Tourism companies' risk exposures on text disclosure

Jianping Li[a,c], Yuyao Feng[a,b], Guowen Li[a,b], Xiaolei Sun[a,b,*]

[a] *Institutes of Science and Development, Chinese Academy of Sciences, Beijing, China*
[b] *School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing, China*
[c] *School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China*

A B S T R A C T

Tourism is a risk-prone industry. But most studies focus on tourist risk perception while ignoring company risk exposure. As service providers, the companies play an important role in tourism activities, and systematically identifying the risks they face is vital to the development of the tourism industry. This paper attempts to identify tourism companies' risk exposures based on textual risk disclosure of financial statements. Using 51,008 risk headings of 255 public companies, we adopt Sentence-Latent Dirichlet Allocation (Sent-LDA) method to discover 30 risk exposures of the tourism industry. Further, we discuss the universality and industry representativeness of these risk exposures, as well as risk differences between different sub-industries and years. Findings can help stakeholders develop reasonable and timely risk management strategies.

Introduction

Tourism is a risk-sensitive industry. The intricate relationship between risk and tourism makes it necessary to have a thorough understanding of tourism risk factors. Although these factors cannot be completely eliminated, we can implement risk early warning through risk identification to reduce the risk perception (Sheng-Hshiung et al., 1997).

In the existing related studies, the vast majority focus on the risk perception of tourists, especially after the 9/11 attacks (Blake & Sinclair, 2003; Fuchs & Reichel, 2011; Kim et al., 2015; Mansfeld & Pizam, 2006; Wolff et al., 2019), and they have provided a reliable reference for tourism service optimization and tourist risk management. However, relatively few studies have focused on identifying the risks for tourism companies. Among the few studies that take companies as research samples, the topics are more about risk management, assessing a specific risk related to an event, or the influence of company characteristics (such as corporate social responsibility and advertising spending) on their risk status (Kim et al., 2013; Lee & Jang, 2011; Paraskevas & Quek, 2019; Park et al., 2017). While these studies provide specific recommendations for business management, they still lack a theoretical framework, resulting in a fragmented understanding of company risks in general.

As an important participant in tourism activities, research on the risk exposure of tourism companies is even more urgent. The financial crisis of 2008 showed that markets usually underestimate risk, which always results in a severe economic recession or bankruptcy (Hope et al., 2016). Therefore, identifying risk exposures that may cause business operations to suffer is effective in providing crisis warnings. For companies, risk exposure analysis not only enables them to fully understand the main threats within the industry and adapt their strategy accordingly, but also reduces information asymmetry with stakeholders and lowers the cost of equity (Barry & Brown, 1985; Myers & Majluf, 1984). For investors and tourists, they can have a further understanding of the risks

faced by companies, which are also what they may be subject to when buying products, services, or stocks. In this way, they can develop reasonable and timely risk-hedging strategies (Penela & Serrasqueiro, 2019).

However, due to the obstacle of data sources, this issue has not been addressed until recently. Penela and Serrasqueiro (2019) recognized the importance of risk identification for lodging companies and introduced textual risk disclosure data in Form 10-K, which have been proved to be a feasible and effective data source to evaluate company risk exposure into the tourism sector (Bao & Datta, 2014; Campbell et al., 2014; Wei, Li, Li, & Zhu (2019); Wei et al., 2019c). Based on 40 sample documents from 20 lodging firms, Penela and Serrasqueiro (2019) successfully identified 7 main risk themes in 2008 and 2016 respectively using Leximancer software, and compared the risk difference between these two distinct years. Their work pioneered research on company risk identification in the tourism sector. However, in order to get more representative conclusions, the scope of sample companies, time span and experimental methods still have room for improvement.

In this paper, we attempt to systematically identify the risk exposures of tourism companies from a holistic perspective. To this end, we introduce the Sentence Latent Dirichlet Allocation (Sent-LDA) model, which is an unsupervised clustering method that can effectively identify hidden knowledge from a large amount of text to analyze the textual risk disclosure data in Forms 10-K by all listed tourism companies during 2006–2019. The main contributions of our work are summarized as follows. First, we systematically identify the risk exposures of the tourism industry and compensate for the shortcomings of fragmentation in existing research. Second, through the division of sub-industries and comparison with other industries, we can understand the internal differences and industry representativeness of these risk exposures. Third, we further depict the risk evolution and fluctuation trends in different years, which can convey richer information from the time perspective. Finally, our work emphasizes the importance of risk identification of tourism companies and can provide a valuable reference for stakeholders to develop sound risk management strategies.

Research background

The intricate relationship between risk and tourism

With social progress and economic development, the disposable income and leisure time of tourists are increasing, making tourism activities increasingly popular (Yang & Nair, 2014). According to The World Bank, the number of international inbound and outbound tourists has exceeded one billion each year in the past decade and has maintained an average annual growth rate of approximately 5%. The increase in tourism activities has not only boosted employment and economic growth in the destinations but also exposed more potential risks attached to the tourism process (Cui et al., 2016).

Risk and tourism are intrinsically connected, and their relationship is intricate (Yang et al., 2017). First, tourism refers to traveling from a familiar place to relatively unfamiliar destinations, where people's tacit knowledge will be greatly reduced, which means that they are more likely to be targeted by risk (Boksberger & Craig-Smith, 2006; Williams & Baláž, 2015). Second, as a service product, tourism is intangible and heterogeneous (Roehl & Fesenmaier, 1992; Yang et al., 2017), making it necessary for visitors to participate personally without a reliable expectation, which indicates that the decision to travel is a risk in itself (Evans et al., 2012). Third, tourism is strictly time-bound. Because there is no concept of storing, service providers (tourism companies) are required to sell as many tickets or rooms as possible within a limited time (Williams & Baláž, 2015). Besides, since many attractions are evidently seasonal (Cuccia & Rizzo, 2011; Jang, 2004), service providers have to make full use of resources within a few months, and once this time passes, the industry enters a relatively depressed phase. For service recipients (tourists), their reservations are also time-bound, so they can and must use them within a fixed period (Dickinson & Peeters, 2014; Haldrup, 2004). In case of unforeseen circumstances, making changes to their tickets (or allowing the tickets to expire) results in huge monetary loss for tourists (Park & Jang, 2014). Finally, tourism is highly dynamic, and its development is interlinked with other relevant fields such as tourism marketing, transportation, accommodation, attractions, and tourism regulations (Leiper, 1979). Errors in any part will affect the normal operation of the entire chain, making tourism risk-sensitive.

A new way to tourism company risk identification

The company is one of the most important participants in economic activities. It is significant to comprehensively identify the risks it is facing. In the past, the main obstacle to company risk identification was data acquisition. Taking tourism research as an example, in the study of tourist risk perception, most data comes from questionnaires administered to tourists, but this method often fails when the research objects are companies. The disclosure of company risks directly affects the perceived reliability of stakeholders, and hence, few companies are willing to accept surveys from researchers.

In recent years, financial statements have garnered significant research attention. Scholars have discovered that these financial statements contain important information for stakeholders to have a deep understanding of the development and potential risks of companies (Zhu et al., 2016). Since financial statements are mandated by regulators and are issued by the company itself, the reliability can be guaranteed. In these reports, not only quantitative financial data but also textual data are included, the latter (named "Item 1A. Risk Factors") of which is relatively new content that the U.S. Securities and Exchange Commission (SEC) has requested from all listed companies in their annual financial statements on Forms 10-K since 2005. This part discusses "the most significant factors that make the offering speculative or risky" (U.S. Securities and Exchange Commission, 2005). It usually follows the description structure shown in Fig. 1—each risk disclosure is a description of a specific risk topic and consists of a risk heading and detailed explanation.

In practical applications, because the heading describes the same risk as the subsequent explanation, it is used for analysis more
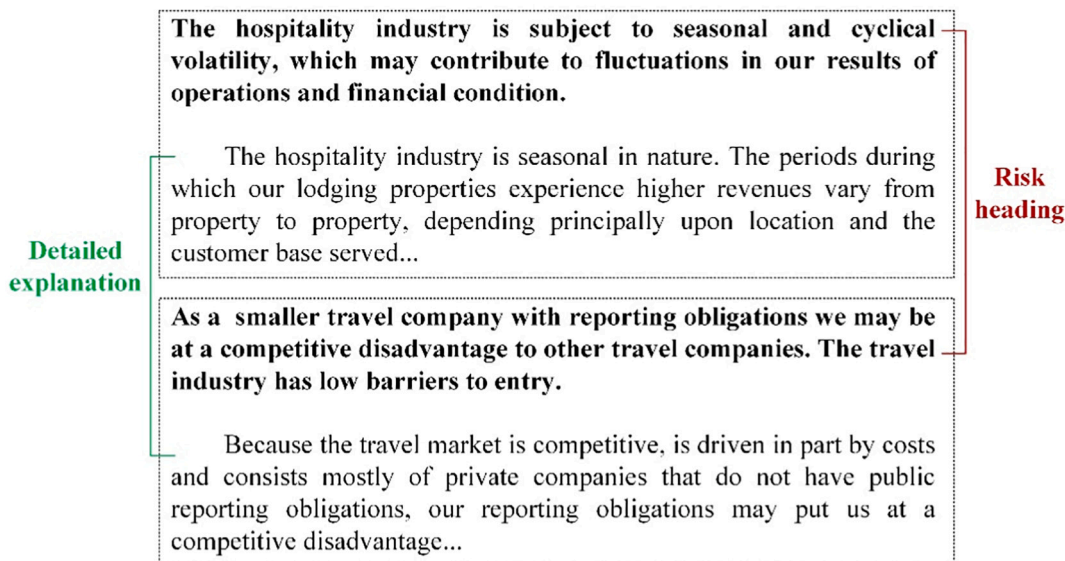
The hospitality industry is subject to seasonal and cyclical volatility, which may contribute to fluctuations in our results of operations and financial condition.

The hospitality industry is seasonal in nature. The periods during which our lodging properties experience higher revenues vary from property to property, depending principally upon location and the customer base served...

As a smaller travel company with reporting obligations we may be at a competitive disadvantage to other travel companies. The travel industry has low barriers to entry.

Because the travel market is competitive, is driven in part by costs and consists mostly of private companies that do not have public reporting obligations, our reporting obligations may put us at a competitive disadvantage...

Detailed explanation   *Risk heading*

Fig. 1. Description structure of risk factors in Form 10-K.

widely. The effectiveness of it has been well verified in many other works. For example, Campbell et al. (2014) used publicly listed companies in the U.S. to confirm that risk disclosure can effectively reflect the risks faced by companies. Immediately afterward, Bao and Datta (2014) used similar samples and innovatively introduced a model to successfully quantify 30 types of risks from company risk disclosure and verify that effective risk disclosure will significantly affect investors' risk perception.

As each industry has different characteristics and is affected by different business environments, their risk exposures are also different (Borde, 1998; Kim et al., 2012). In recent years, more professional risk exposure identification for different industries has begun to appear. For example, Wei, Li, Zhu, and Li (2019b)) and Wei et al. (2019c) identified and quantified the risk exposures of the energy industry and its nine sub-sectors and constructed a comprehensive and systematic risk system in the energy field. Similarly, Wei et al. (2019b) analyzed the banking industry, and Wang et al. (2019) analyzed the insurance industry. Their research verifies that there are obvious differences between risk exposures in different industries, which implies the necessity of company risk identification in the tourism industry.

Conceptual foundation

Tourism, hospitality, and tourism company

Tourism and hospitality are the most widely used terms in tourism and related fields. According to the definition of the *Oxford English Dictionary*, hospitality places more emphasis on serving guests with a sincere attitude, while tourism focuses on the elements of tourists, entertainment, and leisure (Simpson & Weiner, 1989). In a sense, the tourism industry can be regarded as the "subset" of the hospitality industry, because not all customers of hotels and restaurants are tourists (Keiser, 1998). However, in practical applications, neither is a single industry, and it is difficult to strictly distinguish them. First, in academic research, these two words are often interpreted as a whole (Kim et al., 2018; Litvin et al., 2008; Viglia & Dolnicar, 2020). Singal (2015) even deemed that hospitality industry and tourism industry had the same structural characteristics and named them as HT industry. Second, in the industry classification system, there is no separate tourism industry and hospitality industry. For example, the federal government's Standard Industry Classification (SIC) system divides each sub-sector involved in these two industries into different categories, where "retail-eating places" falls in the office of Trade & Services, whereas "hotels & motels" is classified as Real Estate & Construction (Keiser, 1998).

In the work of Leiper (1979), the author defined the tourist industry as "all those firms, organizations and facilities which are intended to serve the specific needs and wants of tourists". Draw lessons from his definition, we also use a broader perspective to define tourism companies in this paper—all HT companies with tourism as one of their main business, which can provide service to tourists. When conducting experiments, we adopt the industry categories provided by the Global Industry Classification Standard (GICS) and select all companies with "GIC Industries" equal to 253010, which are named "Hotels, Restaurants & Leisure" as a sample.

Risk, tourism risk, and tourism company risk

The future is unknown, this unpredictability can be understood as risk (Williams & Baláž, 2015). For tourists, tourism involves a destination that they are relatively unfamiliar with, so the absence of tacit knowledge of a relatively new environment will make

them feel more at risk than when they are shopping near their home. For tourism companies, the lack of tacit knowledge also plagues business development, especially for those trying to explore the international market (Liesch et al., 2002).

Another term with a similar definition is uncertainty, which is usually discussed in the same context as risk (Knight, 1921; Williams & Baláž, 2015). Both are caused by limited tacit knowledge, but a material difference exists between them (Polanyi, 2009). Knight (1921) defined risk as decision-making with available probabilities, and uncertainty as decision situations in which information is too blurred to be calculated by probabilities (Kravet & Muslu, 2013). Similar views exist in behavioral economics, where risks are probabilities for different outcomes (Tversky & Kahneman, 1974), implying that risks are uncertainty with known outcome probabilities. In other words, the decision-maker knows both the possible outcome and the probability of the outcome when making a decision (Zinn, 2004). While uncertainty is completely unknown, and the decision-maker cannot predict the probabilities of the outcomes, or even grasp all the possible outcomes (Knight, 1921).

With the development of the knowledge revolution and the advancement of information technology, people are forced to face a new dilemma of information overload (Lee & Lee, 2004), and unable to extract valid information from a large amount of information, which cause risk to be gradually replaced by uncertainty (Beck et al., 1992; Taylor-Gooby & Zinn, 2006). In most cases, risks and uncertainties are intertwined, and the same goes for tourism. While traveling, tourists really cannot know exactly whether the destination will have a riot, but they can estimate a reliable probability of "tomorrow will be a sunny day" based on the weather forecast. Hence, following Williams and Baláž (2015), Cui et al. (2016) and Sheng-Hshiung et al. (1997), this paper defines tourism risk as a general term for all the risks and uncertainties that arise from the uncertainty of the negative influence that the tourism subjects may have after making a decision, and the possibility of encountering various unexpected conditions while traveling. According to different participants, tourism risk can be further divided into the tourist risk, tourism company risk, and destination risk (Williams & Baláž, 2015). Accordingly, tourism company risk can be defined as the unpredictability of the negative impact on decision-making outcomes and the adverse market conditions that they may encounter in the daily production and operation process due to limited tacit knowledge.

Real risk, risk perception, and risk exposure

In terms of risk categories, Haddock (1993) divided risk into real (absolute) risk and perceived (subjective) risk. Real risk refers to risks that exist objectively in reality and does not change with the wishes of the subject. Therefore, strictly speaking, regardless of the attributes of subjects, the risks they face are the same and cannot be eliminated (Wong & Yeh, 2009).

The objective existence of risk cannot be denied, but it must be clear that risks mean different things to different people (Slovic et al., 1982). In most cases, people will not pay attention to all real risks but only focus on risks that they define as real and where the results are beyond their tolerance (Bauer, 1960; Budescu & Wallsten, 1985; Thomas & Thomas, 1928). For example, health risks are the most common form of risk. Yet, one does not constantly worry about falling ill. However, once a serious infectious disease occurs, one may become disturbed and perceive a strong health risk, that is, health becomes very important. Table 1 shows the definition of risk perception in several high-impact studies. Through these definitions, three elements of risk perception can be summarized: subjective feeling, objective assessment, and negative consequences (Cui et al., 2016; Glaesser, 2006). Therefore, in this paper, we define risk perception as the subjective feelings of tourists about the negative results beyond a tolerable range, caused by the various potential uncertainties in their traveling.

The level of risk faced by tourism companies is determined not only by the volatility of objective risks but also by the exposure and vulnerability to these risks (Cardona et al., 2012). Risk exposure is a quantitative concept in business management (Harvey, 1995). However, in our paper, we are concerned about facts rather than the extent, and we define risk exposure as the "risk perception" of a specific company, that is, the risk environment that the company believes may seriously affect future operations based on their subjective judgment of real risks.

Table 1
Related definitions of risk perception.

Reference	Concept	Definition
Cox and Rich (1964)	Perceived risk	The nature and amount of risk perceived by a consumer in contemplating a particular purchase decision
Dowling and Staelin (1994)	Perceived risk	The consumer's perceptions of the uncertainty and adverse consequences of buying a product (or service)
Sheng-Hshiung et al. (1997)	Perceived risk	The consumer's perceptions both of the uncertainty and the magnitude of the possible adverse consequences
Reisinger and Mavondo (2005)	Perceived risk	The individual's perceptions of the uncertainty and negative consequences of buying a product (or service), performing a certain activity, or choosing a certain lifestyle
Huang et al. (2008)	Perceived risk	The experience of anxiety or psychological discomfort resulting from tourist's spiritual and/or supernatural beliefs associated with the purchase and consumption of travel-related services for the destination
Adam (2015)	Perceived risk	A subjectively determined expectation of a potential loss in which some measure of probability can be attached to each possible outcome
Williams and Baláž (2015)	Risk perception	How predetermined notions about particular places, objects or activities, influence tourist behaviour
Kapuściński and Richards (2016)	Perceived risk	The processing of physical signals and/or information about potentially harmful events or activities, and the formation of a judgment about seriousness, likelihood and acceptability of the respective event or activity
Wolff et al. (2019)	Perceived risk	The subjective understanding of outcome severity weighted by outcome probability

In tourism risk research, the research on real risk is generally an after-the-fact study, and more inclined to analyze the impact of a specific risk event on the industry (including companies and tourists) and the countermeasures. The majority of risk perception researches focus on tourist risk perception, which is mainly conducted through questionnaires to assess the subjective judgments of tourists on the occurrence probability and severity of a certain risk. In contrast, the research on company risk exposure is relatively thin, which is exactly what this paper wants to make up for.

Methodology

Companies disclose risk factors through unstructured text data, which makes it difficult to identify risk types using traditional econometrics methods. To address this issue, we turn to text mining technology which can automatically obtain valuable information and knowledge from text data. In specific, a kind of text mining technology named topic model is employed, which is a statistical model used for discovering a set of topics that describes text data (risk headings in this study). By using the topic model, risk factors telling different topics can be classified into different topic clustering, which can be the final risk exposures after labeled based on feature words of it (Li, Li, Liu, Zhu, & Wei, 2020).

To identify the risk exposures disclosed in financial statements of tourism companies, this paper applies a topic model named Sentence Latent Dirichlet Allocation (Sent-LDA) proposed by Bao and Datta (2014), which is an extension of the original Latent Dirichlet Allocation model (LDA). LDA is a typical bag-of-words model developed by Blei et al. (2003), which specializes in processing a large amount of text data and is widely used in identifying potential topics in many research fields. The generation process of LDA can be shown below:

(1) For each topic $k \in \{1, ..., K\}$, draw a distribution over vocabulary words $\beta_k \sim Dirichlet(\eta)$, where η is the hyperparameter of the Dirichlet distribution.

(2) For each document $d \in \{1, ..., D\}$, draw a vector of topic proportions $\theta_d \sim Dirichlet(\alpha)$.

(3) For each word $w_{d, n}$ in document d, draw a topic assignment $z_{d, n} \sim Multinomial(\theta_d)$, and a word $w_{d, n} \sim Multinomial(\beta_{zd, n})$.

The concept of LDA is to infer the topic distribution and the word distribution based on the observed documents. Sent-LDA inherits the basic concept of LDA and further adjust its bag-of-words assumption as the rule that the boundary of sentence matters and each sentence discusses only one topic. In our research, the risk factor disclosure in Form 10-K usually follows the structure as shown in Fig. 1—every risk heading describes one specific risk type (topic), which is significantly aligned with the "one topic per sentence" rule of Sent-LDA. Therefore, we adopt Sent-LDA to identify the risk types of the tourism industry.

Another critical issue in using Sent-LDA is to choose the appropriate training algorithm to obtain the optimal parameters based on the observed documents. The widely used algorithms are Collapsed Gibbs Sampling (CGS) (Griffiths & Steyvers, 2004) and Variational Expectation Maximization (VEM) algorithm (Blei et al., 2003). In this paper, we employ VEM to train the Sent-LDA model, which has been proved to have a better performance than CGS in Bao and Datta (2014).

Before applying Sent-LDA, we need to extract and store the risk headings from the same financial statement as a document d, and all documents form a document set D. Then Sent-LDA will take the document set D as input and output a set of risk topics and further map each risk headings to a certain risk topic. Besides, Sent-LDA can also quantify the topic by output variable θ_d, showing the proportion of different topics in the document d. The greater the risk proportion, the more times the risk is disclosed, which can represent the universality of risk to a certain extent. As shown in Eq. (1), we can express the importance of the risk factor i by calculating the ratio of the number of risk headings under a certain risk topic to the total number of risk headings (Wei et al., 2019b; Wei et al., 2019c).

$$Importance_i = \sum_{d=1}^{D} \theta_{d_i}$$

(1)

Experiment analysis

Data preparation

To obtain as much data as possible, we first retrieve a list of sample companies with Central Index Key (CIK) from the *Compustat* database based on the selection criteria "GIC Industries" equal to 253010. Then, we further screen out the listed companies from the Electronic Data Gathering and Retrieval (EDGAR) database on the SEC's website according to CIK and locate the Forms 10-K they released over the years. In the following process, based on the text characteristics shown in Fig. 1 (the risk headings are represented by a special font and the different risk disclosures are segmented), a Python program is written to complete the task of heading extraction. Because there is no uniform template and prescribed format for risk disclosures, there will be considerable interference in the process of data crawling. To ensure the integrity and reliability of the data, we manually re-examine the documents that the program can recognize, and documents that cannot be processed are also manually classified. But there are still many "outliers". As shown in Table 2, we detail the process of sample selection and abnormal causes of these "outliers". Finally, we obtain 51,008 risk headings from 1870 Forms 10-K of 255 listed tourism companies during 2006–2019.

Parameter settings

While applying Sent-LDA, two necessary parameters—topic number k and hyperparameter α need to be set. Following Griffiths

Table 2
The process of sample selection.

Sample selection		Count
Company	All sample companies (GICS Industries = 253010, Year = [2006, 2019])	313
	Unlisted companies and small listed companies without risk disclosure	(58)
	The number of selected listed companies	**255**
Form 10-K	All Forms 10-K released by sample listed companies	2042
	Repeated release in the same year	(20)
	No Item 1A for some small companies	(143)
	Unable to distinguish heading from the explanation	(9)
	The number of Forms 10-K in the final sample	**1870**
Risk heading	The number of sentences (risk headings) in the final sample	**51,008**

and Steyvers (2004), α in this paper is set to $50/k$. The setting of topic number k is going to be discussed in detail later, since it has a great impact on clustering results. Besides, the convergence criteria of VEM also needs to be set. Following Bao and Datta (2014), this paper sets the convergence bound for Variational inference as 10^{-8}, the maximum number of iterations of EM as 1500 and the convergence bound of EM as 10^{-5}. The results of 30 times training on the same topic number show that these parameter settings of VEM can make the clustering results stable.

To determine the topic number, an evaluation index named perplexity is introduced. Perplexity can be easily understood as the uncertainty of the clustering results (Azzopardi et al., 2003). The smaller the perplexity, the higher is the accuracy of the model. In general, for a test set D_{test} of M documents, the perplexity can be expressed as follows, where N_d is the number of words in document d.

$$Perplexity(D_{test}) = exp\left(-\sum_{d=1}^{M} logp(w_d)/\sum_{d=1}^{M} N_d\right)$$

(2)

Perplexity is a monotone decreasing function with the number of topics, so it will continue to decrease as the number of topics increases. When k is set to the total number of sentences, we can get the minimum of the perplexity, but the classification at this time is meaningless. Therefore, when using perplexity to determine the value of k, we are not focusing on the optimal solution but the stable point—the number of topics greater than or equal to the point where the perplexity begins to converge is preferred. Fig. 2 shows the perplexity of our model by varying the number of topics from 2 to 200 with the step of 1, which is obtained based on 10-fold cross validation as described in Blei and Lafferty (2007). As seen in Fig. 2, when the number of topics reaches 120, perplexity tends to be steady.

Depending on the perplexity, it can roughly know the range of the optimal number of topics but still cannot determine an exact value. Grimmer and Stewart (2013) pointed out that to determine the number of topics, not only the statistical fit (perplexity in this study) but also the substantive fit need to be considered. Therefore, this paper further intends to introduce a word intrusion task to further optimize the model.

The purpose of the word intruder 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 other words. When the experimenters think that removing a certain word will make the remaining words make sense together, then this word can be an intruder. Although the intruder task can help us determine the number of topics, we do not need to conduct it on all models with $k \geq 120$. Because the intruder task is a labor-cost process, and when the number of topics increases little, the clustering result does not have a significant change. To depict the clustering similarity of two models with different topic numbers, we define the similarity index.
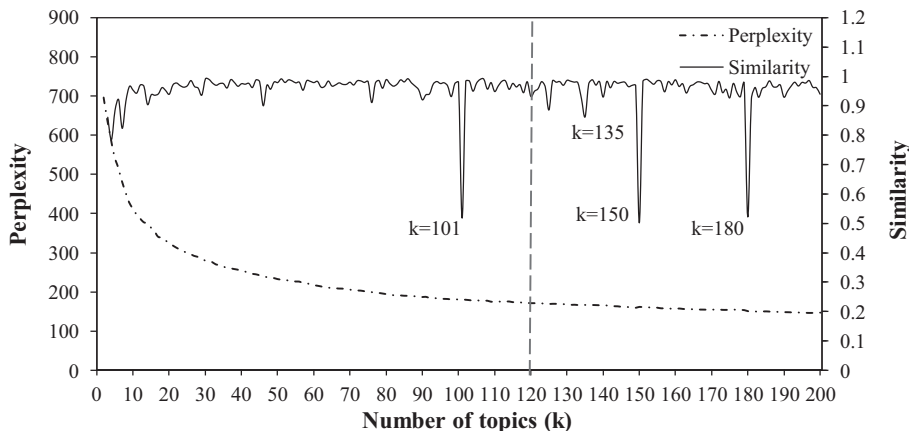

Fig. 2. The perplexity and similarity as the topic number grows.

Denoting the model with k topics as m_k, the set of risk topics generated by m_k as T_k, and the topic i of T_k as $T_{k,\ i} \in \{1, ..., k\}$, the model similarity of m_k and m_{k-1} can be expressed as:

$$MS(m_{k-1}, m_k) = \frac{1}{k-1} \sum_{j=1}^{k-1} max(Similarity(T_{k-1,j}, T_{k,i}^{j-1}))$$
$$i \in \{1, ..., k - j + 1\}$$
(3)

Among them, T_k^{j-1} represents the topic set subtracting $j - 1$ topics that maximize $Similarity(T_{k-1,\ a}, T_{k,i}^{a-1})$, $a \in \{1, ..., j - 1\}$ from T_k. $Similarity(T_{k-1,\ j}, T_{k,i}^{j-1})$ is obtained by calculating the cosine value of the word distribution vector $T_{k-1,\ j}$ and $T_{k,i}^{j-1}$ through the Vector Space Model (VSM), where the word distribution vector consists of the top 30 words with the highest frequency. Fig. 2 shows the model similarity with k ranging from 3 to 200. As shown in the figure, after the perplexity tends to be stable ($k \geq 120$), only when the values of k are 101, 130, 150, and 180, there are obvious differences between the clustering results of adjacent models. Therefore, we only conduct intruder task on models within a certain interval after the perplexity starts to stabilize and several models with similarity changes obviously, that is $k \in [120, 130] \cup \{101, 135, 150, 180\}$.

To construct the experiment, we follow the process designed by Chang et al. (2009). First, we randomly select a topic and five words with the highest frequency within this topic. At the same time, a word with low probability in this topic but high frequency in other topics is randomly selected as an intruder. Then, all six selected words are presented to the experimenters after shuffling the order. The accuracy of the model can be defined by Eq. (4), which indicates the cognitive consistency between the experimenters and classification accuracy of the model.

$$MP_k^m = \frac{1}{S} \sum_s I(i_{k,s}^m = w_k^m)$$
(4)

where the total number of people participating in the experiment is defined as S, $i_{k,s}^m$ and w_k^m represent the intruder word selected by the experimenter s and the real intruder word, respectively. When the two words are the same, the indicator function $I()$ is equal to 1, otherwise, it is equal to 0. The precision of the model m can be expressed as the average of MP_k^m over topics.

In our work, the word intruder tasks are performed by four professionals. Fig. 3 shows the boxplot of model precision. We can clearly see that the model has the highest precision when the number of the topic is set to 125. So, we finally choose 125 as the optimal value of the topic number k.

Experimental result

By running Sent-LDA with topic number as 125, we can get 125 risk exposure topics. To clarify what each topic expresses, these 125 topics need to be labeled.

Although there are some automatic labeling methods (Mei et al., 2007), when the research focuses on a specific area that requires professional knowledge, these methods might be powerless. In general, in most works with the topic model, manual labeling methods are more popular for ensuring higher accuracy. To execute our task, we have designed a manual labeling process. First, we use the 32 risk factors defined by Huang and Li (2011) and Bao and Datta (2014) as our tag list. The experimenters need to use their expertise to match the 125 identified topics with the 32 categories. When the experimenters feel that a topic does not meet any of these 32 categories, they should label it as "other risk types", of which label will be determined later. To complete this experiment, we select four experts in the field of risk management as participants. Before the experiment, they are provided with a thorough explanation of our study. Then, they independently complete their work based on the high-frequency words under each topic and discuss and
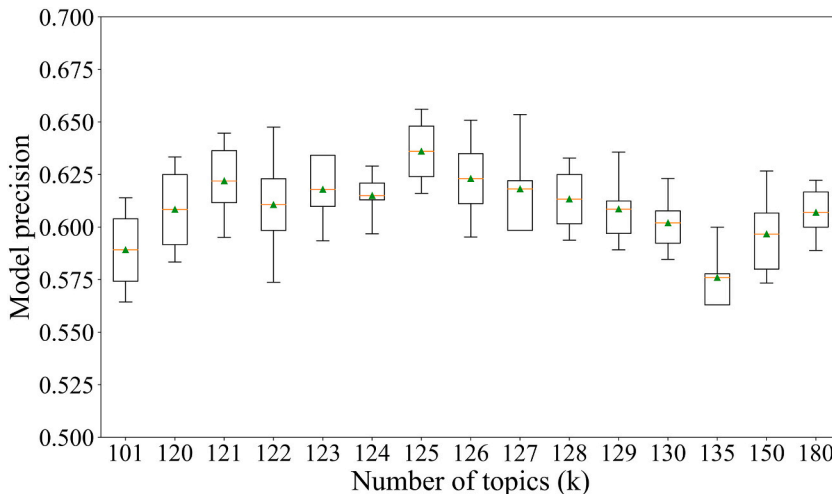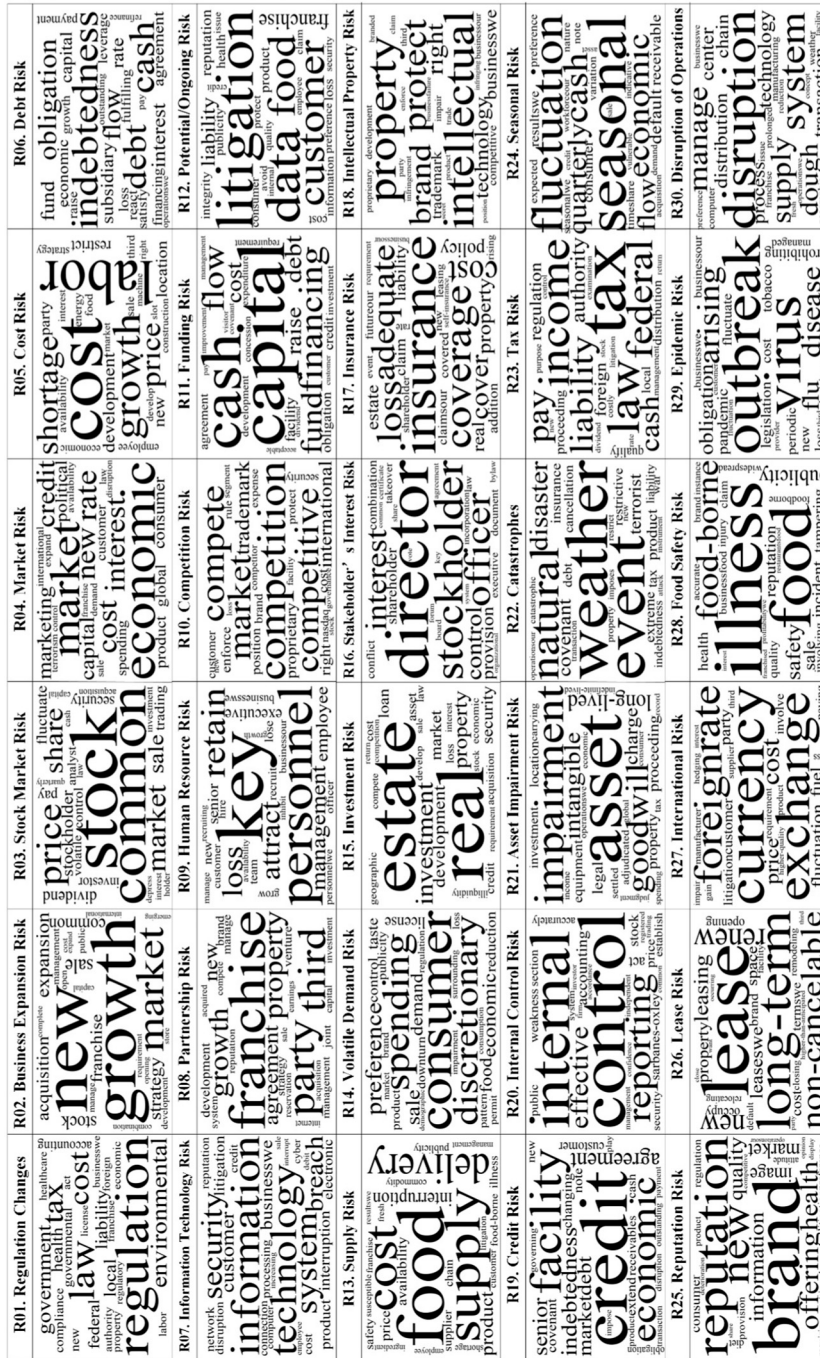

Fig. 3. The model precision of word intruder task.

Fig. 4. Word clouds of identified risks.

Table 3

Description of tourism company risk exposures.

No.	Risk exposure categories	Proportion	Whether it appeared in other works	
			Bao and Datta (2014)	Huang and Li (2011)
R01	Regulation changes	11.53%	Y	Y
R02	Business expansion risk	10.73%	Y	Y
R03	Stock market risk	9.79%	Y	Y
R04	Market risk	6.27%	Y	Y
R05	Cost risk	4.79%	Y	N
R06	Debt risk	4.50%	Y	N
R07	Information technology risk	3.89%	Y	Y
R08	Partnership risk	3.72%	N	N
R09	Human resource risk	3.62%	Y	Y
R10	Competition risk	3.51%	Y	Y
R11	Funding risk	3.27%	Y	Y
R12	Potential/ongoing lawsuits	3.22%	Y	Y
R13	Supply risk	2.41%	Y	Y
R14	Volatile demand risk	2.38%	Y	Y
R15	Investment risk	2.34%	Y	N
R16	Stakeholder's interest risk	2.18%	Y	Y
R17	Insurance risk	1.98%	N	N
R18	Intellectual property risk	1.79%	Y	Y
R19	Credit risk	1.77%	Y	N
R20	Internal control risk	1.69%	N	N
R21	Asset impairment risk	1.64%	N	N
R22	Catastrophes	1.51%	Y	Y
R23	Tax risk	1.44%	Y	N
R24	Seasonal risk	1.24%	N	N
R25	Reputation risk	1.22%	N	N
R26	Lease risk	1.05%	N	N
R27	International risk	0.92%	Y	Y
R28	Food safety risk	0.79%	N	N
R29	Epidemic risk	0.58%	N	N
R30	Disruption of operations	0.43%	Y	Y
R31	Others	3.82%	/	/

optimize the experimental results together. Finally, by comparing the results of the three participants, we find that their results are highly consistent, proving that risks identified by the Sent-LDA are interpretative.

Finally, 125 topic labels are determined. Because it is difficult to describe 125 topics in detail, following Dyer et al. (2017) and Li et al. (2020), we further divide them into 30 broader risk categories, which are shown in Fig. 4 with labels. Each graph contains 25 words with the highest frequency, and the larger font indicating a higher probability of occurrence within this category. For example, in R02, the three most common words are "new", "growth" and "market". Two points need to be explained. First, to reduce the noise, we have added some common words that have high frequency but no practical meaning (such as preposition, noun related to tourism or risk, degree adverb, etc.), punctuation, and some special characters to stop word list before the experiment. Second, due to the existence of some sentences that do not clearly describe a certain type of risk and describe multiple types of risks, it is difficult for us to accurately label some clustering results, which are classified as "others" categories (Bao & Datta, 2014). So, we finally get 31 risk categories, and the description of them is shown in Table 3.

Focusing on the tourism industry allows us to discover some new risk exposures, which have been ignored in previous works using all listed companies as samples (Bao & Datta, 2014; Huang & Li, 2011). These emerging risk exposures can reflect the characteristics of tourism to a certain extent. For example, seasonal risk can reflect the obvious seasonality of tourism activities, which further verifies "tourism is strictly time-bound" mentioned in the research background. Epidemic risk can also present the fact that tourism as a consumer cyclical industry is particularly sensitive to a public health emergency, which has been verified by COVID-19 (Duan et al., 2020; Yang et al., 2020). The definitions and examples of these new risk exposures are shown in Table 4, and others can refer to the works of Huang and Li (2011) and Bao and Datta (2014).

Discussion

To provide more specific information, we will discuss the identified risk exposures from the internal, external, and time perspective in this section.

The industry representativeness of identification risk exposures

To verify whether our risk identification results are representative of the tourism industry, we compare the top 5 and industry representative risk exposures with other industries, including energy and banking. As shown in Table 5, regulation risk is in an

Table 4

Definitions and examples of new identified risk exposures.

No.	Name	Definitions and examples
1	Partnership risk	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., "Investing through partnerships or joint ventures decreases our ability to manage risk."
2	Insurance risk	The amount of insurance cannot cover company losses, or losses are not covered by insurance, which could reduce anticipated profits. e.g., "Our current insurance coverage may not be adequate, our insurance premiums may increase, and we may not be able to obtain insurance at acceptable rates or at all."
3	Internal control risk	Defects in internal management and control. e.g., "If we fail to maintain an effective system of internal controls, we may not be able to accurately report our financial results."
4	Asset impairment risk	Risks associated with a sudden decline in asset value, where assets include both fixed assets and intangible assets. e.g., "Impairments of assets or goodwill may increase the risk of default under our debt obligations."
5	Seasonal risk	Seasonal fluctuations in results. e.g., "Our business is highly seasonal, and unfavorable weather conditions can adversely affect our business."
6	Reputation risk	The risk of negative evaluation of the company due to omissions in their daily operations, such as customer information leakage. e.g., "Improper conduct of our employees could harm our reputation and adversely affect our business operations."
7	Lease risk	Financial and contractual issues related to leases. e.g., "We may be locked into long-term and non-cancelable leases that we want to cancel, and may be unable to renew leases that we want to extend at the end of their terms."
8	Food safety risk	Adverse effects related to food safety affairs. e.g., "Food safety and food-borne illness concerns may have an adverse effect on our business."
9	Epidemic risk	Obstacles to the normal operations caused by health concerns arising from outbreaks of diseases. e.g., "A regional or global health pandemic could severely affect our business."

Table 5

Risk exposures comparison between industries.

Industry	Top 5 risk exposures		Industry representative risk	
	Risk topics	Proportion	Risk topics	Proportion
Energy	Regulation	12.38%	Drill success	5.07%
	Energy price	7.30%	Pipeline	2.67%
	Partner and shareholder interest	6.78%	Exploration	2.51%
	Merger & acquisition	5.96%	Derivatives	1.92%
	Stock market	5.76%	Power transmission	1.55%
Banking	Regulation	16.64%	Asset value fluctuation	7.22%
	Strategy	14.89%	Development of borrowers	2.71%
	Management operation	13.02%	Financial expense	3.21%
	Loan loss	9.38%	Financial institutions interaction	1.38%
	Asset value fluctuation	7.22%	Country credit rating	0.16%
Tourism	Regulation changes	11.53%	Partnership risk	3.72%
	Business expansion risk	10.73%	Demand fluctuation risk	2.38%
	Stock market risk	9.79%	Seasonal risk	1.24%
	Market risk	6.27%	Food safety risk	0.79%
	Cost risk	4.79%	Epidemic risk	0.58%

Note: The results of the energy and banking industry are quoted separately from the work of Wei et al. (2019c), and Wei et al. (2019b).

absolute leading position in all these three industries, and stock market risk is also a relatively common one. The differences between industries are also very obvious. For the energy industry, risk factors related to energy price, exploration, and transmission are significant. For the banking industry, risk factors are more relevant to strategy, fund operations, and capital loan. For the tourism industry, the risk profile is completely different—business expansion risk has a higher proportion, cost risk, partnership risk, demand fluctuation risk, and epidemic risk can reflect the particularity of the tourism industry. Taking partnership risk and demand fluctuation risk as examples, we can explain this phenomenon from the following two aspects: 1) Tourism is an intangible product and there are many types of services involved. The sales of tourism products rely on third-party booking and payment platforms, and the perfect tourism services also depend on franchises (Guo et al., 2013); 2) "Tourism is best defined on the demand side" (Keller, 2006: 18). So, unlike the energy and banking industries, tourism is not necessary for daily life, and its demand is extremely sensitive to international political, economic, and climatic conditions. The consumer preference also has strong volatility, which results in significant demand fluctuation (Gautam, 2012; Ridderstaat et al., 2014). All the facts prove the necessity and rationality of risk identification in the tourism industry.

Table 6

Top 10 risk exposures of four tourism sub-industries.

No.	Casinos & gaming		Hotels, resorts & cruise lines		Leisure facilities		Restaurants	
1	Regulation changes	12.55%	Regulation changes	9.75%	Stock market risk	13.14%	Business expansion risk	12.25%
2	Stock market risk	10.19%	Stock market risk	9.49%	Business expansion risk	11.32%	Regulation changes	12.22%
3	Business expansion risk	8.57%	Business expansion risk	9.40%	Regulation changes	9.51%	Stock market risk	8.93%
4	Debt risk	7.06%	Market risk	9.07%	Market risk	6.74%	Cost risk	5.67%
5	Market risk	5.42%	Partnership risk	8.45%	Debt risk	6.23%	Market risk	5.41%
6	Funding risk	5.20%	Cost risk	5.11%	Investment risk	6.16%	Information technology risk	4.99%
7	Competition risk	4.59%	Investment risk	4.64%	Human resource risk	4.36%	Supply risk	4.79%
8	Human resource risk	4.30%	Stakeholder's interest risk	4.46%	Funding risk	4.05%	Potential/ongoing lawsuits	4.70%
9	Information technology risk	4.24%	Debt risk	4.37%	Catastrophes	3.74%	Competition risk	3.95%
10	Credit risk	4.10%	Funding risk	3.48%	Cost risk	2.79%	Demand fluctuation risk	3.38%

Different risk performance among sub-industries

We further divide the tourism industry into four sub-industries—casinos & gaming, hotels, resorts & cruise lines, leisure facilities, and restaurants—in accordance with GICS. Then, we calculate the top 10 risk exposures in each sub-industry and further compare them to distinguish the main threats of different sub-industries.

As shown in Table 6, regulation changes, stock market risk, business expansion risk, and market risk have higher disclosure frequency in all sub-industries. There are also some visible differences between them. For casinos & gaming, debt risk, funding risk, and credit risk are relatively high compared to others due to its important economic attribute. The regulation on whether it operates legally will also be stricter. For leisure facilities, there is a certain similarity with the gaming industry, which are mainly reflected in the higher debt risk, human resource risk, and funding risk compared with the other two industries. Furthermore, due to a large number of outdoor activities, the probability of catastrophes is relatively high. The larger floor space makes its real estate investment risk very significant, the same is true of hotels, resorts & cruise lines. For restaurants, the success of their growth strategy upon their ability to expand new restaurants and franchise operations, which led to a higher business expansion risk. As a sub-industry with the closest contact with consumers, its sale is easily affected by consumer preferences or discretionary spending, which brings strong fluctuations in demand. Moreover, the restaurant industry is strongly dependent on the supply of raw materials, making supply risk, and cost risk relatively significant. Litigation risk caused by food safety is also an important characteristic risk. For hotels, resorts & cruise lines, in addition to higher cost risk due to the similar reason with that of the restaurant industry, partner risk is also very prominent, because an increase in the use of third-party travel website and internet reservation channels may negatively impact their revenues (Law, 2006). These differences can reflect their distinct business characteristics, which have important implications for providing more targeted decision-making suggestions and risk hedging strategies between different sub-industries.

The performance of risk exposure over years

To understand the performance of risk exposures in different years, we draw the annual changing trend of three types of risk exposure in Fig. 5, where Fig. 5(a) is the top 5 risks with the largest cumulative proportion during the sample period, Fig. 5(b) is the top 5 risks with highest geometric growth rate, and Fig. 5(c) is the top 5 risks with the highest decay rate.

As can be seen from Fig. 5(a), risk exposures with high disclosure proportion have been relatively stable over the years. Among these risks, macro market risk and stock market risk have increased significantly after the financial crisis. The results in Fig. 5(b) show that information technology risk not only has a higher exposure proportion but also shows a higher growth trend (4.98%), implying the significance of cyber security for business operations with the advent of the digital age and the popularity of online transactions. Although the annual exposure proportion of epidemic risk and food safety risk is very low, it shows an obvious growth trend, with a growth rate of 5.86% and 4.36%, respectively. So, the emergencies related to visitor health also need to attract the attention of stakeholders. Besides, the growth rate of tax risk and asset impairment risk is 5.98% and 5.25%, respectively. The exposure proportion of risks presented in Fig. 5(c) shows a downward trend over the years. Among them, seasonal risk has the highest decay rate (4.47%). With the diversification of tourism products and the increasing universality of tourism activities, the season obstacles to the tourism industry will gradually decrease. The decay rate of other four risks—competition risk, stakeholder's interest risk, credit risk, and funding risk is 3.95%, 2.90%, 2.80%, and 2.57%, respectively.

Conclusions

In this paper, we successfully use the text mining technology to identify the risk exposures of tourism companies from the textual risk disclosure in financial statements, which has compensated for the lack of research on company risk identification. By analyzing 51,008 risk headings from 255 public companies' financial statements, we discover 30 risk exposures for tourism companies. In which, the regulation changes, business expansion risk, and stock market risk are the most common risk exposures faced by tourism companies. Business expansion risk, partnership risk, demand fluctuation risk, seasonal risk, food safety risk, and epidemic risk are the most representative risk of the tourism industry. From the internal perspective, each sub-industry has its unique business
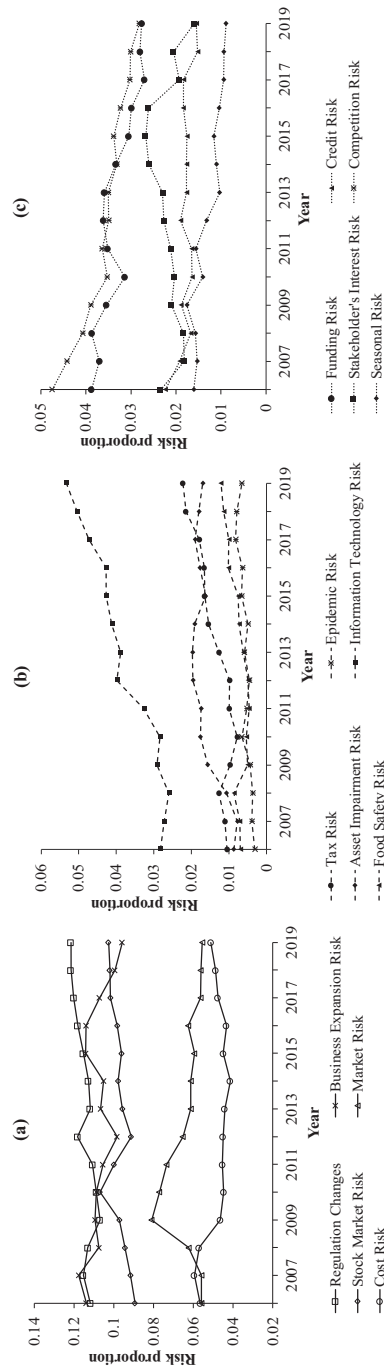
Fig. 5. Risk proportion trend over the years.

characteristics, and the risk exposures it faces are also different. High debt risk, funding risk, and credit risk can reflect the economic attribute of casinos & gaming industry; Partnership risk is the representative risk of hotels, resorts & cruise lines industry; High investment risk and catastrophe risk is the effect of many outdoor activities and larger floor space in the leisure facilities industry; Supply risk, potential/ongoing lawsuits, and demand fluctuation risk can reflect the characteristics of restaurants industry. From the time perspective, most risk exposures maintain a relatively stable disclosure proportion. Risk exposures with average fluctuation rate exceeding 5% include Tax risk (5.98%), epidemic risk (5.86%), asset impairment risk (5.25%), information technology risk (4.98%), food safety risk (4.36%), and seasonal risk (−4.47%).

The findings of our research have some practical implications. First, our findings can help investors better understand the risk status of the tourism industry, especially the newly identified risk exposures with industry representativeness in Table 4 and Table 5, which usually has a more serious impact on the tourism industry. Besides, when making more detailed investment choices, the risk differences between different sub-sectors can provide a reference for formulating a reasonable risk hedging strategy. For managers, they should pay more attention to the risk exposures with high disclosure frequency, and constantly optimize the business structure in the daily operation process to improve the replying ability. Besides, although the epidemic risk and food safety risk have lower disclosure proportions, they are extremely volatile and uncertain. Once it happens, the impact is devastating in most cases, which should cause great attention. To this end, managers need to strengthen food safety management, strictly control all aspects of procurement-production-sales, and carry out innovation to improve the intelligence level and diversity business types, which can play a key role in responding to public safety events.

The current work is just a preliminary attempt. Although we have solved the dilemma of tourism company risk identification in terms of data sources and analysis methods with the help of financial statements and topic model. There are also some limitations need to be solved in the future. First, Form 10-K restricts sample selection to the scope of listed companies with a certain size, so, the information of SMEs is still unable to be effectively captured. In the future, risk identification for small and medium-sized tourism companies will be a new challenge worthy of attention. Second, to obtain richer research conclusions, sentiment analysis can be considered to measure the severity and urgency of risk exposures. Besides, as shown in the discussion section, although there are fewer disclosures of epidemic risk, its volatility is extremely strong, and once it happens, it is usually accompanied by a serious negative shock. The past outbreaks of SARS and avian flu had a material and severe impact on the tourism industry (Kuo et al., 2008). The recent worldwide outbreak of COVID-19 has brought the world to a standstill, and tourism has been one of the hardest-hit of all major economic sectors. It is estimated that in 2020 global international tourist arrivals could decline between 20% to 30%, and tourism could lose up to 18% of their usual output (World Tourism Organization, 2020). In the future, risk perception of COVID-19 and its full impact on international tourism will become hotspots.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.annals.2020.102986.

References

Adam, I. (2015). Backpackers' risk perceptions and risk reduction strategies in Ghana. *Tourism Management, 49*, 99–108.
Azzopardi, L., Girolami, M., & Risjbergen, K. V. (2003, July). Investigating the relationship between language model perplexity and IR precision-recall measures. *Proceedings of the 26th annual international ACM SIGIR conference on research and development in informaion retrieval* (pp. 369–370). ACM.
Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science, 60*(6), 1371–1391.
Barry, C. B., & Brown, S. J. (1985). Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis, 20*(4), 407–422.
Bauer, R. A. (1960). *Consumer behavior as risk taking.* Chicago, IL384–398.
Beck, U., Lash, S., & Wynne, B. (1992). *Risk society: Towards a new modernity. Vol. 17.* sage.
Blake, A., & Sinclair, M. T. (2003). Tourism crisis management: US response to September 11. *Annals of Tourism Research, 30*(4), 813–832.
Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics, 1*(1), 17–35.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.
Boksberger, P. E., & Craig-Smith, S. J. (2006). Customer value amongst tourists: A conceptual framework and a risk-adjusted model. *Tourism Review, 61*(1), 6–12.
Borde, S. F. (1998). Risk diversity across restaurants: An empirical analysis. *Cornell Hotel and Restaurant Administration Quarterly, 39*(2), 64–69.
Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes, 36*(3), 391–405.
Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in company filings. *Review of Accounting Studies, 19*(1), 396–455.
Cardona, O. D., van Aalst, M. K., Birkmann, J., Fordham, M., McGregor, G., & Mechler, R. (2012). *Determinants of risk: Exposure and vulnerability.*
Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems,* 288–296.
Cox, D. F., & Rich, S. U. (1964). Perceived risk and consumer decision-making—The case of telephone shopping. *Journal of Marketing Research, 1*(4), 32–39.
Cuccia, T., & Rizzo, I. (2011). Tourism seasonality in cultural destinations: Empirical evidence from Sicily. *Tourism Management, 32*(3), 589–595.
Cui, F., Liu, Y., Chang, Y., Duan, J., & Li, J. (2016). An overview of tourism risk perception. *Natural Hazards, 82*(1), 643–658.
Dickinson, J. E., & Peeters, P. (2014). Time, tourism consumption and sustainable development. *International Journal of Tourism Research, 16*(1), 11–21.
Dowling, G. R., & Staelin, R. (1994). A model of perceived risk and intended risk-handling activity. *Journal of Consumer Research, 21*(1), 119–134.

Duan, H., Wang, S., & Yang, C. (2020). Coronavirus: Limit short-term economic damage. *Nature, 578*(7796), 515.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics, 64*(2–3), 221–245.

Evans, N., Stonehouse, G., & Campbell, D. (2012). *Strategic management for travel and tourism.* Taylor & Francis.

Fuchs, G., & Reichel, A. (2011). An exploratory inquiry into destination risk perceptions and risk reduction strategies of first time vs. repeat visitors to a highly volatile destination. *Tourism Management, 32*(2), 266–276.

Gautam, V. (2012). An empirical investigation of consumers' preferences about tourism services in indian context with special reference to state of Himachal Pradesh. *Tourism Management, 33*(6), 1591–1592.

Glaesser, D. (2006). *Crisis management in the tourism industry.* Routledge.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(Suppl. 1), 5228–5235.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267–297.

Guo, X., Ling, L., Dong, Y., & Liang, L. (2013). Cooperation contract in tourism supply chains: The optimal pricing strategy of hotels for cooperative third party strategic websites. *Annals of Tourism Research, 41*, 20–41.

Haddock, C. (1993). *Managing risks in outdoor activities.* New Zealand Mountain Safety Council.

Haldrup, M. (2004). Laid-back mobilities: Second-home holidays in time and space. *Tourism Geographies, 6*(4), 434–454.

Harvey, C. R. (1995). The risk exposure of emerging equity markets. *The World Bank Economic Review, 9*(1), 19–50.

Hope, O. K., Hu, D., & Lu, H. (2016). The benefits of specific risk-factor disclosures. *Review of Accounting Studies, 21*(4), 1005–1045.

Huang, J. H., Chuang, S. T., & Lin, Y. R. (2008). Folk religion and tourist intention avoiding tsunami-affected destinations. *Annals of Tourism Research, 35*(4), 1074–1078.

Huang, K. W., & Li, Z. (2011). A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. *ACM Transactions on Management Information Systems (TMIS), 2*(3), 18.

Jang, S. S. (2004). Mitigating tourism seasonality: A quantitative approach. *Annals of Tourism Research, 31*(4), 819–836.

Kapuściński, G., & Richards, B. (2016). News framing effects on destination risk perception. *Tourism Management, 57*, 234–244.

Keiser, J. D. (1998). Hospitality and tourism: A rhetorical analysis and conceptual framework for identifying industry meanings. *Journal of Hospitality & Tourism Research, 22*(2), 115–128.

Keller, P. (2006). *Innovation and tourism policy.*

Kim, C. S., Bai, B. H., Kim, P. B., & Chon, K. (2018). Review of reviews: A systematic analysis of review papers in the hospitality and tourism literature. *International Journal of Hospitality Management, 70*, 49–58.

Kim, H., Kim, J., & Gu, Z. (2012). An examination of US hotel companies' risk features and their determinants of systematic risk. *International Journal of Tourism Research, 14*(1), 28–39.

Kim, W., Jun, H. M., Walker, M., & Drane, D. (2015). Evaluating the perceived social impacts of hosting large-scale sport tourism events: Scale development and validation. *Tourism Management, 48*, 21–32.

Kim, Y., Kim, M., & O'Neill, J. W. (2013). Advertising and company risk: A study of the restaurant industry. *Journal of Travel & Tourism Marketing, 30*(5), 455–470.

Knight, F. H. (1921). *Risk, uncertainty and profit.* Boston: Houghton Mifflin.

Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies, 18*(4), 1088–1122.

Kuo, H. I., Chen, C. C., Tseng, W. C., Ju, L. F., & Huang, B. W. (2008). Assessing impacts of SARS and Avian Flu on international tourism demand to Asia. *Tourism Management, 29*(5), 917–928.

Law, R. (2006). Internet and tourism—Part XXI: TripAdvisor. *Journal of Travel & Tourism Marketing, 20*(1), 75–77.

Lee, B. K., & Lee, W. N. (2004). The effect of information overload on consumer choice quality in an on-line environment. *Psychology & Marketing, 21*(3), 159–183.

Lee, S. K., & Jang, S. S. (2011). Foreign exchange exposure of US tourism-related companies. *Tourism Management, 32*(4), 934–948.

Leiper, N. (1979). The framework of tourism: Towards a definition of tourism, tourist, and the tourist industry. *Annals of Tourism Research, 6*(4), 390–407.

Li, J., Li, G., Liu, M., Zhu, X., & Wei, L. (2020). A novel text-based framework for forecasting agricultural futures using massive online news headlines. *International Journal of Forecasting.* https://doi.org/10.1016/j.ijforecast.2020.02.002.

Liesch, P. W., Welch, L. S., Welch, D., McGaughey, S. L., Petersen, B., & Lamb, P. (2002). Evolving strands of research on company internationalization: An Australian-Nordic perspective. *International Studies of Management & Organization, 32*(1), 16–35.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management, 29*(3), 458–468.

Mansfeld, Y., & Pizam, A. (Eds.). (2006). *Tourism, security and safety.* Routledge.

Mei, Q., Shen, X., & Zhai, C. (2007, August). Automatic labeling of multinomial topic models. *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 490–499). ACM.

Myers, S. C., & Majluf, N. S. (1984). Company financing and investment decisions when companies have information that investors do not have. *Journal of Financial Economics, 13*(2), 187–221.

Paraskevas, A., & Quek, M. (2019). When Castro seized the Hilton: Risk and crisis management lessons from the past. *Tourism Management, 70*, 419–429.

Park, J. Y., & Jang, S. S. (2014). Sunk costs and travel cancellation: Focusing on temporal cost. *Tourism Management, 40*, 425–435.

Park, S., Song, S., & Lee, S. (2017). Company social responsibility and systematic risk of restaurant companies: The moderating role of geographical diversification. *Tourism Management, 59*, 610–620.

Penela, D., & Serrasqueiro, R. M. (2019). Identification of risk factors in the hospitality industry: Evidence from risk factor disclosure. *Tourism Management Perspectives, 32*, Article 100578.

Polanyi, M. (2009). *The tacit dimension.* University of Chicago press.

Reisinger, Y., & Mavondo, F. (2005). Travel anxiety and intentions to travel internationally: Implications of travel risk perception. *Journal of Travel Research, 43*(3), 212–225.

Ridderstaat, J., Oduber, M., Croes, R., Nijkamp, P., & Martens, P. (2014). Impacts of seasonal patterns of climate on recurrent fluctuations in tourism demand: Evidence from Aruba. *Tourism Management, 41*, 245–256.

Roehl, W. S., & Fesenmaier, D. R. (1992). Risk perceptions and pleasure travel: An exploratory analysis. *Journal of Travel Research, 30*(4), 17–26.

Sheng-Hshiung, T., Gwo-Hshiung, T., & Kuo-Ching, W. (1997). Evaluating tourist risks from fuzzy perspectives. *Annals of Tourism Research, 24*(4), 796–812.

Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). *The Oxford English dictionary*(2nd ed.). Oxford, UK: Clarendon.

Singal, M. (2015). How is the hospitality and tourism industry different? An empirical test of some structural characteristics. *International Journal of Hospitality Management, 47*, 116–119.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Why study risk perception? *Risk Analysis, 2*(2), 83–93.

Taylor-Gooby, P., & Zinn, J. O. (Eds.). (2006). *Risk in social science.* Oxford University Press.

Thomas, W. I., & Thomas, D. S. (1928). *The child in America: Behavior problems and programs.* New York: Knopf.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

U.S. Securities and Exchange Commission. (2005, December 1). Secutities offering reform. Retrieved from http://www.sec.gov/rules/final/33-8591.pdf.

Viglia, G., & Dolnicar, S. (2020). A review of experiments in tourism and hospitality. *Annals of Tourism Research, 80*, Article 102858.

Wang, Y., Li, B., Li, G., Li, J., Zhu, X., & Li, J. (2019). Risk factors identification and evolution analysis from textual risk disclosures for insurance industry. *Procedia Computer Science, 162*, 25–32.

Wei, L., Li, G., Li, J., & Zhu, X. (2019a). Bank risk aggregation with forward-looking textual risk disclosures. *The North American Journal of Economics and Finance, 50*, Article 101016. https://doi.org/10.1016/j.najef.2019.101016.

Wei, L., Li, G., Zhu, X., & Li, J. (2019b). Discovering bank risk factors from financial statements based on a new semi-supervised text mining algorithm. *Accounting & Finance, 59*(3), 1519–1552.

Wei, L., Li, G., Zhu, X., Sun, X., & Li, J. (2019c). Developing a hierarchical system for energy company risk factors based on textual risk disclosures. *Energy Economics, 80*, 452–460.

Williams, A. M., & Baláž, V. (2015). Tourism risk and uncertainty: Theoretical reflections. *Journal of Travel Research, 54*(3), 271–287.

Wolff, K., Larsen, S., & Øgaard, T. (2019). How to define and measure risk perceptions. *Annals of Tourism Research, 79*(November), 102759.

Wong, J. Y., & Yeh, C. (2009). Tourist hesitation in destination decision making. *Annals of Tourism Research, 36*(1), 6–23.

World Tourism Organization. (2020). Tourism and COVID-19. Retrieved from https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2020-04/COVID19_NewDS_.pdf.

Yang, E. C. L., Khoo-Lattimore, C., & Arcodia, C. (2017). A systematic literature review of risk and gender research in tourism. *Tourism Management, 58*, 89–100.

Yang, E. C. L., & Nair, V. (2014). Tourism at risk: A review of risk and perceived risk in tourism. *Asia-Pacific Journal of Innovation in Hospitality and Tourism (APJIHT), 3*(2), 1–21.

Yang, Y., Zhang, H., & Chen, X. (2020). Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak. *Annals of Tourism Research*Article 102913. https://doi.org/10.1016/j.annals.2020.102913.

Zhu, X., Yang, S. Y., & Moazeni, S. (2016). Company risk identification through topic analysis of textual financial disclosures. *Computational intelligence (SSCI), 2016 IEEE symposium series on* (pp. 1–8). IEEE December.

Zinn, J. (2004). *Literature review: Economics and risk.* Canterbury: University of Kent (SCARR Working Paper 2004/2).