Received 2 June 2020
Revised 14 September 2020
Accepted 4 October 2020

Application of text mining in identifying the factors of supply chain financing risk management

Hao Ying
China Europe International Business School, Shanghai Jiaotong University, Shanghai, China
Lujie Chen
International Business School at Suzhou (IBSS), Xi'an Jiaotong-Liverpool University, Suzhou, China, and
Xiande Zhao
Department of Economics and Decision Sciences, China Europe International Business School, Shanghai, China

Abstract

Purpose – This study aims to clarify the risk management practices of banks as supply chain finance (SCF) service providers.

Design/methodology/approach – Using 4,014 evaluation and approval reports, this study constructed five risk management factors and examined their functions with secondary data. Two text-mining techniques (i.e. word sense induction, TF–IDF) were used to equip the classic routine of dictionary-based content analysis.

Findings – This research successfully identified four important risk management factors: relationship-based assessment, asset monitoring, cash flow monitoring and supply chain collaboration. The default-preventing effect of these factors are different and contingent on the type of financing contexts (i.e. preshipment, postshipment).

Practical implications – The empirical evidences provide practical implications for SCF service providers to manage risk. SCF service providers are suggested to pay more attention to cash flow monitoring when providing postshipment financing services and shift the focus to relationship building and supply chain collaboration when providing preshipment financing services.

Originality/value – The study shows that a large volume of textual materials can provide adequate clues for researches as long as they are mined with suitable analytic techniques and approaches. Based on the results, SCF service providers can identify problems of their operations and directions for improvement. In addition, the risk management vocabulary from the E&A reports can be utilized by SCF service providers to digitize their loan approving process and, further, to facilitate the decision-makings.

Keywords Risk management, Supply chain finance, Computer-aided text analysis

Paper type Research paper

1. Introduction

Supply chain finance (SCF) is a collective term for a series of financial services which aims to resolve the mismatch between cash and physical flows in supply chain (SC) operations (Wuttke *et al.*, 2013). Over the past decade, banks have been the primary initiator of SCF service, and the bank-intermediated mode has become dominant in the market. Despite the fact that this mode is well-established, banks still face two key obstacles when promoting SCFservices. First, the final debtors of SCF are often small- and medium-sized enterprises (SMEs) that suffer from information opaqueness and collateral deficiency (Hofmann and Belin, 2011). Second, banks lack adequate knowledge and experience about SC operations (Ma *et al.*, 2020). Facing with the obstacles, the banking sector ought to have unique business intelligence to work with so that their SCF services can be sustainable and profitable.

Industrial Management & Data Systems
Vol. 121 No. 2, 2021
pp. 498-518

0263-5577
DOI 10.1108/IMDS-06-2020-0325

However, the resource of this business intelligence which has supported the SCF market for years is underexplored. To the best of our knowledge, extant literature discussed operation cash flows and financial service but does not provide many insights about the interplay between the two lines. A big unknown remains in what happens in the period between a SCF application and credit approval: (1) With what themes should banks be concerned when evaluating an SCF application? and (2) To what extent can the combination of these themes prevent the occurrence of defaults? Without clear answers to such questions, the uncertainties in loan approval will hamper the extensive adoption of SCF services.

Partnering with a leading bank in China, this study investigates the decision-making processes by analyzing 4,014 evaluation and approval (E&A) reports. These reports record loan officers' idea regarding prospective SCF borrowers including assessing information and monitoring rules. In order to answer the research questions, a computer-aided text analysis (CATA) methodology is developed to quantify variables of interest (risk management factors) from the unstructured text. Specifically, we constructed a special-purpose dictionary with word sense induction (WSI) method, which combines a word2vecand *k*-means algorithms. Then, TF–IDF model is employed to measure the emphasis on the themes. The follow-up exploratory factor analysis and econometric analysis successively develop risk management factors and examine the function of the factors for loan default prevention in different SCF contexts.

Regarding the characteristics of classical CATA methods (i.e., dictionary, supervised, unsupervised), we believe some existing limitations of are overcome in our methodology. To be specific, the dictionary method which relies on the rules (word categories) created by experts often has a high validity. However, the dictionary construction process is costly; and timely updates are needed whenever a scenario changes. Supervised learning method can identify themes of interest outside the domain the rules were originally developed. However, the performance of the method is determined by the scale of training set. Especially for an underexplored scenario, building a large and high-quality training set is undoubtedly a resource-consuming task.In contrast, the rule creation cost of unsupervised learning method is lower than the above methods. But the absolute unsupervised process makes researchers stay away from data, leading to lower validity. To deal with the research scenario in which there are no well-defined rules or sufficient labeled data, our CATA methodology assigns some tasks of classical dictionary construction routine to unsupervised machine learning algorithm. On one hand, the unsupervised learning can automatically induce the semantic categories from a huge volume of corpus within a short time. On the other hand, the manual interpretation and adjustment designed in an appropriate way is expected to ensure the validity of analytical results.

The remainder of this paper is structured as follows: in Section 2, we document the literature on credit risk management and the particularity of SCF programs; in Section 3 , we introduce the CATA methodology for analyzing the textual content of E&A reports; in Section 4, we employ an exploratory factor analysis (EFA) and a regression analysis to obtain additional empirical findings. Finally, in Section 5 we conclude our study and present the practical implications and limitations of our research.

2. Related literature

2.1 Two actions of credit risk management

The large body of literature on credit risk management provides the theoretical framework for this study. As a classical framework, risk management arrangements can be categorized into two umbrella types: assessment and monitoring. Assessment is conducted to prevent adverse selection in loan propositions. Monitoring is carried out to eliminate a borrower's opportunistic behavior or moral hazard.

The literature clarifies two forms of information that underlie assessment practices. Hard information is always quantitative, formal information. Financial information is commonly assessed as "hard" factors that influence the credit availability of SMEs. Through accounting vouchers and financial reports, assessors can understand a company based on the size of its assets, profitability, debt ratio, liquidity ratio, working capital turnover and other indices (Zhu *et al.*, 2019). Focusing on some numerical information about SC partnership, Lyu and Zhao (2019) and Hung *et al.* (2020) demonstrate the potential use of big data analytics to improve the risk assessment of SCF. As a complement, soft information spreads via cognitive procedures in a qualitative form. This information may include data on the ability and experience of company owners, operational performance and even prospects and can be gathered from firm owners, employees and neighboring communities (Degryse and Van Cayseele, 2000). Since soft information is crucial for disclosing the truth, it is considered among the main factors influencing SME credit risk (Grunert *et al.*, 2005).

However, assessors cannot predict all future events. Some tailor-made rules are needed to regulate post-loan behaviors so that borrowers are less likely to harm lenders' interests (Benmelech and Bergman, 2009). As a traditional resource of monitoring, fixed assets in SME business loans can be provided by either a firm or a firm's owner, in forms such as real estate, motor vehicles and equipment (Niinimaki, 2016). As an additional resource for monitoring, inventory and accounts receivable are studied by more researches as working capital collateral (Chen *et al.*, 2017; Li *et al.*, 2020). To monitor collaterals effectively and guarantee their liquidation value ,the practice should rely heavily on contractual regulations that formulate the delivery of goods, information flow and the negotiable instruments of financial flows (Hofmann and Belin, 2011). Second, banks must understand the mechanism of internal control and design integrated process mapping to track and control the exchange of collateral (Jia *et al.*, 2020). Third, efficient collateral monitoring requires comprehensive logistics service and advanced inventory management technology (Chakuu *et al.*, 2020; Wang *et al.*, 2019).

2.2 Bank-intermediated supply chain finance

SCF is an interorganizational optimization service that integrates the processes of financial flow management among SCs to mitigate against the working capital constraints of buyers or suppliers (Wuttke *et al.*, 2013). Bank-intermediated mode is distinct from trade credit mode in that it centers on multilateral agreements that include at least two SC members (e.g., a focal firm and a supplier or buyer), and one bank (Chen, 2015). As clarified in Figure 1, the bank can provide low-cost funds for traditional SC transactions by anchoring on the high credit score of the focal firm. By matching the physical and cash flows in the two contexts, SCF services with bank credit can be classified into two types: postshipment finance and preshipment finance. Corresponding to Wuttke *et al.* (2013)'s concept of supply chain financial context, these two types of bank credit arrangements cover different SC processes and elements; thus they need different strategies for risk management.

In postshipment finance, funds are remitted to suppliers once their goods have been delivered and approved by the focal firm. The bank can grant credit to weak suppliers based either on pricing or, partially, on the credit rate of specific creditworthy buyers who are more transparent and creditworthy than the others. Among the typically used instruments are factoring and reverse factoring, which are based on accounts outstanding between a weak supplier and a strong buyer. Factoring is a two-side agreement via which the supplier can independently sell its accounts receivables to a bank (Klapper, 2006). Reverse factoring is initiated by a creditworthy buyer, who guarantees the credit if the bank makes early payment of a buyer's trade obligation (Vliet *et al.*, 2015). Song *et al.* (2019) investigate the innovative mode of accounts receivable pool and leverage real-time invoicing information for improving the risk management capability of banks.

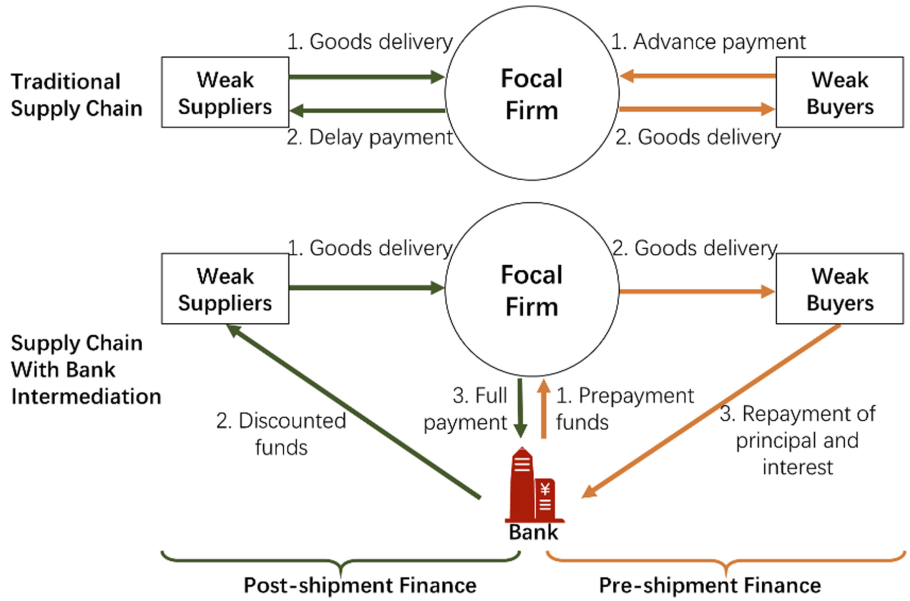


Figure 1.
Two types of bank SCF program

Conversely, preshipment finance is arranged in an early period, prior to goods delivery. Zhao and Huchzermeier (2019) argue that the purchase order is the only eligible collateral. In the ordering process, a bank and the transaction counterparties enter into a tripartite agreement so that the bank credit facilitates prepayments or deposits (Cao *et al.*, 2019). Considering a dyadic SC with a strong supplier and a weak buyer, Chen *et al.* (2017) introduce the concept of the buyback guarantee, with which the core manufacturer promises to buy back unsold goods to help its capital-constrained retailers rapidly secure loans. Lin and He (2019) compare buy-back contract with price discount contract, and conclude that the former is a better financing strategy under uncertain market demand. As a supplement, Lin *et al.* (2018) investigate the method of confirming warehouse financing. They claim that the bank can require the ownership of goods with a provision in the agreement for buyback of unsold goods to effectively prevent the risk of default.

2.3 Dictionary-based content analysis

Developing in line with information-era changes, modern content analysis has been equipped with computer technology to explore larger volume texts or namely computer-aided text analysis or CATA (Tangpong, 2011). The dictionary approach of CATA has a wide application due to its simplicity, reliability and efficiency (Bao and Datta, 2014). The underlying idea of the approach is that texts can be represented by different semantic categories which distribute in the body in varying levels of significance.

Previous studies have generated various well-established dictionaries, such as LIWC (Chung and Pennebaker, 2012), SentiWordNet (Agarwal and Mittal, 2016) and SenticNet (Hung, 2017). However, the sharing of dictionaries across disciplines always leads to misleading results. In order to avoid the problem caused by dictionary mismatch, some scholars tailor existing dictionaries for specific research purposes (e.g., Somers and Casal, 2017; Strauß *et al.*, 2016). However, a special-purpose dictionary is sometimes necessary to fit the specific research purpose because analysis results are restricted by the predefined categories of the dictionary.

In order to arrange a special-purpose dictionary, prior literature has examined different methods in different contexts. These works coincidentally show that the WSI method outperforms other alternative methods in terms of the accuracy of theme distinguishing. For example, Lin *et al.* (2016) use WSI method with two theme models (i.e., PLSA and LDA) to discover the themes from news comments. With WSI method, Zhao *et al.* (2019) construct a financial sentiment dictionary which is specific to Chinese market analytics. Kim *et al.* (2019) utilize the similar method to train sentences classifiers and generate a dictionary specific to SWOT and PEST analytics. More recently, Kim *et al.* (2020) propose a systematic procedure of WSI method (W2V-LSA) and identify the thematic evolution in blockchain field. According to these studies, the algorithms used in WSI (i.e., skip-gram, *k*-means) quantifies not only lexical features but also syntactic and semantic features, thus fit the requirement of a high-quality dictionary. More importantly, the technique is unsupervised which means that a large-scale training corpus is not replied to learn the underlying rules of theme identification. These advantages make the word vectoring a pivotal method for special-purpose dictionary construction. However, despite the algorithm and analytical process, the studies overlook the potential value of the special-purpose dictionary and fail to exploit research opportunities behind the meaningful themes.

3. Methodology

As Figure 2 illustrates, ourCATA methodology comprises seven frames: (1) data describing and collecting, (2) automatic wordsense grouping, (3) theme screening and labeling, (4) validity checks, (5) theme quantifying, (6) empirical analysis and (7) transforming the results into knowledge.

3.1 Data preparation

In data preparation phase, we got access to a bank's loan application system and analyzed all E&A reports generated from 2011 to 2016. Our research partner, which is a leading service provider in Chinese SCF market, provided an excellent research setting. First, the E&A reports are stored in an independent system in electronic form. Thus, we could locate the investigation scenario in the SCF business. Then, the E&A process during which the reports are written should involve four loan officer members: the client manager, business unit manager, approval officer and compliance regulator. A survey report will become an E&A report only after iterative investigation and revision by the members. Third, the lending business and responsible loan officers are supervised by the China Banking Regulatory Commission. According to this commission's rules, E&A reports should clearly record all information regarding the financing object (*Credit Due Diligence Guidelines* [CDDG], Chapter 2), professional

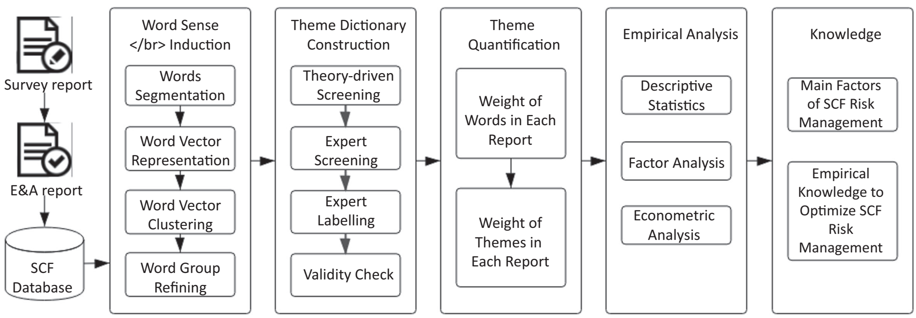


Figure 2.
Research design

judgment (CDDG, chapter 3), regulatory oversight (CDDG, chapter 4), key actions post loan (CDDG, chapter 5) and due diligence (CDDG, chapter 6).

To preprocess the data, we modulated the Chinese syntax format of the E&A reports, which have no space between words, by a word segmentation technique. A well-programmed word segmentation package (namely, Jieba) was used to eliminate stop words and convert the continuous sequence of character strings into a discrete sequence of meaningful words. Then, we measured the word counts of all reports. The average scenario in reporting was to find about 450 words. By reading report samples with word counts obviously lower than average, we found that some reports were either not fully acquired or had formatting problems. We therefore discarded 599 reports which had less than 150 words and finally retained 4,014 reports. Panel A of Table 1 classifies the reports according to their financing contexts and reporting years. Panels B and C illustrate the size and age structures of the approved borrowers.

3.2 WordSense induction

The dictionary construction process is initiated by grouping frequent words sharing same meanings. We selected words as the analytical units, instead of sentences or paragraphs, based on an examination of the syntactical nature of idea expression. To define word meanings, traditional content analyzers read the article artificially for a precise

Panel A: Accumulation of E&A reports for different financing contexts by years

Category	2011	2012	2013	2014	2015	2016	Total
Pre-shipment	8	8	36	517	1,497	990	3,056
Post-shipment	5	13	47	114	427	352	958
Total	13	21	83	631	1,924	1,342	4,014

Panel B: Firm size (employee number)

	Freq.	Percent	Cum.
Less than 20	643	16.02	16.02
20–40	775	19.31	35.33
40–60	825	20.55	55.88
60–100	754	18.79	74.67
100–200	503	12.53	87.2
200–500	245	6.04	93.24
500–1,000	47	1.13	94.37
1,000–10,000	85	2.1	96.47
more than 10,000	137	3.41	100
Total	4,014	100	

Panel C: Firm age (year)

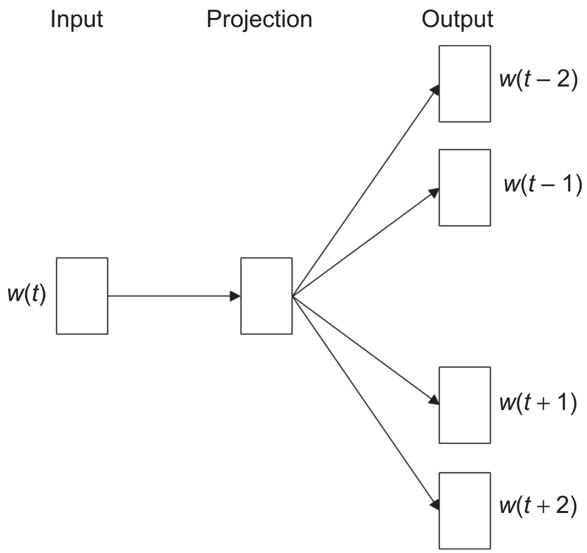
	Freq.	Percent	Cum.
Less than 2	223	5.56	5.56
2–4	640	15.94	21.5
4–7	1,093	27.24	48.74
7–10	731	18.21	66.95
10–15	978	24.36	91.31
15–30	326	8.1	99.41
More than 30	23	0.57	100
Total	4,014	100	

Table 1. Composition of the E&A reports and borrowers

understanding of word meaning. The progress in machine learning makes it possible to replace the workload of human cognition by the process of machine cognition. As a good example of the current study, the WSI algorithms can automatically glean the meaning of a word based on its contextual features, represents the word meaning by a vector and locates all words by grouping their vector representations.

To vectorize the word sense, we use word2vec, which is a set of open-source machine learning models released by Google (Mikolov *et al.*, 2013). Specifically, a skip-gram model is employed. As clarified in Figure 3, the model uses each object word, $w(t)$, as the input to a projection layer and predicts its nearby words within a certain range before and after a certain position t. The underlying intuition is that the meaning of a certain word always depends on its surroundings. Since its introduction byMikolov and colleagues, the model has been repeatedly applied and has proved itself an outstanding tool for symbolizing word meanings. Following Mikolov's advice, we set the upper limit of surrounding words at five and the vector dimensionality at 300. We also dismissed some low-frequency words which occur less than five times throughout the corpus. The vectorization process successfully results in the vector representations of 17,967 words.

Then, k-means clustering is conducted to group the words whose vector representations are close on the Euclidean distance. The suitable number of clusters (k) is very important for the performance of k-means algorithm. We decided the optimal k value based on the measure of average Silhouette value. The Silhouette value of each word group is in the range of -1 to 1. The value would approach to 1 if the belonging word vectors of a cluster are centralized and far from other clusters, and the value would close to -1 otherwise. The average silhouette value therefore can reflect the suitability of the choice of k value. We repeated the algorithm for ten times with k equals to the values ranging 10 to 100, spaced at 10 and finally set the optimal k value at 40. By observing the results of clustering, we found that some word groups have obvious similarity with respect to their word meanings and should be combined into



Source(s): Adopted from "Efficient Estimation of Word Representations in Vector Space" by 1. T. Mikolov, K. Chen, G. Corrado, J. Dean (2013)

Figure 3.
The Skip-gram model architecture

one. The word vector clustering consequently identified 34 anonymous themes from all frequent words of the E&A reports.

3.3 Dictionary construction

Based on the undefined categories, procedures for expert judgement should be conducted (Bengston and Xu, 1995). Theory-driven and expert-driven processes are operationalized in parallel to define a word list in each theme. In the theory-driven process, a screening task is assigned to two researchers who are knowledgeable regarding risk management actions. They determine whether each thematic word is meaningful according to aspects in the empirical literature. To further verify the results, a group discussion is organized with five subject matter experts: three from the business community and two from academia. The expert-driven process aimed to utilize the participants' SCF knowledge for theme screening and labeling. Backtracking reading is recommended throughout the procedures.

The experts are encouraged to provide additional comments to further refine the initial word groups. In the theme screening stage, the word group is rated "1" if recognized as a relevant theme in SCF risk management and "0" otherwise. The ratings from seven independent raters, including researchers and subject matter experts, delivered a Cronbach's alpha of 0.9424, which indicates high inter-rater reliability in the theme screening tasks. The word groups that obtain four or more votes are defined as relevant and are therefore retained in the dictionary. Further, the label can directly show what a theme refers to.

The theme labeling stage also follows a manual procedure to achieve high labeling quality. The five subject matter experts annotate all relevant themes following their own perception and domain knowledge and then come together to decide consensus labels for all themes. The final dictionary contains 16 relevant themes and 8,002 words. All identified themes, both relevant and irrelevant, are visualized by the word clouds in Figure 4. The size of each word cloud is determined by its frequency in the text corpus.

3.4 Quality of dictionary

The quality of the dictionary is determined by the extent to which our defined themes can precisely present and develop the actual meanings of expression. To check the validity of the themes, we examine semantic validation, theme labeling validation and predictive validation.

3.4.1 Semantic validation. To statistically test semantic coherence, we develop a word intrusion task that can help examine whether the themes correspond to natural groupings for humans.The subject of the word intrusion task is presented with a set of six randomly ordered words. The subject is asked to identify an intruding word that is out of place or does not belong with the others. If the remaining five words make sense together, the subject can readily identify the intruder. In contrast, if a set of words lacks semantic coherence, the intruder cannot be easily identified and the subject makes a choice at random.

Similar to Chang *et al.* (2009)'s practice, a word set is constructed following a procedure with four steps. First, a theme is randomly selected from the dictionary. Then, five high-frequency words are selected from the theme. Third, an intruding word is selected randomly from a pool of high-frequency words from some other theme. Fourth, all six words are shuffled and presented to the subject. To indicate how well the theme matches human concepts, we refer to model precision (MP_k), which is the fraction of subjects who select the intruder from words of theme k:

$$MP_k = \sum_s \mathbb{1}(i_{k,s} = \omega_k)/S \tag{1}$$

Source(s): Cronbach's alpha between the seven raters in screening process equals to 0.9424. Theme labels defined as relevant to risk management are bold, and new theme labels recognized as risk management relevance by the domain subject experts are bold and preceded by an asterisk "*" Product and industry theme labels are bold and followed by (P&I)

Figure 4.
Theme definition and labeling

where S is the total number of subjects, $i_{k,s}$ is the word selected by subject s on the words from the kth theme and ω_k is the true intruder among the thematic words of the kth theme. The indicator function equals 1 if the intruder is correctly selected by each subject.

Eight subjects are assigned to select intruders from the word sets generated from the 16 relevant themes. Figure 5 presents the frequency distribution of themes in terms of their model precision. Shown by the labels inparticular bins, the themes with high model precision

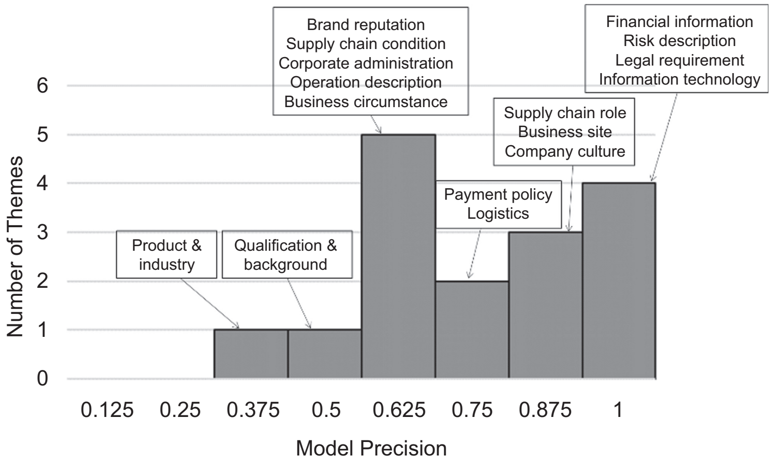


Figure 5.
The model precision of relevant themes

are more coherent semantically. The average model precision of all relevant themes is 75.78%, indicating that more than three-quarters of subjects choose the intruder correctly. In addition, for all themes, model precision is higher than the benchmark of random selection (16.67%). Therefore, our dictionary construction method can yield high-quality themes in terms of semantic coherence.

3.4.2 Theme labeling validation. The labeling validation tests whether the label of each word groups makes sense. It is noteworthy that ourWSI method follows the same principle to align with human practice of language reading and comprehension. Specifically, word meanings are not decided by words themselves, but rather, by how words relate to their surroundings in the article. This principle is followed throughout the processes of word vector training, word grouping, theme refinin and theme labeling; it thus enables continuous improvement of the labeling validation.

We operationalize an iteration of qualitative case verification, particularly in the theme refining and labeling stages. In each round of verification, two researchers read 10 sampled reports and examine the context of all mentioned words to understand their meanings. If the meaning of a word is found to be obviously inconsistent with the particular thematic label, it is recorded. A group discussion is organized at the end of every round to decide whether to relabel the word or dismiss it. The verification rounds continue until 80% of the words are found to be correctly defined by their thematic labels. Appendix 1 illustrates that words sharing the same risk management idea are well defined by one label.

3.4.3 Predictive validation. Predictive validation refers to how well the inferred themes can explain the exogenous variables. As contingency theory states, decisionmakers must consider environmental circumstances when putting different methods into action. Postshipment financiers provide funds to capital-constrained suppliers after goods delivery. To ensure money collection, more focus is given to buyer-side behaviors related to meeting their commitment, for example, payments. Theoretically speaking, the possibility of default will be dramatically reduced if the bank can collect money directly from the focal firms. In contrast, preshipment financiers are inclined to emphasize physical factors since funds are provided at an early stage before goods delivery. The logistic arrangements for intermediate or finished goods are crucial to successfully perform the contract of stock ownership exchange or unsold goods buyback. Figure 6 describes the word count ratio of "payment policy" to "logistics description" in postshipment and preshipment contexts.

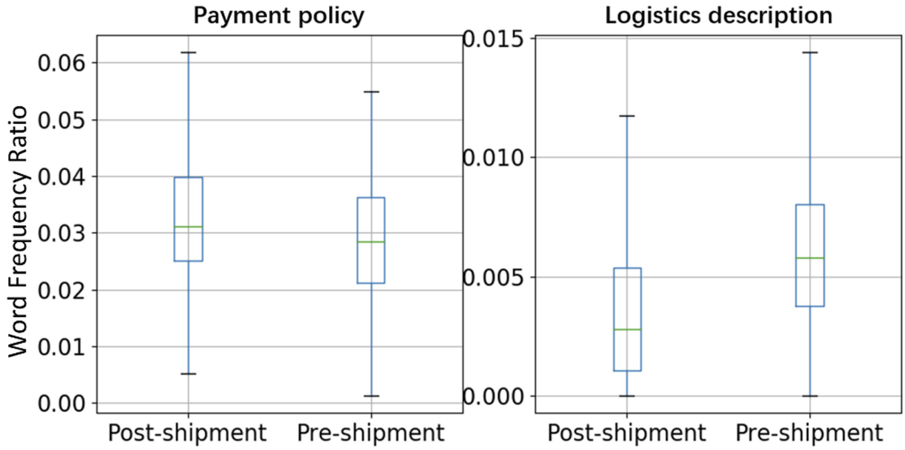


Figure 6.
The frequency of "payment policy" and "logistics description" by financing contexts

The ratio is determined by the count of words about the themes among all words in each E&A report. It is obvious that payment policy is more frequently mentioned for managing cash flows in postshipment than preshipment context. Meanwhile, logistics is more likely to be emphasized to safeguard the value of some physical elements (e.g., raw materials, finished goods) in preshipment context. The observations are statistically significant since the hypothesized equality of means is rejected at $p < 0.001$. This shows that the contingent risks in two SCF contexts predict the highlights in risk management arrangements.

3.5 Emphasis measurement

A word that is mentioned more frequently in one article than in others carries more article-specific emphasis. In this context, TF–IDF is employed in a two-step computation to measure the emphasis of themes in each article (Zobel and Moffat, 1998).

In the first step, the TF–IDF model is administered as in equation (2):

$$TF - IDF_{t,d} = TF_{t,d} \times IDF_t, \tag{2}$$

which yields importance metrics at the word level. Then, in the second step, the cosine-combining function of (Salton and Buckley, 1988) is applied to convert the word-level metrics to theme-level metrics (equation (3)):

$$S_{q,d} = \left(\sum_{t \in d \cap q} TF - IDF_t \right) \Big/ (W_d \times W_q) \tag{3}$$

In equation (2), $TF_{t,d}$ refers to the frequency of a word (t) in a document (d). IDF$_t$ equals $\log \frac{N}{DF_t} + 1$, where N is the total number of documents and DF_t indicates the number of documents that contain the word (t).

In equation (3), t, d and q refer to a word, a document and a theme, respectively. W_d refers to the weight of a document (d) and equals $\sqrt{\sum_{t \in d} (TF\text{–}IDF_t^2)}$ W_q refers to the weight of a theme (q). Although the model can be applied with different weights if needed, the weights of all themes are set as 1 in the current model.

The final score ($S_{q,d}$) reflects the weight of theme (q) in the linguistic context of document (d), which we refer to as the degree of emphasis. The emphasis weighting method allows for adjustments for some particular words which occur in boilerplate. In simple terms, a high score indicates that the theme is emphasized in the E&A report. Conversely, a low or zero score indicates negligible mention of the theme in the report. Descriptive statistics for the relevant themes are presented in Table 2.

3.6 Exploratory factor analysis

The themes, alone or in combination, are the keys to construct the factors. In this part of the study, EFA is employed to identify the latent dimensionality and construct risk management factors from the relevant themes.

3.6.1 Dimensionality and construct validity. Table 3 provides statistical evidence with respect to the factors and the themes to which they belong. The Kaiser–Meyer–Olkin measure (0.800) and the Bartlett sphericity test ($p < 0.001$) indicate that the variables are well-prepared for EFA in terms of their statistical correlations. The level of eigenvalues and the cumulative variance percentage are considered to determine dimensionality. From the four dimensions, distinct risk management factors are identified: relationship-based assessment (RBA), cash flow monitoring (CFM), SC collaboration (SCC) and assets monitoring (AM). As a result, all factors have eigenvalues above 1 and explained 57.95% of the variance. In terms of convergent validity, all themes have their loading values above 0.50, after excluding the variables *product and industry* and *risk description*. Each factor has average variance extracted higher than 0.4. These results provide evidence of the convergent validity of the themes for each factor. Finally, discriminant validity is evaluated using the squared correlations of each factor with the remaining factors. The factors are orthogonal (correlation = 0), suggesting sufficient discriminant validity. Therefore, the constructed factor could be used for hypotheses testing.

3.6.2 Risk management factors. In terms of factor 1, six variables have high loadings: *operation description*, *business circumstance*, *corporate administration*, *market reputation*, *company culture* and *qualification and background*. Hence, the themes center on soft information. To understand their qualifications and background, a loan officer needs to maintain a relationship with debtors — either publicly or personally. Then, to obtain an

No.	Relevant themes	Reports	Mention rate	Distribution of emphasis metrics: Mean	Standard deviation	Minimum	Maximum
1.	Payment policy	4,014	100.00%	0.66	0.16	0.06	0.96
2.	Product and industry	4,014	100.00%	0.35	0.24	0.01	1.42
3.	Financial information	4,014	100.00%	0.3	0.13	0.02	0.67
4.	Legal requirement	4,014	100.00%	0.08	0.04	0.01	0.39
5.	Supply chain condition	4,014	100.00%	0.03	0.02	0	0.25
6.	Operation description	4,013	99.98%	0.06	0.04	0	0.28
7.	Business circumstance	4,011	99.93%	0.05	0.05	0	0.36
8.	Risk description	4,008	99.85%	0.04	0.02	0	0.23
9.	Logistics description	3,926	97.81%	0.07	0.05	0	0.63
10.	Market reputation	3,645	90.81%	0.03	0.03	0	0.43
11.	Supply chain role	3,633	90.51%	0.02	0.02	0	0.26
12.	Corporate administration	3,563	88.76%	0.05	0.04	0	0.33
13.	Business site	3,464	86.30%	0.02	0.02	0	0.44
14.	Information technology	3,369	83.93%	0.02	0.02	0	0.36
15.	Company culture	2,955	73.62%	0.01	0.01	0	0.12
16.	Qualifications and background	2,342	58.35%	0.01	0.01	0	0.2

Table 2.
Descriptive statistics of relevant themes in E&A reports

Factor and construct	Factor 1 Relationship-based assessment	Factor 2 and 3 Cash flow monitoring / - financial analysis	Factor 4 Supply chain collaboration	Factor 5 Asset monitoring
Operation description	*0.76*	−0.07	0.34	0.02
Business circumstance	*0.73*	0.01	0.28	0.04
Corporate administration	*0.71*	−0.12	0.10	0.23
Market reputation	*0.68*	−0.13	0.21	−0.06
Company culture	*0.66*	0.25	−0.14	−0.21
Qualification and background	*0.57*	−0.04	−0.07	0.27
Legal requirement	−0.04	*0.90*	0.00	0.01
Payment policy	0.10	*0.81*	0.09	0.14
Financial information	0.35	*−0.59*	0.16	0.29
Supply chain role	0.10	−0.07	*0.69*	0.11
Supply chain condition	0.35	0.17	*0.67*	0.00
Information technology	0.22	−0.03	*0.52*	0.01
Logistic description	−0.18	0.27	0.20	*0.70*
Business site	0.38	−0.17	−0.15	*0.62*
Eigenvalue	3.32	2.05	1.57	1.18
Average Variance Extracted	0.47	0.60	0.40	0.44
% of Variance	23.70	14.63	11.19	8.43
% of Variance explained in total	23.70	38.33	49.52	57.95

Note(s): Bartlett test of sphericity: Chi-square = 14,897.708; p-value = 0.0000. Kaiser–Meyer–Olkin measure of sampling adequacy: KMO = 0.8000. The loadings which have absolute value larger than 0.5 are reported in italics

Table 3.
Five SCF risk management factors identified from risk management themes

insight into a debtor's corporate administration, company culture and operation situation, a loan officer may undertake field visits to the company to obtain first-hand information. A company's market reputation and business environment are more likely to be perceived according to their industry expertise. We believe the bank–debtor relationship is the principal source of this information. These themes in combination are named relationship-based assessment (*RBA*).

Then, *legal requirement*, *payment policy* and *financial information* characterize factor 2. A trade-off exists between financial information and the other two variables regarding their high, but reversed, loadings. On one hand, the payment policy combined with the legal regulations in SCF projects can mitigate against loan uncertainty by scheduling the financial flow. For example, a bank could ask applicants to create escrow accounts so that some transaction incomes can be monitored efficiently and legally. On the other hand, the theme derived from static financial statements has high but negative loading in the factor. The opposite signs within the factor reflect that loan officers attempt to dynamically monitor interorganizational financial flows rather than relying solely on financial statements. Therefore, factor 2 is labeled cash flow monitoring (CFM).

Factor 3 has high loadings in *SC role*, *SC condition* and *information technology*. By jointly discussing the SC role and the SC condition in its E&A reports, a bank can clarify an applicant's

SC position, duty, obligation and relationship with SC partners. The two themes complement each other to portray the status of SC circumstances. In addition, the theme of information technology involves the programmed procedure arranged for information collection, verification and transmission. An inference is that the bank adopts this factor to connect more SC partners, so that it can better capture the features and risks of debtors. Hence, the factor is named supply chain collaboration (SCC) for both assessment and monitoring purposes.

Finally, *logistic description* and *business site* comprise factor 4. The portfolio of this dimensional arrangement is based primarily on two themes: the description of locations where the production elements are formed and the description of transportation matters for supervising, managing or disposing of the pledged assets. The loan officers who highlight this factor attempt to take control of the assets. In theory, the existence of high-quality tangible assets (e.g., real estate, equipment or inventory) will extend the available credit as long as they are leased or pledged as collateral. Moreover, asset monitoring increases a debtor's default cost and helps the bank to recover from credit failures. Thus, this factor is titled assets monitoring (AM).

4. Empirical analysis of risk management factors

4.1 Hypotheses development

To manage loan portfolios, financial institutions employ various indicating models, among which the default event is always the target of prediction. In this study, we are interested in exploring risk management factors for preventing loan default in the SCF context. Corresponding to the two actions of credit risk management, an effective risk management factor should significantly enhance a bank's ability to evaluate the parties involved; screen good final debtors and limit the opportunity for fraudulent activities. Therefore, we hypothesize that an emphasis on risk management factors will reduce the occurrence of default events (H1). Moreover, the financing context — preshipment or postshipment — includes different SC operations processes and some heterogeneous risks that may influence the performance of risk management. The risk management factors are the contingent strainer through which some specific risks are isolated from funds granted while some other risks are missed. Thus, we further hypothesize that the effects of the risk management factors are contingent on the type of financing context (H2).

We employ loan charge-off as the measure of loan default since it is an observable behavior of banks which allows loan loss provision. For our research bank, the loans for the SCF service are charged off based on a number-of-days-past-due rule, which applies 90 days as past due. The rule setting ensures that there is less discretion in our default identification.

4.2 Econometric model

Logistic regressions are used to estimate the relationship between the emphasis level of risk management factors and the probability of loan charge-off, conditional on some characteristics of final debtor, responsible organization and operation year. The definitions and data resources of all variables involved are presented in Table 4. The econometric model is designed as follows:

$$\begin{aligned}\log - \text{odds}(\text{CHGOFF}) = \beta_0 &+ \beta_1\text{RBA} + \beta_2\text{CFM} + \beta_3\text{SCC} + \beta_4\text{AM} + \beta_5\text{lnEMPNUM} \\ &+ \beta_6 ln\text{AGE} + \beta_7\text{lnPATTNUM} + \beta_8\text{lnLWSTNUM} + \beta_9\text{CORP} \\ &+ \sum_1^{30}\gamma_i\text{BRANCH}_i + \sum_1^{5}\delta_i\text{YEAR}_i + \varepsilon.\end{aligned} \tag{4}$$

Table 4.
Variable definitions and data sources

Variable	Definition	Source
CHGOFF	A dummy variable to indicate charge-off event	Customer information system
RBA	Emphasis level of relationship-based assessment factor in E&A report	E&A report
CFM	Emphasis level of cash flow monitoring factor in E&A report	E&A report
SCC	Emphasis level of supply chain collaboration factor in E&A report	E&A report
AM	Emphasis level of asset monitoring factor in E&A report	E&A report
lnEMPNUM	The natural logarithm of on-site employee number	Customer information system
lnAGE	The natural logarithm of business duration	Customer information system and Public information
lnPATTNUM	The natural logarithm of patent number	Public information
lnLWSTNUM	The natural logarithm of lawsuit number	Public information
CORP	A dummy variable to indicate the legal form as corporation	Public information
$BRANCH_i$	A series of dummy variables to indicate the responsible branches	Customer information system
$YEAR_i$	A series of dummy variables to indicate the approval years	Customer information system

According to equation (5), a variable is more likely to increase the probability of charge-off if its odds ratio is greater than 1, and vice versa:

$$\log - \text{odds}(\text{CHGOFF}) = \ln\left(\frac{\Pr(\text{CHGOFF} = 1)}{1 - P(\text{CHGOFF} = 1)}\right) \tag{5}$$

In equation (4), RBA, CFM, SCC and AM indicate the emphasis levels of corresponding risk management factors in each E&A report. They are the key variables to assess our hypotheses, and β_1 *to* β_4 are the model coefficients of interest. In addition, key exogenous variables are measured to capture firm size (lnEMPNUM), business duration (lnAGE), innovation capability (lnPATTNUM), litigation pressure (lnLWSTNUM) and legal form (CORP) of final debtors. These organizational and operational issues are suggested in empirical studies for predicting the credit capability of SMEs. The remaining variables in the equation control for operational difference between 30 responsible branches of our research bank ($BRANCH_i$) and other unknown changes by six years ($YEAR_i$).

The descriptive statistics and correlation matrix for all employed variables are presented in Table 5. The intercorrelations between independent variables are low. Variance inflation factor (VIF) scores are also low, at 1.02–1.29. Therefore, collinearity is not a concern for subsequent regressions.

4.3 Results and discussions

The four columns in Table 6 present the main estimators. Column 1 presents the estimators of the model including only control variables for log-odds of charge-off. In columns 2–4, the statistics are respectively estimated for full observations, postshipment specific observations and preshipment specific observations. The values of pseudo R-square reported at the bottom of the columns are relatively large (0.321–0.411), indicating a good model fit.

In general context, the statistics in column 2 show that three out of four factors do not have significant effects on default events. Although the coefficient of AM has a significant magnitude, its sign is not negative as expected. Therefore, a high emphasis on asset

Table 5. Descriptive statistics and correlation matrix

Variable	N	Mean	Standard deviation	Minimum	Maximum	1	2	3	4	5	6	7	8	9
1. CHGOFF	4,014	0.03	0.17	0.00	1.00	1								
2. RBA	4,014	0.00	1.00	−1.81	4.59	0.06***	1							
3. CFM	4,014	0.00	1.00	−2.28	3.73	−0.01	0	1						
4. SCC	4,014	0.00	1.00	−6.69	6.40	−0.01	0	0	1					
5. AM	4,014	0.00	1.00	−6.46	10.67	0.06***	0	0	0	1				
6. lnEMPNUM	4,014	3.90	1.12	1.10	11.16	0.03*	0.03*	−0.02	0.003	0.02	1			
7. lnAGE	4,014	0.72	1.14	0.00	8.52	0.31***	0.19***	−0.05***	0.03*	0.08***	0.04**	1		
8. lnPATTNUM	4,014	1.89	0.70	−1.93	3.56	0.02	0.10***	−0.05***	0.03**	−0.05**	0.01	0.21***	1	
9. lnLWSTNUM	4,014	0.51	1.16	0.00	8.52	0.06***	0.19***	0.05***	0.06***	−0.04***	0.03**	0.32***	0.28***	1
10. CORP	4,014	0.97	0.16	0.00	1.00	0.01	0.10***	−0.20***	−0.06***	0.01	0.02	0.08***	0.05***	0.05***

Note(s): Significance level for each correlation: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

Variables	Control (1)	Log-odds of charge-off (CHGOFF) Overall (2)	Post-shipment (3)	Pre-shipment (4)
Relationship-based assessment (RBA)		0.8982	1.0127	0.5435***
		(0.1033)	(0.1665)	(0.1086)
Cash flow monitoring (CFM)		0.8311	0.6104**	0.7615
		(0.1097)	(0.1337)	(0.1743)
Supply chain collaboration (SCC)		0.8951	0.9601	0.6563**
		(0.0993)	(0.1348)	(0.1278)
Assets monitoring (AM)		1.1862*	1.0049	1.4115
		(0.1179)	(0.1023)	(0.4274)
Ln employee number (lnEMPNUM)	1.1251	1.1204	1.0110	1.3514***
	(0.0981)	(0.0987)	(0.1202)	(0.1525)
Ln business duration (lnAGE)	0.9001	0.8839	1.4155	0.4699***
	(0.1452)	(0.1429)	(0.3263)	(0.1092)
Ln patent number (lnPATTNUM)	0.8474*	0.8597*	0.6886***	0.9165
	(0.0722)	(0.0746)	(0.0885)	(0.1660)
Ln lawsuit number (lnLWSTNUM)	2.4606***	2.4610***	2.1026***	3.6800***
	(0.2163)	(0.2323)	(0.3120)	(0.4918)
Corporate dummy (CORP)	0.5405	0.4226	0.3381	0.4792
	(0.4173)	(0.3382)	(0.3783)	(0.6080)
INTERCEPT	−4.5190***	−4.3185***	0.2173	−3.2893
	(1.6261)	(1.6045)	(2.1741)	(2.2504)
Branch dummies ($BRANCH_i$)	Included	Included	Included	Included
Year dummies ($YEAR_i$)	Included	Included	Included	Included
Model statistics				
Pseudo R-square	0.321	0.327	0.354	0.411
Observations	3,685	3,685	881	2,337

Note(s): Odds ratios are presented. Huber–White robust standard errors in parentheses with *, **, and *** indicating significance at 10%, 5%, and 1%, respectively

Table 6.
Estimation results

monitoring will catalyze defaults. Such a finding lends support to the argument that good firms are unwilling to offer assets as security because doing so can limit their operational flexibility (Myers and Rajan, 1998).The insignificant coefficients of RBA, CFM and SCC from the context-free estimation indicate that the three risk management factors are unable to play an effective role in general.

For the estimators inpost-shipment context, the odds ratio on CFM is statistically significant and less than 1 (column 3). The postshipment financiers are less likely to suffer loan default if they can regulate payment behaviors among SC partners. The odds ratios on the other risk management factors are not statistically significant. Therefore, if the bank can collect direct payments from focal firm buyers and limit private cash settlements between transaction partners, some supply-side uncertainties are avoided. As per Hofmann (2005), opportunism is substantially eliminated by arrangements designed to plan, steer and control cash flows among organizations.

Column 4 shows the regression with observations in preshipment context. Statistically, the odds ratios on RBA and SCC are significantly less than 1 and the odds ratios on the other two factors are insignificant. Therefore, lender–debtor relationship and integrative SC management are more likely to prevent default in the financing context which covers operational activities including ordering, production, shipment, warehousing and selling. Unlike postshipment financing, which is based primarily on the creditworthiness of focal firms, preshipment financing necessarily requires weak buyers to perform their duty, redeeming and selling specific products.The observed effect of RBA suggests that preshipment financiers should collect more first-hand information (e.g., owners, employees and neighboring communities) about the functional units along SC to manage risk associated with the operations. In addition, SCC also represents value in preventing loan defaults. An ideal SC management system integrates the isolated SC functional units and thus can provide visibility and timely information for risk management. Specifically, for preshipment context, the collaboration can facilitate the execution of implicit warranty agreements (e.g., unsold goods buyback, warehouse confirmation) which involve multiple SC members.

In summary, since AM cannot reduce the occurrence of defaults as the other factors do, H1 is partially supported. Then, there is no one factor that can significantly prevent defaults in both financing contexts, indicating support for H2.

To check the robustness, we examine the magnitudes of the estimated effects on the occurrence of loan overdue as an alternative measurement of defaults (Appendix 2). Overdue occurs immediately whenever a loan is not settled in time, which is a more random event. The same model specification is used to predict the new dependent variable. The estimation results are consistent except for the coefficient significance of SCC in preshipment context. Therefore, interorganization collaboration cannot prevent accidental overdues. We also rerun the regressions with different segments of our dataset, which respectively include the loans approved from 2015 to 2016 and the loans assigned to the firms who are with less than 300 employees. The coefficients on the four risk management factors are unchanged across the data segments.

5. Conclusions and implications

The experience and knowledge of risk management from banking secto ris valuable for the extensive adoption of SCF service. However, existing literature fails to explore the intelligent resource. One major reason is that the decision-making process for risk management is difficult to observe, code, measure and test.

Our methodology shows its potential to overcome the difficulties. First, constructing special-purpose dictionary is no longer time-intensive or relying on human resources. Second, subjective bias and extraneous noise are reduced during the course of replacing manual practices by programmed scripts. Contrast to previous text-mining literature which mainly focus on theme discovery, this study rests on the well-developed techniques (i.e., WSI and TF–IDF) and demonstrates a CATA methodology which can contribute to our ability to explore more insights about the role of the themes for specific research purpose. The whole processes from the data preprocessing of textual data to the application of empirical testing for hypothesis can offer the case about how the textual resource in SCF field should be analyzed and utilized in practice.

This study also contributes to the credit risk management literature. Our empirical results suggest that SC-oriented factors (i.e., CFM, SCC) are implemented in parallel with corporate finance factors (i.e., RBA, AM) for managing risks of SCF services. The following analysis presents that the function of the four risk management factors for default prevention is unique to specific financing contexts (i.e., postshipment, preshipment). Based on these findings, researchers of SCF field can further resolve the particularities of different financing contexts and provide valuable guidance to better prevent credit risks of SCF.

The business intelligence presented in this study can also be adapted among SCF service providers. According to the findings, it is not necessary to establish the comprehensive capability on all risk management factors. They are suggested to pay more attention to cash flow monitoring when providing postshipment financing services and shift the focus to relationship building and interorganizational collaboration when providing preshipment financing services. This economically optimal arrangement of risk management is supposed to prevent potential risks with lower investments and improve the profitability of SCF service. Moreover, the vocabulary of risk management themes can help them to digitize conventional loan approving process and, further, to facilitate the decision-makings. Based on the digital system, SCF service providers can also identify potential issues of their past operations and try to improve.

One of the key limitations of this study is that our dataset contains only approved debtors. Since approval decisions include loan officers' risk management reasoning, some biases may exist in our results. Future studies should compare the decision process of both approved and rejected cases, which will provide useful knowledge to help optimize the E&A process and reduce defaults.

References

Agarwal, B. and Mittal, N. (2016), "Prominent feature extraction for review analysis: an empirical study", *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 28 No. 3, pp. 485-498, Taylor Francis.

Bao, Y. and Datta, A. (2014), "Simultaneously discovering and quantifying risk types from textual risk disclosures", *Management Science*, Vol. 60 No. 6, pp. 1371-1391.

Bengston, D.N. and Xu, Z. (1995), "Changing national forest values: a content analysis", *Forest Service*, Vol. NC No. 323, pp. 1-30.

Benmelech, E. and Bergman, N.K. (2009), "Collateral pricing", *Journal of Financial Economics*, Vol. 91 No. 3, pp. 339-360.

Cao, Y., Zhang, J.-H. and Ma, X.-Y. (2019), "Optimal financing and production decisions for a supply chain with buyer-backed purchase order financing contract", IEEE Access, *IEEE*, Vol. 7, pp. 119384-119392.

Chakuu, S., Masi, D. and Godsell, J. (2020), "Towards a framework on the factors conditioning the role of logistics service providers in the provision of inventory financing", *International Journal of Operations and Production Management*.

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C. and Blei, D.M. (2009), "Reading tea leaves: how humans interpret topic models", *Advances in Neural Information Processing Systems*, Curran Associates, New York, Vol. 22, December, pp. 288-296.

Chen, X. (2015), "A model of trade credit in a capital-constrained distribution channel", *International Journal of Production Economics*, Vol. 159 No. 1, pp. 347-357, Elsevier.

Chen, J., Zhou, Y.W. and Zhong, Y. (2017), "A pricing/ordering Model for a dyadic supply chain with buyback guarantee financing and fairness concerns", *International Journal of Production Research*, Vol. 55 No. 18, pp. 5287-5304, Taylor & Francis.

Chung, C.K. and Pennebaker, J.W. (2012), "Linguistic inquiry and word count (LIWC): pronounced 'Luke,' . . . and other useful facts", *Applied Natural Language Processing: Identification, Investigation and Resolution*, pp. 206-229.

Degryse, H. and Van Cayseele, P. (2000), "Relationship lending within a bank-based system: evidence from european small business data", *Journal of Financial Intermediation*, Vol. 9 No. 1, pp. 90-109.

Grunert, J., Norden, L. and Weber, M. (2005), "The role of non-financial factors in internal credit ratings", *Journal of Banking and Finance*, Vol. 29 No. 2, pp. 509-531.

Hofmann, E. (2005), "Supply chain finance: some conceptual insights", *Logistik Management-Innovative Logistikkonzepte*, Vol. 80 No. 1, pp. 203-214.

Hofmann, E. and Belin, O. (2011), *Supply Chain Finance Solutions*, Springer, Berlin. doi: 10.1007/978-3-642-17566-4.

Hung, C. (2017), "Word of mouth quality classification based on contextual sentiment lexicons", *Information Processing and Management*, Vol. 53 No. 4, pp. 751-763, Elsevier.

Hung, J.L., He, W. and Shen, J. (2020), "Big data analytics for supply chain relationship in banking", *Industrial Marketing Management*, Vol. 86, October 2019, pp. 144-153.

Jia, F., Blome, C., Sun, H., Yang, Y. and Zhi, B. (2020), "Towards an integrated conceptual framework of supply chain finance: an information processing perspective", *International Journal of Production Economics*, Vol. 219, April 2019, pp. 18-30, Elsevier B.V.

Kim, Y.S., Rim, H.C. and Lee, D.G. (2019), "Business environmental analysis for textual data using data mining and sentence-level classification", *Industrial Management and Data Systems*, Vol. 119 No. 1, pp. 69-88.

Kim, S., Park, H. and Lee, J. (2020), "Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: a study on blockchain technology trend analysis", *Expert Systems with Applications*, Vol. 152, No. 1, pp. 1-12.

Klapper, L. (2006), "The role of factoring for financing small and medium enterprises", *Journal of Banking and Finance*, Vol. 30 No. 11, pp. 3111-3130.

Li, T., Fang, W., Dash Wu, D. and Zhang, B. (2020), "Inventory financing a risk-averse newsvendor with strategic default", *Industrial Management and Data Systems*, Vol. 120 No. 5, pp. 1003-1038.

Lin, Q. and He, J. (2019), "Supply chain contract design considering the supplier's asset structure and capital constraints", *Computers and Industrial Engineering*, Vol. 137, September, p. 106044, Elsevier.

Lin, J., Zhou, Y., Yang, A. and Wang, W. (2016), "Analysis of topic evolution on news comments based on word vectors", *Computer Engineering and Science*, Vol. 38 No. 11, pp. 464-475.

Lin, Q., Su, X. and Peng, Y. (2018), "Supply chain coordination in confirming warehouse financing", *Computers and Industrial Engineering*, Vol. 118 No. 1, pp. 104-111, Elsevier.

Lyu, X. and Zhao, J. (2019), "Compressed sensing and its applications in risk assessment for internet supply chain finance under big data", *IEEE Access, IEEE*, Vol. 7, pp. 53182-53187.

Ma, H.L., Wang, Z.X. and Chan, F.T.S. (2020), "How important are supply chain collaborative factors in supply chain finance? A view of financial service providers in China", *International Journal of Production Economics*, Vol. 219, July 2019, pp. 341-346, Elsevier B.V.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013), "Distributed representations of words and phrases and their compositionality", *Advances in Neural Information Processing Systems*, Vol. 26, pp. 1-9.

Myers, S.C. and Rajan, R.G. (1998), "The paradox of liquidity", *Quarterly Journal of Economics*, Vol. 113 No. 3, pp. 733-771.

Niinimaki, J.P. (2016), "Spatial competition in the banking system: alternative lending technologies and collateral assets", *Journal of Multinational Financial Management*, Vol. 35 No. 1, pp. 93-109, Elsevier B.V.

Salton, G. and Buckley, C. (1988), "Term-weighting approaches in automatic text retrieval", *Information Processing and Management*, Vol. 24 No. 5, pp. 513-523.

Somers, M.J. and Casal, J. (2017), "Introducing neural computing in governance research: applying self-organizing maps to configurational studies", *Corporate Governance: An International Review*, Vol. 25 No. 6, pp. 440-453.

Song, X., Ying, H., Zhao, X. and Chen, L. (2019), "Improving assessment capability for accounts receivable pool financiers: an empirical investigation", *Industrial Management and Data Systems*, Vol. 120 No. 3, pp. 547-566.

Strauß, N., Vliegenthart, R. and Verhoeven, P. (2016), "Lagging behind? Emotions in newspaper articles and stock market prices in The Netherlands", *Public Relations Review*, Vol. 42 No. 4, pp. 548-555, Elsevier.

Tangpong, C. (2011), "Content analytic approach to measuring constructs in operations and supply chain management", *Journal of Operations Management*, Vol. 29 No. 6, pp. 627-638, Elsevier B.V.

Vliet, K. Van Der, Reindorp, M.J. and Fransoo, J.C. (2015), "The price of reverse factoring: financing rates vs . payment delays", *European Journal of Operational Research*, Vol. 242 No. 3, pp. 842-853, Elsevier.

Wang, R., Yu, C. and Wang, J. (2019), "Construction of supply chain financial risk management mode based on Internet of Things", *IEEE Access, IEEE*, Vol. 7, pp. 110323-110332.

Wuttke, D.A., Blome, C. and Henke, M. (2013), "Focusing the financial flow of supply chains: an empirical investigation of financial supply chain management", *International Journal of Production Economics*, Vol. 145 No. 2, pp. 773-789, Elsevier.

Zhao, L. and Huchzermeier, A. (2019), "Managing supplier financial distress with advance payment discount and purchase order financing", *Omega (United Kingdom)*, Vol. 88 No. 1, pp. 77-90, Elsevier.

Zhao, F., Yao, Z., Luan, J. and Liu, H. (2019), "Inducing stock market lexicons from disparate Chinese texts", *Industrial Management and Data Systems*, Vol. 120 No. 3, pp. 508-525.

Zhu, Y., Zhou, L., Xie, C., Wang, G.J. and Nguyen, T.V. (2019), "Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach", *International Journal of Production Economics*, Vol. 211, September 2018, pp. 22-33, Elsevier B.V.

Zobel, J. and Moffat, A. (1998), "Exploring the similarity space", *ACM SIGIR Forum*, Vol. 32, pp. 18-34.

Appendix

The appendix files are available online for this article.

Corresponding author

Hao Ying can be contacted at: jogayh@sjtu.edu.cn