2015; Beaver et al. 2019). Therefore, considering only the basic features of SMEs when evaluating their credit risk is insufficient.

Interfirm networks and relational risk

The network approach has become an important tool not only for modeling and describing complex systems of interacting entities but also for facilitating understanding and analyzing complex system structure, trust propagation, and contagion risk (Buldyrev et al. 2010; Zha et al. 2020). Accordingly, using a network approach to study relational risk in an interfirm network is common (i.e., analyzing how a firm's neighbor firms' risk affects the risk of that firm). Currently, the network approach primarily demonstrates the existence of relational risk in a network or identifies mechanisms in different networks. For example, Elliott et al. (2014) construct a network of organizations based on cross-holding relationships and then identify how the network propagates discontinuous changes in asset values triggered by failures. Wang et al. (2019) construct a dynamic risk contagion model of interfirm credit guarantee network and study the dynamic risk contagion mechanism in the network. Grant and Yung (2021) infer the firm networks by estimating a vector autoregression of daily equity returns and find that network connections can function as conduits for contagion and elevated systemic risk.

Several studies use the network approach to incorporate relational risk features into the credit risk evaluation of SMEs. Vinciotti et al. (2019) construct an interfirm network based on transaction data and incorporate some features related to the network into credit risk evaluation of SMEs (e.g., number of companies from which transactions are received) and percentage of failed companies in the first-order neighborhood. Similarly, Letizia and Lillo (2019) construct an interfirm network based on payment data and only use network properties (e.g., degree, community, etc.) for credit risk evaluation of SMEs. However, these features only depict the distribution of neighbor firms but rarely capture how a firm's neighbor firms' risk affect the risk of that firm. Moreover, these features are based on an explicit transaction network and do not incorporate other network types. Only a few of the recent studies have focused directly on how to incorporate relational risk features into credit risk evaluation of SMEs. Calabrese et al. (2017) incorporate interdependence among London small business defaults into a credit risk analysis framework and demonstrate that incorporating the interdependence could improve the performance of credit risk analysis. Óskarsdóttir and Bravo (2021) develop a personalized multilayer PageRank centrality measure that ranks firms in the multilayer network according to a set of defaulted firms to understand relational risk and demonstrate that incorporating the measure could improve the performance of credit risk evaluation in

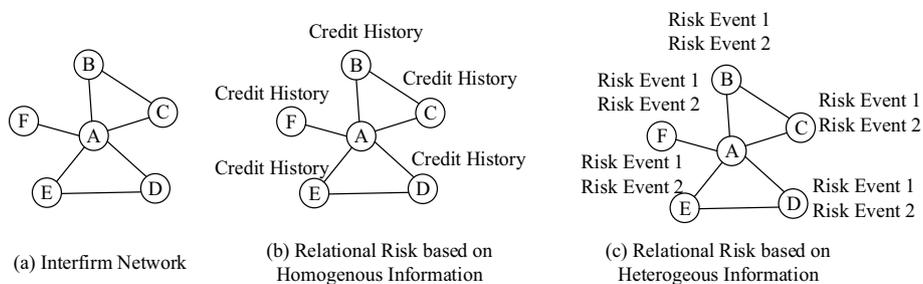


Fig. 1 Interfirm network and relational risk. **a** shows an example of an inter-firm network, in which node A represents a target firm, and nodes B–F represent neighbor firms of A. **b** and **c** show how to quantify A’s relational risk using credit histories of neighbor firms and risk events of neighbor firms, respectively

agricultural lending. Notably, these studies assume that the credit history information of each firm in the interfirm network is known. As accessing the credit history information of each firm in the network incurs access and cost issues, the applicability of these studies is limited.

Overall, previous studies have mainly used basic features for credit risk evaluation of SMEs or only demonstrate the existence of relational risk in different interfirm networks. Studies have rarely focused on incorporating relational risk features in credit risk evaluation of SMEs. Moreover, owing to access and cost issues, existing studies on incorporating relational risk features into credit risk evaluation of SMEs have limited applicability. Hence, our study presents a novel methodology for quantify relational risk (i.e., quantifying relational risk based on heterogeneous information) and proposing a framework to incorporate quantified relational risk into credit risk evaluation of SMEs.

Research design

In this section, we begin with a brief introduction into our conceptual framework and then go into detail about each part in the framework.

Conceptual framework

We propose a framework for quantifying relational risk based on heterogeneous information for credit risk evaluation of SMEs. Figure 1 illustrates the difference between the existing methods of quantifying relational risk based on homogeneous information and our method of quantifying relational risk based on heterogeneous information. Figure 1b shows that existing methods quantify relational risk (i.e., quantifying the impact of credit history information of firms (B–F) on the credit risk of firm A). We quantify relational risk (i.e., quantifying the impact of risk event information of firms (B–F) on the credit risk of firm A) as indicated in Fig. 1c. This difference raises our three research questions. Accordingly, our framework considers key considerations related to the relational risk quantified based on heterogeneous information (i.e., network type and risk event type). Furthermore, as risks events are dependent on time, our framework also considers time dependence of risk events.

Figure 2 shows that the framework includes: network construction, risk identification, relational risk quantification, and relational risk application. We first construct two types of interfirm networks (i.e., shareholding and governance networks). Second,

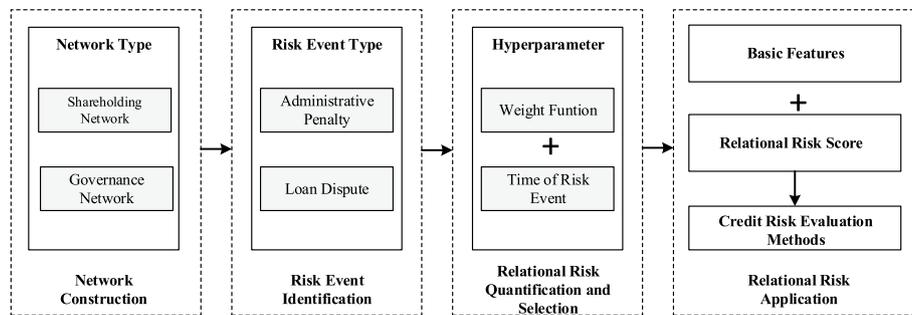


Fig. 2 Framework for quantifying relational risk based on heterogeneous information for credit risk evaluation of SMEs

each firm in the network is assigned a risk score based on its involvement with publicly available risk events (e.g., administrative penalty or loan dispute). Third, as interfirm network is unstructured and cannot be directly added, we use a smoothed version of the wvRN algorithm to generate a score that quantifies relational risk by weighting the time-dependent risk scores of firms based on the network structure (Tobback et al. 2017). Particularly, the wvRN algorithm includes several types of weight functions. Finally, we combine the relational risk score with basic features to evaluate the credit risk of SMEs. The following is a summary of the four phases in the framework:

1. Construct interfirm networks.
2. Identify the risk of each firm in the interfirm network depending on its involvement in administrative penalties or loan disputes.
3. Produce a score to quantify relational risk using the smoothed version of wvRN algorithm, with the selection of weight function and time of risk events.
4. Add the relational risk score into the credit risk evaluation methods of SMEs.

Network construction

We use a bipartite graph to construct interfirm networks. The bipartite graph is a graph whose nodes can be divided into two disjointed and independent sets in that every link connects a node in one set to a node in another set. Resources and firms can form a bipartite graph, where a link connects a firm and a resource if the firm has this resource. Figure 3a indicates that two firms are linked if they share at least one resource, including a manager, employee, buyer, address.

To construct an interfirm network, the bipartite graph is compressed into a projection graph, wherein two firms are linked if and only if they have at least one common resource, as shown in Fig. 3b. To incorporate information about shared resources into the edge between firms in the projection graph, we determine the weight of the edges between two firms by combining weights of the shared resources. We adopt several weight functions commonly used to weight a resource (Tobback et al. 2017) (Table 1). Most weight functions are based on the concept of micro-affinity, that is, down weighing nodes (i.e., resources) with a high degree. A firm’s manager (as a common “resource” among a small number of firms) has a greater impact on the linked firms

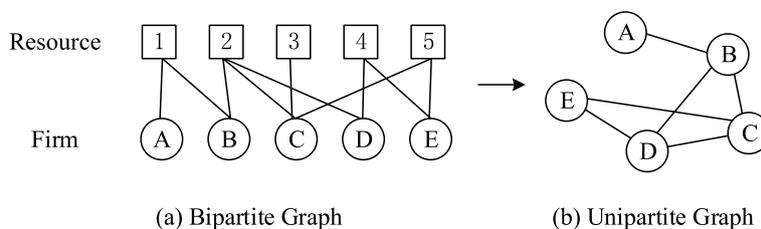


Fig. 3 A bipartite graph and a unipartite graph. The circular nodes A–E represent firms and the square nodes 1–5 represent resources

Table 1 Weight functions

Weight function	Formula
Inverse degree	$s_k = 1/d_k$
Inverse frequency	$s_k = \log_{10}(N/d_k)$
Hyperbolic tangent	$s_k = \tanh(1/d_k)$
Adamic and Adar	$s_k = 1/\log_{10} d_k$
Class-degree ratio	$s_k = d_k^c/d_k$

k represents the resource, s_k represents the weight of the resource k , d_k refers to the degree of the resource k (i.e., the number of links), N represents the number of firms in the network, and d_k^c refers to the degree of resource k linking to different classes.

than if the manager is shared among a larger number of firms, as a manager shared between many firms has limited time and energy for any one individual firm. Consequently, based on the weight function of resources, weight w_{ij} between firms i and j is determined by aggregating the weights, s_k , of the shared resources, i.e., $w_{ij} = \sum_{k \in N(i) \cap N(j)} s_k$. Multiple candidates for s_k (see Table 1) and $N(i)$ is the set of neighbor nodes for node i . The weight, w_{ij} , is considered in the method of quantifying relational risk (see “Risk identification” section for more details).

As firms are linked through shared resources, different resources can form different network structures. For example, if two firms share an address, a geographical relationship exists between the two firms. Hence, a geographical network is formed when more firms are linked by geographical relationships. In our framework, we consider the shareholding and governance relationships to construct and define two types of interfirm network—shareholding and governance networks. Consequently, we elaborate on each network in greater detail.

The *shareholding network* has edges representing common shareholders between two firms—an edge between two firms exists if they have at least one common shareholder. The shareholding network can be used to predict cascading financial distress between firms (Giesecke and Weber 2006). Essentially, the financial distress of a firm can lead to neighbor firms incurring financial liquidity problems or credit risk through the ownership relationship.

The *governance network* has edges representing the governance relationship between firms, which is characterized by shared managers (or directors) among firms, including an interlocking directorate. The governance network can reflect a manager’s ability and a firm’s performance. Specifically, the manager impacts firm performance, and an incompetent or fraudulent manager may lead a firm into default or bankruptcy. Two firms are

more likely to have similar performance if they share a manager, and if they share more managers, the performance will become even more similar.

Risk identification

To study the effect of the relational risk based on heterogeneous information on credit risk evaluation (i.e., whether the credit risk of a firm could be influenced by risk events of its neighbor firms), we must define risk events. We select two types of publicly available risk event (i.e., administrative penalty and loan dispute) to identify the risk of each firm in the network and then use the propensity of risk events of a firm's neighbor firms, in either the shareholding or governance network, to determine the firm's relational risk. We now discuss two risk events in greater detail.

Loan dispute constitutes a valuable repository of firm risk information. When a firm is designated as a debtor in a loan dispute and must compensate all associated debts, the situation reflects the firm's problems of financial liquidity and its ability to repay. Additionally, a firm can be found guilty of not honoring contractual financial obligations, reflecting its willingness or ability to repay. Hence, a firm indicating signs of financial distress is more likely to default if it is related, as defined in the interfirm network structures we consider, to several firms that have loan disputes.

Administrative penalty includes disciplinary warnings, fines, confiscation of illegal gains, orders for suspension of production or use, rescission of permits or other certificates of similar character, and other types of administrative penalties. A firm could be subject to an administrative penalty owing to environmental pollution, misleading advertising, faulty products, and so on. Hence, administrative penalties are usually quite negative. For instance, environmental penalties may result in fines and factory closures pending compliance with environmental regulations, which affect the firm's ability to repay any debts. Some penalties reflect on or result in a firm's bad reputation (e.g., misleading advertising or faulty products). As a bad reputation affects consumers' and partners' awareness of the firm—thus having a negative effect on the economic benefit of the firm—a firm that has signs of financial distress is more likely to default on its financial obligations if related to several firms with administrative penalties and thus bad reputations.

Relational risk quantification

To add relational risk into the credit risk evaluation methods, unstructured interfirm network data must be transformed to structured data, which can quantify relational risk with a single scalar value. Additionally, interpretability of relational risk is instrumental in the credit risk evaluation of SMEs. Therefore, the process of quantifying relational risk must be comprehensible and easily implementable. The wvRN algorithm is a simple collective inference method that utilizes network structure to estimate how a node is affected by its neighbor nodes in a network (Stankova et al. 2020). The smoothed version of the wvRN algorithm fits the credit risk evaluation methods better (Tobback et al. 2017). Hence, we adopt the smoothed version of the wvRN algorithm to calculate a score quantifying the relational risk for each firm in the network. We now briefly describe the original wvRN algorithm and then the smoothed version.

Let w_{ij} denote the weight associated with the link between firms i and j . Let Z denote the normalization factor, equal to $\sum_{j \in N(i)} w_{ij}$. Let $p(L_i = c|N(i))$ denote the probability score that the label L_i of firm i equal the class c given its neighbors $N(i)$ in the network. For example, in our work, class c depends on whether risk events are involved, that is, $L_i = \{Risk\ Event\ Occurs, No\ Risk\ Event\ Occurs\}$ and,

$$p(L_i = c|N(i)) = \frac{\sum_{j \in N(i)} w_{ij} \cdot p(L_j = c|N(j))}{Z} \tag{1}$$

According to Eq. (1), the resulting probability score $p(L_i = c|N(i))$ (i.e., the probability score that the firm i belongs to class c) is equal to a weighted average of the probability scores of its neighbor firms belonging to class c .

The weight of the link between firms i and j as the aggregation of the weights, s_k , of all shared resources, k , between firms i and j . Hence, $w_{ij} = \sum_{k \in N(i) \cap N(j)} s_k$. Table 1 reports weight functions of resource k . The neighbor firm’s probability score, that is, $p(L_j = c|N(j))$, is set to 0 or 1 depending on whether this neighbor firm j has a historical risk event. Each firm in the network will have an updated probability score based on its neighbor firms’ probability scores. This score captures our supposition that the probability of a risk event at a firm is determined by the weighted average of the equivalent probabilities of that firm’s neighbors. Our framework uses the resulting probability score to quantify relational risk. Additionally, we consider the following three causes:

1. If a firm has no neighbor firms, the weight of the link between firms is set to zero. Hence, the normalizing constant is zero, and the fraction in Eq. (1) is undefined.
2. If a firm only has one neighbor firm, and the neighbor firm has a risk event, then that one neighbor firm with the risk event will potentially skew the firm’s probability too high as the population propensity of risk events is low.
3. Similarly, if a firm only has one type of neighbor firm, if all the neighbor firms are involved in risk events, the neighbor firms’ failures will result in too much penalization for the target firm when using the wvRN (Eq. (1)). Similarly, a firm’s relational risk is undetected if all of its neighbor firms (of one type) have not been involved in risk events according to Eq. (1).

The smoothed version of the wvRN introduces μ_c , which is equal to the incidence rate of class c in the entire dataset, to assign a boundary value μ_c to the firm with no neighbor firms, only one neighbor firm, or only one type of neighbor firm. Equation (2) provides the smoothed version of the wvRN. In the following empirical evaluation, we use the smoothed version of the wvRN to calculate a resulting probability score for each firm in the network, which is used to quantify the relational risk of each firm:

$$p(L_i = c|N(i)) = \frac{\sum_{j \in N(i)} w_{ij} \times p(L_j = c|N(j)) + 2\mu_c}{Z + 2} \tag{2}$$

To quantify relational risk, a hyperparameter other than the weight function (i.e., the time of the risk event) must be considered. Given that a firm may have multiple risk

events that may influence a firm over a period, and to our knowledge, the ideal observation time is not known, conducting a pre-test for different values of the observation time is necessary. Relational risk is finally determined with the selection of weight function and time of risk event.

Empirical evaluation

Data for empirical evaluation were collected from a Chinese regional commercial bank and an enterprise information inquiry website called Qichacha.² We extracted and chose features from borrowing firms historically used to evaluate the credit risk of SMEs, including demographic and financial features. We collected the information required for constructing the interfirm network and quantifying relational risk from Qichacha.

Data

We evaluated our framework using two datasets from China (i.e., manufacturing SMEs dataset and NEEQ SMEs dataset). The large number of SMEs in China, especially default SMEs, provides substantial samples to reflect differences between default and non-default SMEs, allowing to learn reliable knowledge of relational risk. According to the *Law of the People's Republic of China on the Promotion of Small and Medium-Sized Enterprises (2017)*, the definition of SME is based on the industry, operating income, and number of employees. A manufacturing firm in China is defined as an SME when its own operating income is between 2 and 400 million RMB and the number of its own employees is between 20 and 1000. Besides, according to the default definition in the Basel II Capital Accords, if a credit loan has been in arrears for more than 90 days, this is identified as a default credit loan. Ultimately, the first Chinese manufacturing SMEs dataset comprises a sample of credit loans from 2136 manufacturing SMEs during 2016 and 2017, of which 136 default loans, with a default rate of 6.7%. Table 7 in Appendix A presents the summary statistics of SMEs in terms of both operating income and the number of employees.

The NEEQC SMEs dataset is from the National Equities Exchange and Quotations (NEEQ) market, which serves mostly innovative and growth SMEs. NEEQ firms eligible for listing can be transferred directly to the main board market (Wang et al. 2020a, b). "Default" in the NEEQ SMEs dataset is in accordance with the definition of special treatment firms in NEEQ market (i.e., a firm is defined as a special treatment firm when its negative net asset at the end of a fiscal year is negative) (Jiang et al. 2021). Accordingly, this dataset covers 7943 SMEs between 2018 and 2020, of which 406 firms are identified as default, with a 5.1% default rate.

Our study shows that basic features for credit risk evaluation of SMEs include demographic features (i.e., firm age, district, and registered capital), risk events of firms themselves (i.e., loan disputes and administrative penalties), and financial features (i.e., financial ratios). Age and district are usually considered when determining SME credit risk (Fernandes and Artes 2016). For risk events, Altman and Wilson (2010) and Yin et al.

² Qichacha (<http://www.qichacha.com>) displays publicly available information about firms (e.g., board structure, managerial information, administrative penalties, and loan disputes). Qichacha obtains all collected information from the state administration for industry and commerce of China using API technology, and thus can be considered authoritative and trustworthy.

(2020) have demonstrated that a legal judgment can effectively evaluate the credit risk of SMEs. We considered several financial ratios commonly used in the credit risk evaluation of SMEs, including the current, debt to assets, receivable turnover, inventory turnover, assets turnover, quick ratio, and operating profit ratios; return on equity (ROE); and return on assets (ROA) (Abdou et al. 2016). As discussed earlier, SMEs' financial information is of lower quality compared to large companies. Hence, we have performed winsorization for all financial ratios. Specifically, we establish the values of the top and bottom 1% of each financial ratio to the means of their k -nearest neighbors, which are identified by the k -nearest neighbor algorithm. Risk events of target firms themselves are equal to 0 or 1 depending on whether the target firm has been involved in a loan dispute or administrative penalty. Table 8 in Appendix 1 and Table 13 in Appendix 2 summarize the descriptive statistics of basic features on the two datasets.

For relational risk, we used the firm's shareholders, chief executive officer (CEO), and/or directors, supervisors, and senior management (DSS) as resources to form a bipartite graph and then construct interfirm network. We then followed three steps to collect information to quantify relational risk. For the manufacturing SMEs dataset,

1. At Qichacha, we did a reverse lookup on the 2136 SMEs, to find their CEO, shareholders, and DSS;
2. Then, at Qichacha, we crawled each person's (i.e., CEO, shareholder, and DSS) web page and parsed the web page to retrieve information about each person's employment at various firms. We finally included 9672 neighbor firms other than 2136 SMEs.³
3. We examined risk events of the 9672 neighbor firms from Qichacha.

Ultimately, we obtained a network of firms consisting of 2136 target SMEs and 9672 other neighbor firms; hence, 11,808 firms are in the interfirm network. Similarly, for the NEEQ SMEs dataset, we obtained a network of firms comprising 7943 target SMEs and 119,131 other neighbor firms (or a total of 127,074 firms in the interfirm network). Table 9 in Appendix 1 and Table 14 in Appendix 2 summarize the descriptive statistics of interfirm networks on the two datasets. Notably, in our empirical evaluation, for practicality (i.e., default prediction), we used cross-sectional data (i.e., the data at the time point of loan application) to quantify relational risk. Hence, changes of relationships among firms or macroeconomic factors may not influence the stability of the effect of the relational risk.

Relational risk

Our empirical evaluation used the shareholder network to represent the shareholding network and used the CEO network and the DSS network to represent the governance network. Subsequently, we used loan disputes and severe administrative penalties to identify risk events of all firms, both the 2136 target SMEs and the 9672 companies found via Qichacha. For administrative penalty, we considered severe administrative

³ The directors (or shareholders or the CEO) of each SME could also serve on boards of non-SMEs. In our empirical study, we incorporate SMEs and non-SMEs to construct an interfirm network and treat them equally. However, we use only the target 2,136 SMEs for credit risk evaluation.

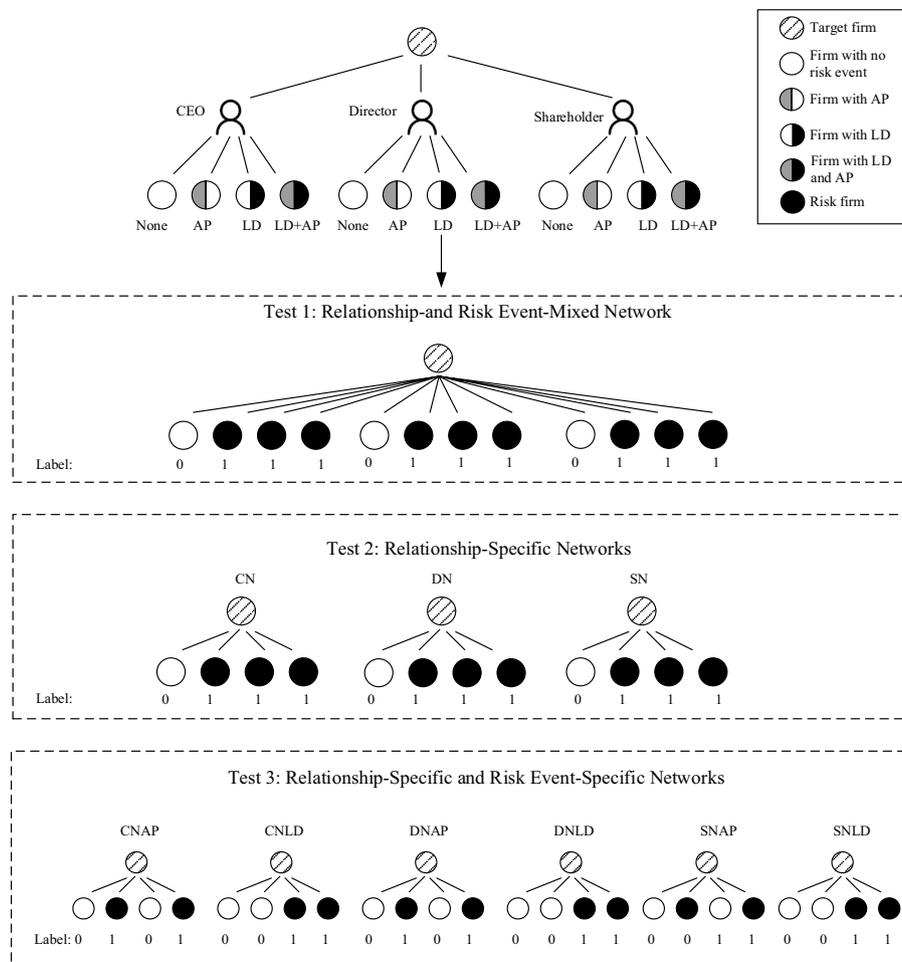


Fig. 4 Tests design for examining research questions

penalties that may have obvious effects on firm reputation and production, which we defined as fines over 10,000 RMB, forfeiture of illegal proceeds property, suspending of a production business, or temporary withholding or suspension of a license, according to the Law of the People’s Republic of China on Administrative Punishments. For loan disputes, we focused on the court judgments of loan disputes owing to the availability of data. Considering that the impact of a loan dispute will be affected by lawsuit status and the result of the judgment, we considered only loan disputes wherein the firm is on the losing end of the lawsuit and must pay compensation or is designated as a debtor in a losing case stipulating that the firm compensates all associated debts. We did not consider loan disputes that were later withdrawn. Finally, we used the smoothed version of the wvRN algorithm based on the network structure and risk events to produce a score to quantify relational risk, reflecting how a firm is affected by its neighbor firms’ risk events. Particularly, two hyperparameters (i.e., the weight function and the time of the risk event) in quantifying relational risk exist, and “[Selection of hyperparameters](#)” section provides details about the selection of the two hyperparameters.

Model building and evaluation

To examine three research questions, we implemented three sets of tests. Figure 4 presents a description of three sets of tests provided below, along with an example.

The first test examines whether relational risk quantified based on heterogeneous information can improve the performance of credit risk evaluation of SMEs. Specifically, we quantified relational risk based on a mixed network, wherein all the relationship types (CEO, DSS, and shareholder) can link two firms, and both types of risk event can trigger neighbor firms with risk. The target firm is linked to all neighbor firms through shared persons, and each neighbor firm is provided with a probability score of 1 or 0 depending on whether it was involved in risk events (AP, LD, or AP + LD).

The second test examines whether network type affects the effectiveness of relational risk in credit risk evaluation of SMEs. Specifically, the CEO-specific relationship forms the CEO network (CN), wherein the target firm links to a neighbor firm if they share a CEO. Similarly, the DSS relationship forms the DSS network (DN), and the shareholder relationship forms the shareholder network (SN). We considered risk events equally, and each firm is provided with a probability score of 1 or 0 depending on whether it has been involved in risk events (AP, LD, or AP + LD).

The third test examines whether network and risk event type simultaneously affect the effectiveness of relational risk in credit risk evaluation of SMEs. Specifically, we constructed six networks, according to different combinations of relationships and risk events. To combine the CEO network and administrative penalties (CNAP), the target firm is linked to a neighbor firm if they share a CEO, and each firm is given a probability score of 1 or 0 depending on whether it has been involved in an administrative penalty (AP). The remaining five networks are the CEO network and loan disputes (CNLD), DSS network and administrative penalties (DNAP), DSS network and loan disputes (DNLD), shareholder network and administrative penalties (SNAP), and shareholder network and loan disputes (SNLD), and are defined analogously.

For each test, we compared two types of feature sets (basic features, and combination of basic features and the relational risk) and tested their performance to determine whether and to what extent relational risk affects the performance of credit risk evaluation methods. Specifically, we selected three models for credit risk evaluation, namely, LR, RF, and XGB. LR is a typical linear model and commonly used for credit risk evaluation owing to its strong interpretability (Figini et al. 2017). RF and XGB, as representative models of ensemble learning, are good at handling nonlinear relationships among data. Specifically, RF is an ensemble of multiple decision trees based on bagging and is usually used to balance errors for unbalanced data sets (Veganzones and Séverin 2018). XGB is a powerful and effective implementation of the gradient boosting ensemble algorithm and widely used owing to its high precision and efficiency (Sigrist and Hirnschall 2019). Table 18 in Appendix 3 presents meta-parameters for the three classifiers.

Credit risk evaluation methods incorporating different features were compared based on discrimination and granting performances. Discrimination performance refers to the ability to distinguish bad loans from good loans. The area under the receiver operating characteristic curve (AUC), Kolmogorov–Smirnov (KS) statistic, and H measure are typical measures for measuring discrimination performance (Kotsiantis et al. 2006). Moreover, three measurements can measure the discrimination performance in learning

from imbalanced data, which solves the problem wherein two classes in loan default prediction are largely unbalanced (Hand 2009). The higher the AUC, KS, and H measure, the better the model's discrimination performance. Granting performance refers to the proportion of defaulters under different granting ratios of loans (Wang et al. 2020a, b). Under a given granting ratio, lower default rate means less losses and more benefit for financial institutions. Notably, the economic benefit of additional features in credit risk evaluation could also be evaluated by other measures (Maldonado et al. 2017; Garrido et al. 2018).

Selection of hyperparameters

As discussed earlier, we need to optimize two hyperparameters (i.e., weight function and the time of the risk event). Table 1 presents five types of weight function. Considering that loan disputes and administrative penalties cannot be predicted and should be disclosed in a timely manner, we collected the information of the mentioned risk events in the short term to capture the temporal influence of risk events. Examining whether there is a temporal effect of risk events when dividing different intervals by year, while the rate of risk events is too small if dividing different intervals by months. Therefore, we divided the time of risk events by quarters and used 17 different intervals (i.e., within T months with $T = \{3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48\}$ and more than 48 months). Notably, T is the duration from the time that the risk event occurred to the date of the loan application. Considering that the date of a loan application is usually 1 year before the performance point (i.e., the date of loan default), in our sample, the data did not include a firm's neighbor firms' risk event information occurring near the default date of that firm.

To estimate the discrimination performance of each model while two hyperparameters are optimized, we employed nested cross-validation (NCV) (Varma and Simon 2006). For the NCV, an inner cross-validation splitter is embedded within an outer cross-validation splitter. The inner cross-validation splitter is used to select hyperparameters. Outer cross-validation splitter averages the test error over multiple training–testing splits with the chosen hyperparameter. In our study, considering the manufacturing SMEs loan dataset as an example, the NCV is applied (Fig. 5). There are five steps as follows:

1. We transformed the interfirm network information using the wvRN into the relational risk score for each target SME. Basic features, the relational score, and the label of the credit loan of the target 2136 SMEs constitute the original dataset.
2. We select a tenfold cross-validation as the outer cross-validation splitter for the original dataset. In credit risk evaluation, the tenfold cross-validation technique is commonly used to split data and test model effectiveness (e.g., Van Vlasselaer et al. 2016; Yin et al. 2020). During the outer tenfold cross-validation, the original dataset is randomly split into ten equally sized datasets, called folds. Of the ten folds, a single fold is taken as the testing data. Conversely, the remaining nine folds are considered training data.

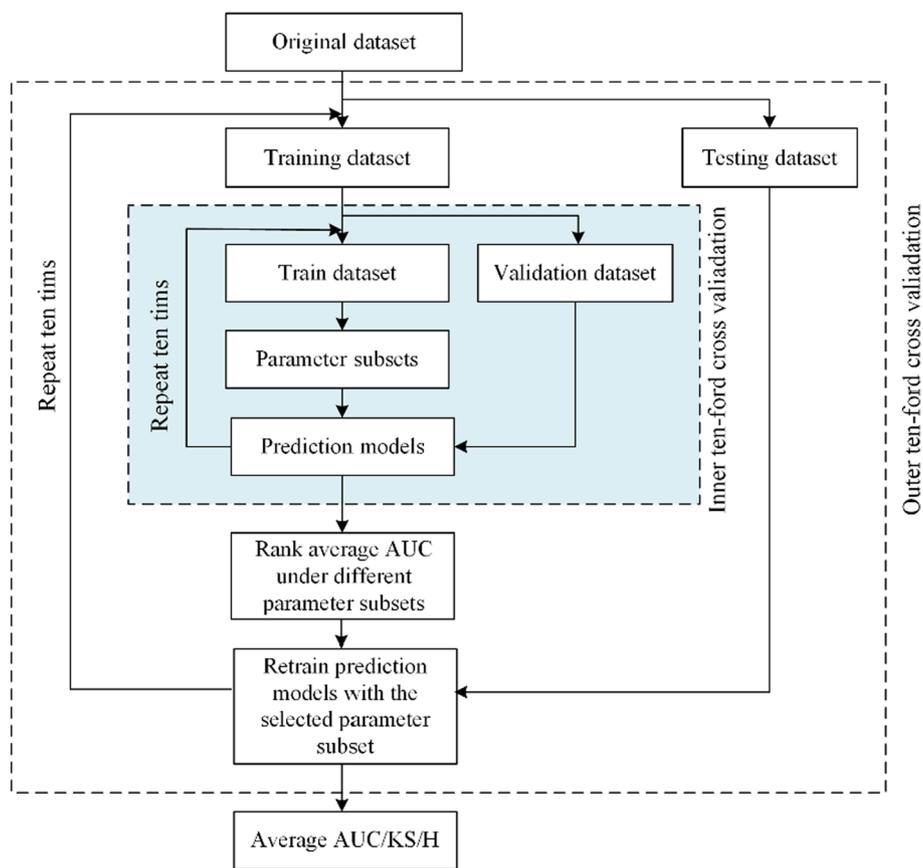


Fig. 5 Flowchart of the nested cross-validation. An inner cross-validation splitter chooses hyperparameters first, followed by an outer splitter that averages the test error over multiple training–testing splits

3. We selected a tenfold cross-validation as the inner cross-validation splitter on the training dataset. Similarly, during the inner tenfold cross-validation, the training dataset is randomly split into ten equal folds. Of the ten folds, a single fold is considered as the validation dataset, while the remaining nine folds are taken as training dataset. One fits a model on the training dataset with different hyperparameter subsets (i.e., different combinations of weight functions and the time of risk events) and evaluates it on the validation data. In our empirical study, 5 types of weight function and 17 types of time interval of risk events exist, resulting in 85 (17 × 5) hyperparameter subsets. To obtain a robust result of discrimination performance, we repeated the inner tenfold cross-validation procedure ten times, with a different split of folds each time, resulting in 100 performance estimates. Hence, the discrimination performance of each model with different hyperparameter subsets is determined by the average of 100 performance estimates.
4. We ranked the discrimination performance of each model with different hyperparameter subsets and selected the hyperparameter subset with the best performance as the optimal hyperparameter subset.
5. We retrained the model on the training dataset in the outer cross-validation with the optimal hyperparameter subset and evaluated this on the testing data. Similarly,

we repeated the outer tenfold cross-validation procedure ten times, resulting in 100 performance estimates. Discrimination performance (“Results” section), is estimated as the mean of 100 estimates and 95% confidence interval of the outer tenfold cross-validation.

Tables 10, 11, and 12 in Appendix 1 and Tables 15, 16, and 17 in Appendix 2 present summary statistics of risk events during different time intervals in different interfirm networks on two datasets. Tables 19 and 20 in Appendix 4 present the result of the optimal hyperparameter subset in different tests.

Results

We first evaluated the utility of the relational risk in credit risk evaluation of SMEs by implementing three sets of tests outlined in “Model building and evaluation” section on the manufacturing SMEs dataset. We then compared the granting performance of the baseline method with that of our own. Subsequently, to validate the utility of relational risk based on heterogeneous information in credit risk evaluation of SMEs, we selected another method (i.e., personalized PageRank algorithm) to quantify the relational risk and test its predictive power. Finally, we used the NEEQ SMEs dataset to further evaluate the generalizability of our proposed framework.

Relational risk in a mixed network

In the first evaluation, we examined the following:

Does relational risk quantified based on heterogeneous information improve the performance of credit risk evaluation of SMEs?

We assessed the discrimination performance of relational risk in a mixed network, wherein all relationship types (CEO, DSS, and shareholder) can link two firms, and both the risk event types (loan disputes and administrative penalties) can trigger neighbor firms with risk.

Table 2 summarizes the discrimination performance (by AUC, KS, and H measures) of the three classification models (LR, RF, and XGB) using two types of feature sets: only basic features (BF) and combination of basic features and the relational risk score (BF + RS). Baseline model uses only BF. When comparing two types of feature sets, the combination of basic features and relational risk score always performed better than the baseline model that uses only basic features in every performance measure. This suggests that relational risk is not a noisy feature. Furthermore, we tested the statistical significance of the effect of two types of feature sets (BF vs. BF + RS) on discrimination performance using a nonparametric test (Table 21, Appendix 5). Overall, differences between two types of feature set were statistically significant ($\chi^2 = 97.018$, $p < 0.001$, after the Bonferroni correction).

We further reported results of the coefficients of basic features and the relational risk score based on the mixed network in the LR model and importance (e.g., mean decrease in Gini) of features in the RF model (as shown in Table 32 and Fig. 7 in Appendix 6). The coefficient of the relational risk score in the LR model reflects a significant positive correlation between the credit risk and the relational risk score ($p < 0.05$). Mean decrease

Table 2 Discrimination performance of different feature sets (BF and BF + RS)

Model	Metrics	BF	BF + RS
LR	AUC	0.715	0.726
	KS	0.408	0.419
	H	0.251	0.270
RF	AUC	0.725	0.738
	KS	0.426	0.445
	H	0.265	0.279
XGB	AUC	0.695	0.715
	KS	0.387	0.416
	H	0.219	0.252

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times. In Appendix 6, Tables 29 and 31 show the 95% and 90% confidence intervals of each mean, and Table 33 shows the mean of 50 estimates obtained by repeating the ten-fold cross-validation procedure five times

in Gini of the relational risk score in the RF model demonstrates that relational risk is important in the credit risk evaluation of SMEs.

In summary, results show that the relational risk based on heterogeneous information indeed has discrimination ability in the credit risk evaluation of SMEs. Specifically, the relational risk score can significantly improve discrimination performance of the default prediction model when combined with basic features, again relative to the basic features alone.

Relational risk in relationship-specific networks

In the second evaluation, we examined the following:

Does network type affect the effectiveness of relational risk in credit risk evaluation of SMEs?

Compared with the mixed network, a fine-grained network, constructed by specific relationship type, involves a higher degree of specificity (firms with a specific relationship are linked together and may reflect relational risk more accurately). However, these networks inevitably have a lower degree of sensitivity (i.e., relatively fewer related firms are identified in each relationship-specific network). To investigate the impact of relationship type on the performance of relational risk score, we constructed three relationship-specific networks—CEO-specific network, DSS-specific network, and SN from the shareholder-specific network—and extracted three relational risk scores, CN, DN, and SN, respectively.

Table 3 summarizes the discrimination performance of relational risk scores extracted from relationship-specific networks. Combining relational risk score (CN, DN, or SN) with basic features significantly improves the discrimination performance of default prediction methods. However, the enhancing effects varied across relational risk scores, depending on the type of relationship-specific network they were extracted from. As mentioned, the results may echo those in “Relational risk in a mixed network” section, if the propensity of risk events is higher in our data. Across the three relational risk scores (CN, DN, and SN), relational risk score based on the DSS-specific network (DN) led to the greatest improvement.

Table 3 Discrimination performance of relational risk in relationship-specific networks

Model	Metric	BF	BF + CN	BF + DN	BF + SN
LR	AUC	0.715	0.722	0.729	0.725
	KS	0.408	0.414	0.431	0.417
	H	0.251	0.266	0.283	0.269
RF	AUC	0.725	0.738	0.741	0.735
	KS	0.426	0.444	0.450	0.439
	H	0.265	0.289	0.294	0.281
XGB	AUC	0.695	0.718	0.722	0.720
	KS	0.387	0.415	0.434	0.424
	H	0.219	0.253	0.257	0.245

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times. In Appendix 6, Tables 29 and 31 show the 95% and 90% confidence intervals of each mean, and Table 33 shows the mean of 50 estimates obtained by repeating the ten-fold cross-validation procedure five times

We also tested the statistical significance of the effect of the four types of feature set (i.e., BF, BF + CN, BF + DN, and BF + SN) on discrimination performance using a non-parametric test (Table 22, Appendix 5). Overall, differences among the four types of feature sets were statistically significant ($\chi^2 = 174.014, p < 0.001$, after Bonferroni correction). Further pairwise comparisons demonstrate that the performance of BF + DN is significantly better than that of BF and BF + CN.

Overall, these results demonstrate that the DSS-based governance network outperformed all alternatives, including the CEO-based governance and shareholder networks. Firms with the same DSS have a more similar level of management. Therefore, the more similar their profitability, the more similar their risk will be. This result provides strong evidence that the DSS relationship is the most effective type of relationship in constructing a network for capturing relational risk.

Relational risk in relationship-specific and risk event-specific networks

In the third evaluation, we examined the following:

Does network type and risk event type simultaneously affect the effectiveness of relational risk in credit risk evaluation of SMEs?

To further investigate the impacts of relationship type and risk event type on the performance of relational risk, we used a full-factorial design. We constructed six relationship-specific and risk event-specific networks and correspondingly calculated six relational risk scores (as shown in Table 4).

Table 5 summarizes the discrimination performance of the six relational risk scores. Risk event type of loan dispute demonstrates a robust performance across performance measures and models. Specifically, using the CEO relationship to link firms, CNLD outperformed CNAP in all performance measures (i.e., BF + CNLD vs. BF + CNAP). Using DSS to link firms, DNLD outperformed DNAP in terms of all performance measures (i.e., BF + DNLD vs. BF + DNAP). Using shareholders to link firms, SNLD outperformed SNAP in terms of all performance measures (i.e., BF + SNLD vs. BF + SNAP). Additionally, using DSS to link firms and loan disputes to identify related firms with risk (DNLD) outperformed all alternatives in terms of all performance measures.

Table 4 Relationship- and risk event-specific networks and corresponding relational risk scores

Relationship	Risk event	Network	Relational risk score
CEO	Administrative penalty	CEO + administrative penalty	CNAP
	Loan dispute	CEO + loan dispute	CNLD
DSS	Administrative penalty	DSS + administrative penalty	DNAP
	Loan dispute	DSS + loan dispute	DNLD
Shareholder	Administrative penalty	Shareholder + administrative penalty	SNAP
	Loan dispute	Shareholder + loan dispute	SNLD

Table 5 Discrimination performance of relational risk in relationship-specific and risk event-specific networks

Model	Metric	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP	BF + SNLD
LR	AUC	0.723	0.729	0.730	0.734	0.725	0.731
	KS	0.417	0.429	0.427	0.442	0.417	0.425
	H	0.268	0.282	0.269	0.288	0.269	0.278
RF	AUC	0.736	0.744	0.743	0.747	0.737	0.741
	KS	0.442	0.457	0.456	0.455	0.444	0.454
	H	0.283	0.296	0.291	0.296	0.282	0.295
XGB	AUC	0.713	0.715	0.721	0.732	0.716	0.725
	KS	0.410	0.415	0.423	0.439	0.415	0.436
	H	0.250	0.239	0.244	0.276	0.244	0.251

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times. In Appendix 6, Tables 30 and 32 show the 95% and 90% confidence intervals of each mean, and Table 34 shows the mean of 50 estimates obtained by repeating by repeating the ten-fold cross-validation procedure five times

We tested the statistical significance of the effect of the seven types of feature set (i.e., BF, BF + CNAP, BF + CNLD, BF + DNAP, BF + DNLD, BF + SNAP, and BF + SNLD) on discrimination performance (Table 23, Appendix 5). Overall, differences among the seven types of feature set were statistically significant ($\chi^2 = 433.640, p < 0.001$, after Bonferroni correction). Further pairwise comparisons demonstrate that the performance of BF + DNLD is significantly better than that of the others. Results suggest that the DSS relationship and loan disputes as risk event are most informative in determining a firm’s credit risk.

We further tested the statistical significance of the effect of the three types of feature set (i.e., BF + RS, BF + DN, and BF + DNLD) on discrimination performance using a nonparametric test (Table 24, Appendix 5). Overall, differences among the three types of feature set were statistically significant ($\chi^2 = 94.957, p < 0.001$, after Bonferroni correction). Further pairwise comparisons demonstrate that the performance of BF + DNLD is significantly better than that of BF + DN and BF + RS ($p < 0.001$, after Bonferroni correction).

Overall, these results demonstrate that network and risk event types can simultaneously affect the effectiveness of relational risk in credit risk evaluation of SMEs. Specifically, relational risk score in the combination of DSS network and loan dispute achieved the best incremental effect on the credit risk evaluation of SMEs. The results provide a noteworthy rationale for financial institutions to construct a fine-grained network with a specific relationship (i.e., DSS) and specific risk event (i.e., loan dispute) to maximize the discrimination performance of relational risk in the credit risk evaluation of SMEs.

Granting performance

To evaluate whether adding the relational risk could bring profits to financial institutions, we compared the granting performance of two models: (1) baseline model only using basic features and (2) model using both basic features and the relational risk score DNLD. First, we used a prediction model to estimate the default probability of a loan through a tenfold cross validation and then ranked the default probability from smallest to largest. Second, according to rank, we selected different cut-off values of the granting proportion of loans (i.e., percentage of loan applications approved) and calculated the default rate (i.e., proportion of defaulters) under different granting proportions.

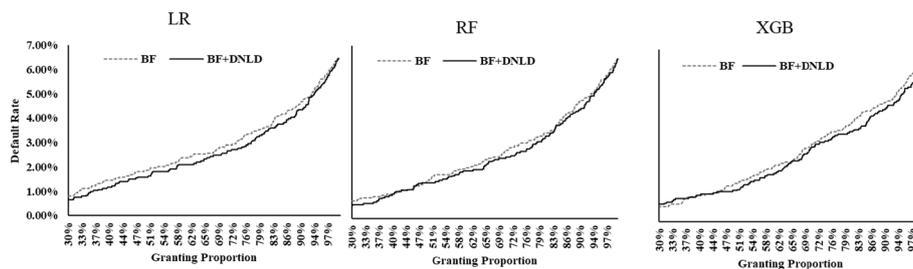


Fig. 6 Granting performance of two models. The horizontal axis shows various loan granting ratios ranging from 30 to 97%. The vertical axis shows the default rate, which ranges from 0 to 6.4%, with the maximum default rate equaling the overall default rate

Figure 6 visualizes the granting performance of the two models. We find that for the three classification methods, modeling using both basic features and relational risk score DNLD to scan loan applications produced a lower default rate than using only basic features. The result shows that relational risk could bring economic profits to banks.

Results of the personalized PageRank

To sufficiently validate that using publicly available risk events of each firm in an inter-firm network could be used to quantifying relational risk, in addition to the smoothed version of wvRN, we quantified relational risk using another method (i.e., Personalized PageRank algorithm) (Van Vlasselaer et al. 2016). First, we calculated the relational risk score using the Personalized PageRank algorithm, wherein the setting is identical in quantifying relational risk using wvRN in the mixed network (i.e., considering both the loan disputes and administrative penalties within 18 months in the mixed network). We then compared the discrimination performance of three types of feature set: (1) basic features (BF); (2) a combination of basic features and relational risk score using the wvRN (BF + RS); and (3) a combination of basic features and relational risk score using the Personalized PageRank algorithm (BF + prRS). Both relational risk scores RS and prRS are useful for credit risk evaluation (Table 6). We further tested the statistical

Table 6 Discrimination performance of different feature sets

Model	Metric	BF	BF + RS	BF + prRS
LR	AUC	0.715	0.726	0.723
	KS	0.408	0.419	0.417
	H	0.251	0.270	0.269
RF	AUC	0.725	0.738	0.735
	KS	0.426	0.445	0.440
	H	0.265	0.279	0.286
XGB	AUC	0.695	0.715	0.705
	KS	0.387	0.416	0.403
	H	0.219	0.252	0.228

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times

significance of the effect of the three types of feature set on discrimination performance using a nonparametric test (Table 25, Appendix 5). Overall, differences among three types of feature set were statistically significant ($\chi^2 = 104.004, p < 0.001$, after Bonferroni correction). Further pairwise comparisons demonstrate that the relational risk using both the relational risk scores RS and prRS significantly improves the discrimination performance of baseline models. Conversely, the incremental effect of the relational risk score RS is not significantly better than prRS ($p = 0.231$, after Bonferroni correction). The result shows that the relational risk quantified based on publicly available risk events of each firm in an interfirm network can indeed improve discrimination performance of credit risk evaluation. Moreover, quantification methods contain no restrictions.

Robustness tests

We conducted three robustness tests to further ensure that our results are not affected by (1) dataset selection; (2) data coverage; and (3) exclusion biases. In this subsection, we provided a summary discussion of the main tests and findings; the main test results are shown in Appendixes.

The first test examines whether our results are robust to potential dataset selection bias. We only evaluated the utility of our framework using samples on the manufacturing SMEs dataset. Hence, results and findings from samples on different datasets may differ. We then tested the robustness of results on a new dataset (i.e., NEEQ SMEs dataset) with the three sets of tests introduced in “Model building and evaluation” section. Tables 33, 34 and 35 and Fig. 8 in Appendix 8 present the results. Overall, in the NEEQ SMEs dataset, main results are similar to those in the manufacturing SMEs dataset. However, the performance of the LRs is lower. In our perspective, the main reason for the difference can be attributed to differences in sample size. With the increase of samples, the nonlinear relationship between features becomes more obvious; however, LR cannot effectively capture nonlinear relationships among features.

The second test examines whether our results are robust to potential data coverage bias. In the first robustness analysis, we only consider non-listed SMEs in the NEEQC SMEs dataset, while significant differences may be found on results between listed and unlisted SMEs. Hence, to alleviate data coverage bias, we added a dummy variable—whether an SME is listed or non-listed—to the basic features of NEEQC SMEs to test

the robustness of our results. Tables 36 and 37 in Appendix 8 indicate that our results are consistent with the first robustness analysis.

The third test examines whether our results are robust to potential exclusion bias. Loan disputes that were later withdrawn were not considered, as discussed in “[Relational risk](#)” section. To relieve exclusion bias, we included loan disputes that were later withdrawn when quantifying relational risk. Tables 38 and 39 in Appendix 8 present the results. Overall, our results are consistent with those in “[Relational risk in a mixed network](#)”—“[Relational risk in relationship-specific and risk event-specific networks](#)” sections, except in the LR model, wherein the relational risk score improves relatively poorly.

Conclusion

Previous studies on the credit risk evaluation of SMEs have mostly considered the intrinsic risk arising from SMEs, while only few studies have considered both the intrinsic and relational risks arising from SMEs’ neighbor firms (e.g., Vinciotti et al. 2019; Kou et al. 2021). Furthermore, existing studies typically quantify relational risk based on the credit history information of each firm in an interfirm network (e.g., Calabrese et al. 2017; Vinciotti et al. 2019). However, owing to access and cost issues, these studies have limited applicability. Hence, our study presents a novel methodology to quantify relational risk (i.e., quantifying relational risk based on publicly available risk events of each firm in an interfirm network) and proposes a framework to incorporate relational risk into credit risk evaluation of SMEs. Subsequently, we evaluate our framework relative to two Chinese SMEs datasets. Our empirical results indicate that incorporating our proposed relational risk statistically improves both discrimination and granting performances of credit risk evaluation of SMEs. Moreover, network type, risk event type, and time dependence of risk events influence the effectiveness of relational risk in credit risk evaluation of SMEs. These results can provide vital managerial implications for financial institutions in terms of quantifying relational risk and incorporating relational risk in credit risk analysis framework.

Our study provides the following contributions to research and practice. Considering the research perspective, we propose a framework for quantifying relational risk based on heterogeneous information for credit risk evaluation of SMEs. This framework can be used in other financial scenarios (e.g., bankruptcy prediction) to exploit relational risk hidden in interfirm networks. Moreover, our empirical evaluation suggests that relational risk quantified based on heterogeneous information can be used to evaluate credit risk for SMEs, broadening the existing literature on quantifying relational risk that quantifies relational risk based on homogeneous information (e.g., Calabrese et al. 2019; Óskarsdóttir and Bravo 2021).

From the practical perspective, our study may help financial institutions better evaluate SMEs’ credit risk. Specifically, our study provides a noteworthy rationale for financial institutions to incorporate relational risk based on heterogeneous information when granting credit to SMEs—especially in markets where the financial information of SMEs is insufficient and opaque. Additionally, we provide a potential new tool to enhance the current credit risk assessment process by incorporating relational risk score and, hence, direct more capital at a lower cost to the

least-risky SMEs. This promotes efficient allocation of resources, economic growth, and employment. Moreover, considering information collection and storage costs, we provide valuable guidance for financial institutions on how to collect information for studying relational risk, including quantifying relational risk using publicly available risk events, constructing a fine-grained network with a specific relationship (i.e., DSS) and specific risk event (i.e., loan dispute), and considering time dependence of risk events.

Our study contains the following limitations, which can serve as future research directions. First, the sample size for our empirical analysis is relatively small, which prevents us from using deep learning models for studying relational risk, including graph convolutional neural and graph neural networks (Lee et al. 2021). Second, owing to data privacy protection, we cannot obtain the credit history information of each firm in an interfirm network. Accordingly, we cannot compare the effectiveness of relational risk based on homogeneous information with that based on heterogeneous information. Third, as samples from different times are not available, we cannot use out-of-time validation to test model robustness. Fourth, we only evaluated our proposed framework on Chinese samples. In the future, we can conduct a more comprehensive evaluation involving samples from other countries and a cross-country study to test the external validity of our proposed framework. Additional data can address these shortcomings, warranting further study.

Appendix 1: Summary statistics of manufacturing SMEs dataset

The definition of SMEs in China is based on the industry, operating income, and number of employees. A manufacturing SME in China is defined as having an operating income of less than 400 million RMB and employees who number fewer than 1000. In our empirical study, the average operating income of the target 2136 manufacturing SMEs is 22,943,305 RMB, the maximum operating income is 399,280,700 RMB, and the minimum operating income is 2,006,597 RMB. The distribution of the number of employees of the target 2136 manufacturing SMEs is as shown in Table 7. We know from the frequency distribution histogram that there are no firms with more than 1000 employees. All firms in our dataset, are qualified as SMEs (the operating income range from 2 to 400 million RMB, the number of employees range from 20 to 1000), except one firm whose operating income was 1,006,597 RMB at the loan application time (Table 8).

Table 7 shows the descriptive statistics of basic features. Table 9 show the descriptive statistics of different interfirm networks on manufacturing SMEs dataset.

Table 10 shows the number and the rate of firms involved in risk events (both loan disputes and administrative penalties) over different intervals in different networks, including the mixed network (MN), the DSS network (DN), the shareholder network (SN), and the CEO network (CN). Table 11 shows the number and the rate of firms involved in administrative penalties over different intervals in different networks. Table 12 shows the number and the rate of firms involved in loan disputes over different intervals in different networks.

Table 7 Descriptive statistics of the number of employees

Number of employees	Number of firms
[20, 120)	1427
[120, 220)	615
[220, 320)	27
[320, 420)	12
[420, 520)	38
[520, 620)	2
[620, 720)	1
[720, 820)	3
[820, 920)	2
[920, 1000)	9

Table 8 Descriptive statistics of basic features on manufacturing SMEs dataset

No	Feature	Min	Mean	Max	S.D
1	Age (years)	1.000	9.033	36.000	5.141
2	Registered Capital (RMB) (Log)	2.790	6.978	12.180	0.736
3	Current Ratio (%)	8.800	95.700	192.400	37.131
4	Debt to Assets Ratio (%)	1.400	70.980	97.900	38.427
5	Receivable Turnover Ratio (times)	1.000	2.572	17.739	3.054
6	Inventory Turnover ratio (times)	1.000	4.517	17.801	3.301
7	Assets Turnover Ratio (times)	0.709	1.206	2.873	0.669
8	Quick Ratio (%)	1.300	70.770	189.200	41.344
9	OPR (%)	− 7.500	4.449	19.900	12.359
10	ROE (%)	− 18.900	5.292	20.800	4.965
11	ROA (%)	− 7.700	1.015	30.300	4.578
12	District		{3402–3417}, 15 different districts		
13	Risk Event of Firms Themselves		{0,1}, 0:1837;1:299		

Debt to assets ratio refers to the ratio of debt to assets, in which the total debt of a firm is the sum of its short-term and long-term debts. OPR refers to the proportion of profit on sales among sales revenue; ROE refers to the proportion of after-tax profits among the average value of the net assets at the beginning and end of the period; and ROA refers to the proportion of net profits among total assets. A district refers to a prefecture-level city that is an administrative division of China, ranking below a province. We coded these districts for privacy

Table 9 Descriptive statistics of different interfirm networks

	MN	DN	SN	CN
Average number of individuals per firm	3.591	2.617	2.685	1.000
Average number of neighbor firms per firm	5.635	3.766	3.511	2.800
Average number of firms per individual	3.172	2.803	2.672	2.773

The individuals in the mixed network (MN) refer to the CEO, shareholder and DSS, in the DSS network (DN) refer to the DSS, in the shareholder network (SN) refer to the shareholders, in the CEO network (CN) refer to the CEO

Table 10 Descriptive statistics of LD + AP in different networks

Risk event	MN	DN	SN	CN
MRE_M3	409 (3.458%)	190 (1.960%)	192 (2.397%)	135 (2.342%)
MRE_M6	619 (5.237%)	249 (2.575%)	254 (3.182%)	183 (3.165%)
MRE_M9	738 (6.247%)	292 (3.017%)	300 (3.764%)	215 (3.737%)
MRE_M12	834 (7.058%)	327 (3.379%)	331 (4.152%)	241 (4.182%)
MRE_M15	926 (7.835%)	418 (4.321%)	413 (5.167%)	301 (5.217%)
MRE_M18	1017 (8.605%)	464 (4.803%)	461 (5.770%)	334 (5.802%)
MRE_M21	1073 (9.085%)	451 (4.664%)	452 (5.663%)	330 (5.720%)
MRE_M24	1129 (9.565%)	464 (4.795%)	464 (5.807%)	340 (5.904%)
MRE_M27	1186 (10.045%)	531 (5.487%)	526 (6.588%)	384 (6.672%)
MRE_M30	1232 (10.426%)	555 (5.737%)	552 (6.907%)	404 (7.006%)
MRE_M33	1280 (10.839%)	483 (4.988%)	490 (6.142%)	362 (6.290%)
MRE_M36	1333 (11.295%)	678 (7.004%)	658 (8.237%)	482 (8.358%)
MRE_M39	1356 (11.485%)	590 (6.100%)	584 (7.322%)	430 (7.475%)
MRE_M42	1376 (11.650%)	534 (5.520%)	540 (6.759%)	397 (6.892%)
MRE_M45	1398 (11.841%)	561 (5.800%)	564 (7.069%)	415 (7.209%)
MRE_M48	1410 (11.940%)	578 (5.971%)	580 (7.267%)	427 (7.409%)
MRE_M49	1472 (12.470%)	580 (6.000%)	583 (7.300%)	432 (7.500%)

MRE_M3 refers to loan disputes and administrative penalties within 3 months. MRE_M6 through MRE_M48 are similar. MRE_M49 refers to loan disputes and administrative penalties in greater than 48 months

Table 11 Summary statistics of AP in different networks

Risk event	MN	DN	SN	CN
AP_M3	200 (1.687%)	161 (1.660%)	168 (2.097%)	118 (2.042%)
AP_M6	293 (2.482%)	237 (2.455%)	244 (3.062%)	176 (3.045%)
AP_M9	347 (2.943%)	282 (2.907%)	291 (3.654%)	209 (3.627%)
AP_M12	378 (3.202%)	308 (3.179%)	315 (3.952%)	229 (3.982%)
AP_M15	414 (3.508%)	341 (3.521%)	349 (4.367%)	255 (4.417%)
AP_M18	392 (3.323%)	319 (3.303%)	341 (4.270%)	248 (4.302%)
AP_M21	409 (3.463%)	335 (3.464%)	356 (4.463%)	260 (4.520%)
AP_M24	422 (3.569%)	348 (3.595%)	368 (4.607%)	271 (4.704%)
AP_M27	435 (3.685%)	357 (3.687%)	382 (4.788%)	281 (4.872%)
AP_M30	440 (3.734%)	362 (3.737%)	392 (4.907%)	289 (5.006%)
AP_M33	563 (4.766%)	463 (4.788%)	474 (5.942%)	351 (6.090%)
AP_M36	452 (3.831%)	369 (3.809%)	402 (5.037%)	297 (5.158%)
AP_M39	450 (3.806%)	368 (3.800%)	401 (5.022%)	298 (5.175%)
AP_M42	440 (3.731%)	360 (3.720%)	396 (4.959%)	293 (5.092%)
AP_M45	556 (4.714%)	455 (4.700%)	477 (5.969%)	352 (6.109%)
AP_M48	554 (4.688%)	452 (4.671%)	477 (5.967%)	352 (6.109%)
AP_M49	590 (5.002%)	464 (4.800%)	487 (6.100%)	363 (6.300%)

AP_M3 refers to administrative penalties within 3 months. AP_M6 through AP_M48 are similar. AP_M49 refers to administrative penalties in greater than 48 months

Table 12 Summary statistics of LD in different networks

Risk event	MN	DN	SN	CN
LD_M3	194 (1.642%)	170 (1.760%)	176 (2.197%)	123 (2.142%)
LD_M6	307 (2.595%)	242 (2.505%)	248 (3.112%)	179 (3.095%)
LD_M9	372 (3.147%)	283 (2.927%)	293 (3.674%)	210 (3.647%)
LD_M12	423 (3.582%)	317 (3.279%)	323 (4.052%)	235 (4.082%)
LD_M15	475 (4.017%)	350 (3.621%)	357 (4.467%)	260 (4.517%)
LD_M18	520 (4.402%)	377 (3.903%)	389 (4.870%)	282 (4.902%)
LD_M21	557 (4.720%)	402 (4.164%)	412 (5.163%)	301 (5.220%)
LD_M24	590 (5.004%)	426 (4.395%)	432 (5.407%)	317 (5.504%)
LD_M27	622 (5.272%)	444 (4.587%)	454 (5.688%)	332 (5.772%)
LD_M30	651 (5.506%)	459 (4.737%)	472 (5.907%)	346 (6.006%)
LD_M33	672 (5.690%)	473 (4.88%)	482 (6.042%)	357 (6.190%)
LD_M36	692 (5.858%)	485 (5.009%)	498 (6.237%)	366 (6.358%)
LD_M39	705 (5.975%)	493 (5.100%)	505 (6.322%)	373 (6.475%)
LD_M42	707 (5.992%)	495 (5.120%)	508 (6.359%)	374 (6.492%)
LD_M45	721 (6.109%)	503 (5.200%)	517 (6.469%)	381 (6.609%)
LD_M48	733 (6.209%)	510 (5.271%)	524 (6.567%)	387 (6.709%)
LD_M49	796 (6.744%)	532 (5.500%)	543 (6.800%)	403 (7.000%)

LD_M3 refers to loan disputes within 3 months. LD_M6 through LD_M48 are similar. LD_M49 refers to loan disputes in greater than 48 months

Appendix 2: Summary statistics of NEEQ SMEs dataset

See Tables 13, 14, 15, 16 and 17.

Table 13 Descriptive statistics of basic features on NEEQ SMEs dataset

No	Feature	Min	Mean	Max	SD
1	Age (years)	4.000	12.730	53.000	5.090
2	Registered capital (RMB) (Log)	0.699	3.590	6.176	0.414
3	Current ratio (%)	0.025	3.844	670.428	16.419
4	Asset-liability ratio (%)	0.070	40.98	99.82	21.261
5	Receivable turnover ratio (times)	1.930	22.814	5701.690	181.560
6	Inventory turnover ratio (times)	1.980	50.23	2865.59	203.629
7	Assets turnover ratio (times)	0.471	0.921	23.497	0.927
8	Quick ratio (%)	− 18.100	3.026	1323.810	20.651
9	OPR (%)	− 355.439	− 0.300	5.295	6.149
10	ROE (%)	− 749.860	0.851	178.870	35.4221
11	ROA (%)	− 319.480	2.834	90.000	17.396
12	District	{3401–3419}, 18 different districts			
13	Risk event of firms themselves	{0,1}, 0:6431;1:1512			

Table 14 Descriptive statistics of different interfirm networks on NEEQ SMEs dataset

	MN	DN	SN	CN
Average number of individuals per firm	15.732	8.874	12.365	1.000
Average number of neighbor firms per firm	14.991	10.412	6.78	2.461
Average number of firms per individual	1.120	1.291	0.75	3.462

Table 15 Descriptive statistics of LD + AP in different networks on NEEQ SMEs dataset

Risk event	MN	DN	SN	CN
MRE_M3	1017 (0.800%)	635 (8.278%)	432 (7.258%)	110 (2.944%)
MRE_M6	2033 (1.600%)	1360 (17.729%)	864 (14.516%)	192 (5.138%)
MRE_M9	3050 (2.400%)	1994 (25.994%)	1358 (22.816%)	274 (7.332%)
MRE_M12	3939 (3.100%)	2538 (33.086%)	1666 (27.991%)	329 (8.804%)
MRE_M15	4829 (3.800%)	3173 (41.364%)	2036 (34.207%)	411 (10.998%)
MRE_M18	5718 (4.500%)	3807 (49.628%)	2468 (41.465%)	466 (12.470%)
MRE_M21	6608 (5.200%)	4442 (57.906%)	2839 (47.698%)	548 (14.664%)
MRE_M24	7370 (5.800%)	4895 (63.812%)	3147 (52.873%)	603 (16.136%)
MRE_M27	8133 (6.400%)	5348 (69.717%)	3517 (59.089%)	658 (17.608%)
MRE_M30	8895 (7.000%)	5892 (76.809%)	3826 (64.281%)	713 (19.079%)
MRE_M33	9658 (7.600%)	6436 (83.900%)	4134 (69.456%)	767 (20.524%)
MRE_M36	10,293 (8.100%)	6798 (88.619%)	4381 (73.606%)	795 (21.274%)
MRE_M39	10,801 (8.500%)	7161 (93.352%)	4628 (77.755%)	850 (22.746%)
MRE_M42	11,183 (8.800%)	7433 (96.897%)	4813 (80.864%)	850 (22.746%)
MRE_M45	11,437 (9.000%)	7614 (99.257%)	4937 (82.947%)	877 (23.468%)
MRE_M48	11,818 (9.300%)	7886 (102.803%)	5060 (85.013%)	904 (24.191%)
MRE_M49	13,597 (10.700%)	9155 (119.346%)	5739 (96.421%)	1014 (27.134%)

Table 16 Summary statistics of AP in different networks on NEEQ SMEs dataset

Risk event	MN	DN	SN	CN
AP_M3	381 (0.300%)	272 (3.546%)	185 (3.108%)	27 (0.723%)
AP_M6	890 (0.700%)	544 (7.092%)	370 (6.216%)	82 (2.194%)
AP_M9	1271 (1.000%)	816 (10.637%)	555 (9.325%)	110 (2.944%)
AP_M12	1652 (1.300%)	997 (12.997%)	740 (12.433%)	137 (3.666%)
AP_M15	2160 (1.700%)	1360 (17.729%)	987 (16.583%)	192 (5.138%)
AP_M18	2669 (2.100%)	1632 (21.275%)	1234 (20.733%)	219 (5.860%)
AP_M21	3050 (2.400%)	1813 (23.634%)	1419 (23.841%)	274 (7.332%)
AP_M24	3431 (2.700%)	2085 (27.180%)	1604 (26.949%)	301 (8.055%)
AP_M27	3812 (3.000%)	2266 (29.540%)	1728 (29.032%)	329 (8.804%)
AP_M30	4066 (3.200%)	2538 (33.086%)	1913 (32.140%)	356 (9.526%)
AP_M33	4448 (3.500%)	2719 (35.445%)	2098 (35.249%)	384 (10.276%)
AP_M36	4702 (3.700%)	2901 (37.818%)	2221 (37.315%)	411 (10.998%)
AP_M39	4956 (3.900%)	2991 (38.991%)	2283 (38.357%)	411 (10.998%)
AP_M42	4956 (3.900%)	2991 (38.991%)	2283 (38.357%)	439 (11.747%)
AP_M45	4956 (3.900%)	2991 (38.991%)	2283 (38.357%)	439 (11.747%)
AP_M48	4956 (3.900%)	2991 (38.991%)	2283 (38.357%)	439 (11.747%)
AP_M49	4956 (3.900%)	2991 (38.991%)	2345 (39.399%)	439 (11.747%)

Table 17 Summary statistics of LD in different networks on NEEQ SMEs dataset

Risk event	MN	DN	SN	CN
LD_M3	635 (0.500%)	453 (5.905%)	247 (4.150%)	55 (1.472%)
LD_M6	1271 (1.000%)	816 (10.637%)	555 (9.325%)	137 (3.666%)
LD_M9	1779 (1.400%)	1178 (15.357%)	802 (13.474%)	164 (4.389%)
LD_M12	2287 (1.800%)	1450 (18.902%)	987 (16.583%)	192 (5.138%)
LD_M15	2669 (2.100%)	1813 (23.634%)	1172 (19.691%)	219 (5.860%)
LD_M18	3304 (2.600%)	2175 (28.354%)	1419 (23.841%)	274 (7.332%)
LD_M21	3812 (3.000%)	2447 (31.899%)	1666 (27.991%)	301 (8.055%)
LD_M24	4193 (3.300%)	2810 (36.631%)	1913 (32.140%)	329 (8.804%)
LD_M27	4702 (3.700%)	3082 (40.177%)	2098 (35.249%)	356 (9.526%)
LD_M30	5210 (4.100%)	3444 (44.896%)	2345 (39.399%)	384 (10.276%)
LD_M33	5718 (4.500%)	3807 (49.628%)	2592 (43.548%)	411 (10.998%)
LD_M36	6227 (4.900%)	4079 (53.174%)	2715 (45.615%)	439 (11.747%)
LD_M39	6608 (5.200%)	4351 (56.720%)	2900 (48.723%)	466 (12.470%)
LD_M42	6989 (5.500%)	4623 (60.266%)	3085 (51.831%)	466 (12.470%)
LD_M45	7370 (5.800%)	4895 (63.812%)	3270 (54.940%)	493 (13.192%)
LD_M48	7752 (6.100%)	5076 (66.171%)	3394 (57.023%)	521 (13.942%)
LD_M49	9912 (7.800%)	6617 (86.260%)	4196 (70.497%)	658 (17.608%)

Appendix 3: Meta-parameters for each classifier

See Table 18.

Table 18 Meta-parameters for each classifier

Classifier	Meta-parameters
LR	L2 penalty
RF	Base learner = CART with maximum depth of trees = 5
XGB	Nrounds = 30, learning rate = 0.1

Appendix 4: Selection of the optimal hyperparameter subset in different methods

We used the NCV to train models while selecting the optimal hyperparameter subset. In our empirical study, there are 5 kinds of weight function and 17 kinds of time interval of risk events, resulting in 85 (17 × 5) hyperparameter subsets. To get a robust result of discrimination performance, we repeated the inner ten-fold cross-validation procedure ten times, with a different split of folds each time, resulting in 100 performance estimates. As such, the discrimination performance of each model with different hyperparameter subsets is determined by the average of 100 performance estimates. Then, for each model, we ranked their discrimination performance (in terms of AUC) with different hyperparameter subsets, and selected the hyperparameter subset with the best discrimination performance as the optimal hyperparameter subset, as shown in Tables 19 and 20.

Table 19 Optimal hyperparameters of different methods on manufacturing SMEs dataset

Model	LR		RF		XGB	
	Time	Weight function	Time	Weight function	Time	Weight function
BF + RS	M24	SKAA	M3	HT	M48	CDR
BF + CN	M3	SKAA	M42	CDR	M48	HT
BF + DN	M42	SKAA	M36	SKAA	M42	INF
BF + SN	M42	IND	M42	CDR	M18	HT
BF + CNLD	M42	IND	M36	CDR	M48	SKAA
BF + CNAP	M12	SKAA	M12	CDR	M42	SKAA
BF + DNLD	M36	CDR	M36	SKAA	M48	IND
BF + DNAP	M3	INF	M12	SKAA	M3	INF
BF + SNLD	M42	CDR	M18	SKAA	M48	SKAA
BF + SNAP	M42	CDR	M42	CDR	M8	CDR

CDR refers to the weight function class-degree ratio, HT refers to the hyperbolic tangent, IND refers to the inverse degree, INF refers to the inverse frequency, and AA refers to the Adamic and Adar

Table 20 Optimal hyperparameters of different methods on NEEQ SMEs dataset

Model	LR		RF		XGB	
	Time	Weight function	Time	Weight function	Time	Weight function
BF + RS	M27	CDR	M15	CDR	M21	INF
BF + CN	M33	HT	M33	INF	M27	SKAA
BF + DN	M33	INF	M36	CDR	M42	CDR
BF + SN	M33	INF	M27	CDR	M15	HT
BF + CNLD	M33	SKAA	M27	INF	M30	SKAA
BF + CNAP	M33	CDR	M30	CDR	M33	IND
BF + DNLD	M33	SKAA	M33	IND	M42	CDR
BF + DNAP	M33	SKAA	M27	CDR	M33	IND
BF + SNLD	M33	CDR	M30	INF	M33	IND
BF + SNAP	M15	CDR	M24	IND	M18	IND

CDR refers to the weight function class-degree ratio, HT refers to the hyperbolic tangent, IND refers to the inverse degree, INF refers to the inverse frequency, and AA refers to the Adamic and Adar

Appendix 5: Results of non-parametric tests on manufacturing SMEs dataset

We used a non-parametric test to test the statistical significance of the effect of different types of feature set on discrimination performance. Since the Friedman test is rank sum test, the results across classification models (i.e., LR, RF, and XGB) and performance measures (i.e., AUC, KS, and H measure) were put together (i.e., sample size is $100 \times 3 \times 3 = 900$). Tables 21, 22, 23, 24 and 25 show the difference across different feature sets.

Table 21 Results of pairwise comparison (BF vs. BF + RS)

	Average rank	p-value of pairwise comparison BF
BF	1.66	–
BF + RS	1.34	< 0.001
Friedman χ^2	97.018 (< .001)	

Table 22 Results of full pairwise comparison (BF, BF + CN, BF + DN, and BF + SN)

	Average rank	p-value of pairwise comparison		
		BF	BF + CN	BF + DN
BF	2.98	–	–	–
BF + CN	2.43	<0.001	–	–
BF + DN	2.34	<0.001	0.012	–
BF + SN	2.25	<0.001	0.850	0.625
Friedman χ^2	174.014 (<0.001)			

Table 23 Results of full pairwise comparison (BF, BF + CNAP, BF + CNLD, BF + DNAP, BF + DNLD, BF + SNAP, and BF + SNLD)

	Average rank	p-value of pairwise comparison					
		BF	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP
BF	5.07	–	–	–	–	–	–
BF + CNAP	4.40	<0.001	–	–	–	–	–
BF + CNLD	3.76	<0.001	<0.001	–	–	–	–
BF + DNAP	3.60	<0.001	<0.001	1.000	–	–	–
BF + DNLD	3.28	<0.001	<0.001	<0.001	0.037	–	–
BF + SNAP	4.24	<0.001	1.000	<0.001	<0.001	<0.001	–
BF + SNLD	3.65	<0.001	<0.001	1.000	<0.001	0.005	<0.001
Friedman χ^2	433.640 (<0.001)						

Table 24 Results of full pairwise comparison (BF + RS, BF + DN, and BF + DNLD)

	Average rank	p-value of pairwise comparison	
		BF + RS	BF + DN
BF + RS	2.18	–	–
BF + DN	2.07	0.046	–
BF + DNLD	1.74	<0.001	<0.001
Friedman χ^2	94.957 (<0.001)		

Table 25 Results of full pairwise comparison (BF, BF + RS, and BF + prRS)

	Average rank	p-value of pairwise comparison	
		BF	BF + RS
BF	2.27	–	–
BF + RS	1.82	<0.001	–
BF + prRS	1.91	<0.001	0.231
Friedman χ^2	104.004 (<0.001)		

Appendix 6: Different confidence intervals and estimate of each mean of each metric

Tables 26 and 27 show the 95% confidence intervals of each mean of each metric. Tables 28 and 29 show the 90% confidence intervals of each mean of each metric. The metrics in Tables 30 and 31 are estimated based on the mean of 50 estimates, which is derived by repeating the outer ten-fold cross-validation procedure five times.

Table 26 Discrimination performance of different feature sets

Model	Metric	BF	BF + RS	BF + CN	BF + DN	BF + SN
LR	AUC	0.715 ± 0.015	0.726 ± 0.014	0.722 ± 0.014	0.729 ± 0.014	0.725 ± 0.014
	KS	0.408 ± 0.022	0.419 ± 0.022	0.414 ± 0.022	0.431 ± 0.022	0.417 ± 0.021
	H	0.251 ± 0.020	0.270 ± 0.020	0.266 ± 0.020	0.283 ± 0.021	0.269 ± 0.020
RF	AUC	0.725 ± 0.014	0.738 ± 0.013	0.738 ± 0.013	0.741 ± 0.013	0.735 ± 0.013
	KS	0.426 ± 0.023	0.445 ± 0.020	0.444 ± 0.020	0.450 ± 0.021	0.439 ± 0.022
	H	0.265 ± 0.020	0.279 ± 0.018	0.289 ± 0.018	0.294 ± 0.019	0.281 ± 0.019
XGB	AUC	0.695 ± 0.013	0.715 ± 0.013	0.718 ± 0.014	0.722 ± 0.012	0.720 ± 0.014
	KS	0.387 ± 0.021	0.416 ± 0.022	0.415 ± 0.021	0.424 ± 0.018	0.434 ± 0.022
	H	0.219 ± 0.018	0.252 ± 0.019	0.253 ± 0.019	0.245 ± 0.017	0.257 ± 0.021

Each metric is measured in mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times (mean ± standard deviation from a 95% confidence interval)

Table 27 Discrimination performance of different feature sets under 95% confidence intervals

Model	Metric	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP	BF + SNLD
LR	AUC	0.723 ± 0.014	0.729 ± 0.014	0.730 ± 0.013	0.734 ± 0.014	0.725 ± 0.014	0.731 ± 0.014
	KS	0.417 ± 0.021	0.429 ± 0.022	0.427 ± 0.020	0.442 ± 0.022	0.417 ± 0.022	0.425 ± 0.022
	H	0.268 ± 0.020	0.282 ± 0.020	0.269 ± 0.018	0.288 ± 0.021	0.269 ± 0.020	0.278 ± 0.020
RF	AUC	0.736 ± 0.013	0.744 ± 0.013	0.743 ± 0.013	0.747 ± 0.012	0.737 ± 0.014	0.741 ± 0.013
	KS	0.442 ± 0.021	0.457 ± 0.022	0.456 ± 0.021	0.455 ± 0.021	0.444 ± 0.022	0.454 ± 0.023
	H	0.283 ± 0.019	0.296 ± 0.020	0.291 ± 0.018	0.295 ± 0.018	0.282 ± 0.020	0.296 ± 0.021
XGB	AUC	0.713 ± 0.013	0.715 ± 0.012	0.721 ± 0.012	0.732 ± 0.013	0.716 ± 0.013	0.725 ± 0.012
	KS	0.415 ± 0.019	0.410 ± 0.020	0.423 ± 0.019	0.439 ± 0.021	0.415 ± 0.020	0.436 ± 0.020
	H	0.239 ± 0.017	0.250 ± 0.018	0.244 ± 0.016	0.276 ± 0.020	0.244 ± 0.018	0.251 ± 0.018

Each metric are measured in mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times (mean ± standard deviation from a 95% confidence interval)

Table 28 Discrimination performance of different feature sets under 90% confidence intervals

Model	Metric	BF	BF + RS	BF + CN	BF + DN	BF + SN
LR	AUC	0.715 ± 0.012	0.726 ± 0.012	0.722 ± 0.012	0.729 ± 0.012	0.725 ± 0.012
	KS	0.408 ± 0.018	0.419 ± 0.018	0.414 ± 0.018	0.431 ± 0.019	0.417 ± 0.018
	H	0.251 ± 0.017	0.270 ± 0.017	0.266 ± 0.017	0.283 ± 0.017	0.269 ± 0.016
RF	AUC	0.725 ± 0.012	0.738 ± 0.011	0.738 ± 0.011	0.741 ± 0.011	0.735 ± 0.011
	KS	0.426 ± 0.019	0.445 ± 0.017	0.444 ± 0.017	0.450 ± 0.018	0.439 ± 0.018
	H	0.265 ± 0.017	0.279 ± 0.015	0.289 ± 0.015	0.294 ± 0.016	0.281 ± 0.016
XGB	AUC	0.695 ± 0.011	0.715 ± 0.011	0.718 ± 0.011	0.722 ± 0.010	0.720 ± 0.012
	KS	0.387 ± 0.018	0.416 ± 0.018	0.415 ± 0.017	0.424 ± 0.015	0.434 ± 0.018
	H	0.219 ± 0.015	0.252 ± 0.016	0.253 ± 0.016	0.245 ± 0.014	0.257 ± 0.017

Each metric are measured in mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times (mean ± standard deviation from a 90% confidence interval)

Table 29 Discrimination performance of different feature sets under 90% confidence intervals

Model	Metric	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP	BF + SNLD
LR	AUC	0.723 ± 0.012	0.729 ± 0.012	0.730 ± 0.011	0.734 ± 0.012	0.725 ± 0.012	0.731 ± 0.011
	KS	0.417 ± 0.018	0.429 ± 0.019	0.427 ± 0.017	0.442 ± 0.018	0.417 ± 0.018	0.425 ± 0.019
	H	0.268 ± 0.017	0.282 ± 0.017	0.269 ± 0.015	0.288 ± 0.017	0.269 ± 0.017	0.278 ± 0.017
RF	AUC	0.736 ± 0.011	0.744 ± 0.011	0.743 ± 0.011	0.747 ± 0.010	0.737 ± 0.011	0.741 ± 0.011
	KS	0.442 ± 0.018	0.457 ± 0.018	0.456 ± 0.018	0.455 ± 0.017	0.444 ± 0.019	0.454 ± 0.019
	H	0.283 ± 0.016	0.296 ± 0.016	0.291 ± 0.015	0.295 ± 0.015	0.282 ± 0.017	0.296 ± 0.017
XGB	AUC	0.713 ± 0.011	0.715 ± 0.010	0.721 ± 0.010	0.732 ± 0.011	0.716 ± 0.011	0.725 ± 0.010
	KS	0.415 ± 0.016	0.410 ± 0.017	0.423 ± 0.016	0.439 ± 0.017	0.415 ± 0.017	0.436 ± 0.017
	H	0.239 ± 0.014	0.250 ± 0.015	0.244 ± 0.014	0.276 ± 0.017	0.244 ± 0.015	0.251 ± 0.015

Each metric are measured in mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times (mean ± standard deviation from a 90% confidence interval)

Table 30 Discrimination performance of different feature sets

Model	Metric	BF	BF + RS	BF + CN	BF + DN	BF + SN
LR	AUC	0.717	0.727	0.723	0.731	0.726
	KS	0.410	0.420	0.416	0.435	0.419
	H	0.250	0.271	0.265	0.286	0.268
RF	AUC	0.726	0.739	0.741	0.746	0.737
	KS	0.432	0.449	0.452	0.455	0.439
	H	0.274	0.282	0.452	0.302	0.287
XGB	AUC	0.699	0.717	0.718	0.727	0.724
	KS	0.395	0.415	0.415	0.431	0.442
	H	0.228	0.253	0.253	0.248	0.260

Each metric is measured in mean of 50 estimates obtained by repeating the outer ten-fold cross-validation procedure five times

Table 31 Discrimination performance of different feature sets

Model	Metric	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP	BF + SNLD
LR	AUC	0.725	0.731	0.732	0.736	0.727	0.728
	KS	0.421	0.431	0.427	0.443	0.420	0.422
	H	0.271	0.283	0.271	0.288	0.270	0.273
RF	AUC	0.739	0.747	0.739	0.749	0.738	0.745
	KS	0.448	0.467	0.442	0.467	0.446	0.463
	H	0.288	0.303	0.292	0.305	0.283	0.299
XGB	AUC	0.718	0.717	0.726	0.739	0.719	0.723
	KS	0.420	0.420	0.440	0.449	0.420	0.431
	H	0.249	0.252	0.252	0.292	0.252	0.257

Each metric is measured in mean of 50 estimates obtained by repeating the outer ten-fold cross-validation procedure five times

Appendix 7: Analysis of results in LR and RF

Table 32 shows the coefficients of features in LR. Figure 7 shows the Mean Decrease in Gini of the relational risk score in RF. The meta-parameters for each classifier in our empirical study are shown in Table 33.

Table 32 Coefficients of features in the logistic regression model

	Estimate	Std. error	z-value	Pr(> z)
(Intercept)	-4.220	1.250	-3.376	0.001***
Current ratio (%)	-0.084	0.384	-0.218	0.828
Debt to assets ratio (%)	0.016	0.321	0.049	0.961
Receivable turnover ratio (days)	-0.097	0.054	-1.782	0.075
Inventory turnover ratio (days)	-0.035	0.043	-0.804	0.421
Assets turnover ratio (days)	-0.548	0.185	-2.966	0.003**
Quick ratio (%)	-0.373	0.338	-1.102	0.270
OPR (%)	-1.415	0.735	-1.926	0.054
ROE (%)	3.983	4.653	0.856	0.392
ROA (%)	-9.398	5.985	-1.57	0.116
Log (registered capital)	0.374	0.123	3.037	0.002**
District 3402	-0.009	0.696	-0.013	0.990
District 3403	-1.760	0.740	-2.377	0.017*
District 3404	-0.492	0.702	-0.7	0.484
District 3405	-0.805	0.789	-1.021	0.307
District 3406	-0.664	0.774	-0.857	0.391
District 3408	-0.105	0.754	-0.139	0.889
District 3409	-0.256	0.752	-0.34	0.734
District 3410	-1.311	0.868	-1.509	0.131
District 3411	-0.891	0.781	-1.14	0.254
District 3412	-0.794	0.722	-1.099	0.272
District 3413	-1.094	0.740	-1.478	0.139
District 3414	-1.599	0.849	-1.883	0.060
District 3415	-0.967	0.788	-1.228	0.219
District 3416	-1.953	0.874	-2.233	0.026*
District 3417	-1.636	0.869	-1.882	0.060
Age	0.015	0.021	0.718	0.473
Risk events of firms themselves	0.653	0.252	2.591	0.010**
Relational risk	6.430	2.152	2.987	0.003**

***refers to the significance at the 0.001 level; **refers to the significance at the 0.005 level; *refers to the significance at the 0.05 level; refers to the significance at the 0.01 level

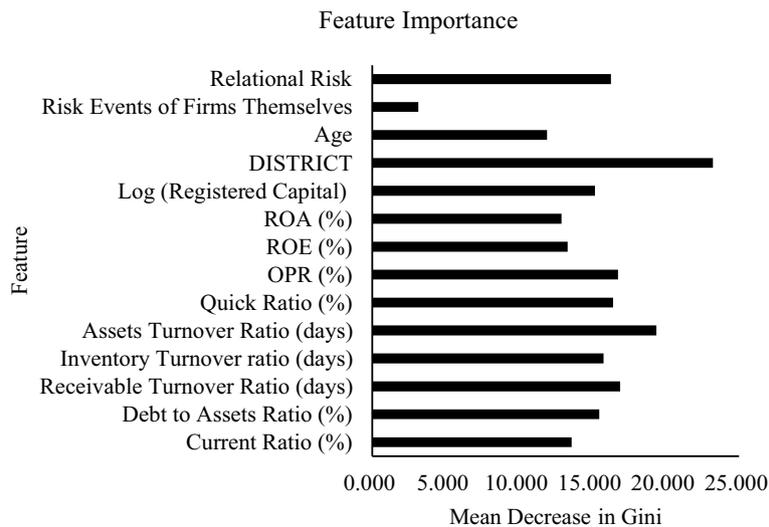


Fig. 7 Random forest feature importance. A greater mean decrease in Gini indicates a more important feature

Appendix 8: Robustness analyses

We conducted three sets of robustness tests to ensure that our results are not affected by (1) dataset selection bias; (2) data coverage bias; and (3) exclusion bias. Below are main results of robustness tests. Interested readers can contact us for more detailed descriptions of test procedures and results.

For the first robust test, Tables 33 and 34 summarize the discrimination performance (in terms of AUC, KS, and H measure) of the three classification models (LR, RF, and XGB) using 11 types of feature sets: (1) BF, (2) BF + RS, (3) BF + CN, (4) BF + DN, (5) BF + SN, (6) BF + CNAP, (7) BF + CNLD, (8) BF + DNAP, (9) BF + DNLD, (10) BF + SNAP, and (11) BF + SNLD. Table 35 shows the result of full pairwise comparisons between different methods on NEEQ SMEs dataset. Figure 8 shows the granting performance of the three classification models using two types of feature sets: (1) BF; (2) BF + DNLD.

Table 33 Discrimination performance of different feature sets on NEEQ SMEs dataset

Model	Metric	BF	BF + RS	BF + CN	BF + DN	BF + SN
LR	AUC	0.801	0.804	0.803	0.806	0.805
	KS	0.533	0.541	0.537	0.540	0.538
	H	0.419	0.428	0.421	0.422	0.423
RF	AUC	0.840	0.846	0.847	0.855	0.849
	KS	0.565	0.579	0.573	0.599	0.583
	H	0.451	0.466	0.467	0.487	0.475
XGB	AUC	0.826	0.835	0.831	0.846	0.838
	KS	0.563	0.572	0.565	0.599	0.577
	H	0.436	0.446	0.445	0.482	0.451

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times

Table 34 Discrimination performance of different feature sets on NEEQ SMEs dataset

Model	Metric	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP	BF + SNLD
LR	AUC	0.801	0.805	0.803	0.808	0.802	0.807
	KS	0.535	0.543	0.534	0.546	0.534	0.545
	H	0.421	0.432	0.421	0.432	0.420	0.430
RF	AUC	0.841	0.849	0.846	0.865	0.843	0.861
	KS	0.566	0.582	0.579	0.618	0.571	0.611
	H	0.454	0.463	0.468	0.512	0.456	0.497
XGB	AUC	0.829	0.832	0.834	0.852	0.832	0.840
	KS	0.566	0.572	0.573	0.598	0.566	0.581
	H	0.442	0.449	0.462	0.488	0.444	0.460

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times

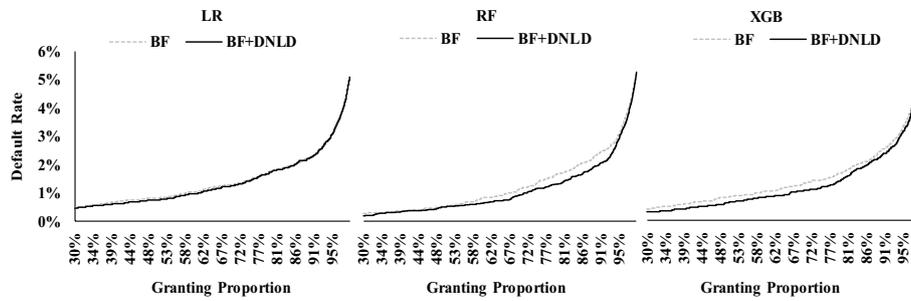


Fig. 8 Granting performance of two models on NEEQ SMEs Dataset. The horizontal axis represents different granting proportions of loans ranging from 30 to 95%. The vertical axis is the default rate ranging from 0 to 5.1%

For the second robust test, Tables 36 and 37 summarize the discrimination performance (in terms of AUC, KS, and H measure) of the three classification models (LR, RF, and XGB) using 11 types of feature sets: (1) BF, (2) BF + RS, (3) BF + CN, (4) BF + DN, (5) BF + SN, (6) BF + CNAP, (7) BF + CNLD, (8) BF + DNAP, (9) BF + DNLD, (10) BF + SNAP, and (11) BF + SNLD.

Table 36 Discrimination performance of different feature sets

Model	Metric	BF	BF + RS	BF + CN	BF + DN	BF + SN
LR	AUC	0.792	0.796	0.797	0.801	0.799
	KS	0.526	0.526	0.529	0.535	0.534
	H	0.424	0.420	0.423	0.429	0.425
RF	AUC	0.842	0.849	0.847	0.853	0.850
	KS	0.568	0.586	0.577	0.591	0.588
	H	0.455	0.468	0.461	0.477	0.480
XGB	AUC	0.826	0.837	0.839	0.846	0.843
	KS	0.563	0.582	0.583	0.591	0.590
	H	0.437	0.454	0.462	0.474	0.472

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times

Table 37 Discrimination performance of different feature sets

Model	Metric	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP	BF + SNLD
LR	AUC	0.797	0.801	0.800	0.805	0.798	0.803
	KS	0.528	0.541	0.529	0.539	0.528	0.540
	H	0.421	0.432	0.422	0.427	0.422	0.433
RF	AUC	0.844	0.862	0.851	0.868	0.846	0.861
	KS	0.573	0.607	0.587	0.622	0.572	0.606
	H	0.458	0.496	0.469	0.517	0.459	0.495
XGB	AUC	0.830	0.848	0.832	0.858	0.828	0.850
	KS	0.569	0.592	0.569	0.612	0.561	0.598
	H	0.443	0.475	0.442	0.498	0.437	0.478

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times

For the third robust test, Tables 38 and 39 summarize the discrimination performance (in terms of AUC, KS, and H measure) of the three classification models (LR, RF, and XGB) using 11 types of feature sets: (1) BF, (2) BF + RS, (3) BF + CN, (4) BF + DN, (5) BF + SN, (6) BF + CNAP, (7) BF + CNLD, (8) BF + DNAP, (9) BF + DNLD, (10) BF + SNAP, and (11) BF + SNLD.

Table 38 Discrimination performance of different feature sets

Model	Metric	BF	BF + RS	BF + CN	BF + DN	BF + SN
LR	AUC	0.716	0.720	0.720	0.724	0.717
	KS	0.410	0.419	0.424	0.427	0.416
	H	0.252	0.267	0.269	0.277	0.261
RF	AUC	0.725	0.734	0.735	0.740	0.732
	KS	0.429	0.439	0.439	0.446	0.435
	H	0.263	0.280	0.284	0.290	0.276
XGB	AUC	0.708	0.719	0.713	0.726	0.718
	KS	0.407	0.426	0.411	0.431	0.420
	H	0.232	0.252	0.246	0.262	0.253

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times

Table 39 Discrimination performance of different feature sets

Model	Metric	BF + CNAP	BF + CNLD	BF + DNAP	BF + DNLD	BF + SNAP	BF + SNLD
LR	AUC	0.722	0.727	0.723	0.729	0.720	0.724
	KS	0.424	0.432	0.424	0.428	0.417	0.421
	H	0.268	0.280	0.279	0.267	0.265	0.267
RF	AUC	0.734	0.741	0.738	0.746	0.734	0.737
	KS	0.438	0.454	0.445	0.457	0.441	0.442
	H	0.284	0.293	0.283	0.304	0.278	0.282
XGB	AUC	0.709	0.719	0.721	0.732	0.718	0.723
	KS	0.407	0.421	0.422	0.440	0.430	0.424
	H	0.232	0.241	0.260	0.272	0.247	0.258

Each metric is measured by the mean of 100 estimates obtained by repeating the outer ten-fold cross-validation procedure ten times

Abbreviations

- SMEs Small and medium-sized enterprises
- wvRN Weighted-vote relational neighbor
- LDA Linear discriminant analysis
- LR Logistic regression
- MDA Multivariate discriminant analysis
- RF Random forest
- XGB EXtreme gradient boosting
- ROE Return on equity
- ROA Return on assets
- CEO Chief Executive Officer
- DSS Directors, Supervisors, and Senior Management
- CN CEO network
- DN DSS network
- SN Shareholder network
- AP Administrative penalty
- LD Loan disputes
- CNAP CEO network and administrative penalties
- CNLD CEO network and loan disputes

DNAP	DSS network and administrative penalties
DNLD	DSS network and loan disputes
SNAP	Shareholder network and administrative penalties
SNLD	Shareholder network and loan disputes
NCV	Nested cross-validation

Acknowledgements

We would like to thank the National Natural Science Foundation of China, which financially supported this research (Grant Nos. 71731005 and Nos.72101073).

Author contributions

JL contributes to the acquisition and analysis of data, and drafted the work. CJ contributes to the conception of the work and the interpretation of data. SD contributes to the design of the work and drafted the work. ZW contributes to the design of the work and the creation of new software used in the work. All authors read and approved the final manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (Grant Nos. 71731005 and Nos.72101073).

Availability of data and materials

The data that support the findings of this study are available from a local commercial bank but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the local commercial bank.

Declarations

Competing interests

The authors declared that they have no competing interests.

Received: 3 February 2022 Accepted: 14 September 2022

Published online: 24 October 2022

References

- Abdou HA, Tsafack MDD, Ntim CG, Baker RD (2016) Predicting creditworthiness in retail banking with limited scoring data. *Knowl Based Syst* 103:89–103. <https://doi.org/10.1016/j.knosys.2016.03.023>
- Abellán J, Castellano JG (2017) A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Syst Appl* 73:1–10. <https://doi.org/10.1016/j.eswa.2016.12.020>
- Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J Financ* 23(4):589–609. <https://doi.org/10.2307/2978933>
- Altman EI, Sabato G, Wilson N (2010) The value of non-financial information in small and medium-sized enterprise risk management. *J Credit Risk* 6(2):95–127. <https://doi.org/10.1074/jbc.275.13.9563>
- Angilella S, Mazzù S (2015) The financing of innovative SMEs: a multicriteria credit rating model. *Eur J Oper Res* 244(2):540–554. <https://doi.org/10.1016/j.ejor.2015.01.033>
- Bai C, Shi B, Liu F, Sarkis J (2019) Banking credit worthiness: evaluating the complex relationships. *Omega* 83:26–38. <https://doi.org/10.1016/j.omega.2018.02.001>
- Beaver WH, Cascino S, Correia M, McNichols MF (2019) Group affiliation and default prediction. *Manag Sci* 65(8):3559–3584. <https://doi.org/10.1287/mnsc.2018.3128>
- Berger AN, Udell MS (2007) Small business credit scoring and credit availability. *J Small Bus Manag* 45(1):5–22. <https://doi.org/10.1111/j.1540-627X.2007.00195.x>
- Buldirev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464(7291):1025–1028. <https://doi.org/10.1038/nature08932>
- Calabrese R, Andreeva G, Ansell J (2019) “Birds of a feather” fail together: exploring the nature of dependency in SME defaults. *Risk Anal* 39(1):71–84. <https://doi.org/10.1111/risa.12862>
- Cassar G, Ittner CD, Cavalluzzo KS (2015) Alternative information sources and information asymmetry reduction: evidence from small business debt. *J Acc Econ* 59(2–3):242–263. <https://doi.org/10.1016/j.jacceco.2014.08.003>
- Dierkes M, Erner C, Langer T, Norden L (2013) Business credit information sharing and default risk of private firms. *J Bank Finance* 37(8):2867–2878. <https://doi.org/10.1016/j.jbankfin.2013.03.018>
- Djeundje VB, Crook J, Calabrese R, Hamid M (2021) Enhancing credit scoring with alternative data. *Expert Syst Appl* 163:113766. <https://doi.org/10.1016/j.eswa.2020.113766>
- do Prado JW, de Castro Alcântara V, de Melo Carvalho F, Vieira KC, Machado LKC, Tonelli DF (2016) Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968–2014). *Scientometrics* 106(3):1007–1029. <https://doi.org/10.1007/s11192-015-1829-6>
- Elliott M, Golub B, Jackson MO (2014) Financial networks and contagion. *Am Econ Rev* 104(10):3115–3153. <https://doi.org/10.1257/aer.104.10.3115>
- Fernandes GB, Artes R (2016) Spatial dependence in credit risk and its improvement in credit scoring. *Eur J Oper Res* 249(2):517–524. <https://doi.org/10.1016/j.ejor.2015.07.013>
- Figini S, Bonelli F, Giovannini E (2017) Solvency prediction for small and medium enterprises in banking. *Decis Support Syst* 102:91–97. <https://doi.org/10.1016/j.dss.2017.08.001>

- Fosu S, Danso A, Agyei-Boapeah H, Ntim C, Adegbite E (2020) Credit information sharing and loan default in developing countries: the moderating effect of banking market concentration and national governance quality. *Rev Quant Financ Acc* 55:55–103. <https://doi.org/10.1007/s11156-019-00836-1>
- Garrido F, Verbeke W, Bravo C (2018) A robust profit measure for binary classification model evaluation. *Expert Syst Appl* 92:154–160. <https://doi.org/10.1016/j.eswa.2017.09.045>
- Giesecke K, Weber S (2004) Cyclical correlations, credit contagion, and portfolio losses. *J Bank Finance* 28(12):3009–3036. <https://doi.org/10.1016/j.jbankfin.2003.11.002>
- Giesecke K, Weber S (2006) Credit contagion and aggregate losses. *J Econ Dyn Control* 30(5):741–767. <https://doi.org/10.1016/j.jedc.2005.01.004>
- Grant E, Yung J (2021) The double-edged sword of global integration: Robustness, fragility, and contagion in the international firm network. *J Appl Econom* 36(6):760–783. <https://doi.org/10.1002/jae.2839>
- Gunnarsson BR, vanden Broucke S, Baesens B, Óskarsdóttir M, Lemahieu W (2021) Deep learning for credit scoring: Do or don't? *Eur J Oper Res* 295(1):292–305. <https://doi.org/10.1016/j.ejor.2021.03.006>
- Hand DJ (2009) Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach Learn* 77(1):103–123. <https://doi.org/10.1007/s10994-009-5119-5>
- Henninger M (2013) The value and challenges of public sector information. *Cosmopolit Civ Soc Interdiscip J* 5(3):75–95. <https://doi.org/10.5130/ccs.v5i3.3429>
- Jiang C, Lyu X, Yuan Y, Wang Z, Ding Y (2021) Mining semantic features in current reports for financial distress prediction: empirical evidence from unlisted public firms in China. *Int J Forecast*. <https://doi.org/10.1016/j.ijforecast.2021.06.011>
- Kotsiantis S, Kanellopoulos D, Pintelas P (2006) Handling imbalanced datasets: a review. *Science* 30(1):25–36. https://doi.org/10.1007/978-0-387-09823-4_45
- 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 multi-objective feature selection. *Decis Support Syst* 140:113429. <https://doi.org/10.1016/j.dss.2020.113429>
- Lee JW, Lee WK, Sohn SY (2021) Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Expert Syst Appl* 168:114411. <https://doi.org/10.1016/j.eswa.2020.114411>
- Lessmann S, Baesens B, Seow H, Thomas LC (2015) Benchmarking state-of-the-art classification algorithms for credit. *Eur J Oper Res*. <https://doi.org/10.1016/j.ejor.2015.05.030>
- Letizia E, Lillo F (2019) Corporate payments networks and credit risk rating. *EPJ Data Sci* 8(1):21. <https://doi.org/10.1140/epjds/s13688-019-0197-5>
- Liang D, Lu CC, Tsai CF, Shih GA (2016) Financial ratios and corporate governance indicators in bankruptcy prediction: a comprehensive study. *Eur J Oper Res* 252(2):561–572. <https://doi.org/10.1016/j.ejor.2016.01.012>
- Maldonado S, Pérez J, Bravo C (2017) Cost-based feature selection for support vector machines: an application in credit scoring. *Eur J Oper Res* 261(2):656–665. <https://doi.org/10.1016/j.ejor.2017.02.037>
- Medvedeva M, Vols M, Wieling M (2020) Using machine learning to predict decisions of the European Court of Human Rights. *Artif Intell Law* 28(2):237–266. <https://doi.org/10.1007/s10506-019-09255-y>
- Murro P, Peruzzi V (2019) Family firms and access to credit. Is family ownership beneficial? *J Bank Finance* 101:173–187. <https://doi.org/10.1016/j.jbankfin.2019.02.006>
- OECD (2016) Financing SMEs and Entrepreneurs 2016: an OECD Scoreboard. OECD Publishing, Paris. https://doi.org/10.1787/fin_sme_ent-2016-en
- OECD (2020) Coronavirus (COVID-19): SME policy responses. OECD Publishing, Paris. <https://doi.org/10.1787/04440101-en>
- Orton P, Ansell J, Andreeva G (2015) Exploring the performance of small-and medium-sized enterprises through the credit crunch. *J Oper Res Soc* 66(4):657–663. <https://doi.org/10.1057/jors.2014.34>
- Óskarsdóttir M, Bravo C (2021) Multilayer network analysis for improved credit risk prediction. *Omega* 105:102520. <https://doi.org/10.1016/j.omega.2021.102520>
- Saruni LA, Koori J (2020) Credit information sharing and default rate of loans issued by commercial banks listed at the Nairobi securities exchange. *J Field Act* 4(5):1–24
- Sigrist F, Hirsenschall C (2019) Grabit: gradient tree-boosted Tobit models for default prediction. *J Bank Finance* 102:177–192. <https://doi.org/10.1016/j.jbankfin.2019.03.004>
- Stankova M, Praet S, Martens D, Provost F (2020) Node classification over bipartite graphs through projection. *Mach Learn*. <https://doi.org/10.1007/s10994-020-05898-0>
- Stiglitz JE, Weiss A (1981) Credit rationing in markets with imperfect information. *Am Econ Rev* 71(3):393–410
- Van Vlasselaer V, Eliassi-Rad T, Akoglu L, Snoeck M, Baesens B (2016) GOTCHA! Network-based fraud detection for social security fraud. *Manag Sci* 63(9):3090–3110. <https://doi.org/10.1287/mnsc.2016.2489>
- Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinform* 7(1):1–8. <https://doi.org/10.1186/1471-2105-7-91>
- Veganzones D, Séverin E (2018) An investigation of bankruptcy prediction in imbalanced datasets. *Decis Support Syst* 112:111–124. <https://doi.org/10.1016/j.dss.2018.06.011>
- Wang L, Li S, Wu C (2019) Risk contagion in inter-firm credit guarantee network. *Phys A* 526:120842. <https://doi.org/10.1016/j.physa.2019.04.078>
- Wang Z, Jiang C, Zhao H, Ding Y (2020a) Mining semantic soft factors for credit risk evaluation in peer-to-peer lending. *J Manag Inf Syst* 37(1):282–308. <https://doi.org/10.1080/07421222.2019.1705513>
- Wang K, Chen Y, Liu Y, Tang Y (2020b) Board secretary's financial experience, overconfidence, and SMEs' financing preference: evidence from China's NEEQ market. *J Small Bus Manag*. <https://doi.org/10.1080/00472778.2020.1838177>
- Wojewodzki M, Poon WPH, Shen J (2018) The role of credit ratings on capital structure and its speed of adjustment: an international study. *Eur J Finance* 24(9):735–760. <https://doi.org/10.1080/1351847X.2017.1354900>
- Wu J, Zhang Z, Zhou SX (2022) Credit rating prediction through supply chains: a machine learning approach. *Prod Oper Manag* 31(4):1613–1629. <https://doi.org/10.1111/poms.13634>

- Yin C, Jiang C, Jain HK, Wang Z (2020) Evaluating the credit risk of SMEs using legal judgments. *Decis Support Syst* 136:113364. <https://doi.org/10.1016/j.dss.2020.113364>
- Zha Q, Kou G, Zhang H, Liang H, Chen X, Li CC, Dong Y (2020) Opinion dynamics in finance and business: a literature review and research opportunities. *Financ Innov* 6(1):1–22. <https://doi.org/10.1186/s40854-020-00211-3>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
