



ARTICLE



<https://doi.org/10.1057/s41599-022-01341-y>

OPEN

What drives reputational risk? Evidence from textual risk disclosures in financial statements

Xiaoqian Zhu^{1,2}, Yinghui Wang^{1,3} & Jianping Li^{1,2}✉

The drivers of reputational risk are still far from explicit, making proactive risk management and quantitative research rather difficult. The Basel Committee on Banking Supervision encourages financial institutions to systematically identify reputational risk drivers; however, such drivers still represent an unsolved problem. Therefore, the objective of this paper is to systemically identify reputational risk drivers from textual risk disclosures in financial reports. We find that textual risk disclosures in financial reports contain abundant information about the causes of reputational risk, thus indicating the possibility of systematically identifying the reputational risk drivers. To accurately extract reputational risk drivers from massive and unstructured textual risk disclosure data, we modify a text mining method to make it more suitable for this type of textual data with noise words. Based on 352,326 risk headings extracted from 11,921 annual reports released by 1570 U.S. financial institutions from 2006 to 2019, a total of 13 reputational risk drivers are identified to extend upon existing studies. The importance of reputational risk drivers and their dynamic evolutions are also quantified to discover the drivers of greatest concern. This paper can clarify the sources of reputational risk to help companies realize proactive reputational risk management and provide a theoretical basis for further quantitative studies, especially the measurement of reputational risk.

¹School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China. ²MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS, Beijing, China. ³Institutes of Science and Development, Chinese Academy of Sciences, Beijing, China.
✉email: ljp@ucas.ac.cn

Introduction

The reputation of a company is its most important asset but the most difficult asset to recover once it is lost (Scandizzo, 2011). The necessity of managing reputational risk is especially important for financial institutions whose business models are based on trust (Gatzert, 2015; Heidinger and Gatzert, 2018; Scholtens and Klooster, 2019). For example, the rogue trader scandal that impacted the United Bank of Switzerland (UBS) in 2011 led to an operational loss of ~2 billion dollars; moreover, the company's reputation deteriorated, and it eventually lost 6.3 billion dollars' in market value (Eckert and Gatzert, 2017). Due to the very large amplification effect reputation damage has on losses, reputational risk has received increasing attention in recent years by managers, regulators, and academics (Gatzert et al., 2016; Vig et al., 2017; Cornejo et al., 2019). The effect is especially prominent based on the increasing impact of the internet and social media, where bad news in particular spreads quickly (Gatzert, 2015).

However, even as discussions of reputational risk have intensified, relevant studies on reputational risk are still at a preliminary stage (Fiordelisi et al., 2013; Eckert and Gatzert, 2017; Heidinger and Gatzert, 2018). The Basel Committee on Banking Supervision (BCBS) and Comité Européen des Assurances (CEA) have defined reputational risk as the risk of negative perception by other market participants (BCBS, 2009; CEA, 2007). However, the definition is quite broad and can be considered an unobservable "cognition" (Gatzert et al., 2016); in addition, the risk drivers underlying negative perception are still far from explicit (Eckert and Gatzert, 2017). Researchers have also generally considered that reputational risk is a "risk of risks" and has many sources (Gatzert and Schmit, 2016; Heidinger and Gatzert, 2018), but the literature has not yet agreed on the risk drivers that may destroy a reputation. Currently, the lack of clarity about reputational risk drivers is rather similar to the situation that inspired the preliminary studies on operational risk. At that time, quantitative studies on operational risk were very intractable because all activities of financial institutions are in some way exposed to operational risk (Rosenberg and Schuermann, 2006). Until the Basel Committee on Banking Supervision (BCBS) clarified the mechanism of operational risk, especially the causes and loss types; an increasingly systematic management framework and in-depth quantitative studies on operational risk followed. This inspired us to explore the drivers of reputational risk.

The lack of systematic risk driver identification studies increases the difficulty of performing proactive risk management and quantitative research on reputational risk. Successful risk management requires the anticipation of events that have not yet happened and that are active rather than reactive (Gatzert and Schmit, 2016). However, most studies have primarily contributed to reputation building and repair after a crisis event (see, e.g., Rhee and Valdez, 2009; Gow et al., 2018), and fewer studies have focused on proactive reputational risk management approaches (Scandizzo, 2011). One important reason is that the drivers or antecedents of reputational risk are still inexplicit (Gatzert, 2015). Quantitative studies related to reputational risk measurement start by identifying underlying risk sources. Most empirical studies are based on the assumption that financial institutions suffer reputational losses following operational risk events (see, e.g., Gillet et al., 2010; Heidinger and Gatzert, 2018). However, Eckert and Gatzert (2017) specifically mention the limitations of existing studies in which other risk drivers are neglected, which may lead to an underestimation of reputational risk. In addition, Gatzert et al. (2016) found that insurers also face the challenge of identifying the drivers of reputational risk because such information is not only relevant for general risk management purposes but is especially crucial for insurers attempting to accept the risk of loss due to a damaged reputation of a policyholder.

Therefore, comprehensive recognition of the drivers of reputational risk is urgently needed. The BCBS also encourages academics and financial institutions to "identify potential sources of reputation risk to which it is exposed" (BCBS, 2009). However, previous studies have either been based on expert experiences or summarized the drivers mentioned in prior studies (Scandizzo, 2011; Gatzert et al., 2016); thus, the derived risk factors are usually incomplete and subjective, and a comprehensive and objective list of reputational risk drivers has not been formed. A method of systematically identifying these drivers has not been developed.

This study finds that the textual risk disclosures provided in financial reports, which have not been fully utilized compared to quantitative data, contain valuable information about reputational risk drivers. Beginning in 2005, the US Securities and Exchange Commission (SEC) required firms to include a new "risk factor" section in their Form 10-K reports to discuss "the most significant factors that make the company speculative or risky", and the standardization and effectiveness of these factors are subject to strict supervision (SEC, 2005; Hope et al., 2016; Dyer et al., 2017). These risk disclosures by each financial institution are disclosed based on senior managers' risk perceptions from the actual operating conditions. We find that reputational risk is widely disclosed by financial institutions, and most of them appear in the form of "something will damage our reputation" or "if we fail to do something, our reputation will be damaged", which can directly reflect the reputational risk drivers. Therefore, by analysing the risk disclosures related to reputational risk in the whole financial industry, reputational risk drivers that aggregate senior managers' risk perceptions of the entire industry can be systematically identified.

However, the risk disclosure section in an annual report appears as a free-form textual segment, i.e., as completely unstructured text. Moreover, the amount of textual risk disclosure data for all financial institutions is enormous. Thus, using manual methods to identify the reputational risk drivers from the massive unstructured textual risk disclosures is almost impossible, which leads to subjective and incomplete results. Therefore, this paper innovatively introduces a text mining method, the Sentence-latent Dirichlet allocation (Sent-LDA) model proposed by Bao and Datta (2014), to systematically extract the topics that reflect the reputational risk drivers from massive risk disclosures in Form 10-K reports. To improve the accuracy of the identified risk drivers, we further modify the Sent-LDA model and demonstrate that the results of the improved Sent-LDA model in this paper are superior.

Overall, the drivers of reputational risk are still far from explicit, which seriously hinders proactive risk management and quantitative research. Therefore, the objective of this paper is to systematically identify the reputational risk drivers from the textual risk disclosures in financial reports by modifying a text mining approach. This paper contributes to the literature in three ways. First, we innovatively introduce the text mining method to extract the reputational risk drivers from the textual risk disclosures in financial reports. Compared with identifying risk drivers based on experts' judgements or by summarizing existing drivers in previous studies, this new method is more objective and effective because it can aggregate the risk perceptions of all senior managers in the financial industry. Second, we modified the Sent-LDA text mining method to increase its ability to handle reputational risk-related textual risk disclosures in financial reports. The improved Sent-LDA is verified to be much better than the original Sent-LDA model and can also be used to extract some specified information from other types of short texts. Third, we comprehensively identify the reputational risk drivers from large

amounts of textual risk disclosure data to largely extend the driver list of reputational risk with 7 newly discovered drivers.

The remainder of this paper is organized as follows. Section “Literature review” presents the literature review. Section “Methodology” provides the details of the methodology. Section “Empirical data” and section “Empirical results” present the empirical data and results. Section “Discussion” discusses the empirical findings, and section “Conclusion” concludes the paper.

Literature review

Reputational risk has become a subject of increasing importance for both managers and academics in recent years. Effective risk management and quantitative studies of reputational risk should be based on the identification of underlying risk sources (risk drivers in this paper) (Gatzert and Schmit, 2016). However, relevant studies on the identification of reputational risk drivers still lack theoretical frameworks, which has resulted in a fragmented understanding in general.

Most empirical studies generally consider that reputational risk usually follows operational risk events, which is verified by examining the market reactions to the announcement of operational loss events (Fiordelisi et al., 2013, 2014; Gillet et al., 2010; Heidinger and Gatzert, 2018). For example, Gillet et al. (2010), Fiordelisi et al. (2013), and Sturm (2013) verified that financial institutions suffer reputational losses following operational risk events and found that financial ratios (e.g., the price-to-book ratio, level of liabilities, and level of intangibles) influence the degree of reputational damage suffered. Some researchers have further studied the driving effects of different types of operational risk events on reputational risk. Specifically, Gillet et al. (2010) find that the most negative impact on returns occurs following the “internal fraud” operational risk type. Fiordelisi et al. (2014) and Zhu et al. (2021) also focused on the different event types of operational risk but found that the “external fraud” event type has the greatest impact on a company’s reputation. In addition, Confente et al. (2019) and Asthana et al. (2021) empirically studied the consequences of poorly managed data breaches on corporate reputation. Radanliev et al. (2021) focus on new types of data protection issues and cyber risks triggered by the Internet-of-Things. As relevant laws and regulations are still in their infancy, the leakage of consumer’s privacy without effective legal protection may also lead to severe reputational risks.

In addition, a small number of studies have realized that other drivers besides operational risk events may cause reputational risk. Therefore, some fragmented information about reputational risk drivers is mentioned, especially when discussing future research prospects. Specifically, Csiszar and Heidrich (2006) stated that reputation risks may be caused by associations with other parties’ misconduct. Sturm (2013) observed that rating downgrades should be considered when examining damage to banks’ reputations. Vig et al. (2017) proposed that fraud, corruptive activity or discrimination, security risks (including cyber risks, protection of personal data), product and service risks, and third-party risks should be considered reputational risk drivers. Barakat et al. (2019) note that future research should consider money laundering cases, product recalls, downsizings, and layoffs as risk drivers that might damage a reputation.

Only a few studies have attempted to construct a system of reputational risk drivers based on the information in the prior literature. Scandizzo (2011) proposed that reputational risk drivers can be classified into internal risk drivers (including corporate governance, human, human resources, community involvement, environment, and business behaviour) and external risk drivers (including project, counterparty, country, and sector risks). Gatzert et al. (2016) summarized the reputational risk

drivers from the literature to embed reputational risk in a holistic enterprise risk management (ERM) framework, and they adopted the results of Scandizzo (2011) and added risk drivers from prior literature, including changes in technology and social norms, layoffs, and downsizing (Love and Kraatz, 2009).

In summary, the literature has not yet reached a consensus on the risk drivers that may destroy a firm’s reputation. Systematically identifying reputational risk drivers remains an unresolved issue. Most studies generally consider operational risk events as the main drivers while ignoring damage from other risks. Although some studies mention risk drivers beyond the limits of operational risk, these analyses are incomplete and lack empirical evidence. Only a few studies have attempted to construct a system of reputational risk drivers; however, they can only summarize part of the fragmented information from prior studies. This paper finds that risk disclosures reported in Form 10-K reports contain valuable information related to the causes of reputational risks, so they can be used to address this issue from a new perspective. These risk factors are disclosed based on a company’s daily operating conditions and are subject to the strict supervision of the SEC (Hope et al., 2016; Dyer et al., 2017); thus, authenticity and reliability can be ensured to some extent. By collecting the risk disclosures of all Form 10-K reports, all risk perceptions of financial institutions’ senior managers can be aggregated. Therefore, a relatively comprehensive and authentic reputational risk driver system can be constructed, which is fundamental for further reputational risk management and quantitative studies.

Methodology

Overview. This paper uses a text mining method to identify the drivers of reputational risk from the textual risk disclosures in financial reports. An overview of the methodology is provided in this section. In 2005, the SEC started to require all listed companies to disclose their risk factors in the newly created Item 1A in the annual financial report called Form 10-K (SEC, 2005). In this section, each company discloses important risk factors that the company faces based on the senior managers’ experiences from the actual operating conditions. Each risk factor usually consists of a risk heading and a detailed explanation of the risk heading, with the risk heading representing an accurate summary of this risk factor.

Reputational risk is one of the risks faced by financial institutions, and it has been gradually valued and disclosed. We find that the risk headings related to reputational risk are usually in the format of “something will damage our reputation” or “if we fail to do something, our reputation will be damaged”, which precisely reflects the causes of reputational risk. Examples of risk headings related to reputational risk are shown in Table 1, and these four risk headings imply that “information security”, “legal and regulatory action”, “employee misconduct” and “fraudulent activity” are drivers of reputational risk. Thus, by collecting all risk headings related to the reputational risk of the whole financial industry, the drivers of reputational risk can be systematically identified based on the aggregated actual risk perceptions of all financial institutions’ senior managers.

Extracting drivers of reputational risk from large amounts of unstructured textual risk disclosure data based on manual methods is a nontrivial task since it is difficult and indeed infeasible to perform exhaustive text reviews, even of a moderately sized corpus (Bao and Datta, 2014). To address this issue, a text mining method named Sent-LDA was introduced to automatically obtain valuable information from text data. Sent-LDA is an unsupervised machine learning method and a topic model used to automatically discover a set of topics from text data

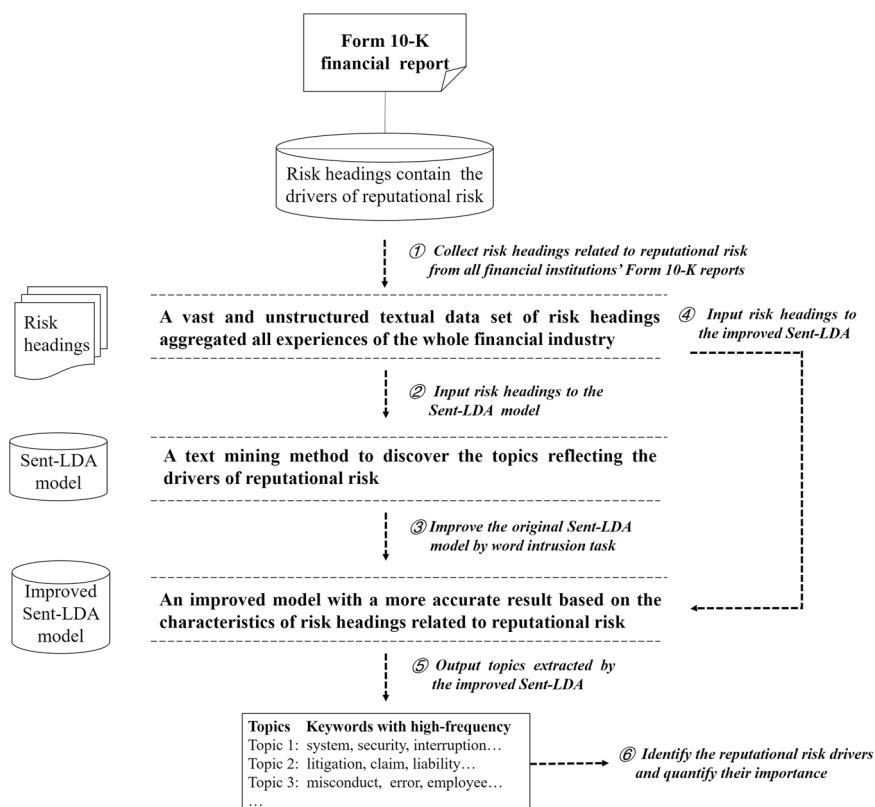


Fig. 1 The framework of the methodology. This paper uses a text mining method named “Improved Sent-LDA” to identify the drivers of reputational risk from the textual risk disclosures in Form 10-K financial reports. The five key steps in the methodology are presented in the figure.

(risk headings in this paper). By analysing the high-frequency words of each topic, the topic can be labelled, and the reputational risk drivers in the risk headings can be further identified.

However, we find that not all the topics extracted by the original Sent-LDA model can accurately indicate the drivers of reputational risk in the empirical study. Because most risk headings contain words such as “reputation”, “reputational”, “risk”, and “condition”, the high-frequency words of almost every topic extracted by the model are meaningless words that cannot reflect the drivers, while the keywords that truly reflect the reputational risk drivers are covered. To obtain a more accurate identification result, we further modify the original Sent-LDA based on the characteristics of reputational risk headings by designing a *word intrusion* task to recognize and remove these high-frequency but meaningless words and the modified approach is named the improved Sent-LDA. Thus, by inputting risk headings to improve the Sent-LDA method, the topics that more accurately represent the drivers of reputational risk can be outputted.

The framework of the methodology is shown in Fig. 1. Detailed descriptions of Sent-LDA and the improved Sent-LDA model are presented in sections “Sent-LDA model” and “Improved Sent-LDA model”.

Sent-LDA model

Principle of Sent-LDA. This paper adopts Sent-LDA, a topic model, to identify the drivers of reputational risk from textual risk disclosures. The topic model is an unsupervised machine learning technique for discovering the latent topics in text data by clustering the same semantic structures together (Bao and Datta, 2014; Glynatsi and Knight, 2021; Radanliev and Roure, 2021). The latent Dirichlet allocation (LDA) model proposed by Blei

et al. (2003) is one of the most popular topic models. The basic idea of the LDA model is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words (Blei et al., 2003). The LDA model assumes that each word in a document is generated by “selecting a certain topic with a certain probability and then selecting a certain word from this topic with a certain probability”, and it then uses a three-layer Bayesian probability model that corresponds to the “document–topic–word” structure. Based on a large amount of text data, the LDA model can automatically cluster words generated from one topic together (Brown et al., 2019). As a result, we can quickly identify latent topics (the reputational risk drivers in this paper) in a large amount of textual data by analysing the keywords of each topic.

The LDA model can be used to extract the latent main risk topics in Form 10-K. However, as shown in Table 1, one risk heading usually only describes one reputational risk driver in Form 10-K, which means that all words in one risk heading have a high probability of being extracted from the same topic. However, the LDA model assumes that a sentence contains multiple topics (Bao and Datta, 2014). To better fit the characteristics of risk factors in Form 10-K reports, Bao and Datta (2014) proposed Sent-LDA to improve the traditional LDA model. Sent-LDA inherits the basic concept of LDA and further adjusts its bag-of-words assumption as the rule that each sentence discusses only one topic (Bao and Datta, 2014). The empirical results show that Sent-LDA has a more accurate topic extraction effect for short text data, in which “a sentence usually contains only one topic” (Li et al., 2022).

Figure 2 presents a graphical model of the Sent-LDA model, which adds a sentence layer to the original hierarchy of the LDA model. Let M , N , K , V and S represent the number of documents in a corpus, the number of words in a document, the number of

Table 1 Examples of risk headings related to reputational risk in Section 1A of Form 10-K.

No.	Risk headings of risk factors relating to reputational risk	Implied risk drivers
1	We may be liable for invasion of privacy or misappropriation by others of our users' information, which could adversely affect our reputation and financial results. (From THINKORSWIM GROUP INC's Form 10-K in 2006)	Information security
2	Our exposure to legal liability is significant, and damages that we may be required to pay and the reputational harm that could result from legal or regulatory action against us could materially adversely affect our businesses. (From EDELMAN FINANCIAL GROUP INC's Form 10-K in 2010)	Legal and regulatory action
3	Employee misconduct, which is difficult to detect and deter, could harm us by impairing our ability to attract and retain clients and subjecting us to significant legal liability and reputational harm. (From LAZARD LTD's Form 10-K in 2016)	Employee misconduct
4	Fraudulent activity on our platform could negatively impact our operating results, brand and reputation and cause the use of our loan products and services to decrease. (From SENMIAO TECHNOLOGY's Form 10-K in 2019)	Fraudulent activity

topics, the vocabulary size, and the number of sentences in a document, respectively. The notations Dirichlet(.) and Multinomial(.) represent Dirichlet and multinomial distributions with parameter (·), respectively. The notation β_k is the V -dimensional word distribution for topic k , and θ_d is the K -dimensional topic proportion for document d . The notations η and α represent the hyperparameters of the corresponding Dirichlet distributions. Table 2 summarizes the meanings of the parameters. The generative process of Sent-LDA is shown below:

- (1) For each topic $k \in \{1, \dots, K\}$, draw a distribution over vocabulary words $\beta_k \sim \text{Dirichlet}(\eta)$.
- (2) For each document d ,
 - (a) draw a vector of topic proportions $\theta_d \sim \text{Dirichlet}(\alpha)$; and
 - (b) for each sentence s in document d ,
 - i. draw a topic assignment $z_{d,s} \sim \text{Multinomial}(\theta_d)$; and
 - ii. draw a word $w_{d,s,n} \sim \text{Multinomial}(\beta_{z_{d,s}})$ for each word $w_{d,s,n}$.

The Sent-LDA model assumes that a document is generated word by word through the above two steps. First, a topic is selected from the multinomial distribution of topics with the parameter θ_d . Second, for the selected topic, a word is chosen from the multinomial distribution of the words with the parameter β_k . In this paper, the risk factor disclosures related to reputational risk in a certain Form 10-K report are treated as one document, and the financial reports of all companies are consolidated into the document set. After setting the parameters that need to be preset and selecting the appropriate training algorithm, the Sent-LDA model can be trained by repeating the above two steps, and then it can obtain the parameters of the distributions based on the given corpus. The topic assignment z is one of the most important parameters output by the Sent-LDA model; it represents the classification results of risk headings and shows the topic to which each risk heading is classified. By calculating the probability of each word appearing in all sentences that cluster to a certain topic, the high-frequency words of the topic can be obtained. Then, the reputational risk driver reflected by the topic can be determined by analysing the meanings of the high-frequency words of each topic.

Parameter settings and estimations. Before applying Sent-LDA, two necessary parameters, the hyperparameter α and topic number k , need to be preset. α is usually set to $50/k$, and the topic number k is determined by the indicator of perplexity when conducting the Sent-LDA model (Bao and Datta, 2014; Wei et al., 2019). Perplexity is widely used to reflect the precision of the clustering effect for a different number of topics (Wei et al., 2019).

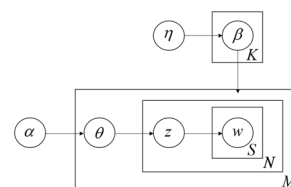


Fig. 2 Graphical model of Sent-LDA. The basic idea of the LDA model is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. Sent-LDA inherits the basic concept of LDA and further adjusts its bag-of-words assumption as the rule that each sentence discusses only one topic.

Denoting M , D_{test} , w_d and N_d as the document number, a test document set, the word w in document d and the total word number of document d , respectively, the *perplexity* is defined as follows:

$$\text{perplexity}(D_{\text{test}}) = \exp\left(-\sum_{d=1}^M \log p(w_d) / \sum_{d=1}^M N_d\right) \quad (1)$$

The values of perplexity for different topic numbers can be calculated via tenfold cross-validation (Blei and Lafferty, 2007). A lower perplexity over a held-out document is equivalent to a higher log-likelihood, which usually indicates better classification results (Bao and Datta, 2014). Perplexity monotonically decreases with the number of topics; therefore, it will continue to decrease as the number of topics increases (Bao and Datta, 2014). Thus, when the topic number is set to the total number of sentences, perplexity has a minimum value. However, the classification results at that time are meaningless. Therefore, when using perplexity to determine a suitable number of topics in the field of text mining, a stable point, i.e., the number of topics greater than or equal to the point where the perplexity value begins to converge, is preferred.

Another critical issue in using the Sent-LDA model is choosing an appropriate training algorithm to estimate the parameters given a certain corpus. The widely used training algorithms are the collapsed Gibbs sampling (CGS) and variational expectation maximization (VEM) algorithms (Blei et al., 2003). This paper chooses VEM to train the Sent-LDA model because it was proven to have a better performance than CGS for short text in Bao and Datta (2014). The idea of VEM is to obtain a lower bound for the log-likelihood of the observed data from a family of approximated distributions. Several parameters related to the VEM training algorithm need to be set in advance. Following Bao and Datta (2014), this paper sets the convergence bound for variational

Table 2 The meanings of parameters.

Parameters	Meanings	Parameters	Meanings
M	The number of documents.	z	Topic.
N	The number of sentences in the document.	θ	The K -dimensional risk factor type proportion for the document d .
S	The number of words in the document.	β	The V -dimensional word distribution for the topic.
K	The number of topics in the document.	α	The hyperparameters of Dirichlet distributions.
w	Word.	η	The hyperparameters of Dirichlet distributions.

inference to 10^{-8} , the maximum number of iterations of VEM to 1500, and the convergence bound of VEM to 10^{-5} .

Quantification of the importance of topics. The topics related to reputational risk drivers can be identified by the Sent-LDA model. In addition, Sent-LDA can quantify the importance of each topic by outputting the variable θ_{di} which shows the proportion of the number of sentences clustered in this topic to the total number of sentences in the document d (Wei et al., 2019). Denoting D as the number of documents, Importance is defined in Eq. (2).

$$\text{Importance}_i = \sum_{d=1}^D \theta_{di} \quad (2)$$

where Importance_i denotes the importance of the reputational risk driver i and is calculated as the proportion of the number of risk headings in this topic to the total number of risk headings. The greater the topic proportion, the more times the risk driver is disclosed. It is worth noting that a greater proportion does not mean that this driver leads to heavier reputational risks or that it is more important for all financial institutions. A higher Importance indicates that more financial institutions regard it as a source of reputational risk, which suggests that companies, especially companies that have not yet realized this driver, and regulators should pay more attention to this driver for more targeted reputational risk supervision and management.

Improved Sent-LDA model. Using the original Sent-LDA model, sentences that discuss the same topic (i.e., risk headings that disclose the same reputational risk driver in this paper) can be clustered into one topic. By analysing the high-frequency words in each topic, we can label each topic and further identify the specific risk driver. However, in the empirical study, some high-frequency words are observed, such as “reputation”, “reputational”, “business”, and “operation”, and they appear frequently on almost every topic. This means that the model has clustered sentences containing these meaningless words together instead of clustering sentences that reflect the same risk driver. Thus, these high-frequency words interfere with the identification of keywords that truly reflect risk drivers, such as “litigation risk”, “fraud”, and “misconduct”. Similar issues were also observed in the studies of Bao and Datta (2014) and Wei et al. (2019), where some noise words that cannot reflect companies’ risk profiles, such as “industry”, “condition” and “operation”, appear frequently in multiple topics. To address this issue, unlike directly applying the original Sent-LDA model in Wei et al. (2019), we improve the Sent-LDA model by designing an experiment to identify these high-frequency but meaningless words.

Specifically, this paper introduces the *word intrusion* task originally designed by Chang et al. (2009) to construct a stop word corpus of reputation risk drivers. The *word intrusion* task is to find an intruder in a given set of words, that is, a word that does not belong to the same category as the other words. In the experiments, when it is determined that the remaining words will make sense together after the removal of a certain word, then that

word is labelled as an intruder. In this article, if the high-frequency words in a topic are all related to the reputational risk, the intruder word can be easily found. However, when the given high-frequency words contain one meaningless word that cannot reflect the driver of reputation risk, it may be mistaken for an intruder word. Therefore, the purpose of this article using a *word intrusion* task is to find words that are mistakenly regarded as intruder words and label these words with high-frequency but that are unrelated to reputational risk as stop words.

To construct the experiment, the original Sent-LDA model is applied to identify the latent topics in the samples. Then, we randomly select a topic from the results and five words with the highest frequency within this topic. In addition, a word with low probability in this topic but a high frequency in other topics is randomly selected as an intruder. All six selected words are presented to the experimenters after shuffling the order. Four experts in the field of risk management are selected as experimenters, and they are asked to select one of the six keywords with the highest frequency for each topic as an “intruder” word that did not belong to it.

The model precision MK_m^k of the k th topic inferred by the model m in the *word intrusion* task is defined as the fraction of subjects that agree with a model:

$$MK_m^k = \frac{1}{S} \sum_s 1(i_{k,s}^m = w_k^m) \quad (3)$$

where $i_{k,s}^m$ is the intruder word selected by subject s among S subjects, w_k^m is the true intruder word, and $1(\cdot)$ is an indicator function that equals 1 if (\cdot) is true and 0 otherwise. To determine stop words, this paper defines the term error TE_t to calculate the rate of mistaken selections for the term t as follows:

$$TE_t = \frac{1}{N} \sum_s \left(1(i_{k,s}^m \neq t | t = w_k^m) + 1(i_{k,s}^m = t | t \neq w_k^m) \right) \quad (4)$$

where N is the total number of instances of term t in the *word intrusion* task. Finally, a term t with a high term error would be regarded as a stop word, which means that it is often mistakenly regarded as an intruder word. Furthermore, the original sentence in which the term t frequently appears in the corpus can be traced back and used to further analyse whether it is a stop word.

Based on the *word intrusion* task, the set of stop words for a specific corpus can be collected, and these stop words are removed when using the Sent-LDA model to analyse the corpus. This is an improvement in the process of executing the Sent-LDA model. The topics reflecting the reputation risk drivers can be more accurately identified by using the improved Sent-LDA model.

Empirical data

The empirical study is based on the textual risk disclosures from Form 10-K annual financial statements of listed financial institutions in the US. The Form 10-K filings are released in the Electronic Data Gathering and Retrieval (EDGAR) database on the SEC website. A certain company’s Form 10-K can be obtained by entering its unique identifier central index key (CIK) code. Thus, we first need to retrieve the CIK list of financial institutions.

Based on the Global Industry Classification Standard (GICS), financial institutions are classified into four subsectors, namely, “banks”, “diversified financials”, “insurance”, and “real estate”, which have GIC codes equal to 4010, 4020, 4030, and 4040, respectively. The CIK list of financial institutions can be obtained from the *Compustat* database with the corresponding GIC codes. In addition, the SEC required companies to disclose their risk factors in 2005 (SEC, 2005), and companies generally started to disclose them in 2006 (Wei et al., 2019). Thus, the period of data is from 2006 to 2019 in this paper. Finally, we collected 13,362 Form 10-K filings released from 1685 financial institutions.

Then, as stated in section “Overview”, the empirical study is based on the risk headings of risk factors disclosed in Item 1A of Form 10-K in the annual reports. Thus, a crawler programme is written to extract the risk headings. In the task of risk heading extraction, the Form 10-K filings of small companies that are not required to disclose risk factors are selected and removed. In addition, because a uniform template for risk disclosures is not available, it is not always possible to distinguish the heading from the explanation in some Form 10-K filings. To ensure the integrity of the data, we further manually examine the documents from which risk headings cannot be extracted by the programme. After removing them, 352,326 risk headings are collected, which

are extracted from 11,921 Form 10-K filings released by 1570 financial institutions from 2006 to 2019.

Finally, the risk headings that may contain reputational risk drivers are selected. In this paper, we consider risk headings containing the two keywords “reputation” and “reputational” to be risk factors related to reputational risk. This step is similar to that of Heidinger and Gatzert (2018), who approximated the awareness of reputational risk based on the frequency of the terms “reputation”, “reputation(al) risk” and “reputation(al) risk management” in financial statements. Finally, 7856 risk headings related to reputational risk from 4590 Form 10-K filings released by 828 U.S. financial institutions from 2006 to 2019 are selected to identify the risk drivers. Compared to Heidinger and Gatzert (2018), who used 820 annual reports from 82 firms over a period of 10 years to analyse the awareness of reputational risk, this paper utilizes a larger sample size and a longer period. The process of sample selection is summarized in Table 3.

In addition, after collecting the empirical data, we gain further insights into the awareness of reputational risk over time, as reflected in the financial institutions’ annual financial statements. From 2006 to 2019, the number of financial institutions that disclose risk factors shows a downwards trend, while the number of financial institutions that disclose risk factors related to reputational risk increased. The more intuitive trend of the proportion of financial institutions that disclose reputational risks over time is shown in Fig. 3. We further calculate the proportion of risk headings that disclose reputational risk to the total risk headings in all Form 10-K samples. These results are also presented in Fig. 3. Compared with other risk types, the awareness of reputational risk significantly increased from 2006 to 2019. Our results are consistent with the findings of Heidinger and Gatzert (2018), who show that an increasing number of financial institutions are paying attention to reputational risk.

Table 3 The process of sample selection.	
Sample selection	Count
<i>Company</i>	
All financial institutions (GICS code = 4010, 4020, 4030, 4040)	1685
(-) Small listed companies without risk disclosure	(-) 115
The number of financial institutions disclosing the risk factor	1570
The number of financial institutions disclosing reputational risk	828
<i>Form 10-K</i>	
All Form 10-K filings released by sample financial institutions	13362
(-) No Item 1A or unable to distinguish heading from the explanation	(-) 1441
The number of Form 10-K filings from which risk headings can be extracted	11,921
The number of Form 10-K filings disclosing reputational risk	4590
<i>Risk heading</i>	
The number of risk headings	352,326
The number of risk headings containing “reputation(al)”	7856

Empirical results

In this section, first, we apply the improved Sent-LDA model and validate its effectiveness over the original Sent-LDA model using both quantitative and qualitative methods. Systematic reputational risk drivers are identified by the improved Sent-LDA method, and the importance of the risk drivers is quantified. Then, this paper further discusses the universality and representativeness of risk drivers among subsectors and determines how the importance of each driver changes over time.

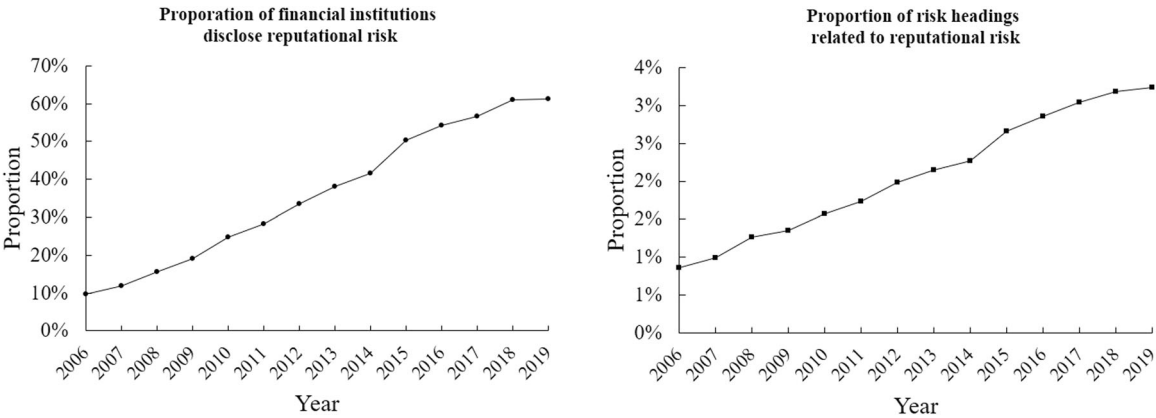


Fig. 3 Proportions of financial institutions and risk headings that disclosed reputational risk over time. The proportion of financial institutions that disclosed reputational risk has increased significantly, meaning that more financial institutions pay attention to reputational risk. The proportion of risk headings related to reputational risk to the total risk headings is also increased, which indicates that the awareness of reputational risk, compared with other risk types, is significantly improved.

Table 4 Stop words list generated by word intrusion experiment.									
No.	Stop words	No.	Stop words	No.	Stop words	No.	Stop words	No.	Stop words
1	may	11	subject	21	reduce	31	suffer	41	increase
2	should	12	additional	22	without	32	incur	42	including
3	could	13	general	23	reputation	33	face	43	become
4	will	14	impact	24	reputational	34	impair	44	operation
5	would	15	negatively	25	damage	35	asset	45	operational
6	result	16	affect	26	value	36	loss	46	impairment
7	adverse	17	certain	27	business	37	cause	47	failure
8	risk	18	related	28	financial	38	harmed	48	ability
9	factor	19	company	29	condition	39	required	49	affected
10	effect	20	cost	30	significant	40	exposed	50	resulting

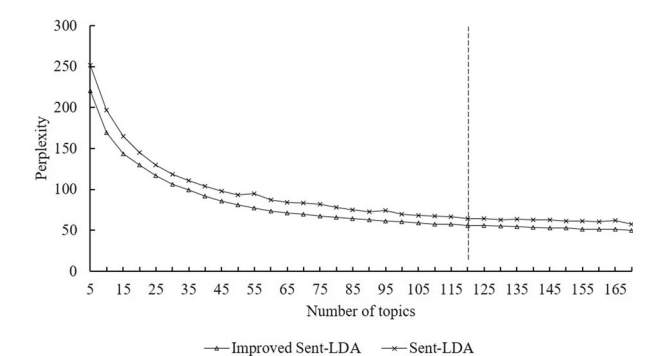


Fig. 4 Perplexity of different topic numbers obtained by the Sent-LDA and improved Sent-LDA models. A lower perplexity value corresponds to a better clustering effect of the model. Figure shows that all perplexity values obtained by the improved Sent-LDA model are smaller than those obtained by the original Sent-LDA model, which demonstrates that the improved Sent-LDA model has better clustering results.

Validation of the improved Sent-LDA model over the Sent-LDA model. This section examines the effectiveness of the improved Sent-LDA model from both quantitative and qualitative perspectives. Due to the existence of some high-frequency words that do not reflect the source of reputation risk, the original Sent-LDA model clusters sentences containing these meaningless words together instead of clustering sentences that reflect the same risk driver. To address this issue, the Sent-LDA model is improved based on the *word intrusion* task in the section “Improved Sent-LDA model” to recognize high-frequency but meaningless words as stop words. Based on the *word intrusion* task, the stop word list for the reputational risk-related textual data is summarized in Table 4. In the improved Sent-LDA model, these stop words are removed when inputting the same data source. Thus, risk headings that reflect the same risk driver can be clustered into one topic with less interference from meaningless words, which is an improvement in the process of the Sent-LDA model.

We validate the effectiveness of the improved Sent-LDA model based on both quantitative and qualitative methods. The quantitative validation method is based on the *perplexity* indicator. As stated in section “Parameter settings and estimations”, perplexity is widely used to reflect the clustering precision under different numbers of topics. A lower perplexity value corresponds to a better clustering effect of the model (Bao and Datta, 2014). Based on the same samples, we use the Sent-LDA model and the improved Sent-LDA model and calculate the values of perplexity under different topic numbers based on Eq. (1). The results are shown in Fig. 4. All perplexity values obtained by the improved Sent-LDA model are smaller than those obtained by the original Sent-LDA model, which demonstrates that our

improved Sent-LDA model has better clustering results according to this quantitative indicator.

The qualitative validation method is based on the word clouds of the identified topics. A word cloud is usually used to show the high-frequency words of each topic to display the identified topics (referring to the drivers of reputational risk in this paper) more intuitively. We examine whether the word clouds of the identified topics from the improved Sent-LDA model can more clearly reflect the reputational risk drivers. Examples of word clouds output by the original Sent-LDA are presented in Fig. 5. Each word cloud contains 25 words with the highest frequency, and a larger font indicates a higher probability of occurrence within this topic. Figure 5 shows that the topics extracted by the original Sent-LDA model usually contain some high-frequency words, such as “reputation”, “reputational”, “operation” or “business”; however, the words that reflect the specific risk drivers of reputation risk, such as “misconduct” and “litigation risk”, are difficult to identify. In contrast, from the word clouds output by the improved Sent-LDA, see Fig. 6 in section “Reputational risk drivers identified by the improved Sent-LDA model”, the reputational risk drivers can be easily recognized because there is no interference from the high-frequency but meaningless words. For example, the second-word cloud in Fig. 6 clearly shows that this topic is related to the reputational risk driver of “system interruption”. Through a comparison of the word clouds, it is also confirmed that the improved Sent-LDA model can remove the noise words to obtain clearer topics than the original Sent-LDA model.

The above quantitative and qualitative analyses both show that the improved Sent-LDA model has better clustering results than the original Sent-LDA model, which demonstrates the superiority of the improved Sent-LDA model. Therefore, the following empirical results are based on the improved Sent-LDA model.

Identification and discussion of reputational risk drivers
Reputational risk drivers identified by the improved Sent-LDA model. This section shows the process and results of reputational risk driver identification. The number of topics needs to be determined before applying the improved Sent-LDA model. The perplexity indicator is used to determine the range of the possible appropriate topic numbers, and then the optimal number of topics is selected and verified through manual inspection, thus ensuring the rationality of the results. The perplexity indicator is presented in Eq. (1) of section “Improved Sent-LDA model”. It is a monotone decreasing function with the number of topics, and the stable point—the number of topics greater than or equal to the point where the perplexity begins to converge—is preferred. When calculating perplexity, existing studies generally consider the time cost and do not calculate the perplexity value for every number of topics. For example, Bao and Datta (2014) and Wei et al. (2019) chose a step size of 10 and calculated the perplexity

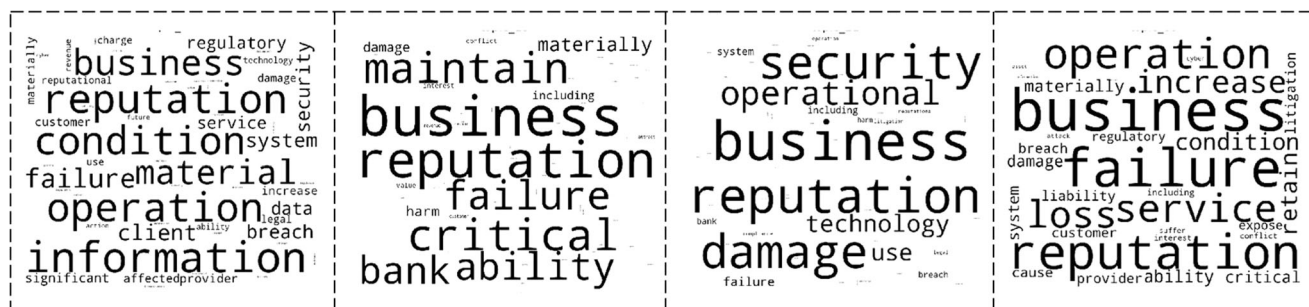


Fig. 5 Examples of word clouds output by the Sent-LDA model. Most topics extracted by the original Sent-LDA model contain some high-frequency words like “reputation”, “operation” or “business”; but the words that reflect the risk drivers of reputation risk cannot be identified.

from 10 to 100. This study goes further and calculates the perplexity of the model by varying the number of topics from 5 to 170 with a step of 5. The results are shown in Fig. 4, which shows that the values of perplexity begin to converge at ~120 and tend to be steady. We then check the clustering results when the number of topics is set to 110, 115, 120, 125, and 130 and find that when the topic number is 120, the clustering results are indeed better. Therefore, the number of topics is set to 120 in this empirical study.

By applying the improved Sent-LDA model, the risk headings reflecting the same reputational risk driver can be clustered into one topic from a semantic analysis perspective, and all the risk headings are finally clustered into 120 topics. For each topic, high-frequency keywords can be obtained by computing the probability of each word appearing in all sentences clustered into this topic. Then, by analysing the meanings of the keywords of each topic, the reputational risk driver that the topic reflects can be recognized and labelled. Although there are some automatic methods for labelling topics, when the research focuses on a specific area that requires professional knowledge, manual labelling methods are usually proven to have higher accuracy in most works with topic models (Bao and Datta, 2014). Thus, in this paper, we also manually determine the specific names of topics by analysing the high-frequency keywords of these topics.

During the labelling process, although the keyword lists of some topics are not exactly the same, they have similar high-frequency words and reflect the same type of reputational risk driver. These risk headings belong to the same topic, but they are assigned to different categories during the model implementation. This issue is very common in topic modelling, and these topics can be labelled the same and merged into one topic. After merging the topics reflecting the same driver, 13 risk drivers (the “others” topic is not included) are identified. It is worth noting that although 120 topics are output by the improved Sent-LDA model and 13 reputational risk drivers are finally obtained, it does not mean that a large amount of identified information was lost. The number of topics is set to 120 to find an appropriate topic number from a semantic analysis perspective so that reputational risk drivers that are not disclosed very frequently can also be identified as much as possible. In addition, some topics cannot clearly describe a certain type of risk driver or reflect multiple reputational risk drivers, and these topics are labelled “others”.

A word cloud is used to intuitively show the identified reputational risk drivers. The word clouds for 13 reputational risk drivers are shown in Fig. 6. Each word cloud contains 25 words with the highest frequencies, and a larger font indicates a higher probability of occurrence within this topic. By using the word clouds, it is easy to recognize the reputational risk drivers that a topic refers to. For example, the first-word cloud in Fig. 6 shows that the words “protect”, “information”, “confidential”, and “data” have larger font sizes, which means that most risk

headings clustered into this topic reflect that the failure to protect the security of information will lead to reputation damage. Therefore, the first risk driver is labelled as an “inadequate information safeguards” risk. In the eighth word cloud, the keywords with the largest font size are “interest” and “conflict”, so it is reasonable to label this risk driver as a “conflicts of interest” risk.

While previous studies mentioned some potential sources of reputational risk, they still lack a theoretical framework, resulting in a fragmented understanding of reputational risk in general. Therefore, after identifying and labelling the 13 reputational risk drivers, to describe them more clearly, we further concretize the meaning of each driver and present a corresponding example, which is shown in Table 5. All the examples are the risk headings selected from the “Risk factor” section in Form 10-K disclosed by the financial institutions.

For the 13 drivers, through detailed literature research, we summarize whether a reputational risk driver has been mentioned in previous studies. The results are also shown in Table 5, and “No” means that it has not been mentioned in previous studies to our knowledge. Specifically, the empirical results provide evidence that the following six drivers mentioned in prior studies are drivers of reputational risk from the perspective of financial institutions’ risk disclosures: “inadequate information safeguards” (Vig et al., 2017; Confente et al., 2019), “human error” (Scandizzo, 2011), “partners’ performance” (Csiszar and Heidrich, 2006; Scandizzo, 2011; Vig et al., 2017), “product and service problems” (Barakat et al., 2019; Vig et al., 2017), “fraud” (Gillet et al., 2010; Fiordelisi et al., 2014; Vig et al., 2017), and “loss of professionals” (Scandizzo, 2011; Gatzert et al., 2016). In addition, we find some risk drivers that are rarely mentioned in prior research on reputational risk, including “system interruptions”, “litigation risk”, “compliance risk”, “conflicts of interest”, “investment risk”, “credit risk” and “liquidity risk”. Our results are used to develop a reputational risk driver system from the perspective of financial institutions’ risk perceptions.

It is worth noting that while the reputational risk is considered a “risk of risks” and it appears that all activities of financial institutions are in some way exposed to it, comprehensive recognition of the drivers of reputational risk is still urgently needed. On the one hand, the systematic identification of reputational risk drivers helps financial institutions evaluate risk more comprehensively to accurately identify the main sources of reputational risk and thus better address proactive reputational risk management approaches. Furthermore, the definition of reputational risk is still quite broad and is considered an unobservable “cognition” (Gatzert et al. 2016), which is rather similar to the situation that inspired preliminary studies on operational risk. After the BCBS clarified the mechanism of operational risk, especially the causes and loss types, an increasingly systematic management framework and in-depth

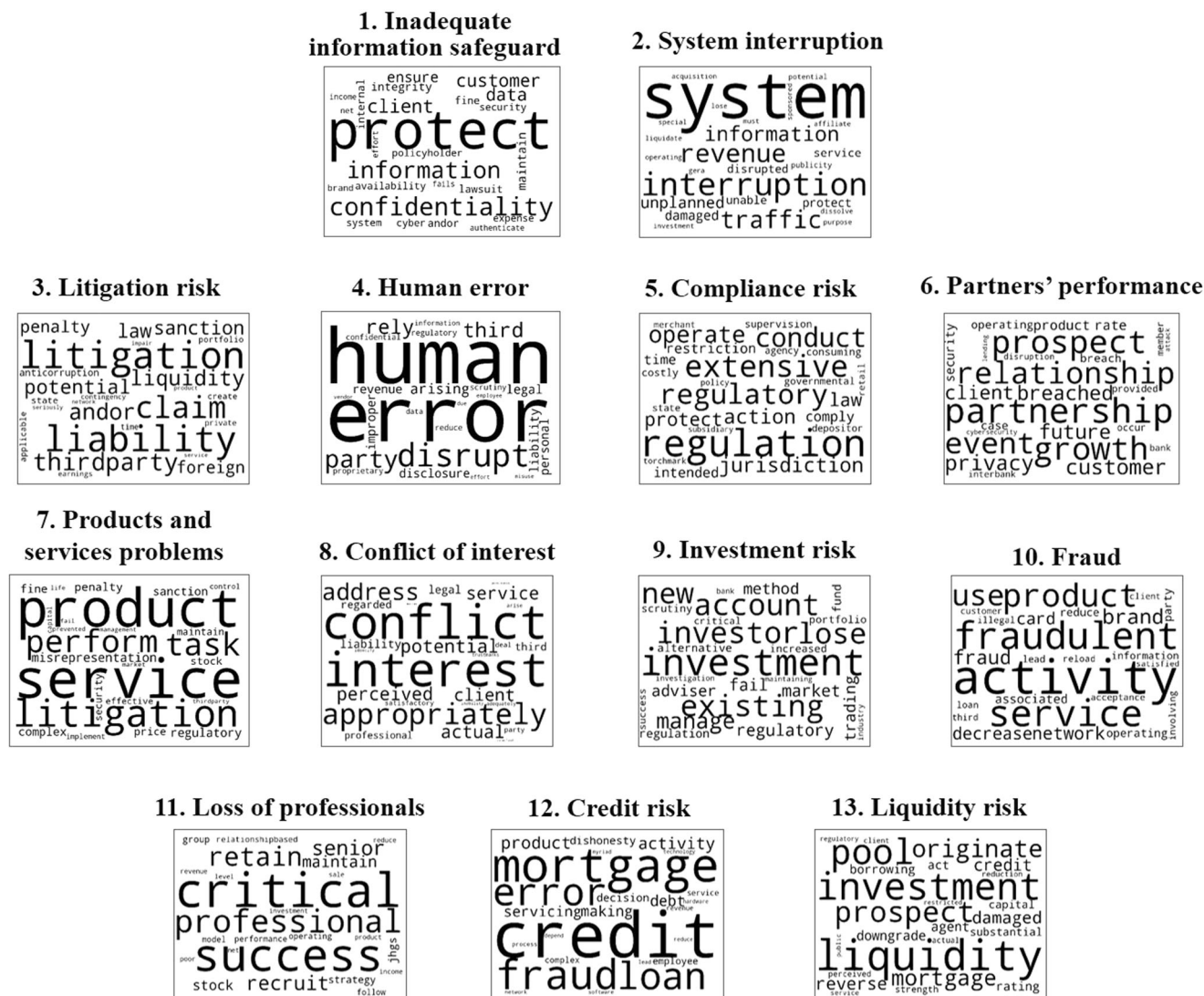


Fig. 6 Word clouds of 13 drivers of reputational risk. Figure presents the 13 topics that reflect the reputational risk drivers identified by the improved Sent-LDA. Each word cloud contains the words with the highest frequency, and the larger font indicating a higher probability of occurrence within this topic. The topics are labelled by analysing the high-frequency keywords in each word cloud.

quantitative studies on operational risk followed. This systematic identification of reputational risk drivers also provides a theoretical basis for further quantitative studies and more accurate measurements of reputational risk.

5.2.2 Quantification of the importance of reputational risk drivers. The reputational risk drivers can be identified by the improved Sent-LDA model. In addition, Sent-LDA can quantify the importance of each risk driver based on Eq. (2) in the section “Quantification of the importance of topics”, which shows the proportion of risk headings clustered in this risk driver to the total number of risk headings. The greater the proportion, the more times the risk driver is disclosed, which indicates that it is considered an important reputational risk driver by more financial institutions. The results are presented in Table 6.

Among the risk drivers, “inadequate information safeguards” is the most important reputational risk driver from the perspective of financial institutions’ risk perceptions; it accounts for 20.15% of the risk driver disclosures. Therefore, a considerable number of financial institutions realize that if they fail to protect their own or customers’ information, they will face reputational risk. The second most frequently disclosed risk driver is “system

interruptions”, which accounts for 15.19% of the risk driver disclosures, followed by “litigation risk”, which accounts for 10.07%. These findings indicate that financial institutions, especially those that have not yet noticed these important risk drivers, should pay more attention to the reputational damage caused by information leaks, system interruptions, and legal actions.

Most drivers of reputational risk are related to operational risk events. According to the definition of “the risk of loss resulting from inadequate or failed internal processes, people and systems, or from external events” (BCBS, 2009), we find that “product and service problems”, “human error”, “fraud”, and “loss of professionals” are related to operational risk events and account for 52.19% of the risk driver disclosures. As stated in the section “Literature review”, previous studies have assumed that financial institutions suffer reputational losses following operational risk events. The empirical results prove that operational risk events are indeed important sources of reputational risk from the perspective of financial institutions’ risk disclosures. The results also indicate that previously overlooked risk drivers are important. Specifically, the proportions of “litigation risk” and “compliance risk”, which are two topics related to legal risk,

Table 5 Definitions and examples of the drivers of reputational risk.

No.	Reputational risk drivers	Definitions and examples	Mentioned in previous studies
1	Inadequate information safeguards	Leakage and loss of internal or customers' data or information. e.g., "Loss of or failure to adequately safeguard confidential or proprietary information may adversely affect our reputation."	Vig et al. (2017), Confente et al. (2019), Radanliev et al. (2021), Asthana et al. (2021)
2	System interruptions	Losses arising from the disruption of business or system failures, examples include hardware and software failures, and telecommunication problems. e.g., "Our information systems may experience an interruption or security breach which could result in serious reputational harm to our business."	No
3	Litigation risk	The process of making or defending a claim in court. e.g., "Litigation is common in our businesses and may result in significant financial losses and harm to our reputation."	No
4	Compliance risk	Penalties for failure to comply with regulations. e.g., "We could be subject to regulatory investigations which could harm our reputation."	No
5	Human error	Employees' errors or misconduct from failed transaction processing or process management like miscommunication, data entry, maintenance or loading error, missing deadline or responsibility, etc. e.g., "Employee misconduct could harm us by subjecting us to significant legal liability regulatory scrutiny and reputational harm."	Scandizzo (2011)
6	Partners' performance	A series of risks encountered in the course of business cooperation with partners, such as dependence on third-party platforms and uncontrollability of franchise operation. e.g., "If partners fail to perform their contractual obligations on a project, we could be exposed to legal liability loss of reputation."	Csiszar and Heidrich (2006), Scandizzo (2011), Vig et al. (2017)
7	Conflicts of interest	The conflict of interest between stakeholders. e.g., "Failure to properly address conflicts of interest could harm our reputation or cause clients to withdraw funds."	No
8	Investment risk	The risk of uncertainties or losses caused by the investment subject to realizing its investment purpose in future operations and financial activities. e.g., "We may enter into new businesses make future strategic investments or acquisitions which may result in additional risks and uncertainties in our business and reputation."	No
9	Product and service problems	The products and services provided are defective or unsatisfactory to the customer. e.g., "If our clients or third parties are not satisfied with our services, we may face additional cost loss of profit opportunities and damage to our reputation."	Vig et al. (2017), Barakat et al. (2019)
10	Fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law, or company policy, excluding diversity/discrimination events by internal or external parties. e.g., "Mortgage fraud could result in significant financial losses and harm to our reputation."	Gillet et al. (2010), Fiordelisi et al. (2014), Vig et al. (2017)
11	Loss of professionals	The brain drain includes a wide range of professional and managerial personnel as well as top executives. e.g., "Our ability to retain our senior professionals and recruit additional professionals is critical to the success of our business and our failure to do so may adversely affect our reputation."	Scandizzo (2011), Gatzert et al. (2016)
12	Credit risk	The counterparties fail to meet their financial obligations under the contract, including default risk, bankruptcy, downgrade risk, and settlement risk. e.g., "Our earnings and reputation may be adversely affected if we fail to effectively manage our credit risk."	No
13	Liquidity risk	Failure to pay down or refinance debts, satisfy any cash obligations to counterparties or fund any capital withdraws, including funding liquidity risk and trading liquidity risk. e.g., "Failure to maintain an adequate level of liquidity could result in an inability to fulfill our financial obligations and could subject us to material reputation and regulatory risk."	No

account for 17.11%. Therefore, legal risk is also an important cause of reputation damage. However, few studies have focused on the relationship between legal and regulatory risk events and reputational risk. In addition, the driving effects of reputation from "partners' performance", "conflicts of interest", "investment

risk", "credit risk" and "liquidity risk" can be further studied in the future.

Differences in reputational risk drivers across different subsectors.
The improved Sent-LDA model can cluster risk headings that

Table 6 The proportions of the drivers of reputational risk.

No.	Risk driver	Proportion (%)	No.	Risk driver	Proportion (%)
1	Inadequate information safeguards	20.15	8	Conflicts of interest	3.44
2	System interruptions	15.19	9	Investment risk	3.28
3	Litigation risk	10.07	10	Fraud	2.34
4	Human error	7.79	11	Loss of professionals	2.19
5	Compliance risk	7.04	12	Credit risk	2.07
6	Partners' performance	4.76	13	Liquidity risk	1.03
7	Product and service problems	4.53	14	Others	15.85

reflect the same risk driver into one topic, and it can also trace back to determine the financial institution that disclosed this risk heading in a specific year. Thus, by analysing the subsector of the company that discloses each risk heading, this paper further discusses the differences in risk drivers across different subsectors. According to the GICS codes, the financial industry includes four subsectors of banks, diversified financials, insurance, and real estate. The numbers of Form 10-K filings disclosing reputational risk in the four subsectors are 3611, 3195, 951, and 99, and the numbers of risk headings disclosing reputational risk are 2205, 1308, 565, and 46, respectively. Due to the small sample size of the real estate industry, we only discuss the differences in the first three subsectors, i.e., banks, diversified financials, and insurance. The importance of each risk driver in each subsector is calculated, and we further compare them to distinguish the main risk drivers, which can help different financial institutions conduct more targeted reputational risk management and provide a warning to companies in this subsector that have not yet realized the importance of these risk drivers.

The proportions of risk drivers across the bank, diversified financial, and insurance subsectors are shown in Table 7. We focused on the top five risk drivers according to importance. The “inadequate information safeguards”, “system interruptions” and “litigation risk” drivers have the highest disclosure frequency in all subsectors. Among them, “inadequate information safeguards” is the risk driver with the largest proportion of disclosures in all subindustries. Therefore, information and system protection is essential to maintaining the reputation of various financial institutions, while certain lawsuits can also lead to negative perceptions of market participants.

Moreover, different subsectors present certain visible differences. Specifically, a considerable number of banks considered “partners’ performance” and “investment risk” to be the main drivers of reputational risk. Banks have close business contacts with other industries and should also pay more attention to the impact of the poor performance of third parties and the uncertainty of investments on their reputation. The diversified financial subsector includes a range of consumer- and commercially oriented companies that offer a wide variety of financial products and services. In addition to the high-frequency risk drivers above, diversified financials should pay more attention to the reputational risk caused by human error and the penalty for regulatory scrutiny.

For insurance, the two risk drivers “inadequate information safeguards” and “system interruptions” account for more than 50% of the risk driver disclosures, which indicates that the protection of data and systems is particularly important for the insurance industry. In addition, “fraud” is one of the important reputational risk drivers for insurance. Insurance fraud is a common risk in the daily operations of insurance companies. Failure to deal with fraud will also damage a company’s reputation.

In summary, “inadequate information safeguards” is the risk driver with the largest proportion in the disclosures of all

subindustries. The differences can reflect their distinct business characteristics, which have important implications for providing more targeted risk management among different subsectors. Specifically, banks should pay more attention to the reputation damage caused by “partners’ performance” and “investment risk”, while diversified financials and insurance should monitor “human error” and “fraud”, respectively.

Evolution of reputational risk drivers over time. As discussed in section “Differences in reputational risk drivers across different subsectors”, the improved Sent-LDA model can cluster risk headings that reflect the same risk driver into one topic, and it can also trace back to identify which year the risk heading was disclosed. Thus, this paper discusses the evolution of risk drivers over time to discover the important drivers that financial institutions should focus on. Figure 7 depicts the trend of the five risk drivers with the largest proportions (refer to Table 6) and four other risk drivers that have had an increasing trend in recent years. From Fig. 7, the trends of “litigation risk”, “human error” and “compliance risk” are stable, while those of “inadequate information safeguards” and “system interruptions” show a significant increase in recent years. Thus, an increasing number of financial institutions have disclosed in recent years that the failure to protect information and systems may damage their reputation. In addition, Fig. 7 also shows the annual changing trends of four other risk drivers with lower proportions but an increasing trend in recent years, including “partners’ performance”, “product and service problems” and “loss of professionals”.

We explain the reasons for the upwards trend of these risk drivers from the perspective of financial technology (fintech) development. In recent years, fintech has developed rapidly. Financial institutions have innovated upon the products and services provided by the traditional financial industry through various technological means (Li et al., 2020). Such progress has injected new vitality into financial development but brought new challenges and risks as well. As the business of financial institutions has shifted from offline to online, financial institutions have accumulated a large amount of customer behaviour and transaction data. However, information system management by these institutions has been unable to address network attacks; therefore, the data and system security measures are inadequate, thus leading to the risk that data will be leaked centrally (Gomber et al., 2018). Therefore, the trend of risk drivers indicates that an increasing number of financial institutions have focused on the reputational risk drivers of “inadequate information safeguards”, “system interruptions”, and “product and service problems”.

Moreover, the development of fintech promotes the increasingly complex connection between transaction entities and the increasing number of business contacts with external cooperative institutions (Li et al., 2020). Thus, the poor performance of third parties will also affect a company’s reputation. In addition, the development of fintech requires financial institutions to attract more technically competent professionals. The lack of staff retention and development means that companies may lag in the

Table 7 Proportions of reputational risk drivers across different subsectors.

No.	Banks	Proportion (%)	Diversified financials	Proportion (%)	Insurance	Proportion (%)
1	Inadequate information safeguards	18.91	Inadequate information safeguards	18.31	Inadequate information safeguards	30.91
2	System interruptions	8.03	Litigation risk	15.27	System interruptions	22.40
3	Partners' performance	7.34	System interruptions	12.11	Compliance risk	16.51
4	Litigation risk	5.93	Human error	12.99	Litigation risk	9.36
5	Investment risk	5.10	Compliance risk	9.01	Fraud	3.36
6	Product and service problems	4.85	Conflicts of interest	6.35	Human error	4.52
7	Compliance risk	2.99	Product and service problems	4.38	Conflicts of interest	1.68
8	Fraud	2.60	Loss of professionals	2.85	Partners' performance	1.68
9	Credit risk	2.16	Credit risk	2.35	Product and service problems	2.10
10	Loss of professionals	1.97	Partners' performance	2.07	Investment risk	1.16
11	Human error	4.07	Investment risk	1.97	Loss of professionals	1.05
12	Conflicts of interest	1.41	Fraud	1.82	Credit risk	1.05
13	Liquidity risk	1.30	Liquidity risk	1.72	Liquidity risk	0.00

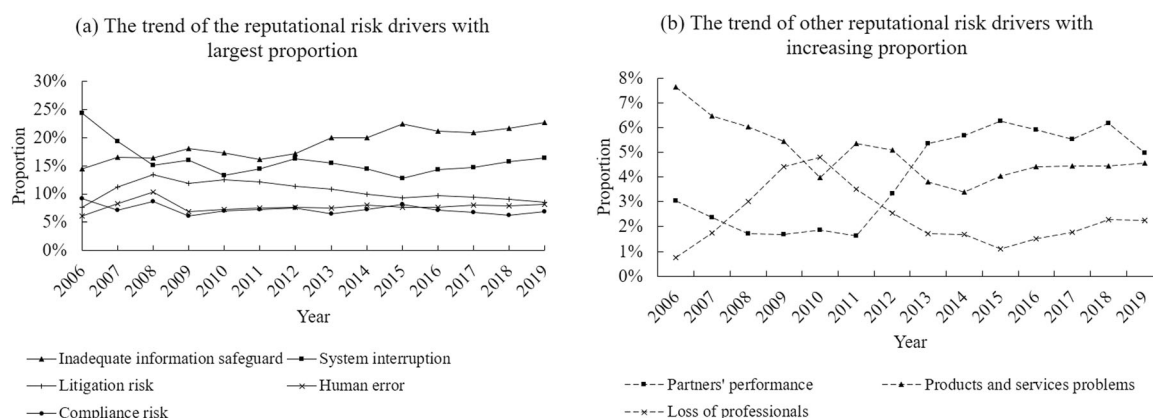


Fig. 7 The trend of risk drivers with the largest or increasing proportion. The trends of the five reputational risk drivers with the largest proportion are depicted. Among them, the “inadequate information safeguards” and “system interruptions” show a significant upward trend in recent years. Other risk drivers with lower proportions but an increasing trend in recent years are also presented, including “partners’ performance”, “product and service problems” and “loss of professionals”.

wake of fintech, which also causes reputational damage to some extent.

In summary, a significant upwards trend in recent years is observed in the risk headings of five risk drivers: “inadequate information safeguards”, “system interruptions”, “partners’ performance”, “product and service problems” and “loss of professionals”. To some extent, this means that the importance of these risk drivers is increasing. Thus, financial institutions should pay more attention to these drivers in future risk management and research, regardless of whether the company has disclosed such risks.

Discussion

The empirical analysis section applies the improved Sent-LDA model to identify the reputational risk drivers from the textual risk disclosures in financial reports. From the four aspects of the empirical analysis in the section “Identification and discussion of reputational risk drivers”, some important and interesting findings are derived. The findings, analysis, and corresponding managerial implications are further discussed and summarized as follows.

First, compared with other risk types, the proportion of disclosures related to reputational risk in Form 10-K reports has

significantly increased from 2006 to 2019, which means that more financial institutions are paying attention to this type of risk. This finding emphasizes the increasing awareness of reputational risk and is consistent with the viewpoints of Vig et al. (2017), Heidinger and Gatzert (2018), and Cornejo et al. (2019); the results in this paper provide new evidence from the perspective of corporate risk disclosures in financial reports. With the increasing attention to and severity of reputational risk, further exploration of reputational risk management strategies and quantitative research are urgently needed.

Second, 13 drivers of reputational risk are identified by the improved Sent-LDA model, and their meanings are defined. Among them, seven drivers were rarely mentioned in prior research, and they extend upon the current understanding of reputational risk drivers. Most of the existing studies discuss the reputational risk caused by “fraud”, “inadequate information safeguards”, “product and service problem” and “partners’ performance” events (Gillet et al., 2010; Scandizzo, 2011; Fiordelisi et al., 2014; Vig et al., 2017; Barakat et al., 2019; Confente et al., 2019). This study identifies them and discovers 3v3n other types of reputational risk drivers, i.e., “system interruptions”, “litigation risk”, “compliance risk”, “conflicts of interest”, “investment risk”, “credit risk”, and “liquidity risk”, which require attention in reputational risk management as well. In addition, prior research

generally quantifies reputational risk based on operational risk alone; thus, our findings provide a theoretical basis for further quantitative studies, especially for more accurate measurements of reputational risk.

Third, by analysing the importance of each risk driver based on the disclosure frequency, the proportion of reputational risk drivers related to operational risk events accounts for 52.19% of risk disclosures. This proves that operational risk events are indeed important sources of reputational risk and validates the rationality of existing studies that usually assume that reputational losses follow operational risk events (Fiordelisi et al., 2013, 2014; Sturm, 2013; Gillet et al., 2010; Heidinger and Gatzert, 2018). Notably, this study goes a step further and finds that the “inadequate information safeguards” and “system interruptions” are the most influential drivers among operational risk events and observes their significant upwards trends in recent years. Although many financial institutions have disclosed the reputational risk drivers related to operational risk, there are still some companies that have not yet noticed them. These companies should place more emphasis on reputational damage from information leaks, system interruptions, etc.

Finally, this study determines the common reputational risk drivers that all financial institutions need to pay attention to, as well as the drivers that different subsectors should focus on. The “inadequate information safeguards”, “system interruption” and “litigation risk” drivers have high disclosure frequencies in the banks, diversified financials, and insurance subsectors. This is consistent with the findings of Confente et al. (2019) that information protection is essential to maintaining the reputation of all financial institutions. Moreover, the results suggest that banks should also pay more attention to the reputation damage caused by “partners’ performance” and “investment risk”, while diversified financials and insurance should consider “human error” and “fraud”, respectively. The differences can reflect their distinct business characteristics and have implications for providing more targeted risk management among different subsectors.

Conclusion

The drivers of reputational risk are still far from explicit and obvious, which seriously hinders proactive risk management and further quantitative research on reputational risk. This paper identifies the reputational risk drivers from the massive textual risk disclosures in financial reports, which can aggregate senior managers’ risk perceptions of the whole financial industry. The Sent-LDA model is modified to make it more suitable for extracting the drivers from the reputational risk-related textual risk disclosure data. Both quantitative and qualitative analyses demonstrate the superiority of the modified Sent-LDA model over the original model.

The empirical analysis finds that the proportion of reputational risk-related disclosures in financial reports has significantly increased from 2006 to 2019, which reveals the increasing awareness of reputational risk in financial institutions. A total of 13 reputational risk drivers are identified and defined, among which 7 drivers were rarely mentioned in prior research and they extend upon the current understanding of reputational risk drivers. It also finds that operational risk events are the most important source of reputational risk, and the two drivers of “inadequate information safeguards” and “system interruptions” show significant upwards trends in recent years and should be highly valued. These findings deepen the knowledge of reputational risk from the aspects of its causes and drivers and can help financial institutions conduct more effective reputational risk management.

This paper is not without limitations. For example, the reputational risk drivers are identified from the textual risk disclosures in financial reports, and their importance is measured by calculating the proportions of their disclosures. This is one reasonable perspective because a higher proportion indicates that more financial institutions regard the driver as a source of reputational risk. However, a possible open question is whether the importance of reputational risk drivers can also be measured from other perspectives, for example, the severity of reputational risk losses caused by the drivers. Future studies can explore the importance of different types of reputational risk drivers by measuring the market participants’ reactions or stock price volatility caused by them.

The success of the methodology in this paper also sheds light on the application of unstructured textual data and text mining techniques in the fields of finance and accounting research. Through cutting-edge big data analysis techniques, such as the topic modelling method in this study, the information that is difficult to capture from traditional quantitative data can be extracted from text, images, or even audio and video data. The utilization of multi-source data in finance and accounting, especially in the financial risk management area, is an important direction that deserves exploration in the future.

Data availability

The dataset used in the current study is from the Electronic Data Gathering and Retrieval (EDGAR) database on the SEC website, available at: <https://www.sec.gov/info/edgar/siccodes.htm>.

Received: 24 November 2021; Accepted: 5 September 2022;

Published online: 16 September 2022

References

- Asthana S, Kalelkar R, Raman KK (2021) Does client cyber-breach have reputational consequences for the local audit office? *Account Horiz* 35(4):1–22. <https://doi.org/10.2308/HORIZONS-2020-018>
- Bao Y, Datta A (2014) Simultaneously discovering and quantifying risk types from textual risk disclosures. *Manage Sci* 60:1371–1391. <https://doi.org/10.1287/mnsc.2014.1930>
- Barakat A, Ashby S, Fenn P, Bryce C (2019) Operational risk and reputation in financial institutions: does media tone make a difference? *J Bank Financ* 98:1–24. <https://doi.org/10.1016/j.jbankfin.2018.10.007>
- Basel Committee on Banking Supervision (2009) Enhancements to the Basel II framework. Bank for International Settlements, Basel, Switzerland, <http://www.bis.org/publ/bcbs157.pdf>
- Blei DM, Lafferty JD (2007) A correlated topic model of science. *Ann Appl Stat* 1:17–35. <https://doi.org/10.1214/07-AOAS114>
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3:993–1022. <https://doi.org/10.5555/944919.944937>
- Brown NC, Crowley RM, Elliott WB (2019) What are you saying? Using topic to detect financial misreporting. *J Account Res* 58(1):237–291. <https://doi.org/10.1111/1475-679X.12294>
- Chang J, Boyd-Graber JL, Gerrish SM, Chong W, Blei DM (2009) Reading tea leaves: how humans interpret topic models. *Adv Neural Inf Process Syst* 32:288–296. <https://doi.org/10.5555/2984093.2984126>
- Comité Européen des Assurances, Groupe Consultatif Actuariel Européen (2007) Solvency II Glossary. Comité Européen des Assurances, Groupe Consultatif Actuariel Européen, Brussels, Belgium. https://www.actuaries.org/CTEES_SOLV/Documents/Solvency2_Glossary_2007_03.pdf
- Confente I, Siciliano GG, Gaudenzi B, Eickhoff M (2019) Effects of data breaches from user-generated content: a corporate reputation analysis. *Eur Manag J* 37:492–504. <https://doi.org/10.1016/j.emj.2019.01.007>
- Cornejo CP, Esther DQP, Delgado G, Juan B (2019) How to manage corporate reputation? The effect of enterprise risk management systems and audit committees on corporate reputation. *Eur Manag J* 37:505–515. <https://doi.org/10.1016/j.emj.2019.01.005>

- Csiszar E, Heidrich GW (2006) The question of reputational risk: perspectives from an industry. Geneva Pap R I-Iss P 31:382–394. <https://doi.org/10.1057/palgrave.gpp.2510096>
- Dyer T, Lang M, Stice-Lawrence L (2017) The evolution of 10-K textual disclosure: evidence from Latent Dirichlet Allocation. J Account Econ 64:221–245. <https://doi.org/10.1016/j.jacceco.2017.07.002>
- Eckert C, Gatzert N (2017) Modeling operational risk incorporating reputation risk: an integrated analysis for financial firms. Insur Math Econ 72:122–137. <https://doi.org/10.1016/j.insmatheco.2016.11.005>
- Fiordelisi F, Soana MG, Schwizer P (2013) The determinants of reputational risk in the banking sector. J Bank Financ 37:1359–1371. <https://doi.org/10.1016/j.jbankfin.2012.04.021>
- Fiordelisi F, Soana MG, Schwizer P (2014) Reputational losses and operational risk in banking. Eur J Financ 20:105–124. <https://doi.org/10.1080/1351847X.2012.684218>
- Gatzert N (2015) The impact of corporate reputation and reputation damaging events on financial performance: empirical evidence from the literature. Eur Manag J 33:485–499. <https://doi.org/10.1016/j.emj.2015.10.001>
- Gatzert N, Schmit J (2016) Supporting strategic success through enterprise-wide reputation risk management. J Risk Financ 17:26–45. <https://doi.org/10.1108/JRF-09-2015-0083>
- Gatzert N, Schmit JT, Kolb A (2016) Assessing the risks of insuring reputation risk. J Risk Insur 83:641–679. <https://doi.org/10.1111/jori.12065>
- Gillet R, Hubner G, Plunus S (2010) Operational risk and reputation in the financial industry. J Bank Financ 34:224–235. <https://doi.org/10.1016/j.jbankfin.2009.07.020>
- Glynatsi N, Knight V (2021) A bibliometric study of research topics, collaboration, and centrality in the iterated prisoner's dilemma. Hum Soc Sci Commun 8:45. <https://doi.org/10.1057/s41599-021-00718-9>
- Gomber P, Kauffman RJ, Parker C (2018) On the Fintech revolution: interpreting the forces of innovation, disruption, and transformation in financial services. J Manage Inform Syst 35:220–265. <https://doi.org/10.1080/07421222.2018.1440766>
- Gow ID, Wahid AS, Yu G (2018) Managing reputation: evidence from biographies of corporate directors. J Account Econ 66:448–469. <https://doi.org/10.1016/j.jacceco.2018.08.009>
- Heidinger D, Gatzert N (2018) Awareness, determinants and value of reputation risk management: empirical evidence from the banking and insurance industry. J Bank Financ 91:106–118. <https://doi.org/10.1016/j.jbankfin.2018.04.004>
- Hope OK, Hu D, Lu H (2016) The benefits of specific risk-factor disclosures. Rev Account Stud 21:1005–1045. <https://doi.org/10.1007/s11142-016-9371-1>
- Li J, Li G, Liu M, Zhu X, Wei L (2022) A novel text-based framework for forecasting agricultural futures using massive online news headlines. Int J Forecasting 38(1):35–50. <https://doi.org/10.1016/j.ijforecast.2020.02.002>
- Li J, Li J, Zhu X, Yao Y, Casu B (2020) Risk spillovers between FinTech and traditional financial institutions: evidence from the U.S. Int Rev Financ Anal 71:101544. <https://doi.org/10.1016/j.irfa.2020.101544>
- Love EG, Kraatz M (2009) Character, conformity, or the bottom line? How and why downsizing corporate reputation. Acad Manage J 52:314–335. <https://doi.org/10.5465/amj.2009.37308247>
- Radanliev P, Roure DD, Burnap P, Santos O (2021) Epistemological equation for analysing uncontrollable states in complex systems: quantifying cyber risks from the internet of things. Rev Socionetwork Str 15:381–411. <https://doi.org/10.1007/s12626-021-00086-5>
- Radanliev P, Roure DD (2021) Review of Algorithms for artificial intelligence on low memory devices. IEEE Access 9:109986–109993. <https://doi.org/10.1109/ACCESS.2021.3101579>
- Rhee M, Valdez M (2009) Contextual factors surrounding reputation damage with potential implications for reputation repair. Acad Manage Rev 34:146–168. <https://doi.org/10.5465/amr.2009.35713324>
- Rosenberg JV, Schuermann T (2006) A general approach to integrated risk management with skewed, fat-tailed risks. J Financ Econ 79:569–614. <https://doi.org/10.1016/j.jfineco.2005.03.001>
- Scandizzo S (2011) A framework for the analysis of reputational risk. J Oper Risk 6:41–63. <https://doi.org/10.21314/JOP.2011.094>
- Scholtens B, Klooster S (2019) Sustainability and bank risk. Hum Soc Sci Commun 5:105. <https://doi.org/10.1057/s41599-019-0315-9>
- Securities and Exchange Commission (2005) Securities and Exchange Commission final rule. 33-8591 (FR-75). <http://www.sec.gov/rules/final/33-8591.pdf>
- Sturm P (2013) Operational and reputational risk in the European banking industry: the market reaction to operational risk events. J Econ Behav Organ 85:191–206. <https://doi.org/10.1016/j.jebo.2012.04.005>
- Vig S, Dumitric K, Klopota I (2017) The impact of reputation on corporate financial performance: median regression approach. Bus Syst Res J 8:40–58. <https://doi.org/10.1515/bsrj-2017-0015>
- Wei L, Li G, Zhu X, Li J (2019) Discovering bank risk factors from financial statements based on a new semi-supervised text mining algorithm. Account Financ 59(3):1519–1552. <https://doi.org/10.1111/acfi.12453>
- Zhu X, Ao X, Qin Z, Chang Y, Liu Y, He Q, Li J (2021) Intelligent financial fraud detection practices in post-pandemic era. The Innovation 2(4):100176. <https://doi.org/10.1016/j.xinn.2021.100176>

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (92046023, 71971207), Fundamental Research Funds for the Central Universities, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at the University of Chinese Academy of Sciences.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Jianping Li.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022